

Gender gaps in primary school achievement. A decomposition into endowments and returns to IQ and non-cognitive factors

Citation for published version (APA):

Golsteyn, B. H. H., & Schils, T. (2014). *Gender gaps in primary school achievement. A decomposition into endowments and returns to IQ and non-cognitive factors*. ROA. ROA Research Memoranda No. 007
<https://doi.org/10.26481/umaror.2014007>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/umaror.2014007](https://doi.org/10.26481/umaror.2014007)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Bart Golsteyn, Trudie Schils

**Gender Gaps in Primary
School Achievement. A
Decomposition into
Endowments and Returns to
IQ and Non-cognitive Factors**

RM/14/017

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O. Box 616
NL- 6200 MD Maastricht
The Netherlands

Gender Gaps in Primary School Achievement

A Decomposition into Endowments and Returns to IQ and Non-cognitive Factors

Abstract

In elementary school, girls typically outperform boys in languages and boys typically outperform girls in math. The determinants of these differences have remained largely unexplored. Using rich data from Dutch elementary schools, we decompose the differences in achievement into gender differences in endowments and returns to IQ and non-cognitive factors. This descriptive analysis is a thought experiment in which we show the consequences for school performance if girls and boys would have similar resources and take similar advantage of these resources. Our findings indicate that gender differences in resources with respect to social and instrumental skills and need for achievement can explain part of the differences in performance. Boys seem to be better equipped with these resources. Additionally, boys and girls employ their skills differently. Girls take more advantage of their IQ than boys. Yet, the largest part of this parameter effect is left unexplained by IQ and non-cognitive factors.

JEL-codes: I21, I24

Keywords: Gender, Education, Achievement, Decomposition, Personality

1. Introduction

A vast literature has reported substantial differences in elementary school performance between boys and girls: boys are better in math, while girls outperform boys in languages. This has important consequences for choices made later on in education and the labor market. From current studies it remains unclear whether the discrepancies are related to differences between boys and girls in resources with respect to IQ and non-cognitive factors (e.g. social skills or personality), or whether they are due to differences in the way boys and girls use these resources for school success. For example, it could be that there are no differences in resources between boys and girls, but that boys take less advantage of their resources when performing certain tasks at school. If so, policies that try to prepare boys for school by increasing their resources might have limited effects in raising their performance and it might be better to target at a more effective use of their resources.

In this paper we decompose gender differences in math and language test scores into (a) differences in resources with respect to IQ and non-cognitive factors (i.e. endowment effects) and (b) differences in the effectiveness of using these resources to obtain higher test scores (i.e. parameter effects). We use unique data from a large sample of Dutch elementary school children aged 11/12, including both test scores in math, language and IQ and a wide range of non-cognitive factors such as social and instrumental skills and personality.

We apply an Oaxaca-Blinder decomposition to separate the effects of gender disparities in resources and returns to resources on school achievement. The main results are that differences in resources with respect to social and instrumental skills and need for achievement can explain part of the differences in performance. Boys seem to be better equipped with these resources. Additionally, boys and girls employ their skills differently. Girls take more advantage of their IQ than boys. Yet, the largest part of the parameter effect is left unexplained by IQ and non-cognitive factors.

Our paper relates to literature on the assessment and explanation of gender differences in school achievement. There is consensus in the literature that girls outperform boys in reading and languages (see, e.g., Cornwell et al. 2013, Husain and Millimet 2009, Lai 2010). This finding is consistent across

countries, different age groups, survey periods and study programmes (see, e.g., European Commission 2010 for an overview). In mathematics, boys generally outperform girls, although gender differences vary more across age groups than for languages. Most studies show that boys' advantage in math emerges during elementary school and becomes more pronounced in high school and at later ages (Bedard and Cho 2010, Fryer and Levitt 2010, Husain and Millimet 2009).¹ Fryer and Levitt (2010) show that the advantage of boys in math is persistent since the 1970s. The size of the gap varies to some extent across mathematical tasks. Doolittle (1989) shows that in high school girls perform worse on geometry and reasoning items, but that they perform similarly on algorithmic, operations-oriented items. With respect to math, Ellison and Swanson (2010) show that the gender gap in skills widens at the top of the distribution. This is not the case for languages. Girls have an advantage in reading and language across the full distribution.

Various explanations for these gender differences have been put forward, including a different translation of background characteristics into test scores, biological and psychological differences. Some studies show that factors such as parental background, school organization, teacher qualifications, teaching methods and classroom factors (e.g. diversity and peer effects) affect boys and girls differently and thereby explain part of the gender disparities in schooling outcomes (Bertrand and Pan 2013, Ellison and Swanson 2012, Tansel 2002, Tinklin 2003).

Another explanation of the gender disparities in school achievement is related to biological differences between boys and girls, such as differences in brain structure and functioning, and differences in hormonal levels (Cahill 2006, Sax 2005, Spelke 2005). For example, girls have a faster development of the right side of the brain, associated with talking, vocabulary, reading and memory, while boys have a faster development of the left side of the brain, associated with visual-spatial-logical skills, perceptual skills, math, problem-solving, building and figuring out puzzles. This is related to the observed gender

¹ There is also some evidence that the gender gap in performance is declining over the years, with higher rates of college entrance and higher GPA for girls (Becker et al. 2010, Fortin et al. 2013). Fortin et al. (2013) show that this changing gender gap in recent years is largely related to higher educational expectations by girls in secondary education.

differences in specific intelligence factors, as shown by a meta-analysis of Hyde (2005): adolescent girls perform better in perceptual speed and verbal tasks, while boys perform better in mental rotation and math tasks.

Partly related to this literature on differential brain development between boys and girls is the literature that explains the gender gap in school achievement by differences in learning-styles, attitude, school motivation and other personality traits between boys and girls. Poropat (2009) argues that academic performance is related to the capacity, opportunity, and willingness to perform. Some studies have suggested that boys develop a 'laddish' culture, encompassing socially disruptive behavior, less self-discipline, anti-school attitudes and interest in traditionally masculine paradigms, which negatively affects boys' school motivation and educational achievement (Cornwell et al. 2013, Epstein et al. 1998, van Houtte 2004, Spinath et al. 2010, Steinmayr and Spinath 2008, Fortin et al. 2013). Moreover, while girls are more serious, diligent, studious and show an increased maturity and have more effective learning strategies at an earlier age (emphasizing collaboration, talking and sharing), boys are less inclined to use cooperative behavior and discussion to enhance learning (Barres 2006, Boaler 1997, Duckworth and Seligman 2006, Gipps 1996). In addition, there are gender disparities in both self-perceived ability and interest in math and reading. Boys are typically more self-confident about their math skills and show a stronger interest in math, while girls have such convictions about languages (Jacobs et al. 2002, Barres 2006).

Cornwell et al. (2013) show that controlling for these differences in personality almost eliminates the estimated gender gap in reading and math grades. Other studies point to gender differences in the returns of personality in terms of test scores, including: (a) a positive relationship between extraversion and grades for girls but a negative one for boys, (b) a larger return for girls of verbal intelligence with respect to scores on language tests (Spinath et al. 2010, Steinmayr and Spinath 2008).

This literature thus reveals that gender differences in school performance are related to differences in IQ and non-cognitive factors between boys and girls. However, it remains unclear whether

the differences in performance are related more to differences in resources of boys and girls, or to differences in the use of such resources.

Our paper contributes to this literature by disentangling these two effects. To the best of our knowledge, such a decomposition has not yet been performed.² Understanding why boys and girls perform differently in school, helps to design school policy that encourages a more efficient use of available talents. It might also foster our understanding of gender differences in later life outcomes, such as wages, career paths and health outcomes.

This paper continues as follows. Section 2 describes the approach. Section 3 discusses the data. Section 4 shows the results and Section 5 concludes.

2. Approach

We apply an Oaxaca-Blinder decomposition to separate the effects of gender differences in resources and differences in returns of resources on school achievement.

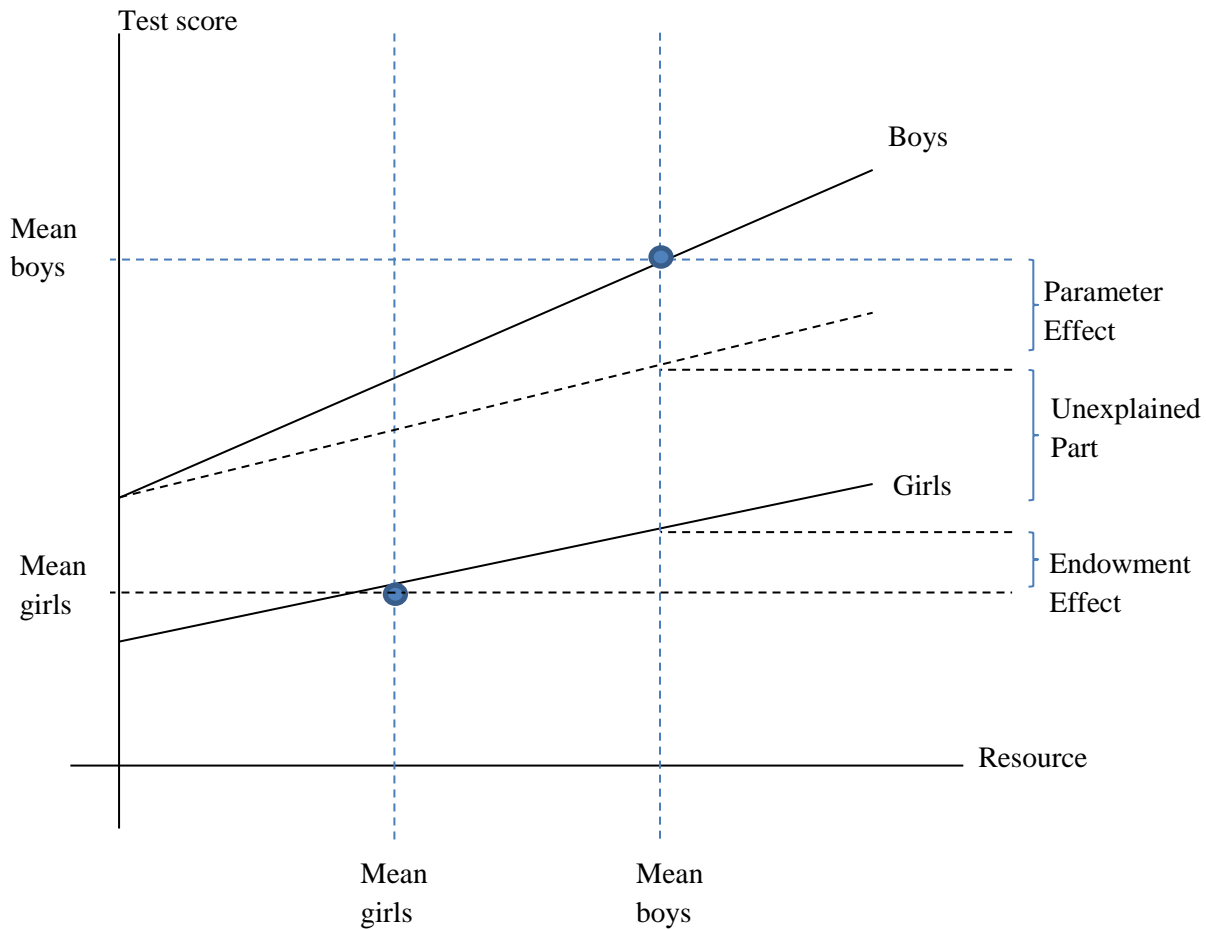
Figure 1 gives a graphical presentation of the decomposition. The figure shows that the mean test score, the mean level of a given resource, and the relationship between the resource and the test score differ between boys and girls. The overall difference in the mean test score can be decomposed into three parts. First, if girls would have the same amount or level of a given resource as boys, their score would increase (i.e. a shift along the girls' curve). This is the endowment effect. Second, if the relationship between the resource and the test score would be the same for boys and for girls, the score of girls would further improve (i.e. a rotation of the girls' curve). This is the parameter effect. The third part of the overall gender gap in the test score is the remaining difference that cannot be explained by the endowment or parameter effect (i.e. a shift of the girls' curve).

² In related work, Bertrand and Pan (2013) explore the role of the home and school environments in the gender gap in disruptive behavior using a decomposition method. In another study in psychology, Steinmayr and Spinath (2008) analyze gender differences in personality and intelligence and their predictive power for achievement tests for 16 year olds at the German gymnasium track. Our study differs from theirs in the sense that we make a formal decomposition using standardized tests, for a representative sample of younger children in elementary education (i.e. before they are tracked).

This exposition reveals several additional points. One is that the perspective of the analysis is of importance. If we would have taken the boys' perspective in the above example, the endowment effect, the parameter effect, and the unexplained part would be different compared to the analysis from the girls' perspective.³ In our analysis, we always take the perspective of the lowest achieving group. So for math test scores, we analyze how the performance of girls could be improved relative to that of boys, while for language test scores, we show how the performance of boys can be improved relative to girls. In this way, our analysis relates best to the policy perspective which is targeted at an improvement of school performance. A second point is that our analysis can be seen as a thought experiment in which boys are assumed to get similar resources and returns of resources as girls (and vice versa) to see how this would affect their test scores. The extent to which resources with respect to IQ and non-cognitive factors such as social skills or personality are malleable is an important discussion in the literature (for an overview of this literature, see Borghans et al. 2008). Thirdly, we use words which may suggest causality (e.g. "effects"), but obviously our results do not allow for a causal interpretation. The estimates should be interpreted as predictive estimates and not as causal effects since there may be important confounders.

³ The figure for instance now draws a curve parallel to that of girls which starts at the intersection of the boys' curve with the vertical axis. If we would take the boys' perspective, we would draw a curve parallel to the boys' curve which starts at the intersection of the girls' curve with the vertical axis. This would alter the indicated effects.

Figure 1. Decomposition of gender gap in test scores in endowment and parameter effects of a specific resource



Taking the perspective of improving the math performance of girls, the endowment effect can be estimated by multiplying the difference in the mean of the resource between boys and girls with the slope of the girls' regression line. The parameter effect is the multiplication of the mean of the resource for boys with the difference in the slopes of the regression lines of boys and girls (i.e. the interaction effect of gender and the resource in a regression of the test score on gender, resource, the interaction effect and

other controls).⁴ Our measures of school performance are math achievement test scores and language achievement test scores. We investigate to what extent differences between boys and girls with respect to IQ and a variety of non-cognitive factors and returns to these resources can explain the variety in performance on the test scores. The non-cognitive factors include: social skills, instrumental skills, the Big-5 personality (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism), need for achievement, and self-control. In the next section we describe these variables in more detail.

In our estimations, we control for the level of education of the parents, information about the composition of the household, the year in which the survey was performed, and school fixed effects. In robustness analyses, we include other controls which are not available in each year of the survey: number of siblings, position of the child in the household (i.e. first born, second born, etc.), and language spoken by the child with the mother. The results of these analyses are similar to our main results.⁵

3. Data

We use a large and unique dataset collected by Maastricht University on the educational development of students in elementary school in the south of the Netherlands. This region has about 609 000 inhabitants, and a population density of 922 inhabitants per square kilometer. This density is almost twice as high as the average population density in the Netherlands, which is 496 inhabitants per square kilometer. The region consists of 18 municipalities. The average disposable household income in the region is about 31500 Euro per year, which is somewhat lower than the national average of 34200 Euro per year (Statistics Netherlands 2013). The data are collected in a cooperative project with schools, schools boards and municipalities to analyze school performance in order to foster educational improvement. The project

⁴ To estimate the effect of many resources at the same time, one should first run a regression of the math test score on all resources for girls only. The endowment effect would then be the multiplication of the difference in the means of the resources between boys and girls with their respective parameters. Second, based on the regression for girls, a prediction for boys can be made. This prediction should be regressed on all the resources for boys only. The resulting parameters for boys give the differences in the slopes between boys and girls. The means of the resources for the boys should then be multiplied with the respective parameters in order to elicit the parameter effects. We use the Stata command “oaxaca” which performs exactly these calculations and reports the standard errors. Jann (2008) describes the procedure in detail.

⁵ Results of the robustness analyses are not presented in this paper, but available upon request.

aims to report on school effectiveness in elementary school, and data is gathered on students' cognitive abilities (IQ and achievement tests), their socio-economic background and a range of non-cognitive factors (through student and parental questionnaires).

Most schools in the region participate in the project, so sample selection problems are expected to be small. The sample consists of 80 to 92 percent of all elementary schools in the given region of the Netherlands. The schools which do not participate are special education schools, schools that have a philosophy not to test children, and schools which were unable to plan the survey activities.

For this study, we use data of four cohorts of students in 6th grade (final year of elementary school in the Netherlands), for the years 2009 through 2012. In total, we have 12891 observations.

Dependent variables: math and language achievement test scores

In the final year of elementary school (6th grade), children take a three-day national achievement test (the CITO test). Scores on this test are very important to determine to which track students are allocated in secondary education. About 85 percent of the schools use this test. The test is developed and assessed by CITO, a non-profit institute which is independent from the schools, so the assessment is not assessed by the teachers themselves. Because of the tracking advice that follows from this test, it is considered to be a high-stakes achievement test. The test contains 290 multiple choice questions testing the students on three main domains: Dutch language, mathematics, and world orientation (e.g. geography, biology and history). For this paper, we use the subscores on Dutch language and mathematics, where we take the relative number of questions answered correctly as the test score. We standardize all scores to have a mean of zero and a standard deviation of one. Figures 2 and 3 show the distributions of the math and language test scores for boys and girls respectively. The figures show that girls score lower on the math test and higher on the language test than boys.

Figure 2. Distributions of math scores for boys and girls

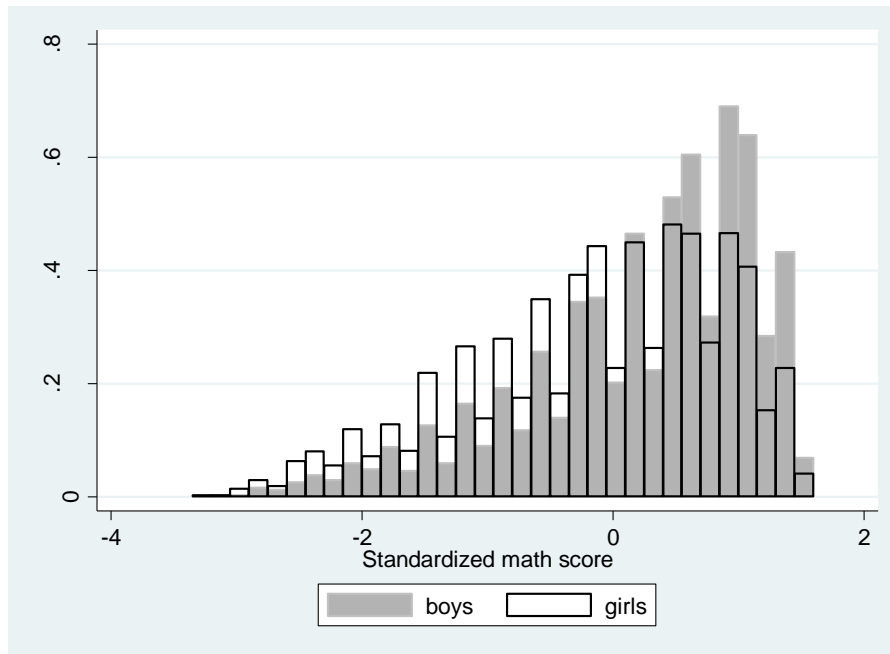
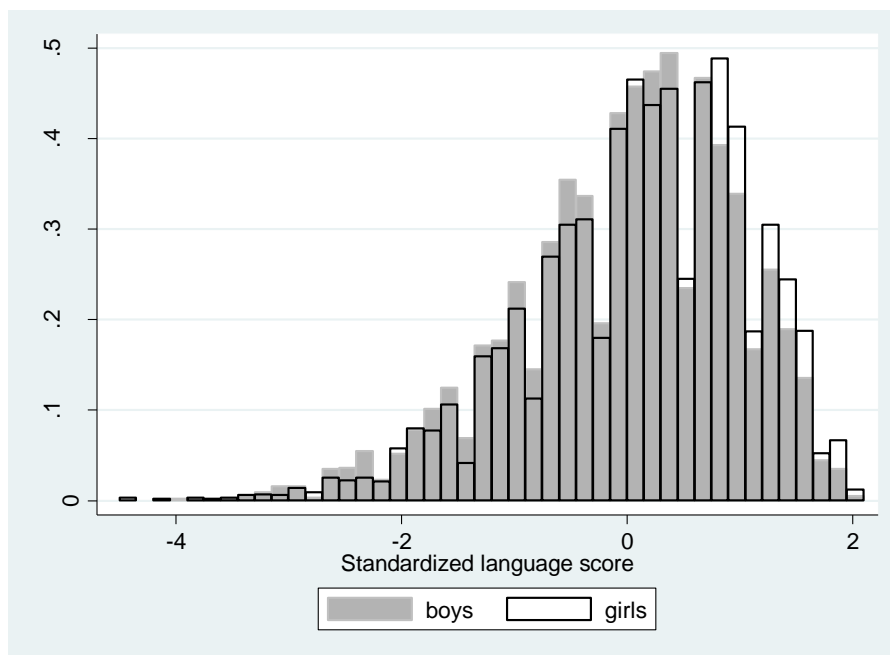


Figure 3. Distributions of language test scores for boys and girls



Explanatory variables: IQ and non-cognitive factors

The students also completed a non-verbal intelligence test.⁶ This is a test of visual-spatial intelligence by means of figures in the two-dimensional field, i.e. a puzzle of geometric forms where students have to identify the missing pieces. Students have to answer 43 questions. We use the relative number of questions answered correctly as the measure of intelligence.

Non-cognitive factors comprise social and instrumental skills and personality traits. The students are confronted with statements to assess their self-perception of social and instrumental skills.⁷ Social or affirmative skills include listening to others, comforting someone, looking nice, and socializing with others. Instrumental skills or assertiveness include giving your own opinion, winning at a verbal fight, being dominant, and getting your way. Table A1 in the appendix shows the statements used to derive these skills. The statements are evaluated on a 4-point scale, ranging from “very bad” to “very good”. We transformed this into a score of 0 to 1. The mean scores are stable over time.

To assess personality we also use statements. Students complete 25 statements with which the Big-5 personality is assessed (see table A1 in the Appendix). For each personality trait, five statements are used (factor analysis confirms the assignment of items to traits). Each statement is measured on a 5-point scale ranging from “fully disagree” to “fully agree”. The scores for each trait are derived using the average of the trait-specific statements, reversing the negatively formulated statements.⁸ In addition to the Big-5 personality traits, students completed some statements on ambition, need for achievement, grit and self-control (for an overview of all statements see table A1 in the appendix). Again, each statement was measured on a 5-point scale ranging from “fully disagree” to “fully agree”. A factor analysis (principal component factor analysis, using varimax rotation) shows two distinct factors. The first factor seems to be

⁶ This test was designed by van Batenburg and van der Werf (Groningen University).

⁷ These skills are all self-assessed. This is standard in the literature but could be a source of bias if girls or boys systematically over- or underestimate their social/instrumental skills, whether internally or based on their perception of norms.

⁸ The scores are relatively stable over time, although in some years some of the personality scores seem to deviate. For example, the 2012 cohort scores lower on openness and extraversion compared to previous cohorts. This is due to the fact that for these traits fewer statements are used (e.g. statements on which students generally score higher are excluded). Conscientiousness slightly increases over the years, whereas neuroticism declines over the years. In our analysis, we use mean scores over the four years since we are not interested in changes over the years. In addition, we include year dummies in the analyses to control for differences in measurement across years.

related to, what is known in the literature as need for achievement and grit (perseverance in reaching longer term goals). Although Duckworth and Quinn (2009) show that these are two distinct factors, excluding the need for achievement items from the factor grit in our sample does not improve the inter-item covariance (see table A2 in the appendix). This is most likely due to the limited number of items we use to measure this. We therefore keep them together and label this “need for achievement”. The second factor relates more to self-control as it measures the perseverance in reaching short-term goals (Duckworth et al. 2007). The scores for both factors are derived using the average of the trait-specific statements, reversing the negatively formulated statements.

Control variables

We use various controls in our estimations. We control for the elementary school which the student attends (dummies), and the survey year (dummies). As a proxy for the socioeconomic background of the students, parental educational attainment is used. We take the level of education of the highest educated parent and distinguish between four levels (dummies for each level): (1) primary or lower secondary education, (2) upper secondary education or lower tertiary education, (3) higher tertiary vocational education, and (4) higher tertiary academic education. In addition, we account for household composition, by measuring whether the student lives with (a) both his mother and his father, (b) his mother and a partner (not being the father), (c) his mother only, or (d) his father only. Again we include dummies for the various family compositions. In robustness analyses we included a wider range of household indicators, such as the number of siblings (continuous variable), or the position of the child in the household (i.e. first born, second born, etc.), and language spoken by the child with the mother. These additional variables are only available in three out of the four years.

4. Results

This section consists of three parts. First, we show the observed gender differences in IQ and non-cognitive factors. Second, we show the decomposition between endowment and parameter effect at an aggregate level. The third part shows the decomposition in more detail for math and language test scores.

4.1 Differences in IQ and non-cognitive factors

Table 1 shows the gender differences in resources, i.e. IQ and non-cognitive factors. We find that girls score slightly but significantly higher on the non-verbal IQ-test. When controlling for the other resources the difference in the IQ-test scores is more pronounced. The IQ-test taken here tests visual-spatial intelligence by means of figures in the two-dimensional field. Gender disparities related to spatial intelligence have been reported in various psychological and educational studies, yet overall the observed gender differences are small and account for no more than 1 to 5 percent of the group variance (see, e.g., Vogel 1990 for an overview of the literature).

In addition, we observe that girls score significantly higher on social skills, whereas boys score significantly higher on instrumental skills. This is not surprising as the items that are labelled social skills seem to reflect feminine characteristics such as being empathetic and caring, while the items that are labelled instrumental skills relate more to power and aggression (see Feingold 1994).

Boys score lower on need for achievement and higher on self-control. The latter is also found by Duckworth and Seligman (2006). Because we use a composite factor of grit and need for achievement, it is more difficult to compare our findings to other findings in the literature. Duckworth and Quinn (2009) find no gender differences across adults on their grit-scale (excluding need for achievement). Gneezy and Rustichini (2004) find that young boys report a higher need for achievement than girls.

With respect to the gender differences in personality, the table shows that boys score higher on openness and extraversion, while girls score higher on conscientiousness, agreeableness and neuroticism. In the literature, there is considerable variation in the observed gender differences in personality, depending on age or nationality and the instruments used to assess personality (Costas et al. 2001; Marsh

et al. 2010). Gender differences in neuroticism and agreeableness seem consistent across different samples, whereas gender differences in conscientiousness, extraversion and openness seem inconsistent. This could explain why the difference in conscientiousness is no longer significant when controlling for all other traits.

Table 1: Differences in resources between boys and girls

| | Mean girls | SD girls | Mean boys | SD boys | t-value of difference (separate) | t-value of difference (simultaneous) |
|----------------------|------------|----------|-----------|---------|----------------------------------|--------------------------------------|
| IQ | 0.761 | 0.105 | 0.757 | 0.112 | -2.45 | -8.11 |
| Social skills | 0.737 | 0.153 | 0.589 | 0.180 | -50.19 | -58.14 |
| Instrumental skills | 0.523 | 0.176 | 0.607 | 0.168 | 27.68 | 36.39 |
| Need for achievement | 0.689 | 0.167 | 0.670 | 0.172 | 15.53 | 17.91 |
| Self-control | 0.769 | 0.149 | 0.809 | 0.141 | -9.35 | -3.75 |
| Openness | 0.513 | 0.180 | 0.548 | 0.185 | 10.91 | 12.18 |
| Conscientiousness | 0.559 | 0.175 | 0.529 | 0.174 | -9.93 | 0.92 |
| Extraversion | 0.524 | 0.187 | 0.537 | 0.180 | 4.08 | 6.22 |
| Agreeableness | 0.801 | 0.127 | 0.730 | 0.139 | -29.91 | -12.12 |
| Neuroticism | 0.412 | 0.217 | 0.374 | 0.212 | -10.04 | -15.66 |
| Observations | 6600 | | 6291 | | | |

Note: All resources are scaled 0-1. In the simultaneous analysis, we run a regression with the respective resource as dependent variable and we include gender and all other resources (IQ and non-cognitive factors) and controls (parental education, household composition, school, and survey year) as independent variables. We report t-values of the gender dummy.

4.2 Results at an aggregate level

Table 2 shows the decomposition of the observed gender gap in math and language scores into endowment and parameter effect on the aggregate level. For math scores, we take the perspective of how to improve the scores of girls relative to those of boys. For language scores, we take the perspective of how to improve the scores of boys relative to those of girls.

The first part of the table shows the overall gaps in math and language test scores: boys score on average 0.35 standard deviations higher on the math test than girls. Girls score on average 0.13 standard deviations higher on the language test. These differences are highly statistically significant. An interesting question is to what extent these differences actually matter in practice. In the Netherlands, the test in 6th grade (that we use for the analyses) is used to determine a student's track in secondary education. There

are three main tracks in secondary education: (1) the lowest track prepares students for vocational education; (2) the middle track prepares students for higher professional education, and (3) the highest track prepares students for academic education. There are no official cutoff points for these tracks with respect to the 6th-grade test, but using the observed cutoff points in practice we can calculate the effects of raising girls' math scores or boys' language scores. Currently 36.2 percent of the girls attend the highest two tracks and 19.5 percent of the girls attend the highest track. If girls would reach similar math scores as boys, these percentages would increase to 44.9 and 27.8 respectively. Alternatively, currently 40.4 percent of the boys attend the highest two tracks and 20.8 percent of the boys attend the highest track. If boys would reach similar language scores as girls, these percentages would increase to 45.6 and 25.5 respectively. *Ceteris paribus*, inducing girls to be as good in math as boys yields the highest returns in terms of percentage of students in the highest track(s).

The total effects row in table 2 reveals that if girls would have similar resources as boys in terms of IQ and non-cognitive factors, the overall gap in math test scores would be 58 percent ($0.204/0.350$) smaller than the current gap (overall endowment effect). In addition, if girls would take similar advantage of these resources (given their own level) as boys do, the overall gap in math test scores would be 40 percent smaller than the current gap (overall parameter effect). So these resources seem very important predictors for math test scores. For language test scores, table 2 shows that if boys have similar resources as girls in terms of IQ and non-cognitive factors, their performance would be even worse: it would decrease by 0.19 standard deviations, so the gap in language test scores would be more than double as large as the current gap. For languages, the overall parameter effect seems more important than the overall endowment effect. If boys would take similar advantage of their resources as girls, the gap in language test scores would be in favor of boys (0.18 standard deviations higher).

In the last part of table 2, we show the endowment and parameter effects for IQ, the non-cognitive factors and the controls. The parameter effects are broken up into an unexplained part (i.e. the intercept in our model) and a part that is explained by the parameters in our model. The results reveal that non-cognitive resources matter most for school achievement. Girls' math scores would improve if they

had similar non-cognitive resources as boys. Boys' language scores would be larger if they had similar non-cognitive resources as girls. In the next section we explore these non-cognitive factors in more detail. Resources with respect to IQ do not affect the gender gaps in math and language to a large extent. This may be due to the fact that the gender difference in IQ is rather small (4 percentage points). Finally, controls also seem somewhat more favorable for boys. Table 2 shows that if girls would have similar controls as boys, their math scores would improve, yet if boys have similar controls as girls, their language scores worsen.

When looking at the parameter effects in table 2, we find that a large part of this effect remains unexplained by the observed variables in our model. This unexplained part seems in favor of boys, since girls would perform better in math when employing these unobserved resources in a similar way as boys. We find that girls take more advantage of their IQ with respect to math and language test scores. If girls would take similar advantage of their IQ as boys, their math test scores would worsen, yet if boys would take similar advantage of their IQ as girls, their language test scores would improve. We show the effects of the endowments and parameters for the separate traits in the sections below.

Table 2. Overall differences between boys and girls ^a

| | Math | Language |
|---|----------------------|---------------------|
| Average group 1 ^b (high performers) | 0.226*** (0.012) | 0.118*** (0.012) |
| Average group 2 ^c (low performers) | -0.123*** (0.013) | -0.007 (0.012) |
| Difference | 0.350*** (0.017) | 0.125*** (0.017) |
| Observations | 12,891 | 12,891 |

| | Endowments | Parameters | Endowments | Parameters |
|-------------------------|---------------------|----------------------|----------------------|---------------------|
| Total effects | 0.204*** (0.020) | 0.141*** (0.021) | -0.186*** (0.018) | 0.309*** (0.021) |
| Partial effects | | | | |
| <i>Explained part</i> | | | | |
| IQ | -0.018** (0.007) | -0.630*** (0.115) | 0.009** (0.004) | 0.251** (0.112) |
| Non-cognitive factors | 0.207*** (0.017) | -0.023 (0.128) | -0.177*** (0.015) | -0.153 (0.126) |
| Controls ^d | 0.015** (0.007) | -0.051 (0.222) | -0.019** (0.008) | 0.719*** (0.253) |
| <i>Unexplained part</i> | | 0.845*** (0.273) | | -0.508* (0.298) |

^a Scores on the math and language tests are standardized to mean zero and standard deviation one. IQ and non-cognitive factors are scaled 0-1. The estimation also includes interaction effects between endowments and parameters. These estimates which are statistically insignificant and/or economically negligible are not reported but available on request. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

^b For the math scores, group 1 are the boys and group 2 are the girls. For language scores, group 1 are the girls and group 2 are the boys.

^c Non-cognitive factors include social and instrumental skills, the Big-5 personality, self-control and need for achievement. We show the coefficients and standard errors for the joint effect.

^d Controls include survey year, school, parental education, and family composition fixed effects. We show the coefficients and standard errors for the joint effect.

4.3 Results at a detailed level

Table 3 gives a more detailed composition of the math and language test scores into endowment and parameter effects. Earlier, we concluded that differences in non-cognitive resources are an important determinant for the gender differences in achievement test scores. Table 3 reveals that the most important non-cognitive factors are social and instrumental skills. If girls would have similar social and instrumental skills as boys, i.e. if they have lower social skills but higher instrumental skills, their math test scores would improve with 0.11 and 0.06 standard deviations respectively. On the contrary, if boys would have

similar social and instrumental skills as girls, their language test scores would deteriorate with 0.08 and 0.05 standard deviations respectively. The non-cognitive factors openness to experience and need for achievement are less important in size, but still significant. If girls would have similar resources in these traits, their math scores would improve with 0.03 and 0.01 standard deviations respectively. If boys would have similar openness to experience and need for achievement as girls, their language scores would deteriorate with 0.04 and 0.02 standard deviations respectively. In addition, for language test scores, agreeableness seems a favorable resource: if boys would have similar levels of agreeableness as girls, their language scores would improve with 0.03 standard deviations. The endowment effects of the other non-cognitive factors are much smaller in size.

Table 3 also shows the parameter effects for math and language test scores, i.e. the multiplication of the difference in parameters with the mean of the given resource for boys.⁹ As mentioned before, the parameter effect is mainly driven by IQ. Girls take more advantage of IQ with respect to both math and language test scores compared to boys. In addition, if girls would take similar advantage of their need for achievement as boys, their math test score would increase with 0.23 standard deviations, but if boys would use their need for achievement similarly to girls their language test score would decrease with 0.41 standard deviations. So both having more need for achievement (endowment effect) and taking more advantage of it (parameter effect) are positively related to achievement test scores and in favor of boys. Additionally, if boys would take similar advantage of their self-control as girls, their language test scores would deteriorate with 0.12 standard deviations.

Interestingly, if girls would take similar advantage of their openness to experience as boys, their math test scores would decrease by 0.13 standard deviations, whereas if boys take similar advantage of their openness to experience as girls, their language test scores would improve with 0.19 standard deviations. Hence, having more openness to experience (endowment effect) is positively related to achievement test scores and in favor of boys, whereas girls take more advantage of it (parameter effect).

⁹ Table A3 in the appendix reveals the difference in parameters. The table contains regression results of test scores on IQ and non-cognitive factors, for women and men separately.

In addition, if boys would take similar advantage of conscientiousness as girls, their language scores would improve with 0.12 standard deviations. The other parameters are less important in size and the effects are statistically insignificant.

Table 3
Decomposition of gender differences regarding math and language test scores ^a

| | Math | | Language | |
|-------------------------|----------------------|----------------------|----------------------|----------------------|
| | Endowments | Parameters | Endowments | Parameters |
| Total effects | 0.204*** (0.020) | 0.141*** (0.021) | -0.186*** (0.018) | 0.309*** (0.021) |
| Partial effects | | | | |
| <i>Explained part</i> | | | | |
| IQ | -0.018** (0.007) | -0.630*** (0.115) | 0.009** (0.004) | 0.251** (0.112) |
| Social skills | 0.114*** (0.013) | -0.017 (0.082) | -0.077*** (0.011) | -0.030 (0.066) |
| Instrumental skills | 0.055*** (0.007) | -0.064 (0.055) | -0.048*** (0.007) | 0.090 (0.064) |
| Self-control | -0.009*** (0.002) | -0.043 (0.073) | 0.007*** (0.002) | -0.118* (0.071) |
| Need for achievement | 0.013*** (0.004) | 0.233** (0.100) | -0.024*** (0.004) | -0.407*** (0.106) |
| Openness to experience | 0.025*** (0.003) | -0.125*** (0.048) | -0.035*** (0.004) | 0.188*** (0.052) |
| Conscientiousness | 0.003 (0.002) | -0.033 (0.055) | -0.018*** (0.003) | 0.124** (0.052) |
| Extraversion | -0.003*** (0.001) | 0.037 (0.050) | 0.005*** (0.002) | -0.048 (0.050) |
| Agreeableness | 0.001 (0.007) | -0.049 (0.106) | 0.026*** (0.007) | -0.004 (0.097) |
| Neuroticism | 0.009*** (0.002) | 0.038 (0.031) | -0.013*** (0.002) | 0.052* (0.028) |
| <i>Controls</i> | | | | |
| Parental education | 0.007** (0.003) | -0.233*** (0.027) | -0.008** (0.003) | 0.253*** (0.027) |
| Survey year | -0.002 (0.002) | -0.112*** (0.029) | -0.002 (0.003) | 0.210*** (0.029) |
| School | 0.010* (0.006) | 0.347* (0.180) | -0.008 (0.006) | 0.262 (0.221) |
| Family composition | -0.000 (0.001) | -0.053 (0.121) | -0.001 (0.001) | -0.006 (0.114) |
| <i>Unexplained part</i> | | 0.845*** (0.273) | | -0.508* (0.298) |

^a Scores on the math test are standardized to mean zero and standard deviation one. IQ and non-cognitive factors are scaled 0-1. Regarding the controls: instead of showing the estimates for each separate dummy, we show the coefficients and standard errors for the joint effects of the dummies. The estimation also includes interaction effects between endowments and parameters. These estimates are statistically insignificant and/or economically negligible and not reported but available on request. The number of observations is 12,891. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

5. Conclusions

There are substantial differences in primary school performance between boys and girls: boys are better in math, while girls outperform boys in languages. This paper decomposes gender differences in test scores into (a) differences in resources with respect to IQ and non-cognitive factors and (b) differences in the translation of such resources into test scores. We use unique Dutch data from a large sample of elementary school children aged 11/12, including both standardized test scores in math, language and intelligence and an assessment of a wide range of non-cognitive factors (including social and instrumental skills, Big-5 personality traits, need for achievement and self-control). Applying an Oaxaca-Blinder decomposition, we separate the effects of gender disparities in resources and their returns on school achievement. The main results are that gender differences in resources with respect to social and instrumental skills and need for achievement can explain part of the differences in performance. Boys seem to be better equipped with these resources. Additionally, boys and girls employ their skills differently. Girls take more advantage of their IQ than boys. Yet, the largest part of this parameter effect is left unexplained by IQ and non-cognitive factors. Boys seem to employ these unobserved resources more efficiently than girls.

These findings cannot be interpreted as causal evidence, but they are informative about the direction in which schools can experiment to increase the math performance of girls or language performance of boys. From the data we learned that the gender gap in math performance is mostly an “endowment difference” in terms of non-cognitive factors. If girls would have similar resources as boys, i.e. less social skills, more instrumental skills, their test scores in math and language would be higher. These types of resources can be trainable: schools can train girls to be less occupied with social life at school or to become more assertive. Similarly, girls’ test scores would improve if they would be as open to experience as boys and have a similar need for achievement. This personality trait is perhaps less malleable (see, e.g., Borghans et al. 2008).

Additionally, in terms of the parameter effect, girls take more advantage of their IQ for tests in math and language. If boys would take equal advantage of IQ, openness to experience and conscientiousness, their test scores would plausibly be higher. Boys could benefit from training on how to raise the returns to these traits, but not so much from having more of these resources.

Acknowledgements

We received valuable comments from Lex Borghans, Eva Feron, Paul Jungbluth, Olivier Marie, the editor, and two anonymous referees.

References

- Barres, B. (2006). Does gender matter? *Nature* (442), 133-136.
- Becker, G.S., W.H.J. Hubbard and K.M. Murphy (2010). Explaining the worldwide boom in higher education of women. *Journal of Human Capital* 4(3), 203-241.
- Bedard, K. and I. Cho (2010). Early gender test score gaps across OECD countries. *Economics of Education Review* 29, 348-363.
- Bertrand, M. and J. Pan (2013). The trouble with boys: social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics* 5(1), 32-64.
- Boaler, J. (1997). Reclaiming school mathematics: the girls fight back. *Gender and Education* 9, 285-305.
- Borghans, L., A. Duckworth, J. Heckman and B. ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources* 43, 972-1059.
- Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews: Neuroscience* 7(6), 477-484.
- Cornwell, C.M., D.B. Mustard and J. van Parys (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school. *Journal of Human Resources* 48(1), 236-264.

- Costa, P. T. Jr., A. Terracciano, and R. R. McCrae (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology* 81, 322–331.
- Doolittle, A. E. (1989). Gender differences in performance on mathematics achievement items. *Applied Measurement in Education* 2(2), 161-177.
- Duckworth, A.L. and M. Seligman (2006). Self-discipline gives girls the edge: gender in self-discipline, grades and achievement test scores. *Journal of Educational Psychology* 98(1), 198-208.
- Duckworth, A. L., C. Peterson, M.D. Matthews and D.R. Kelly (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology* 92, 1087-1101.
- Duckworth, A.L. and P.D. Quinn (2009). Development and validation of the short grit-scale (Grit-S). *Journal of Personality Assessment* 91(2), 166-174.
- Ellison, G. and A. Swanson (2010). The gender gap in secondary school mathematics at high achievement levels: evidence from the American mathematics competitions. *The Journal of Economic Perspectives* 24(2), 109-128.
- Epstein, D., J. Elwood, V. Hey and J. Maw (1998). Schoolboy frictions: feminism and ‘failing’ boys. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.
- European Commission (2010). Gender differences in educational outcomes: study on the measures taken and the current situation in Europe. Brussels, Eurydice.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin* 116 (3), 429-456.
- Fortin, N.M., P. Oreopoulos and S. Phipps (2013), Leaving boys behind: gender disparities in high academic achievement, paper presented at NBER SI 2013, Cambridge: Massachusetts.
- Fryer, R. G. and S. D. Levitt (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics* 2(2), 210-240.

- Gipps, C. (1996). Review and conclusions: a pedagogy or a range of pedagogic strategies? in P. Murphy & C. Gipps (eds) *Equity in the Classroom: towards effective pedagogy for girls and boys*. London: Falmer Press, 260-271.
- Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *American Economic Review* 94(2), 377-381.
- Husain, M. and D. L. Millimet (2009). The mythical boy crisis. *Economics of Education Review* 28, 38-48.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist* 60(6), 581-592.
- Jacobs, J. E., S. Lanza, D.W. Osgood, J.S. Eccles and A. Wigfield (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development* 73, 509-527.
- Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal* 8 (4), 453-479.
- Lai, F. (2010). Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review* 29, 383-399.
- Marsh, H.W., O. Lüdtke, B. Muthén, T. Asparouhov, A.J.S. Morin and U. Trautwein (2010). A new look at the big five factor structure through explanatory structural equation modeling. *Psychological Assessment* 22 (3), 471-491.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin* 135 (2), 322-338.
- Sax, L. (2005). *Why gender matters: What parents and teachers need to know about the emerging science of sex differences*. New York: Doubleday.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist* 60(9), 950-958.

- Spinath, B., H. H. Freudenthaler and A. C. Neubauer (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences* 48, 481-486.
- Statistics Netherlands (2013), Statline online statistics, Heerlen: Statistics Netherlands.
- Steinmayer, R. and B. Spinath (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality* 22, 185-209.
- Tansel, A. (2002). Determinants of school attainment of boys and girls in Turkey: individual, household and community factors. *Economics of Education Review* 21, 455-470.
- Tinklin, T. (2003). Gender differences and high attainment, *British Educational Research Journal* 29 (3), 307-325.
- van Houtte, M. (2004). Why boys achieve less at school than girls: the difference between boys' and girls' academic culture. *Educational Studies* 30, 159-173.
- Vogel, S.A. (1990). Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities: a review of the literature. *Journal of Learning Disabilities* 23, 44-52.

Table A1. Statements used to measure non-cognitive factors and personality

Social skills

How do you rate yourself at:

1 = very bad,

2 = bad,

3 = good,

4 = very good

- working with animals
- looking good and dressing nicely
- dealing and playing with other children
- making friends
- playing music
- playing theatre
- comforting someone who is sad

Instrumental skills

How do you rate yourself at:

1 = very bad,

2 = bad,

3 = good,

4 = very good

- being the boss
- being stronger than others
- giving your own opinion about something
- winning at a fight
- saying what is on your mind
- getting your way
- discussing

Personality: openness

Do you agree with the following statement?

1 = I do not agree at all

2 = I do not agree

3 = neutral

4 = I do agree

5 = I do very much agree

- I use difficult words
- I have many ideas
- I learn quickly
- I have a big imagination
- I like collecting information

Personality: conscientiousness

Do you agree with the following statement?

1 = I do not agree at all

2 = I do not agree

3 = neutral

4 = I do agree

5 = I do very much agree

- I do chores immediately
- I leave my things all around the place
- I stick to my appointments
- I sometimes forget that I have to do things
- I am precise

Personality: agreeableness

Do you agree with the following statement?

1 = I do not agree at all

2 = I do not agree

3 = neutral

4 = I do agree

5 = I do very much agree

- I am interested in other people
- I am friendly
- I try to help other's
- I sometimes insult people
- I try to imagine how other people feel

Personality: extraversion

Do you agree with the following statement?

- 1 = I do not agree at all
- 2 = I do not agree
- 3 = neutral
- 4 = I do agree
- 5 = I do very much agree

- I talk a lot
- I am quiet among a group of strangers
- I am the one who stirs the party
- I like being around many people
- I like to be at the center of attention

Personality: neuroticism

Do you agree with the following statement?

- 1 = I do not agree at all
- 2 = I do not agree
- 3 = neutral
- 4 = I do agree
- 5 = I do very much agree

- I easily feel stressed
- I worry about things
- I easily feel anxious
- My mood changes a lot
- I often have a sad mood

Personality: need for achievement

Do you agree with the following statement?

- 1 = I do not agree at all
- 2 = I do not agree
- 3 = neutral
- 4 = I do agree
- 5 = I do very much agree

- I want to be the best
- Getting good grades at school is important
- I would like to be the number one in some activity
- Trying your best at school is important for later life
- I want to be good in my job later in life
- I want to get good grades at school

Personality: self-control

Do you agree with the following statement?

- 1 = I do not agree at all
- 2 = I do not agree
- 3 = neutral
- 4 = I do agree
- 5 = I do very much agree

- I continue with my task until I succeed
- If something becomes too difficult, I quit
- If I start something, I also finish it
- If something goes wrong, I lose concentration
- I do not care if I succeed

Table A2. Principal component factor analysis and Cronbach's alpha of items related to grit, ambition, self-control and need for achievement

| Item | Factor 1 | Factor 2 | Alpha factor 1 | Alpha factor 2 |
|--|----------|----------|----------------|----------------|
| I always want to be the best | 0.4765 | | 0.8570 | |
| Getting good grades at school is important | 0.6829 | | 0.8481 | |
| I want to be the best at something | 0.5809 | | 0.8564 | |
| Doing your best at school is important for later life | 0.6233 | | 0.8527 | |
| I want to be very good in my future job | 0.6399 | | 0.8434 | |
| The things we learn at school do not matter that much for later life | 0.3394 | 0.2703 | 0.8606 | |
| I want to get high grades | 0.7565 | | 0.8414 | |
| If I want something, I continue until I succeed | | 0.4761 | | 0.6502 |
| If something becomes too difficult, I quit | | 0.7392 | | 0.6075 |
| If I start something, I also finish it | | 0.6723 | | 0.6183 |
| If something turns out to be disappointing, I lose my interest | | 0.6843 | | 0.6564 |
| I do not care that much whether something goes well | 0.2812 | 0.4789 | | 0.6858 |
| I always work hard | 0.3335 | 0.4626 | | 0.6520 |
| Test Scale | | | 0.8698 | 0.6870 |

Table A3.

Regression results of test scores on IQ and non-cognitive factors, for women and men separately.

| | Math Women only | Math Men only | Language Women only | Language Men only |
|------------------------|----------------------|----------------------|------------------------|----------------------|
| IQ | 3.786*** (0.102) | 2.958*** (0.092) | 2.336*** (0.098) | 2.004*** (0.096) |
| Social skills | -0.771*** (0.084) | -0.794*** (0.072) | -0.577*** (0.082) | -0.525*** (0.074) |
| Instrumental skills | 0.652*** (0.072) | 0.529*** (0.074) | 0.717*** (0.070) | 0.568*** (0.077) |
| Self-control | 0.488*** (0.076) | 0.426*** (0.070) | 0.195*** (0.073) | 0.371*** (0.073) |
| Need for achievement | 0.332*** (0.090) | 0.634*** (0.088) | 0.104 (0.087) | 0.607*** (0.092) |
| Openness to experience | 0.721*** (0.066) | 0.478*** (0.063) | 1.330*** (0.064) | 0.987*** (0.066) |
| Conscientiousness | -0.084 (0.070) | -0.142** (0.067) | -0.353*** (0.068) | -0.587*** (0.070) |
| Extraversion | -0.235*** (0.067) | -0.165** (0.065) | -0.501*** (0.065) | -0.411*** (0.068) |
| Agreeableness | -0.011 (0.094) | -0.072 (0.086) | 0.366*** (0.091) | 0.372*** (0.090) |
| Neuroticism | -0.235*** (0.051) | -0.142*** (0.052) | -0.210*** (0.049) | -0.348*** (0.053) |
| Full set of controls | Incl. | Incl. | Incl. | Incl. |
| Constant | -3.313*** (0.584) | -2.469*** (0.781) | -2.801*** (0.565) | -2.293*** (0.811) |
| Observations | 6,600 | 6,291 | 6,600 | 6,291 |
| R-squared | 0.392 | 0.342 | 0.391 | 0.360 |

Note: Scores on the math and language tests are standardized to mean zero and standard deviation one. IQ and all non-cognitive factors are scaled 0-1. Controls include parental education fixed effects, survey year fixed effects, school fixed effects, and family composition fixed effects. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.