The Evolution of Gender Gaps in Numeracy and Literacy Between Childhood and Young Adulthood

Francesca Borgonovi* Álvaro Choi† Marco Paccagnella‡

This version: May 8, 2021

*University College London Institute of Education and OECD. 55-59 Gordon Square, London WC1H 0NU, United Kingdom. E-mail: f.borgonovi@ucl.ac.uk

†Universitat de Barcelona, Faculty of Economics and Business, 690 Diagonal Av. 08034, Barcelona, Spain. E-mail: alvarochoi@ub.edu Tel. +34934021815

‡**Corresponding author.** OECD Directorate for Education and Skills and Bank of Italy. 2, rue André Pascal, 75775 Paris Cedex 16, France. E-mail: marco.paccagnella@oecd.org

Abstract

Numeracy and literacy are important foundation skills which command significant wage premia in modern labour markets. The existence of gender differences in these skills is therefore of potential concern, and has spurred a large amount of research, especially with respect to numeracy skills. Still, little is known about the moment in which such gaps emerge, how they evolve, and if this evolution differs across countries. We use data from large scale international assessments to follow representative samples of birth-cohorts over time, and analyse how gender gaps in numeracy and literacy evolve from age 10 to age 27. We find that the advantage of boys in numeracy is small at age 10, but grows considerably between age 15 and 27. The gender gap in literacy follows a very different pattern: it is small at age 10, large and in favour of girls at age 15, and negligible by age 27.

Keywords: gender gaps, skills, numeracy, literacy, large-scale assessments
JEL Classification: I20, I24, J16, J24

1 Introduction

Human capital is widely recognised as a fundamental ingredient of the economic (and non-economic) success of individuals and societies, and there are reasons to believe that its relevance will grow further in the future (Arntz, Gregory, and Zierahn 2016; Brynjolfsson and McAfee 2014; Castex and Kogan Dechter 2014; Goldin and Katz 2008; OECD 2012).

Gender differences in educational choices and human capital accumulation can therefore be an important determinant of observed gender gaps in labour market outcomes. The historical gender gap in educational attainment in favour of males (Goldin, Katz, and Kuziemko 2006) has now been closed (and often reversed) in most advanced economies, spurring concerns about the degradation of boys' educational outcomes (DiPrete and Buchmann 2013; Fortin, Oreopoulos, and Phipps 2015; Lei and Lundberg 2020; Riordan 1999). However, the academic and career choices of young men and women remain remarkably different. Women are over-represented in the arts and the humanities, while men are over-represented in science, technology, engineering and mathematics (STEM) careers fields (Card and Payne 2021; Flabbi 2012; OECD 2015). This is one reason why the literature on gender differences in labour market outcomes has traditionally focused on gender gaps in numeracy and mathematics skills and on the choice of and returns to fields of study (Altonji, Arcidiacono, and Maurel 2016; Carnevale, Chea, and Strohl 2013; Kirkebøen, Leuven, and Mogstad 2016).

Contrary to school grades, which are normally assigned by teachers and can suffer from bias due to gender stereotypes (Lavy and Sand 2018; Terrier 2020), standardised large-scale assessments are blind-graded, and provide cross-country comparable information on the skills of assessed participants, most commonly in the domains of numeracy (or mathematics) and literacy (or reading). A common finding of international assessments of school-age children is that boys tend to outperform girls in numeracy, while girls perform better in literacy. Additionally, in most countries, the skill distribution is less dispersed for girls than for boys (Machin and Pekkarinen 2008). There are however significant cross-country differences in the magnitude of gender gaps in test scores, which the previous literature has tried to relate to cultural and institutional factors (Bedard and Cho 2010; Breda, Jouini, and Napp 2018; Gevrek, Gevrek, and Neumeier 2020; Guiso, Monte, Sapienza, and Zingales 2008).

What we still know relatively little about is how such gaps evolve as children grow up and progress in their school careers, and if such evolution differs across countries or across subjects. A partial exception is Contini, Tommaso, and Mendolia 2017, who examined the evolution of gender gaps in the mathematics proficiency of Italian students from 2^{nd} to 10^{th} grade and find that the gender gap tends to increase with age. In this paper we perform a conceptually similar exercise, broadening the analysis in three directions. First, we jointly look at the evolution of gender gaps in both literacy and numeracy, which allows to better understand career choices in terms of comparative, rather than absolute, advantage (Aucejo and James 2016; Goulas, Griselda, and Megalokonomou 2020; Stoet and Geary 2018). Indeed, the previous literature, often motivated by the willingness to explain and understand the under-representation of women in more lucrative STEM subjects and occupations, has mostly focused on gender gaps in numeracy or mathematics (Breda, Jouini, and Napp 2018; Gevrek, Gevrek, and Neumeier 2020; Guiso, Monte, Sapienza, and Zingales 2008). Second, we extend the analysis to early adulthood, which is especially important in order to capture what happens during and after the school-to-work transition. Third, we work with internationally comparable data covering a wide range of OECD countries.

In the absence of cross-country comparable longitudinal data, we combine information from existing cross-sectional large-scale international assessments: the Trends in International Mathematics and Science Study (TIMSS), the Programme in Reading and Literacy Skills (PIRLS), the Programme for International Student Assessment (PISA), and the Programme for the International Assessment of Adult Competencies (PIAAC). As all these studies contain representative samples of the same birth cohort at different points in time, we are able to track the evolution of gender gaps for a single cohort of students that participated in different waves of different surveys.

We find that the gender gap in numeracy in favour of boys tends to grow as students age, and that such growth is particularly pronounced after leaving compulsory schooling and entering post-compulsory education or the labour market. By contrast, the gender gap in literacy follows an inverted-U profile, being highest during the teenage years and lowest among young adults. Gender gaps evolve in similar ways at both ends of the skill distribution. The variability of performance is always higher among males, peaking during the teenage years in both numeracy and literacy. Moreover, while the pattern we just described is by and large common to all countries for which we have data, we also find that differences across countries in the magnitude of gender gaps are larger at older ages, for both literacy and numeracy.

Motivated by the fact that between-country heterogeneity is larger among young adults aged 26/27, we look at the extent to which the evolution of gender gaps from age 15 to age 26 is related to common indicators of gender equality at the national level. We find small correlations between these contextual factors and the evolution of gender gaps. Finally, we explore possible mechanisms that could drive our results at the individual level, focusing in particular on educational and career choices and on the frequency of engagement with reading, writing, and numeracy-related tasks. Part of the growth in the gender gap in numeracy can be attributed to educational and career choices, as boys are more likely to pursue STEM-related careers. Unsurprisingly, those who pursue a STEM-related career engage more frequently with tasks requiring the use of numeracy skills. At the same time, they are as likely as people in non-STEM jobs to practice reading and writing, a finding that can contribute to explain why boys are able to catch up with girls and close the gender gap in literacy.

The remainder of the paper is organised as follows. Section 2 describes the data we use. Section 3 presents the empirical strategy we adopt to analyse the evolution of gender gaps. Section 4 presents the main results of the analysis, as well as a number of robustness checks. In Section 5 we provide additional evidence in an attempt to account for the observed variability across countries, as well as for the broad pattern of increasing gender gaps in numeracy and shrinking gender gaps in literacy. Section 6 concludes.

2 Data sources

In the absence of individual-level and cross-country comparable longitudinal data, we combine information from different international cross-sectional assessments of numeracy and literacy. These assessments were mainly designed to monitor the proficiency of students at key developmental stages.

In particular, we use data from TIMSS, PIRLS, PISA, and PIAAC, four international large-scale assessments which, although targeted to different populations, are similar under many respects as they all assess the proficiency of respondents through similar assessments, relying on representative samples of their respective target population at the country level. TIMSS and PIRLS are managed by the International Association for the Evaluation of Educational Achievement (IEA) and survey children aged 9-10. PISA and PIAAC are both promoted and managed by the OECD and cover, respectively, 15-year old students (PISA) and adults aged 16 to 65 (PIAAC). TIMSS assesses numeracy proficiency (mathematics), PIRLS assesses literacy proficiency (reading), while PISA and PIAAC test respondents in both domains.[1]

Different waves of TIMSS, PIRLS and PISA have been administered over the past decades, while only one cycle of PIAAC has so far taken place in 2011/12[2]. We identify a set of countries that participated in all the four studies, and a single birth cohort that we can follow at different points in their life-cycle across the different assessments. When looking at proficiency in numeracy/mathematics, we are able to use data from eleven countries, while for literacy/reading we can rely on ten countries. For five countries (the Czech Republic, England, the Netherlands, Norway and the United States) we are able to use data on both subjects.

The cohort we focus on consists of individuals born in 1984/85. Such individuals were 10 years old when they participated in TIMSS 1995, 15 years old when they participated in PISA 2000, and 27 years old when they participated in PIAAC 2011/12.

Unfortunately, PIRLS was not administered in 1995. For this reason, we resort to data from PIRLS 2001, and we use them as a proxy for the literacy skills that individuals born in 1984/85 had at age 10. We believe this is a reasonable assumption to make, given the lack of major trends in gender gaps in literacy proficiency observed between PIRLS 2001 and PIRLS 2006 and between PISA 2000, 2003, and 2006 (Mullis, Martin, Foy, and Drucker 2012; OECD 2015). We look more closely at cohort effects in Subsection 4.3.

All four studies are conducted on nationally representative samples of their respective target population. For school-based surveys sample sizes normally range between 2,000 and 8,000 students (see Tables 1 and 2). Given that the PIAAC target population is much larger (all adults aged 16-65), the actual number of individuals from the

[1]See Mullis, Martin, Foy, and Arora (2012), Mullis, Martin, Foy, and Drucker (2012), OECD (2011), and OECD (2013b) for technical reports on the different studies.

[2]A second round of data collection took place in 2014 in a different set of countries. All the PIAAC data we use in this paper come from the first round of data collection which took place in 2011/12.

birth cohort we are interested in is much smaller. For this reason, we select from the PIAAC data all individuals born between 1983 and 1986, reaching in this way sample sizes of around 300 individuals in all countries.

2.1 Assessment frameworks

The data we use in this paper come from large-scale international assessments that share many common characteristics and aim at assessing similar skills. However, each assessment is conceptualised in a slightly different manner. Such differences in the underlying conceptual frameworks, together with differences in the test administration procedures, could in principle have a bearing on the estimation of gender gaps and on their observed evolution over time (Lietz 2006; Wu 2009). Assessment frameworks define what is being measured and turn general constructs (such as reading/literacy and mathematics/numeracy) into operational definitions in order to guide item development and test construction.

PIRLS, PISA and PIAAC share very similar assessment frameworks for literacy: the definitions of what it means to be "literate" are almost identical in the three studies. PIRLS defines literacy as "the ability to understand and use those written language forms required by society and/or valued by the individual" (Mullis, Martin, Gonzalez, and Kennedy 2003). Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment". In 2000 PISA defined reading literacy as "understanding, using, and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD 1999). PIAAC defines literacy as "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential" (PIAAC Literacy Expert Group 2009). This is, to a great extent, a reflection of the influence previous adult literacy assessments, such as the International Adult Literacy Survey (IALS) and the Adult Literacy and Lifeskills Survey (ALL), had on PISA and PIAAC (both PISA and PIAAC for example use some test items originally developed in the context of IALS), as well as the mutual influence that the three studies have had on one another.

Although the definition of the underlying construct is similar in the three studies, there are differences in how much the three assessments rely on stimuli which involve different types of texts (narrative vs. informational, continuous vs. non-continuous texts), as well as on response formats (i.e. constructed vs. multiple choice) and the cognitive processes involved in performing a given task (accessing vs. retrieving vs. integrating vs. interpreting information). PISA, for example, makes a more extensive use of constructed responses and continuous texts and the literature suggests that girls' advantage in continuous texts and in constructed responses tends to be wider (Lafontaine and Monseur 2009; Roe and Taube 2003; Routisky and Turner 2003). Similarly, there is evidence that the prevalence of items with a different response format can affect the size of the gender gap (Reardon, Kalogrides, Fahle, Podolsky, and Zarate 2018). Yet, these differences do not seem to have a significant impact on measured gender gaps, as country rankings and estimated gender gaps are similar across surveys in the different sub-domains (Borgonovi, Pokropek, Keslair, Gauly, and Paccagnella 2017).

Larger differences in the assessment frameworks characterise the assessment of mathematics/numeracy in TIMSS, PISA and PIAAC (Gal and Tout 2014; Wu 2009), the main difference being that TIMSS is a curricular assessment, while PISA and PIAAC are based more on the ability to solve real-life problems. In TIMSS 1995 the mathematics assessment was designed to be based, as closely as possible, on what students have actually been taught" (Wu 2009). In 2000 PISA defined mathematical literacy as "an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen" (OECD 1999). In PIAAC, numeracy is defined as "the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life" (PIAAC Numeracy Expert Group 2009).

Moreover, because the mathematics tasks encountered at school are very different from those of everyday life, even the problem-based tasks present in PISA and PIAAC respond to different frameworks and concepts of mathematics/numeracy. Nonetheless, several studies have examined gender differences in TIMSS and PISA, and country rankings for 8^{th} graders in the TIMSS mathematics assessment are highly correlated with the country rankings from the PISA mathematics assessment (Klieme 2016).

2.2 Survey administration

TIMSS, PIRLS and PISA have a very similar survey design: participants are selected through two-stage sampling, whereby schools are selected first and then students are sampled within schools. The main differences lie in the fact that TIMSS and PIRLS adopt a grade-based sampling strategy (effectively surveying 4^{th} grade students), and sample entire classes within schools. The PISA sample is age-based (eligible students are between 15 years and three months and 16 years and two months at the time of testing), and students (rather than classes) are randomly selected within schools.

PIAAC is instead a household survey. The target population is the non-institutionalised adult population aged 16 to 65 living in the country. PIAAC also adopts a complex, stratified (and country-specific) sampling design, whereby households are selected first, and then individuals are sampled within households. PIAAC is administered by trained interviewers, who normally visit participants in their household. Respondents go first through a 45-60 minutes background questionnaire (administered as a computer-assisted personal interview) and then take a computer-based adaptive assessment of literacy, numeracy and problem solving in technology-rich environments[3]. Students participating in TIMSS, PIRLS and PISA sit the assessment in schools, much like normal school examinations. A final key difference is that TIMSS, PIRLS and PISA are designed as timed assessments (with a strict time limit), while PIAAC is not timed. We will discuss in Section 6 how these differences in the assessment design and in test taking conditions can impact our results.

In all the four studies, proficiency is estimated using Item Response Theory (IRT) models. Individual responses to the assessment questions are combined with background information to estimate, for each respondent, a distribution of proficiency, from which a set of "plausible values" is drawn. Scaling depends on the set of countries that take part in the assessment, as well as on the specific IRT models used (a one-parameter Rasch model for PISA, a two-parameters model for PIAAC, and

[3]Participating adults are not administered the assessment of problem solving in technology-rich environments if they fail an assessment of very basic ICT skills (e.g. the ability to operate a mouse or a keyboard) or if they decide to opt-out of the computer-based assessment (in which case they are administered a paper-based version of the literacy and numeracy assessments).

a two-or three-parameter model for TIMSS and PIRLS). Raw scores are not directly comparable across different studies. Throughout our analyses, we standardise all scores by subtracting from individual scores (more precisely, from each individual plausible value) the overall mean score and dividing by the overall standard deviation. Following Jacob and Rothstein (2016), the mean and standard deviation used for the standardisation are computed on the broadest possible population, using all countries that have participated in each assessment.

3 Empirical Strategy

For each assessment, let Y_{ik} be the standardised test score of individual i in country k. Let G_i be a gender dummy (1=female).

Our baseline model is simply

$$Y_{ik} = \alpha + \beta_1 G_i + \epsilon_{ik} \tag{1}$$

which we estimate separately for each assessment. Throghout the paper, we always estimate both a pooled regression that also include country fixed effects, and separate regressions for each country. The estimated $\hat{\beta}_1$ provides a country- and assessment-specific estimate of gender gaps. To properly account for the complex sampling design, standard errors are estimated using Jacknife or Balanced-Repeated Replication (BRR), depending on the study. Survey weights are used throughout.[4]

We have intentionally decided not to include any additional covariate in equation 1 for at least three reasons. First, we believe that the unconditional gender gap is an interesting parameter in itself. Especially from a policy perspective, it is typically more important to know whether the performance of boys and girls differ, rather than whether it differs conditional on a range of other dimensions in which boys and girls also differ. A second and more practical reason is that the focus of this paper is on how gender gaps evolve over time (i.e. as students grow up and enter adulthood). Now, very few background variables are measured in a consistent and comparable way across the various surveys we use, so that including control variables would threaten the

[4]We use the `repest` Stata package (Avvisati and Keslair 2014) in order to handle the presence of plausible values and the different sampling designs.

comparability of the estimates across surveys.[5] Finally, adding additional covariates would be problematic for PIAAC data, given the relatively small sample size (as we only use data for a limited number of cohorts).

In order to investigate gender gaps at different parts of the skill distribution, we also estimate equation 1 using quantile regression, focussing in particular on the first and ninth decile of the skills distribution. Following Machin and Pekkarinen (2008), we also compute the male-female variance ratio and the male-female ratio at the top and at the bottom decile. The variance ratio is defined as the average of the rations of male and female variances, for each plausible value. The male-female ratio is simply the ratio between the number of males and the number of females scoring at the top and at the bottom decile.

We test the robustness of our results in different ways. First, we compare the pooled results from the full set of countries for which we have suitable data (10 for literacy, 11 for numeracy), with the pooled results for the subset of five countries for which we have data on both domains. Second, we compute gender gaps at the median and at the top and bottom deciles using the method proposed in Penney (2017), which delivers estimates that are robust to any monotonic transformation of the dependent variable. This should address concerns deriving from the possible lack of interval property of the scales in which test scores are reported (Bond and Lang 2013). The method simply consists of running an unconditional quantile regression (Firpo, Fortin, and Lemieux 2009) and then divide the estimated coefficient by the standard error of the regression. Third, we use longitudinal data from the Canadian Youth in Transition Survey (YITS), which followed the students who participated in the first PISA assessment in 2000 through to their young adulthood (OECD 2012). In 2009 the PISA reading assessment was administered as part of the YITS, which allows us to compare the evolution of gender gaps in Canada between age 15 and age 24 for the same individuals, with the results of our analysis based on a synthetic-cohorts approach.

Finally, we perform a slightly different analysis using what we label a "moving-cohorts" approach. Using only data from PISA and PIAAC, we look at the evolution

[5]It would be possible, for instance, to control for whether the students are born abroad, but in order to increase the chances that the synthetic cohorts we use are representative of the same underlying population, we decided to exclude foreign-born individuals from our samples.

of gender gaps from age 15/16 to age 26/27 following three different birth cohorts. The youngest cohort was born in 1989/90, and was tested in PISA 2006 at age 15/16 and in PIAAC 2011/12 at age 20/21. The second cohort, born in 1987/88, participated in PISA 2003 at age 15/16 and was aged 23/24 in PIAAC. The oldest cohort is the same we use for our main analysis, and it is formed by individuals born in 1984/85, tested in PISA 2000 at age 15/16 and in PIAAC at age 26/27. Being based on data for multiple cohorts, all assessed at age 15/16, this exercise is also useful to test for the presence of cohort effects. In fact, in the absence of cohort effects across the three different PISA waves, differences in the gender gaps observed in PIAAC for individuals aged 20 to 26 could be more credibly interpreted as age effects.

4 Results

This section is structured in three subsections, in which we present the results of the three different empirical analysis outlined above. In subsection 4.1 we present the evolution of gender gaps in numeracy and literacy between age 10 and age 27, following a single cohort across different assessments. We also describe the evolution of the gaps at the top and bottom of the skills distribution, and look at variance ratios and at differences in the representation of males and females at the top and bottom of the skills distribution. Subsection 4.2 shows that our results are robust to monotonic transformation of the scales in which test scores are expressed. Subsection 4.3 presents the results of what we label "moving-cohorts" analysis, and subsection 4.4 show results from longitudinal data from the Canadian YITS.

4.1 The evolution of gender gaps

Gender gaps in literacy evolve very differently from gender gaps in numeracy between age 10 and age 27, as shown in Figure 1

[INSERT FIGURE 1 ABOUT HERE]

In the case of numeracy, gender gaps are small at age 9/10, with an advantage for boys of about 3 percent of a standard deviation. They grow larger by age 15/16, when they reach 9 percent of a standard deviation, and are largest at age 26/27 (one third of a standard deviation). In the case of literacy, girls have a large advantage at age 9/10 (22 percent of a standard deviation), which grows even larger by age 15/16 (28 percent of a standard deviation). At age 26/27, however, such advantage shrinks to essentially zero, as young men actually have a non-statistically-significant advantage of 13 percent of a standard deviation. This pattern can be observed for the large majority of countries in our sample, as well as when running a pooled regression with the five countries for which we have data on both literacy and numeracy (see Figure 2 and Tables 3 and 4). Furthermore, Figure 2 makes evident that the between-country variability in gender gaps in both literacy and numeracy is greater at age 27 than at age 10 and age 15. However, the greater between-country variability observed at age 27 in both literacy and numeracy appear to stem from different processes. The change in literacy achievement differentials between males and females follows a similar pattern across countries, but it is more pronounced in some contexts than in others. By contrast, in numeracy, greater between-country variability at age 27 is due to widening gender gaps between age 15 and 27 in some countries while in others the gap shrinks over the same period.

[INSERT FIGURE 2 ABOUT HERE]

At age 9/10, the gender gap in numeracy is not statistically different from zero in many countries, and only in the Netherlands and in Korea it is larger than 10% of a standard deviation. Gender gaps in literacy are much larger (and in favour of girls), approaching 30% of a standard deviation in England and Sweden.

By age 15/16, gender gaps in numeracy grow larger everywhere but in England and the Netherlands. The same pattern (even more pronounced) can be observed for gender gaps in literacy, where girls increase their advantage in all countries.

At age 26/27, however, we observe a clear reversal of the gender gap in literacy. In numeracy, on the other hand, the gap in favour of boys tend to widen in most countries (and by a very large amount in England, Ireland, Japan, and the United States); it decreases substantially only in Korea and the Czech Republic, and stays roughly constant in Austria and the Netherlands. In literacy, the gap shrinks everywhere, changing sign in England and in the United States. In almost all countries it is no longer statistically significant (the exception here being Italy).

Results presented in Table 3 indicate that the widening of literacy gaps in favour of girls between age 9/10 and age 15/6 is considerably smaller than the shrinking of such gaps between age 15/16 and age 26/27. In the pooled sample the gender gap in literacy increases by 12 percent of a standard deviation between age 9/10 and age 15/16. However, it shrinks by 33 percent of a standard deviation between age 15/16 and age 26/27. England, Norway and the United States are the countries in which this pattern is most pronounced: in all these countries the gap initially widens slightly but between age 15/16 and age 26/27 it shrinks by over 40 percent of a standard deviation. Italy is the only country in which the initial widening of the gender gap in the earlier age group is more pronounced than the shrinking that occurs at older ages.

[INSERT TABLE 3 ABOUT HERE]

[INSERT TABLE 4 ABOUT HERE]

Table 5 summarises how gender gaps in literacy and numeracy evolve over time across the skills distribution.

[INSERT TABLE 5 ABOUT HERE]

At all ages, males tend to outperform females at the top of the numeracy skill distribution, while gender gaps are much smaller (and not statistically significant) at the 10^{th} percentile. In literacy, females outperform males at both tails of the distribution, but only until age 15/16. Males are over-represented at the bottom of the literacy distribution at all ages, but in numeracy only at age 10. They are always over-represented at the top of the numeracy distribution, and also at the top of the literacy distribution at age 26/27. Consistent with previous results, males' test scores display a higher variability at all ages in both domains, with a peak at age 15/16.

4.2 Robustness to transformations of the scales

Observed test scores are estimates of unobservable latent ability. In most large-scale assessments (including PIRLS, TIMSS, PISA and PIAAC) test scores are estimated using Item Response Theory (IRT). As pointed out, among others, by Bond and Lang (2013), test scores estimated by IRT are not uniquely identified, meaning that any monotonic transformation of the scale would fit the IRT model equally well. Monotonic transformations do preserve the ranking of individual scores, but scales are not guaranteed to have have interval properties: differences of the same magnitude in terms of the underlying unobserved ability would then not be mapped into differences of the same magnitude in estimated test scores (Penney 2017). Bond and Lang (2013) show how (arbitrary) monotonic transformations of the scale in which test scores are expressed affect the observed evolution of the black-white test scores gap over time. Here we implement the transformation of normalised gaps proposed by Penney (2017), who showed that dividing the coefficients estimated through unconditional quantile regressions (RIF-regression, popularised by Firpo, Fortin, and Lemieux (2009)) by the standard error of the regression delivers results that are robust to any monotonic transformation of the dependent variable.

The results of this exercise are presented in Table 6. While the magnitude of the estimated gaps sometimes change (especially at the bottom of the distribution), the main results concerning the evolution of the gaps over time are confirmed.

[INSERT TABLE 6 ABOUT HERE]

4.3 A closer look at the evolution of gender gaps between age 15 and age 27: a "moving-cohort approach"

In this subsection we draw on multiple PISA waves, matching them with the corresponding cohorts that participated in PIAAC 2012. This exercise serves two purposes. First, it allows to investigate whether the gender gap at age 15/16 changes over time (i.e. across PISA waves), which is also a test for the presence of cohort effects. Second, under the assumption of no cohort effects, PIAAC data can be used to look at the evolution of gender gaps between age 20 and age 26/27. Relying only on PISA and PIAAC data also substantially increase the number of countries that can be used in the analysis, from 10/11 to 21.

The pooled results for literacy and numeracy are reported in Table 7, while the country specific results are in Tables A-1 and A-2 in the Appendix.

[INSERT TABLE 7 ABOUT HERE]

The first clear result is the stability of gender gaps, in both literacy and numeracy, across the three different PISA waves, which is evidence for the lack of cohort effects. The second notable result is the disappearance of the gender gap in literacy as early as age 20/21, with men's advantage growing larger up to age 26 (although the gaps here are never statistically significant). In the case of numeracy the pattern is more linear, with gaps increasing from around 10 percent of a standard deviation at age 15, to 14 percent at age 20/21, to 20 percent at age 23/24 and 26/27. This same broad pattern is present in the majority of countries in the sample

4.4 Longitudinal evidence from the Canadian YITS

Following synthetic cohorts through different assessments, as we did in our baseline analysis, is not ideal, because differences in the sampling design and in the assessment frameworks could account for some of our findings. As a robustness check, we use Canadian data from the Youth in Transition Survey (YITS), which followed the students who participated in the first PISA assessment in 2000 through to their young adulthood. At two-year intervals, the original PISA respondents (PISA-15 from now on) were contacted and asked to provide information on their activities related to education and employment, their life choices, and their attitudes. In 2009, on top of standard questionnaire based instruments, they were re-administered a PISA assessment.

The 2009 sample of YITS was representative of the population of 15-year-old Canadian students in 2000 (we will denote them as PISA-24 from now on). A sample of 1,297 respondents took the assessment, which was conducted during May-June 2009 and consisted of a follow-up assessment of readings skills (but not mathematics skills) and a background questionnaire. Assessment results were scored in conjunction with the PISA assessment in 2009. Since the PISA-24 items were also included in the PISA 2009 assessment, qualified coders who scored the PISA 2009 test booklets also scored the PISA-24 test items (OECD 2012).

The PISA-24 longitudinal assessment used a selection of assessment questions known as the PISA link items. This selection of test items was also used for testing reading as a minor domain in PISA 2003 and PISA 2006 and allowed for trend analyses. This ensures that the two tests tap on the same underlying construct. Furthermore, all PISA tests are measured on the same scale, meaning that the raw scores are directly comparable. PISA-24 women outperformed PISA-24 men in reading, but by a smaller margin than they did at age 15. The average gap narrowed from 41 percent to 23 percent of a standard deviation.[6] These results are similar to those we obtain in the moving-cohort analysis for Canada, where the gender gap in literacy decreased from 29 percent of a standard deviation at age 15 (in PISA 2003) to 12 percent of a standard deviation at age 23/24 (in PIAAC 2011/12). Among PISA-24 participants, girls outscored boys in PISA-15 by an average of 32 points; by 2009, that gap had narrowed to 18 points. This is due to a larger growth in test scores of males, which increased by 82 percent of a standard deviation, compared to 64 percent for women.

5 Making sense of the patterns: macro-level factors and micro-level mechanisms

The previous Section has established a set of robust results concerning the evolution of gender gaps in literacy and numeracy in a wide range of countries: (i) gender gaps in literacy peak at age 15 but then virtually disappear by age 26/27; (ii) gender gaps in numeracy grow substantially from age 15 to age 26/27; (iii) while this pattern is common across most countries, substantial heterogeneity emerges in the size of gender gaps between age 15 and age 26/27.

This Section attempts to shed some further light on the possible mechanisms that lie behind these patterns. Unfortunately, the data at our disposal have a number of limitations. Although they all come from similar international large-scale assessments, there are a number of survey-specific characteristics that may have a bearing on the observed size of gender gaps. In particular, beyond the curricular nature of the TIMSS numeracy assessment frameworks, the four studies differ in terms of: administration procedures, test length, mode of delivery, assessment content, response formats, response rates and treatment of missing answers or non reached items in the scaling model used to estimate performance. TIMSS, PIRLS and PISA are administered in schools in a group setting under the supervision of an invigilator, while PIAAC is

[6] We use here the overall standard deviation for PISA 2000 that we use in the rest of the paper.

administered in a one-to-one setting in people's homes under the presence of a trained interviewer. TIMSS, PIRLS and PISA are also timed tests, while PIAAC does not have any formal time limit. Furthermore, while until 2012 PISA considered missing answers and non reached items as wrong, PIAAC considered missing and non reached items as not providing information on the respondent's underlying ability.

Leaving these issues aside, in this Section we provide some further evidence around the major findings of the paper. In particular, we first take a macro-perspective and look at the cross-country heterogeneity in the evolution of gender gaps. Then, we adopt a micro-level focus, trying to better understand the mechanisms that drive the observed evolution of gender gaps. To do so, we focus on the 21 countries for which we have data from both PISA and PIAAC (the same that were used for the so-called "moving-cohort analysis"), disregarding the information on children at age of 9. This choice is motivated by the fact that the largest changes in gender gaps, both within and across countries, occur between age 15 and age 26/27. This choice also allows us to considerably enlarge the set of countries for which we have comparable data.

5.1 Cross-country Heterogeneity

A large literature has used International Large Scale Assessments of schooled populations to examine the role of socio-cultural factors in explaining differences across countries in the size of the gender gaps (Bedard and Cho 2010; Breda, Jouini, and Napp 2018; Gevrek, Gevrek, and Neumeier 2020; Guiso, Monte, Sapienza, and Zingales 2008). An important difference between these studies and our analysis is that while the previous literature has normally looked at gender gaps at a given point in time, our interest is on how gender gaps evolve as a cohort of students moves from adolescence (age 15) to young adulthood (age 26/27). On the other hand, an advantage of previous studies is that they normally rely on a large number of countries, generally more than 50 and sometimes as many as 70 or 80, while we can only use data on 21 countries. The small sample size at the country level means that we only have enough power to identify very strong associations between country-level indicators and between-country differences in how gender gaps evolve over time.

In Figures 3 and 4 we plot associations between the changes in the the gender gap in literacy (Figure 3) and in numeracy (Figure 4) and indicators of societal-level gender equality. Typically, gender equality indicators measure to what degree women can enjoy economic, social and political participation that is similar to what men enjoy. We focus in particular on four indicators that have been used in previous research (Gevrek, Gevrek, and Neumeier 2020; Guiso, Monte, Sapienza, and Zingales 2008; Stoet and Geary 2018): (i) the Gender Equality Index from the World Value Survey, capturing values and beliefs about the role of women in society[7]; (ii) a measure of female representation in Parliament (proportion of seats held by women, World Bank 2012); (iii) the Global Gender Gap Index produced by the World Economic Forum (World Economic Forum 2018); and (iv) the sub-index of the Global Gender Gap Index capturing Economic Participation and Opportunities (World Economic Forum 2018). All indices range, by construction, between 0 and 1, with higher values indicating more gender equality and better outcomes for women. One would expect that in countries where men and women have more equal access to opportunities for further education and training, where they have more equal access to the labour market, more equal returns to their work and where they can play an active role in the political and decision making sphere, women will enjoy more opportunities for skill development. As a consequence, one would expect that in more gender-equal countries the gender gaps in literacy and numeracy evolve in a way that is more favourable to women.

Figures 3 and 4, however, tell a different story.

[INSERT FIGURE 3 ABOUT HERE]

[INSERT FIGURE 4 ABOUT HERE]

Overall, the association between broad indicators of gender equality and how gender gaps in literacy and numeracy evolve between age 15/16 and age 26/27 is rather weak. The strongest association, for gaps in both literacy and numeracy, is with the index of economic participation. However, the correlation coefficients are always negative, indicating that in more gender-equal countries, gender gaps in numeracy increase more (so that men increase their initial advantage), and gender gaps in literacy decrease more (so that women lose their initial advantage). This result, which we don't want to overly emphasise due to the limitations of our data (i.e. a small number of countries)

[7]Unfortunately, the Gender Equality Index from the World Value Survey is only available for 16 out of the 21 countries we use throughout the paper

and methodology (as the analysis is entirely descriptive in nature and only presents correlation), could be seen as a version of the "Gender-Equality Paradox" recently popularised by Stoet and Geary (2018). In our setting, the "paradox" would hold not only for numeracy (participation in STEM in the original article by Stoet and Geary 2018), but also for literacy.

5.2 Potential mechanisms at the individual level

In this last sub-section we try to reflect on plausible mechanisms that can lie behind the observed evolution of gender gaps in literacy and numeracy.

Gender gaps in primary schools and in teenage years can probably be best linked to social gender norms and to differences in psychological traits, in particular related to different timing of cognitive and emotional development. This also rationalises the higher variability of scores displayed by boys relative to girls. In the transition to adulthood, increasing gender gaps in numeracy are consistent with a greater specialization of men in fields of studies and/or occupations that make more intensive use of numeracy skills. It is indeed well-known that women are under-represented in STEM education in most OECD countries (DiPrete and Buchmann 2013). On the other hand, the narrowing gender gap in literacy is more puzzling, but it might be due to the fact that literacy is a more transversal skill that everybody is called to master in order to succeed in education and in the labour market, irrespective of the chosen field. Aucejo and James (2016), for instance, find that verbal skills are more important in explaining university enrolment, and that the comparative advantage of girls in this domain helps explain the gender gap in college and STEM enrolment.

We provide more evidence on these issues by exploiting the information collected as part of the PIAAC Background Questionnaire on respondents' education and occupation (which we use to identify individuals pursuing a STEM career), as well as on the frequency with which they engage with tasks requiring to apply reading, writing and numeracy skills.

Participants were asked questions about the frequency of their engagement with reading, writing and numeracy tasks in two contexts: at work and in their everyday life. Example of the items used to capture this engagement (which we label skills use) are the frequency at which respondents read different types of documents (letters, directions, instructions, books, e-mails), write different types of documents (letters, memos, e-mails), or engage in tasks requiring to use numeracy skills, such as calculating prices or costs, using fractions or decimal numbers, using calculators, preparing graphs or tables. The same set of items were asked for both contexts (work and everyday life). For our analysis we use the indices available in the PIAAC database, which are scaled to have an overall mean of 2 and a standard deviation of 1. A detailed analysis of these indices is contained in Quintini (2014) and OECD (2013a). These and similar questions have also been previously used to construct measures of tasks performed on the job (Agasisti, Johnes, and Paccagnella forthcoming; De La Rica, Gortazar, and Lewandowski 2020; Nedelkoska and Quintini 2018).

In order to identify respondents in STEM-related careers we mainly use the information available in PIAAC on respondents' occupation. We identify as STEM occupations those coded by ISCO-08 as Science and Engineering Professionals and Associate Professionals (codes 21 and 31), and Information and Communication Technology Professionals and Technicians (codes 25 and 35). These are the same criteria adopted by the European Commission (European Commission 2015). Moreover, as a non-negligible share of respondents in the age cohort we are interested in is still in education (and might at the same time be employed in a non-STEM job), we also classify as STEM those that are currently in education and pursuing a STEM-related degree. To identify STEM degrees we rely on the ISCED classification, selecting degrees belonging to the fields of Sciences, mathematics and computing (code 5) or Engineering, manufacturing and construction (code 6) (European Commission 2015; Yao 2019).

Not surprisingly, women are severely under-represented in STEM occupations and fields of study. Across the 21 countries we look at, only 12 percent of women are pursuing STEM-related careers, while 40 percent of men are doing so. Variation across countries is fairly limited; the share of women in STEM is lowest in Japan (6 percent) and highest in Korea (24 percent), but for most countries it lies between 10 and 15 percent.

Of more interest are differences in the Skills Use indices, presented in Table 8. The first column shows that women are less likely to engage in tasks requiring the use of numeracy skills, be it at home or in the workplace (or, to be more precise, they engage in such tasks less frequently than men). No significant gender differences are found in the use of reading and writing skills. The second column shows that the same results hold when comparing people in STEM and people in non-STEM careers. Unsurprisingly, those working in STEM make more frequent use of their numeracy skills (not only at work but also in their everyday life). Perhaps more surprisingly, there are no differences in terms of the use of reading and writing skills: people in STEM careers use these skills as much as people in non-STEM careers. The corresponding results for each country in the sample are presented in Tables A-3 and A-4.

[INSERT TABLE 8 ABOUT HERE]

But to what extent do these findings about career choice and skills use are related to the observed gender gaps in literacy and numeracy? To answer these questions, we extend our baseline regression model (Equation 1) adding, in turn: (i) a dummy for STEM-related careers; (ii) a set of dummies for occupations at the 3-digit ISCO level; (iii) the six indices of skills use at work and in everyday life (labelled "at home" in Tables); and (iv) the STEM dummy and the indices of skills use at home and at work. The results of this exercise on the pooled dataset are presented in Table 9, while country-specific results are in Tables A-5 and A-6.

[INSERT TABLE 9 ABOUT HERE]

When controlling for a STEM-career dummy, the gender gap in numeracy is estimated to be only half as large as in the baseline specification. In literacy, the coefficient changes sign and increases in absolute terms, but remains much smaller than the coefficient for numeracy (half the size) and not statistically significant. Very little happens when controlling for occupation dummies or for the skills use indices. This suggests that, although skills use and skills proficiency are positively related, there is probably much more going on in STEM careers than what the skills use indices (which are after all based on fairly simple and common tasks) capture.

6 Conclusions

The main take-away from this paper is that gender gaps in information-processing skills evolve differently according to the domain tested: in literacy the advantage that girls have at early age peaks during adolescence, but then quickly disappears by early adulthood; in numeracy, the advantage of males increases steadily in an almost linear way. This pattern is observed in a large number of countries that participated in various international skills assessments and is, in our view, much more important and informative than the observed cross-country differences in the size of gender gaps at any given point in time. These results have implications in terms of education policy and practices. They point to the importance of raising the mathematics competency for girls, as gaps that open up during school years are very hard to close later in life, and also suggest further reflections on the optimal timing and the optimal allocation of instruction time in reading and numeracy, in light of the evidence about the ability of men to catch up with women in reading proficiency after completion of secondary education. Although the scores from the different assessment are not vertically linked (meaning that we can only look at the evolution of gaps, and not of levels of proficiencies for the two genders), the evidence from the Canadian longitudinal study does suggest that the catching up is due to larger growth in proficiency for men.

More efforts should be devoted in future research to better understand the underlying causes of these patterns. Here, we have only try to provide some suggestive, but necessarily tentative, interpretations. Gender gaps in primary schools and in teenage years can probably be best linked to social gender norms and to differences in psychological traits, in particular related to different timing of cognitive and emotional development (Pyne 2020). This also rationalises the higher variability of scores displayed by boys relative to girls. The increase in the numeracy gender gap from age 15 to age 26/27 is plausibly related to choices concerning post-compulsory education. We are indeed able to show that controlling for STEM-related careers reduces by half the size of the estimated gap. As far as literacy is concerned, there is evidence that the practice of reading and writing skills does not differ according to whether one pursues or not a career in STEM. In other words, literacy skills are transversal skills that are required and practised in a much broader range of educational pathways and occupations, which could explain why men have a strong incentive to develop them in order to have success in the labour market and are then finally able to catch up with women. Aucejo and James (2016), for instance, find that verbal skills are more important in explaining university enrolment, and that the comparative advantage of girls in this domain helps explain the gender gap in college and STEM enrolment.

Differences in target population, differences in the content of the assessments, differences in response formats and mode of delivery do not appear to explain the very different results on gender gaps estimated in the two PISA and PIAAC surveys (Borgonovi, Pokropek, Keslair, Gauly, and Paccagnella 2017). However, it remains unknown the extent to which these may influence comparisons of gender gaps between TIMSS and PIRLS on the one hand, and PISA and PIAAC on the other. Further work to evaluate the impact of response rates for the estimation of gender gaps should be undertaken given the large difference in response rates between the studies (TIMSS, PIRLS and PISA have, for instance, much higher response rates than PIAAC).

References

Agasisti, T., G. Johnes, and M. Paccagnella (forthcoming): "Tasks, Occupations and Wages in OECD Countries," *International Labour Review*.

Altonji, J., P. Arcidiacono, and A. Maurel (2016): "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, vol. 5 of *Handbook of the Economics of Education*, chap. 7, pp. 305 – 396. Elsevier.

Arntz, M., T. Gregory, and U. Zierahn (2016): "The Risk of Automation for Jobs in OECD Countries. A Comparative Analysis," OECD Social, Employment and Migration Working Papers 189, OECD Publishing, Paris.

Aucejo, E., and J. James (2016): "The Path to College Education: Are Verbal Skills More Important than Math Skills?," Working Papers 1602, California Polytechnic State University, Department of Economics.

Avvisati, F., and F. Keslair (2014): "REPEST: Stata module to run estimations with weighted replicate samples and plausible values," Statistical Software Components, Boston College Department of Economics.

Bedard, K., and I. Cho (2010): "Early gender test score gaps across OECD countries," *Economics of Education Review*, 29(3), 348–363.

Bond, T. N., and K. Lang (2013): "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results," *The Review of Economics and Statistics*, 95(5), 1468–1479.

Borgonovi, F., A. Pokropek, F. Keslair, B. Gauly, and M. Paccagnella (2017): "Youth in Transition: How Do Some of The Cohorts Participating in PISA Fare in PIAAC?," OECD Education Working Papers 155, OECD Publishing, Paris.

Breda, T., E. Jouini, and C. Napp (2018): "Societal inequalities amplify gender gaps in math," *Science*, 359(6381), 1219–1220.

Brynjolfsson, E., and A. McAfee (2014): *The second machine age : work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company, New York.

Card, D., and A. A. Payne (2021): “High School Coices And The Gender Gap In STEM,” *Economic Inquiry*, 59(1), 9–28.

Carnevale, A., B. Chea, and J. Strohl (2013): “College Majors, Unemployment and Earnings: Not all college degrees are created equal,” Discussion paper, Georgetown University Center on Education and the Workforce.

Castex, G., and E. Kogan Dechter (2014): “The Changing Roles of Education and Ability in Wage Determination,” *Journal of Labor Economics*, 32(4), 685–710.

Contini, D., M. L. D. Tommaso, and S. Mendolia (2017): “The gender gap in mathematics achievement: Evidence from Italian data,” *Economics of Education Review*, 58(C), 32–42.

De La Rica, S., L. Gortazar, and P. Lewandowski (2020): “Job Tasks and Wages in Developed Countries: Evidence from PIAAC,” *Labour Economics*, 65, 101845.

DiPrete, T., and C. Buchmann (2013): *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools*. Russell Sage Foundation Press, New York.

European Commission (2015): *Does the EU Need More STEM Graduates?* Publications Office of the European Union, Luxembourg.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): “Unconditional Quantile Regressions,” *Econometrica*, 77(3), 953–973.

Flabbi, L. (2012): “Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD countries,” Discussion paper, World Development Report Background Papers.

Fortin, N. M., P. Oreopoulos, and S. Phipps (2015): "Leaving Boys Behind: Gender Disparities in High Academic Achievement," *Journal of Human Resources*, 50(3), 549–579.

Gal, I., and D. Tout (2014): "Comparison of PIAAC and PISA Frameworks for Numeracy and Mathematical Literacy," OECD Education Working Papers 102, OECD Publishing, Paris.

Gevrek, Z. E., D. Gevrek, and C. Neumeier (2020): "Explaining the gender gaps in mathematics achievement and attitudes: The role of societal gender equality," *Economics of Education Review*, 76(C).

Goldin, C., and L. Katz (2008): *The race between education and technology.* Harvard University Press, Cambridge.

Goldin, C., L. F. Katz, and I. Kuziemko (2006): "The Homecoming of American College Women: The Reversal of the College Gender Gap," *The Journal of Economic Perspectives*, 20(4), 133–156.

Goulas, S., S. Griselda, and R. Megalokonomou (2020): "Comparative Advantage and Gender Gap in STEM," IZA Discussion Papers 13313, Institute of Labor Economics (IZA).

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008): "Culture, Gender, and Math," *Science*, 320(5880), 1164–1165.

Jacob, B., and J. Rothstein (2016): "The Measurement of Student Ability in Modern Assessment Systems," *Journal of Economic Perspectives*, 30(3), 85–108.

Kirkebøen, L., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *The Quarterly Journal of Economics*, 131(3), 1057–1111.

Klieme, E. (2016): "TIMSS 2015 and PISA 2015. How are they related on the country level?," Dipf working paper, German Institute for International Educational Research (DIPF).

Lafontaine, D., and C. Monseur (2009): "Gender Gap in Comparative Studies of Reading Comprehension: To What Extent Do the Test Characteristics Make a Difference?," *European Educational Research Journal*, 8(1), 69–79.

Lavy, V., and E. Sand (2018): "On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases," *Journal of Public Economics*, 167, 263 – 279.

Lei, Z., and S. Lundberg (2020): "Vulnerable Boys: Short-Term and Long-Term Gender Differences in the Impact of Adolescent Disadvantage," HCEO Working Paper Series 2020-08, University of Chicago.

Lietz, P. (2006): "A Meta-Analysis Of Gender Differences In Reading Achievement At The Secondary School Level," *Studies in Educational Evaluation*, 32(4), 317 – 344.

Machin, S., and T. Pekkarinen (2008): "Global Sex Differences in Test Score Variability," *Science*, 322(5906), 1331–1332.

Mullis, I. V., M. O. Martin, P. Foy, and A. Arora (2012): *TIMSS 2011 International Results in Mathematics.* TIMSS & PIRLS International Study Center, Boston College, Chestnut Hills, MA.

Mullis, I. V., M. O. Martin, P. Foy, and K. T. Drucker (2012): *PIRLS 2011 International Results in Reading.* TIMSS & PIRLS International Study Center, Boston College, Chestnut Hills, MA.

Mullis, I. V., M. O. Martin, E. Gonzalez, and A. Kennedy (2003): *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries.* International Study Center, Lynch School of Education, Boston College, Chestnut Hills, MA.

Nedelkoska, L., and G. Quintini (2018): "Automation, skills use and training," OECD Social, Employment and Migration Working Papers 202, OECD Publishing, Paris.

OECD (1999): *Measuring Student Knowledge and Skills: A New Framework for Assessment.* OECD Publishing.

——— (2009): *PISA 2006 Technical Report*. OECD Publishing, Paris.

——— (2011): *PISA 2009 Technical Report*. OECD Publishing, Paris.

——— (2012): *Learning beyond Fifteen*. OECD Publishing, Paris.

——— (2013a): "How Skills Are Used in the Workplace," in *OECD Skills Outlook 2013 : First Results from the Survey of Adult Skills*, chap. 4. OECD Publishing, Paris.

——— (2013b): *Technical report to the Survey of Adult Skills (PIAAC)*. OECD Publishing, Paris.

——— (2015): *The ABC of gender equality in education: Aptitude, behavior, confidence*. OECD Publishing, Paris.

Penney, J. (2017): "Test Score Measurement and the Black-White Test Score Gap," *The Review of Economics and Statistics*, 99(4), 652–656.

PIAAC Literacy Expert Group (2009): "PIAAC Literacy: A Conceptual Framework," OECD Education Working Papers 34, OECD Publishing, Paris.

PIAAC Numeracy Expert Group (2009): "PIAAC Numeracy: A Conceptual Framework," OECD Education Working Papers 35, OECD Publishing, Paris.

Pyne, J. (2020): "Gender Test Score Gaps Under Equal Behavioral Engagement," *Educational Researcher*, 49(6), 459–464.

Quintini, G. (2014): "Skills at Work: How Skills and their Use Matter in the Labour Market," OECD Social, Employment and Migration Working Papers 158, OECD Publishing, Paris.

Reardon, S. F., D. Kalogrides, E. M. Fahle, A. Podolsky, and R. C. Zarate (2018): "The Relationship Between Test Item Format and Gender Achievement Gaps on Math and ELA Tests in Fourth and Eighth Grades," *Educational Researcher*, 47(5), 284–294.

Riordan, C. (1999): "The Silent Gender Gap," *Education Week*, 19(12), 46–49.

Roe, A., and K. Taube (2003): "Reading Achievement and Gender Differences," in *Northern Lights on PISA*, ed. by S. Lie, P. Linnakylä, and A. Roe, pp. 21–37. Department of Teacher Education and School Development, University of Oslo.

Routisky, A., and R. Turner (2003): "Item format types and their influence on crossnational comparisons of sudent performance," Paper presented at the Conference of th American Educational Research Association (AERA), Chicago.

Stoet, G., and D. C. Geary (2018): "The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education," *Psychological Science*, 29(4), 581–593, PMID: 29442575.

Terrier, C. (2020): "Boys lag behind: How teachers' gender biases affect student achievement," *Economics of Education Review*, 77, 101981.

World Bank (2012): "Proportion of seats held by women in national parliaments (%)," .

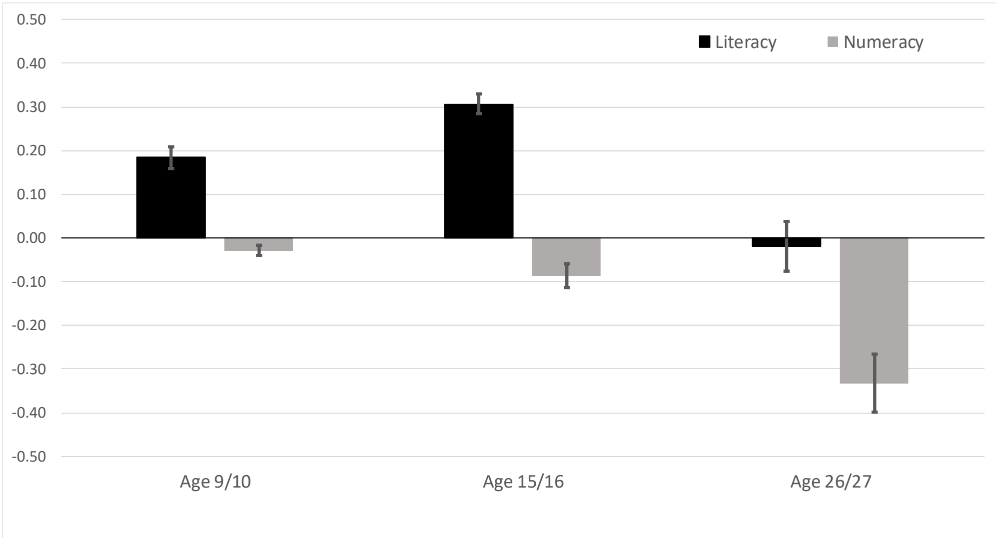
World Economic Forum (2018): "The Global Gender Gap Report," .

Wu, M. (2009): "A comparison of PISA and TIMSS 2003 achievement results in mathematics," *PROSPECTS*, 39(1), 33.

Yao, K. (2019): "Heterogeneous skill distribution and college major: evidence from PIAAC," *Journal of Applied Economics*, 22(1), 504–526.

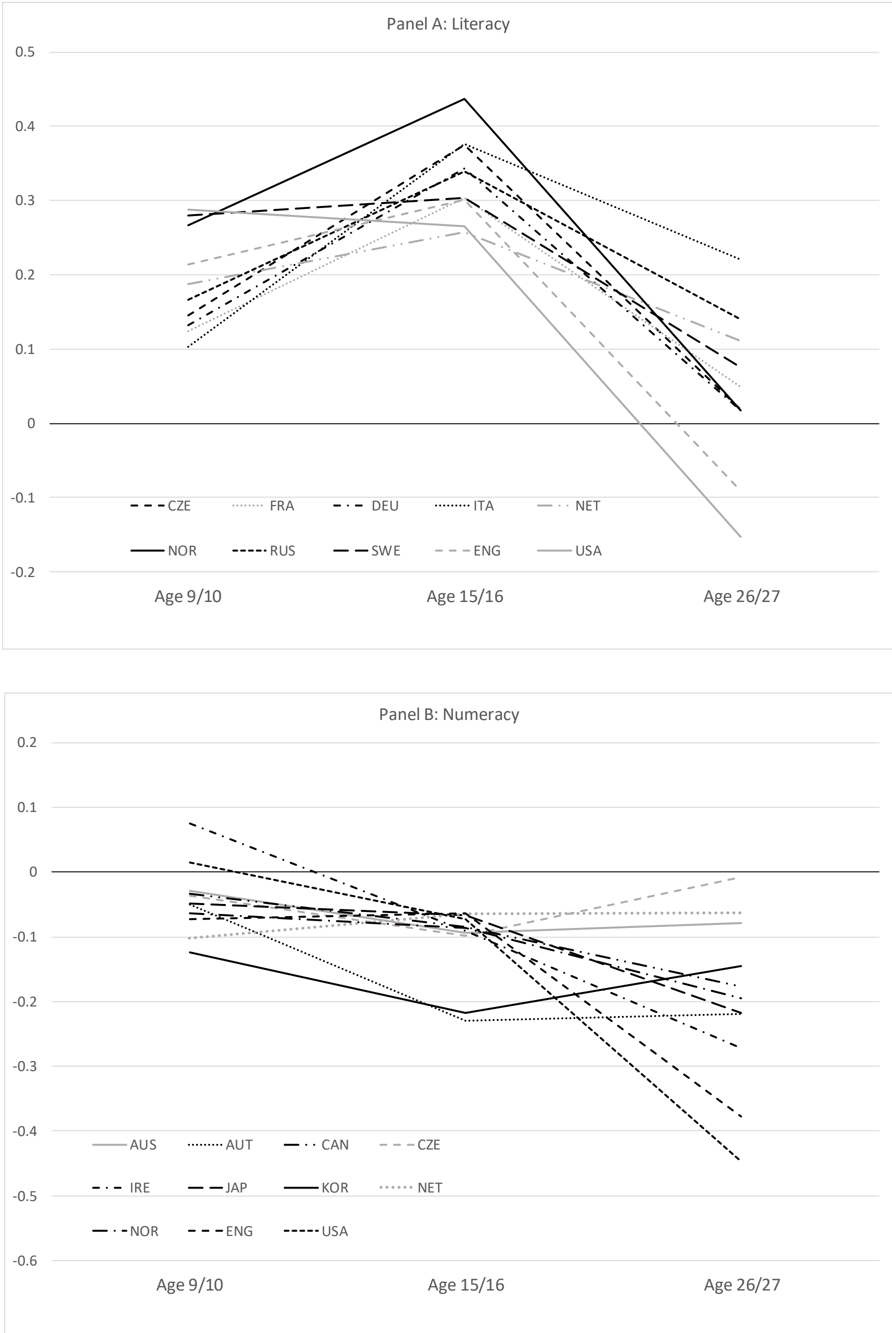
Figures

Figure 1
The Evolution of Gender Gaps in Literacy and Numeracy



Notes: The lines on top of the bars represent 95% level confidence intervals. *Source*: TIMSS (1995), PIRLS (2001), PISA (2000), PIAAC (2011/12).

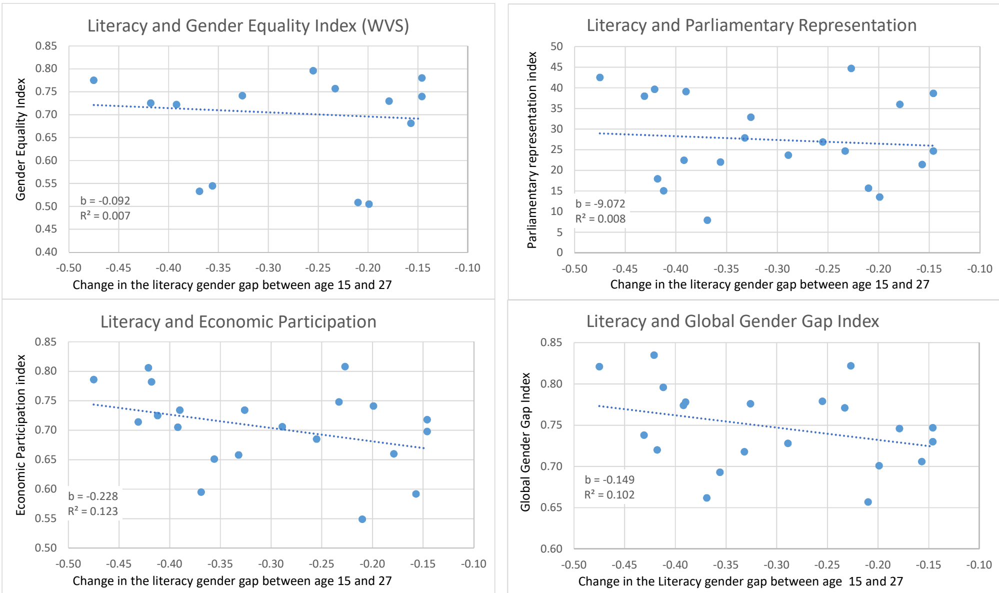
Figure 2
The Evolution of Gender Gaps: Country-Specific Results



Source: TIMSS (1995), PIRLS (2001), PISA (2000), PIAAC (2011/12).

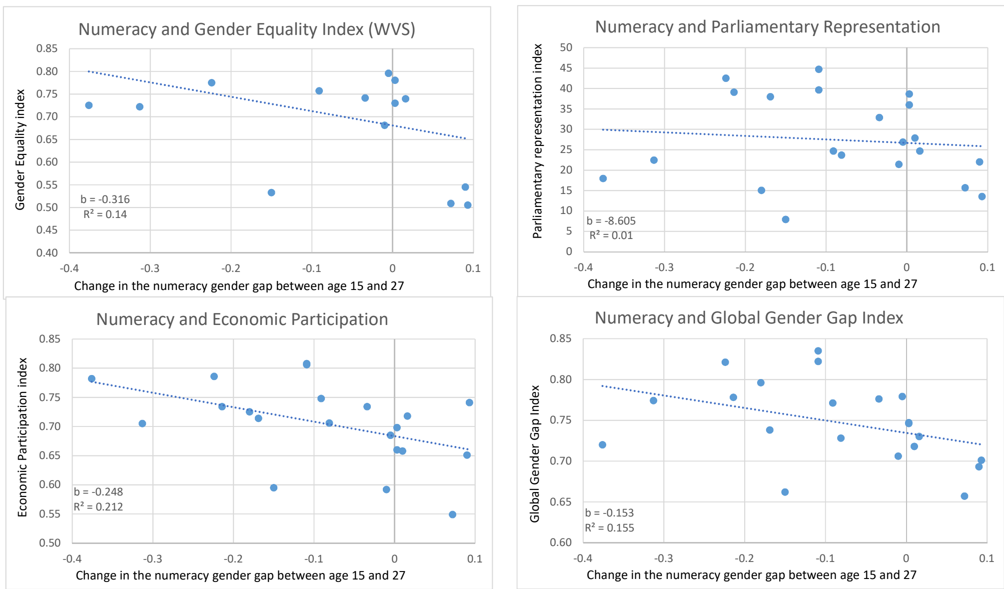
Figure 3

Change in Literacy Gender Gaps and Societal Gender Equality



Source: PISA (2000), PIAAC (2011/12), World Value Survey (Wave 6), World Economic Forum (2018), World Bank (2012).

Figure 4
Change in Numeracy Gender Gaps and Societal Gender Equality



Source: PISA (2000), PIAAC (2011/12), World Value Survey (Wave 6), World Economic Forum (2018), World Bank (2012).

Tables

Table 1

Descriptive statistics - Numeracy assessments

	TIMSS 1995				PISA 2000				PIAAC 2011/12			
	Avg. score Women	Avg. score Men	Z-stat	Sample Size	Avg. score Women	Avg. score Men	Z-stat	Sample Size	Avg. score Women	Avg. score Men	Z-stat	Sample Size
Australia	520	522	-0.79	8,834	529	539	-1.77	2,514	275	279	-0.67	381
Austria	538	543	-1.23	3,204	508	533	-4.73	2,455	286	298	-2.05	309
Canada	527	529	-0.74	8,102	530	539	-5.82	15,451	278	287	-1.94	1,378
Czech Rep.	545	549	-1.10	4,628	492	503	-2.32	3,034	289	290	-0.06	522
Ireland	529	521	1.73	4,237	497	507	-2.16	2,046	258	273	-2.28	318
Japan	571	576	-2.65	7,611	553	561	-1.09	2,918	291	303	-2.36	304
Korea	582	595	-5.40	4,835	532	559	-3.39	2,769	281	289	-1.95	450
Netherlands	547	556	-2.42	3,345	564	572	-1.19	1,301	295	299	-0.58	285
Norway	498	504	-1.53	2,155	498	506	-2.70	2,160	286	297	-1.67	287
England	500	507	-1.73	3,939	527	532	-1.46	2,165	263	284	-2.71	302
United States	532	531	0.65	7,584	491	500	-1.57	1,974	252	277	-3.60	368
Pooled	545	547	-2.24	58,474	516	528	-3.19	38,787	265	285	-4.97	4,904
Int'l Mean	507				478				259			
Int'l SD	100				114				55			

Notes: The table reports average numeracy scores for the cohort born in 1983-1984 (1983-1986 for PIAAC). Survey weights are applied throughout. Z-stats test for differences in average scores of men vs. women. Data for Russia for age 26/27 do not include the Moscow Region. The International Mean and the International Standard Deviation refer to all countries that have participated in the various assessments. *Source*: TIMSS (1995), PISA (2000), PIAAC (2012).

Table 2

Descriptive statistics - Literacy assessments

	PIRLS 2001				PISA 2000				PIAAC 2011/12			
	Avg. score Women	Avg. score Men	Z-stat	Sample Size	Avg. score Women	Avg. score Men	Z-stat	Sample Size	Avg. score Women	Avg. score Men	Z-stat	Sample Size
Czech Rep.	545	534	4.07	2,886	515	473	8.41	3,034	285	284	0.17	522
France	533	522	3.38	3,337	524	490	7.92	2,511	288	285	0.58	425
Germany	554	543	4.02	6,118	509	471	5.75	2,530	289	289	0.15	341
Italy	547	539	3.04	3,341	508	466	5.55	2,709	273	262	1.68	234
Netherlands	563	528	6.02	3,812	555	527	4.24	1,301	307	302	1.04	285
Norway	514	492	5.61	3,154	530	479	9.86	2,160	298	297	0.16	287
Russia	537	524	5.76	3,727	476	438	10.39	3,515	282	275	1.29	374
Sweden	577	556	7.84	5,332	536	502	8.28	2,266	306	302	0.75	260
United States	562	545	3.61	3,014	521	492	5.60	1,974	279	286	-1.29	368
England	573	551	6.78	2,682	543	508	5.12	2,165	282	287	-0.69	302
Pooled	553	538	7.49	37,403	511	476	13.28	24,165	282	283	-0.33	3,398
Int'l Mean	536				478				268			
Int'l SD	77				110				50			

Notes: The table reports average literacy scores for the cohort born in 1983-1984 (1983-1986 for PIAAC). Survey weights are applied throughout. Z-stats test for differences in average scores of men vs women. Data for Russia for age 26/27 do not include the Moscow Region. The International Mean and the International Standard Deviation refer to all countries that have participated in the various assessments. *Source*: PIRLS (2001), PISA (2000), PIAAC (2012).

Table 3
The evolution of gender gaps in literacy

	Age 9/10	Age 15/16	Age 26/27
Czech Republic	0.146***	0.375***	0.019
	(0.036)	(0.044)	(0.113)
France	0.124***	0.304***	0.049
	(0.042)	(0.038)	(0.085)
Germany	0.133***	0.343***	0.017
	(0.035)	(0.059)	(0.115)
Italy	0.104***	0.377***	0.220*
	(0.034)	(0.068)	(0.131)
Netherlands	0.187***	0.257***	0.111
	(0.032)	(0.061)	(0.106)
Norway	0.267***	0.438***	0.017
	(0.050)	(0.044)	(0.105)
Russia	0.167***	0.339***	0.140
	(0.029)	(0.033)	(0.108)
Sweden	0.280***	0.303***	0.076
	(0.035)	(0.036)	(0.101)
England	0.287***	0.301***	-0.091
	(0.042)	(0.054)	(0.133)
United States	0.214***	0.265***	-0.153
	(0.061)	(0.052)	0.119
Pooled	0.185***	0.307***	-0.019
	(0.025)	(0.023)	(0.057)
Pooled 5 common countries	0.221***	0.276***	-0.130
	(0.049)	(0.041)	(0.095)

Notes: The table reports average gender gaps in literacy for the cohort born in 1983/84. Data for Russia for age 26/27 do not include the Region of Moscow. The 5 common countries are the Czech Republic, the Netherlands, Norway, England, and the United States, for which data on boht literacy and numeracy are available. The pooled specifications control for country fixed effects. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIRLS (2001), PISA (2000), PIAAC (2012).

Table 4
The evolution of gender gaps in numeracy

	Age 9/10	Age 15/16	Age 26/27
Australia	-0.029	-0.094*	-0.078
	(0.036)	(0.053)	(0.116)
Austria	-0.051	-0.229***	-0.219**
	(0.039)	(0.049)	(0.106)
Canada	-0.034	-0.085***	-0.176*
	(0.036)	(0.015)	(0.091)
Czech Republic	-0.036	-0.098**	-0.008
	(0.032)	(0.042)	(0.130)
Ireland	0.075*	-0.091**	-0.271**
	(0.043)	(0.042)	(0.119)
Japan	-0.048***	-0.067	-0.217**
	(0.018)	(0.061)	(0.092)
Korea	-0.124***	-0.218***	-0.146*
	(0.023)	(0.064)	(0.075)
Netherlands	-0.102***	-0.065	-0.062
	(0.036)	(0.054)	(0.108)
Norway	-0.064	-0.087***	-0.196*
	(0.041)	(0.032)	(0.117)
England	-0.073*	-0.064	-0.377***
	(0.042)	(0.044)	(0.139)
United States	0.015	-0.072	-0.448***
	(0.023)	(0.046)	(0.124)
Pooled	-0.029**	-0.087***	-0.332***
	(0.012)	(0.027)	(0.067)
Pooled 5 common countries	-0.003	-0.072**	-0.411***
	(0.019)	(0.036)	(0.099)

Notes: The table reports average gender gaps in numeracy for the cohort born in 1983/84. The 5 common countries are the Czech Republic, the Netherlands, Norway, England, and the United States, for which data on both literacy and numeracy are available. The pooled specifications control for country fixed effects. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: TIMSS (1995), PISA (2000), PIAAC (2012).

Table 5
Gender gaps across the skills distribution

	Literacy			Numeracy		
	Age 10	Age 15	Age 26	Age 10	Age 15	Age 26
Average gap	0.185***	0.307***	-0.019	-0.029**	-0.087***	-0.332***
	(0.025)	(0.023)	(0.057)	(0.012)	(0.027)	(0.067)
10th pct	0.274***	0.467***	0.123	0.035	0.028	-0.233
	(0.050)	(0.054)	(0.159)	(0.035)	(0.056)	(0.202)
90th pct	0.138***	0.166***	-0.073	-0.081***	-0.177***	-0.337**
	(0.043)	(0.042)	(0.105)	(0.027)	(0.038)	(0.146)
M/F ratio (10th pct)	1.562***	2.157***	1.229	1.120***	0.991***	0.608***
M/F ratio (90th pct)	0.740***	0.675***	1.146***	1.228***	1.564***	2.426***
Variance Ratio	1.101	1.241	1.128	1.087	1.170	1.024

Notes: Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. For the male/female ratio at the 10th and the 90th percentile, the null hypothesis is that the ratio is equal to 1. *Source*: TIMSS (1995), PIRLS (2001), PISA (2000), PIAAC (2012).

Table 6
Normalised Gaps - Robust to Monotonic Transformations of the Scale

	Literacy			Numeracy		
Variable	Age 10	Age 15	Age 26	Age 10	Age 15	Age 26
Avg. gap	0.185***	0.307***	-0.019	-0.029***	-0.087***	-0.332***
	(0.025)	(0.023)	(0.057)	(0.012)	(0.027)	(0.067)
Normalised (p50)	0.136***	0.264***	-0.038	-0.027*	-0.119***	-0.327***
	(0.025)	(0.025)	(0.077)	(0.016)	(0.036)	(0.089)
10th pct	0.274***	0.467***	0.123	0.035	0.028	-0.233
	(0.050)	(0.054)	(0.159)	(0.035)	(0.056)	(0.202)
Normalised (p10)	0.142***	0.272***	0.071	0.022	0.018	-0.144
	(0.024)	(0.027)	(0.090)	(0.021)	(0.036)	(0.121)
90th pct	0.138***	0.166***	-0.073	-0.081***	-0.177***	-0.337**
	(0.043)	(0.042)	(0.105)	(0.027)	(0.038)	(0.146)
Normalised (p90)	0.099***	0.175***	-0.058	-0.057***	-0.157***	-0.257**
	(0.030)	(0.056)	(0.083)	(0.019)	(0.031)	(0.099)

Notes: Normalised gaps are obtaining by estimating Unconditional Quantile Regressions (RIF-regressions, as in Firpo, Fortin, and Lemieux (2009)), and then dividing the estimated coefficients by the standard error of the regression, as suggested by Penney (2017). Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: TIMSS (1995), PIRLS(2001), PISA (2000), PIAAC (2012).

Table 7
Moving Cohort Approach

	PISA cohort			Age in PIAAC 2011/12		
	1984/5	1987/88	1990/91	20/21	23/24	26/27
Literacy	0.299***	0.293***	0.333***	0.001	-0.005	-0.013
	(0.020)	(0.013)	(0.013)	(0.038)	(0.045)	(0.041)
Numeracy	-0.071***	-0.083***	-0.102***	-0.144***	-0.194***	-0.209***
	(0.018)	(0.013)	(0.012)	(0.040)	(0.042)	(0.043)

Notes: Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PISA (2000, 2003, 2006), PIAAC (2012).

Table 8
Differences in Skills Use

	Differences by gender	Differences by STEM
Numeracy at home	-0.130**	0.307***
	(0.053)	(0.055)
Numeracy at work	-0.146***	0.307***
	(0.056)	(0.044)
Reading at home	0.004	0.130
	(0.005)	(0.108)
Reading at work	0.069	0.026
	(0.055)	(0.057)
Writing at home	0.035	0.032
	(0.046)	(0.048)
Writing at work	0.047	0.039
	(0.047)	(0.049)

Notes: The indices of skills use were constructed using a Generalised Partial Credit Model (OECD 2013b) and were scaled to have a mean of 2 and a standard deviation of 1. Standard errors of the differences in the skills use indices are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).

Table 9
Gender gaps in literacy and numeracy at age 26/27

	Literacy	Numeracy
Baseline	-0.013	-0.209***
	(0.041)	(0.043)
Controlling for:		
STEM	0.062	-0.108**
	(0.045)	(0.044)
ISCO Occupation	-0.073	-0.239***
	(0.050)	(0.051)
Skills Use Indices	0.019	-0.204***
	(0.050)	(0.048)
STEM+Skills Use	0.065	-0.129**
	(0.057)	(0.051)

Notes: Regressions are run on the 21 countries listed in A-1 and A-2. All specifications control for country fixed effects. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).

Appendix

Table A-1

Moving cohort approach: Literacy

	PISA (Age 15/16)			PIAAC		
	1984/85 Cohort	1987/88 Cohort	1990/91 Cohort	Age 20/21	Age 23/24	Age 26/27
Australia	0.309***	0.379***	0.346***	-0.072	0.076	0.163
	(0.065)	(0.030)	(0.031)	(0.146)	(0.158)	(0.123)
Austria	0.270***	0.428***	0.411***	-0.100	-0.098	-0.062
	(0.055)	(0.052)	(0.053)	(0.089)	(0.089)	(0.097)
Belgium	0.360***	0.258***	0.332***	-0.025	0.005	-0.071
	(0.071)	(0.054)	(0.047)	(0.092)	(0.104)	(0.089)
Canada	0.304***	0.290***	0.291***	0.017	-0.120	0.071
	(0.019)	(0.021)	(0.021)	(0.081)	(0.091)	(0.094)
Czech Republic	0.375***	0.289***	0.426***	-0.125	-0.139	0.019
	(0.044)	(0.046)	(0.055)	(0.117)	(0.102)	(0.113)
Denmark	0.249***	0.230***	0.270***	0.043	0.093	-0.141
	(0.042)	(0.028)	(0.029)	(0.092)	(0.113)	(0.125)
Finland	0.519***	0.407***	0.464***	0.088	-0.107	0.044
	(0.035)	(0.026)	(0.025)	(0.103)	(0.090)	(0.095)
France	0.304***	0.349***	0.321***	-0.037	0.064	0.049
	(0.038)	(0.042)	(0.040)	(0.084)	(0.079)	(0.085)
Germany	0.343***	0.413***	0.383***	-0.012	-0.026	0.017
	(0.059)	(0.045)	(0.037)	(0.093)	(0.110)	(0.115)
Ireland	0.278***	0.275***	0.298***	-0.154	-0.082	-0.134
	(0.052)	(0.043)	(0.045)	(0.124)	(0.133)	(0.110)
Italy	0.377***	0.355***	0.374***	0.018	0.269**	0.220
	(0.068)	(0.056)	(0.036)	(0.137)	(0.135)	(0.131)
Japan	0.290***	0.206***	0.280***	-0.137	0.003	-0.079
	(0.065)	(0.050)	(0.070)	(0.084)	(0.094)	(0.089)
Korea	0.143**	0.196***	0.318***	-0.056	-0.104	-0.067
	(0.060)	(0.052)	(0.054)	(0.076)	(0.093)	(0.075)
Netherlands	0.257***	0.202***	0.208***	0.099	-0.013	0.111
	(0.061)	(0.036)	(0.030)	(0.089)	(0.101)	(0.106)
Norway	0.438***	0.449***	0.416***	-0.051	-0.017	0.017
	(0.044)	(0.034)	(0.030)	(0.090)	(0.106)	(0.105)
Poland	0.390***	0.368***	0.364***	0.110***	0.127***	0.101
	(0.075)	(0.034)	(0.027)	(0.038)	(0.047)	(0.082)
Russia	0.339***	0.263***	0.357***	0.069	0.138	0.140
	(0.033)	(0.036)	(0.030)	(0.093)	(0.107)	(0.108)
Spain	0.224***	0.365***	0.322***	0.112	0.079	0.045
	(0.037)	(0.036)	(0.020)	(0.094)	(0.102)	(0.098)
Sweden	0.303***	0.350***	0.376***	0.103	0.180*	0.076
	(0.036)	(0.029)	(0.030)	(0.080)	(0.103)	(0.101)
England	0.301***	0.251***	0.260***	-0.149	0.038	-0.091
	(0.054)	(0.052)	(0.037)	(0.131)	(0.154)	(0.133)
United States	0.265***	0.290***	-	0.022	-0.123	-0.153
	(0.052)	(0.031)		(0.133)	(0.103)	(0.119)
Pooled	0.299***	0.293***	0.333***	0.001	-0.005	-0.013
	(0.020)	(0.013)	(0.013)	(0.038)	(0.045)	0.041

Notes: The pooled specification controls for country fixed effects. Data for Belgium only include Flanders. Because of data quality issues, the results of the Reading assessment for the United States have not been published (see pag. 281 of OECD 2009). Data for Russia from PIAAC do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PISA (2000, 2003, 2006) and PIAAC (2012).

Table A-2

Moving cohort approach: Numeracy

	PISA (Age 15/16)			PIAAC		
	1984/85 Cohort	1987/88 Cohort	1990/91 Cohort	Age 20/21	Age 23/24	Age 26/27
Australia	-0.094*	-0.039	-0.128***	-0.310**	-0.045	-0.078
	(0.053)	(0.029)	(0.030)	(0.156)	(0.153)	(0.116)
Austria	-0.229***	-0.075*	-0.214***	-0.234***	-0.283***	-0.219**
	(0.049)	(0.041)	(0.044)	(0.082)	(0.093)	(0.106)
Belgium	-0.050	-0.136**	-0.092**	-0.207**	-0.146	-0.219**
	(0.059)	(0.053)	(0.047)	(0.085)	(0.097)	(0.096)
Canada	-0.085***	-0.107***	-0.130***	-0.179**	-0.330***	-0.176*
	(0.015)	(0.019)	(0.019)	(0.091)	(0.095)	(0.091)
Czech Republic	-0.098**	-0.133***	-0.094*	-0.154*	-0.105	-0.008
	(0.042)	(0.046)	(0.053)	(0.083)	(0.092)	(0.130)
Denmark	-0.118***	-0.160***	-0.101***	-0.112	-0.073	-0.332***
	(0.032)	(0.029)	(0.027)	(0.095)	(0.105)	(0.124)
Finland	-0.005	-0.062**	-0.110***	-0.184*	-0.340***	-0.229**
	(0.028)	(0.024)	(0.024)	(0.099)	(0.095)	(0.091)
France	-0.114***	-0.080**	-0.057	-0.259***	-0.208**	-0.119
	(0.034)	(0.038)	(0.035)	(0.089)	(0.081)	(0.089)
Germany	-0.099**	-0.065	-0.178***	-0.197**	-0.225*	-0.133
	(0.045)	(0.041)	(0.037)	(0.089)	(0.117)	(0.112)
Ireland	-0.091**	-0.124***	-0.120***	-0.352***	-0.203	-0.271**
	(0.042)	(0.038)	(0.039)	(0.131)	(0.149)	(0.119)
Italy	-0.073	-0.163***	-0.164***	-0.130	0.128	-0.083
	(0.059)	(0.053)	(0.032)	(0.130)	(0.130)	(0.147)
Japan	-0.067	-0.078	-0.185***	-0.179*	-0.124	-0.217**
	(0.061)	(0.052)	(0.068)	(0.096)	(0.098)	(0.092)
Korea	-0.218***	-0.210***	-0.090	-0.103	-0.185**	-0.146*
	(0.064)	(0.060)	(0.060)	(0.083)	(0.087)	(0.075)
Netherlands	-0.065	-0.036	-0.135***	-0.117	-0.156	-0.062
	(0.054)	(0.039)	(0.026)	(0.086)	(0.102)	(0.108)
Norway	-0.087***	-0.063**	-0.059**	-0.187*	-0.177	-0.196*
	(0.032)	(0.029)	(0.030)	(0.100)	(0.112)	(0.117)
Poland	-0.040	-0.049*	-0.087***	-0.038	-0.044	-0.121*
	(0.071)	(0.028)	(0.024)	(0.034)	(0.044)	(0.067)
Russia	0.016	-0.097***	-0.033	0.020	0.006	0.109
	(0.040)	(0.039)	(0.036)	(0.096)	(0.107)	(0.110)
Spain	-0.143***	-0.080***	-0.079***	0.025	-0.044	-0.140
	(0.037)	(0.026)	(0.022)	(0.083)	(0.085)	(0.089)
Sweden	-0.057*	-0.052*	-0.044	-0.061	-0.040	-0.166
	(0.034)	(0.028)	(0.029)	(0.084)	(0.097)	(0.103)
England	-0.064 -	0.071	-0.172***	-0.235*	-0.176	-0.377***
	(0.044)	(0.052)	(0.032)	(0.124)	(0.139)	(0.139)
United States	-0.072	-0.055**	-0.083***	-0.209	-0.381***	-0.448***
	(0.046)	(0.026)	(0.023)	(0.142)	(0.098)	(0.124)
Pooled	-0.071***	-0.083***	-0.102***	-0.144***	-0.194***	-0.209***
	(0.018)	(0.013)	(0.012)	(0.040)	(0.042)	(0.043)

Notes: The pooled specification controls for country fixed effects. Data for Belgium only include Flanders. Data for Russia from PIAAC do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PISA (2000, 2003, 2006) and PIAAC (2012).

Table A-3
Differences in Skills Use, by gender

	At Home			At Work		
	Numeracy	Reading	Writing	Numeracy	Reading	Writing
Australia	-0.214*	0.079	0.009	-0.135	0.118	0.282
	(0.122)	(0.131)	(0.106)	(0.134)	(0.130)	(0.135)
Austria	-0.261**	-0.179**	0.081	-0.253**	-0.130	-0.043
	(0.106)	(0.075)	(0.078)	(0.114)	(0.099)	(0.123)
Belgium	-0.280**	-0.167**	-0.087	-0.288**	0.071	0.136
	(0.112)	(0.079)	(0.072)	(0.150)	(0.094)	(0.094)
Canada	-0.384***	-0.143**	0.078	-0.329***	-0.010	0.085
	(0.077)	(0.063)	(0.093)	(0.121)	(0.089)	(0.106)
Czech Rep.	-0.028	0.080	0.184**	0.382**	0.091	0.327
	(0.114)	(0.078)	(0.087)	(0.164)	(0.114)	(0.174)
Denmark	-0.328**	-0.036	0.061	-0.412***	-0.072	0.206
	(0.149)	(0.116)	(0.127)	(0.155)	(0.109)	(0.112)
England	-0.233*	-0.032	0.131	-0.265**	0.031	-0.109
	(0.133)	(0.082)	(0.106)	(0.150)	(0.164)	(0.152)
Finland	-0.282***	-0.159**	0.010	-0.412***	-0.076	-0.039
	(0.062)	(0.068)	(0.074)	(0.101)	(0.093)	(0.104)
France	-0.186**	0.051	-0.015	-0.300**	-0.029	-0.072
	(0.074)	(0.066)	(0.076)	(0.118)	(0.076)	(0.117)
Germany	-0.467***	-0.166*	-0.001	-0.178	0.033	-0.118
	(0.087)	(0.100)	(0.085)	(0.126)	(0.117)	(0.116)
Ireland	-0.131	0.070	-0.093	0.128	0.111	0.016
	(0.144)	(0.098)	(0.099)	(0.154)	(0.131)	(0.162)
Italy	-0.022	0.329*	0.273*	0.276	0.367*	0.164
	(0.196)	(0.173)	(0.164)	(0.294)	(0.200)	(0.188)
Japan	-0.116	-0.169*	-0.180	-0.400***	-0.073	-0.024
	(0.141)	(0.090)	(0.112)	(0.118)	(0.120)	(0.111)
Korea	-0.294***	-0.094	-0.063	-0.292**	0.050	0.150
	(0.095)	(0.080)	(0.094)	(0.120)	(0.103)	(0.126)
Netherlands	-0.460***	-0.072	0.041	-0.494***	0.139	0.380
	(0.128)	(0.111)	(0.086)	(0.140)	(0.109)	(0.112)
Norway	-0.273***	-0.019	0.130	-0.580***	-0.143*	-0.117
	(0.098)	(0.074)	(0.098)	(0.119)	(0.084)	(0.103)
Poland	0.016	-0.096	-0.015	0.006	0.046	0.137
	(0.060)	(0.075)	(0.073)	(0.095)	(0.096)	(0.100)
Russia	0.089	0.047	0.034	0.287*	0.270*	0.108
	(0.128)	(0.079)	(0.133)	(0.173)	(0.140)	(0.152)
Spain	-0.042	-0.035	0.107	0.125	0.246*	-0.083
	(0.128)	(0.106)	(0.108)	(0.185)	(0.131)	(0.128)
Sweden	-0.237**	-0.138	0.094	-0.295**	-0.090	0.115
	(0.103)	(0.095)	(0.131)	(0.119)	(0.107)	(0.105)
United States	-0.129	0.020	0.065	-0.258	0.038	0.097
	(0.126)	(0.129)	(0.119)	(0.164)	(0.129)	(0.134)

Notes: Data for Belgium only include Flanders. Data for Russia do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).

Table A-4
Differences in Skills Use, by STEM

	At Home			At Work		
	Numeracy	Reading	Writing	Numeracy	Reading	Writing
Australia	0.366***	0.151	0.067	0.274*	-0.091	-0.091
	(0.139)	(0.117)	(0.096)	(0.151)	(0.141)	(0.154)
Austria	0.314***	0.113	0.016	0.087	0.105	-0.151
	(0.102)	(0.088)	(0.105)	(0.115)	(0.119)	(0.109)
Belgium	0.103	0.022	0.022	0.401**	-0.014	-0.070
	(0.126)	(0.082)	(0.084)	(0.166)	(0.116)	(0.135)
Canada	0.341*	0.121	-0.002	0.484***	-0.092	-0.240
	(0.087)	(0.093)	(0.101)	(0.132)	(0.119)	(0.131)
Czech Rep.	0.150*	0.071	0.028	-0.112	-0.254**	-0.163
	(0.085)	(0.092)	(0.104)	(0.136)	(0.119)	(0.145)
Denmark	0.312	0.015	-0.208	0.355**	-0.008	-0.030
	(0.199)	(0.133)	(0.173)	(0.172)	(0.121)	(0.141)
England	0.230	0.030	0.017	0.395**	-0.002	-0.019
	(0.175)	(0.137)	(0.134)	(0.182)	(0.163)	(0.174)
Finland	0.278***	0.120**	-0.105	0.350***	-0.100	-0.115
	(0.068)	(0.059)	(0.077)	(0.099)	(0.089)	(0.107)
France	-0.107	-0.123*	-0.220**	0.296*	-0.078	-0.028
	(0.107)	(0.069)	(0.109)	(0.151)	(0.100)	(0.110)
Germany	0.314***	0.070	-0.105	0.057	-0.196	-0.055
	(0.117)	(0.105)	(0.093)	(0.126)	(0.123)	(0.106)
Ireland	0.372**	0.366**	0.185	0.468**	0.009	-0.023
	(0.184)	(0.153)	(0.142)	(0.220)	(0.169)	(0.245)
Italy	0.623***	0.338**	0.269*	0.450	0.069	0.047
	(0.200)	(0.153)	(0.148)	(0.316)	(0.205)	(0.217)
Japan	0.070	0.287***	-0.063	0.509***	0.052	0.174
	(0.151)	(0.104)	(0.121)	(0.118)	(0.103)	(0.115)
Korea	0.267**	0.087	-0.063	0.380***	0.084	0.121
	(0.107)	(0.078)	(0.123)	(0.132)	(0.101)	(0.133)
Netherlands	0.731***	0.103	-0.058	0.264	0.116	-0.034
	(0.146)	(0.108)	(0.089)	(0.193)	(0.102)	(0.145)
Norway	0.374***	0.135*	-0.158	0.335**	0.141	0.215
	(0.129)	(0.080)	(0.105)	(0.131)	(0.097)	(0.125)
Poland	0.242***	0.081	0.057	0.215**	0.069	-0.056
	(0.065)	(0.086)	(0.078)	(0.107)	(0.090)	(0.101)
Russia	0.046	0.195	0.111	0.234***	0.123	-0.021
	(0.156)	(0.175)	(0.148)	(0.090)	(0.136)	(0.171)
Spain	0.362**	0.296***	0.277**	0.264	-0.062	0.129
	(0.169)	(0.105)	(0.115)	(0.195)	(0.131)	(0.150)
Sweden	0.564***	-0.004	0.010	0.415***	-0.089	-0.183
	(0.113)	(0.101)	(0.139)	(0.127)	(0.110)	(0.112)
United States	0.804***	0.619**	0.327**	0.477***	0.191	0.071
	(0.133)	(0.262)	(0.137)	(0.145)	(0.193)	(0.155)

Notes: Data for Belgium only include Flanders. Data for Russia do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).

Table A-5
Gender Differences in Literacy Skills, extended specifications

	Baseline	Controls for:			
		STEM	ISCO Occupations	Skills Use	STEM+Skills Use
Australia	0.163	0.273	0.079	0.272**	0.343**
	(0.123)	(0.137)	(0.155)	(0.127)	(0.139)
Austria	-0.062	-0.009	-0.232	0.038	0.124
	(0.097)	(0.106)	(0.141)	(0.112)	(0.119)
Belgium	-0.071	-0.010	-0.121	0.017	0.108
	(0.089)	(0.107)	(0.140)	(0.108)	(0.122)
Canada	0.071	0.134	-0.149	0.114	0.132
	(0.094)	(0.101)	(0.113)	(0.105)	(0.112)
Czech Rep.	0.019	0.168	-0.161	0.046	0.143
	(0.113)	(0.136)	(0.159)	(0.158)	(0.172)
Denmark	-0.141	-0.132	-0.316*	-0.138	-0.128
	(0.125)	(0.130)	(0.163)	(0.134)	(0.136)
England	-0.091	-0.040	-0.024	-0.095	-0.093
	(0.133)	(0.148)	(0.139)	(0.161)	(0.161)
Finland	0.044	0.083	-0.045	0.201*	0.170
	(0.095)	(0.104)	(0.112)	(0.104)	(0.116)
France	0.049	0.134	0.050	0.055	0.165*
	(0.085)	(0.083)	(0.097)	(0.101)	(0.099)
Germany	0.017	0.025	-0.140	0.117	0.085
	(0.115)	(0.113)	(0.133)	(0.113)	(0.112)
Ireland	-0.134	-0.075	-0.197	-0.122	-0.088
	(0.110)	(0.111)	(0.141)	(0.130)	(0.132)
Italy	0.220**	0.270**	-0.012	0.165	0.128
	(0.131)	(0.130)	(0.194)	(0.184)	(0.205)
Japan	-0.079	0.003	-0.067	0.076	0.137
	(0.089)	(0.094)	(0.109)	(0.115)	(0.125)
Korea	-0.067	-0.080	-0.130	-0.019	-0.053
	(0.075)	(0.084)	(0.100)	(0.089)	(0.092)
Netherlands	0.111	0.181	0.083	0.354***	0.373***
	(0.106)	(0.120)	(0.159)	(0.109)	(0.113)
Norway	0.017	0.040	-0.067	0.074	0.043
	(0.105)	(0.114)	(0.117)	(0.113)	(0.118)
Poland	0.101	0.105	0.015	0.163	0.142
	(0.082)	(0.084)	(0.093)	(0.109)	(0.122)
Russia	0.140	0.190*	0.316*	0.004	0.135
	(0.108)	(0.102)	(0.175)	(0.120)	(0.117)
Spain	0.045	0.140	-0.029	0.185	0.275*
	(0.098)	(0.100)	(0.133)	(0.147)	(0.154)
Sweden	0.076	0.114	-0.058	0.130	0.143
	(0.101)	(0.103)	(0.132)	(0.123)	(0.123)
United States	-0.153	-0.022	-0.185	-0.122	-0.062
	(0.119)	(0.124)	(0.149)	(0.148)	(0.155)

Notes: Data for Belgium only include Flanders. Data for Russia do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).

Table A-6
Gender Differences in Numeracy Skills, extended specifications

	Baseline	Controls for:			
		STEM	ISCO Occupations	Skills Use	STEM+Skills Use
Australia	-0.078	0.048	-0.100	0.027	0.095
	(0.116)	(0.133)	(0.159)	(0.133)	(0.152)
Austria	-0.219**	-0.094	-0.335**	-0.098	0.023
	(0.106)	(0.114)	(0.144)	(0.120)	(0.123)
Belgium	-0.219**	-0.130	-0.271	-0.131	-0.016
	(0.096)	(0.107)	(0.148)	(0.111)	(0.119)
Canada	-0.176*	-0.047	-0.278**	-0.108	-0.033
	(0.091)	(0.096)	(0.119)	(0.100)	(0.109)
Czech Rep.	-0.008	0.232	-0.179	0.071	0.263
	(0.130)	(0.181)	(0.173)	(0.176)	(0.229)
Denmark	-0.332***	-0.297**	-0.477***	-0.270*	-0.239*
	(0.124)	(0.134)	(0.173)	(0.143)	(0.143)
England	-0.377***	-0.275*	-0.261*	-0.334**	-0.304*
	(0.139)	(0.157)	(0.145)	(0.160)	(0.158)
Finland	-0.229**	-0.136	-0.254**	-0.059	-0.056
	(0.091)	(0.098)	(0.118)	(0.102)	(0.111)
France	-0.119	-0.018	-0.093	-0.022	0.108
	(0.089)	(0.086)	(0.105)	(0.104)	(0.102)
Germany	-0.133	-0.112	-0.226	0.028	-0.005
	(0.112)	(0.116)	(0.139)	(0.113)	(0.116)
Ireland	-0.271**	-0.208*	-0.304**	-0.164	-0.127
	(0.119)	(0.117)	(0.151)	(0.144)	(0.141)
Italy	-0.083	-0.016	-0.218	-0.132	-0.153
	(0.147)	(0.148)	(0.189)	(0.193)	(0.222)
Japan	-0.217**	-0.113	-0.147	-0.040	0.048
	(0.092)	(0.096)	(0.113)	(0.120)	(0.125)
Korea	-0.146*	-0.143	-0.219**	-0.142*	-0.159
	(0.075)	(0.088)	(0.096)	(0.086)	(0.099)
Netherlands	-0.062	0.014	-0.085	0.121	0.151
	(0.108)	(0.124)	(0.143)	(0.112)	(0.122)
Norway	-0.196*	-0.154	-0.201	-0.126	-0.146
	(0.117)	(0.128)	(0.132)	(0.126)	(0.137)
Poland	-0.121*	-0.088	-0.157**	-0.044	-0.036
	(0.067)	(0.073)	(0.072)	(0.086)	(0.099)
Russia	0.109	0.162	0.175	-0.121	-0.019
	(0.110)	(0.111)	(0.146)	(0.120)	(0.118)
Spain	-0.140	-0.025	-0.109	-0.087	0.046
	(0.089)	(0.088)	(0.115)	(0.135)	(0.140)
Sweden	-0.166	-0.088	-0.279**	-0.032	-0.001
	(0.103)	(0.105)	(0.125)	(0.129)	(0.126)
United States	-0.448***	-0.275**	-0.382**	-0.439***	-0.338**
	(0.124)	(0.126)	(0.159)	(0.145)	(0.143)

Notes: Data for Belgium only include Flanders. Data for Russia do not include the Region of Moscow. Standard errors are reported in parenthesis. Asterisks denote statistical significance at the * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels. *Source*: PIAAC (2012).