OECD*publishing*

Please cite this paper as:

Borgonovi, F. *et al.* (2017), "Youth in Transition: How Do Some of The Cohorts Participating in PISA Fare in PIAAC?", *OECD Education Working Papers*, No. 155, OECD Publishing, Paris.
http://dx.doi.org/10.1787/51479ec2-en

OECD Education Working Papers
No. 155

Youth in Transition

HOW DO SOME OF THE COHORTS PARTICIPATING IN PISA FARE IN PIAAC?

Francesca Borgonovi, Artur Pokropek,
François Keslair, Britta Gauly,
Marco Paccagnella

OECD

Unclassified **EDU/WKP(2017)4**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **22-Mar-2017**

English text only

DIRECTORATE FOR EDUCATION AND SKILLS

EDU/WKP(2017)4
Unclassified

YOUTH IN TRANSITION: HOW DO SOME OF THE COHORTS PARTICIPATING IN PISA FARE IN PIAAC?

OECD Education Working Paper No. 155.

Ms. Francesca Borgonovi, OECD; Mr. Artur Pokropek, Educational Research Institute (IBE), Warsaw; Mr. François Keslair, OECD; Ms. Britta Gauly, GESIS, Leibniz Institute for the Social Sciences; Mr. Marco Paccagnella, OECD

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Ms. Francesca Borgonovi (Francesca.Borgonovi@oecd.org); Mr. Artur Pokropek (artur.pokropek@gmail.com); Mr. François Keslair (Francois.Keslair@oecd.org); Ms. Britta Gauly (Britta.Gauly@gesis.org); Mr. Marco Paccagnella (Marco.Paccagnella@oecd.org)

JT03411215

Complete document available on OLIS in its original format

English text only

OECD EDUCATION WORKING PAPERS SERIES

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

----

www.oecd.org/edu/workingpapers

ACKNOWLEDGEMENTS

The authors would like to thank Andreas Schleicher, Montserrat Gomendio, Deborah Roseveare, William Thorn and Glenda Quintini for valuable feedback on earlier drafts of this paper. Francesca Borgonovi led analyses and drafted the paper. Artur Pokropek was responsible for the analytical work developed in section 4. Francois Keslair provided statistical support to section 3. Britta Gauly supported work in section 2. Marco Paccagnella supported the development of section 3. Editorial and administrative support was provided by Jennifer Cannon.

ABSTRACT

This paper uses data from PISA and the OECD Survey of Adult Skills (PIAAC) to examine the evolution of socio-economic and gender disparities in literacy and numeracy proficiency between the ages of 15 and 27 in the sample of countries that took part in both studies. Socio-economic disparities are exacerbated between the age of 15 and 27 and the socio-economic gap in proficiency widens, particularly among low-achievers. Gender disparities in literacy at age 15 are marked across the performance spectrum but are particularly wide among low-performers. However, by age 24 there is no difference in the literacy proficiency of males and females. The gender gap in numeracy at age 15 is quantitatively small when compared with the gap in literacy, although it is more pronounced among high achievers. The paper canvasses possible explanations for the trends observed and discusses implications for policy and practice, including the extent to which the lack of an established link between PISA and PIAAC limits the analytical value of the two studies.

RÉSUMÉ

Le présent document s’appuie sur des données tirées du PISA et de l’Enquête de l’OCDE sur les compétences des adultes (PIAAC) pour examiner l’évolution des écarts de résultats en compréhension de l’écrit et en mathématiques en fonction du niveau socioéconomique et du sexe entre l’âge de 15 ans et de 27 ans dans l’échantillon des pays ayant participé aux deux études. Les disparités socioéconomiques sont accentuées entre l’âge de 15 ans et de 27 ans, et l’écart de résultats en fonction du niveau socioéconomique se creuse, en particulier chez les individus peu performants. L’écart de résultats en compréhension de l’écrit entre les filles et les garçons de 15 ans est globalement marqué, mais il est particulièrement profond chez les élèves peu performants. Toutefois, à l’âge de 24 ans, le niveau en compréhension de l’écrit est le même chez les hommes et les femmes. L’écart de résultats en mathématiques entre les filles et les garçons de 15 ans est quantitativement faible par comparaison avec l’écart de résultats en compréhension de l’écrit, même s’il est plus prononcé chez les élèves très performants. Le document propose des explications possibles aux tendances observées et examine les implications pour la politique et la pratique, y compris en étudiant dans quelle mesure l’absence d’un lien établi entre le PISA et le PIAAC limite la valeur analytique des deux études.

TABLE OF CONTENTS

Figures

Tables

INTRODUCTION

Since its inception in 2000, the OECD's Programme for International Student Assessment (PISA) has become a key source of information regarding the performance of education systems in ensuring that 15-year-old students have acquired key components of the knowledge and skills that are essential for full participation in modern societies. The focus of PISA is on students aged 15 years as, in many OECD countries, this is the age that marks the end of compulsory education.

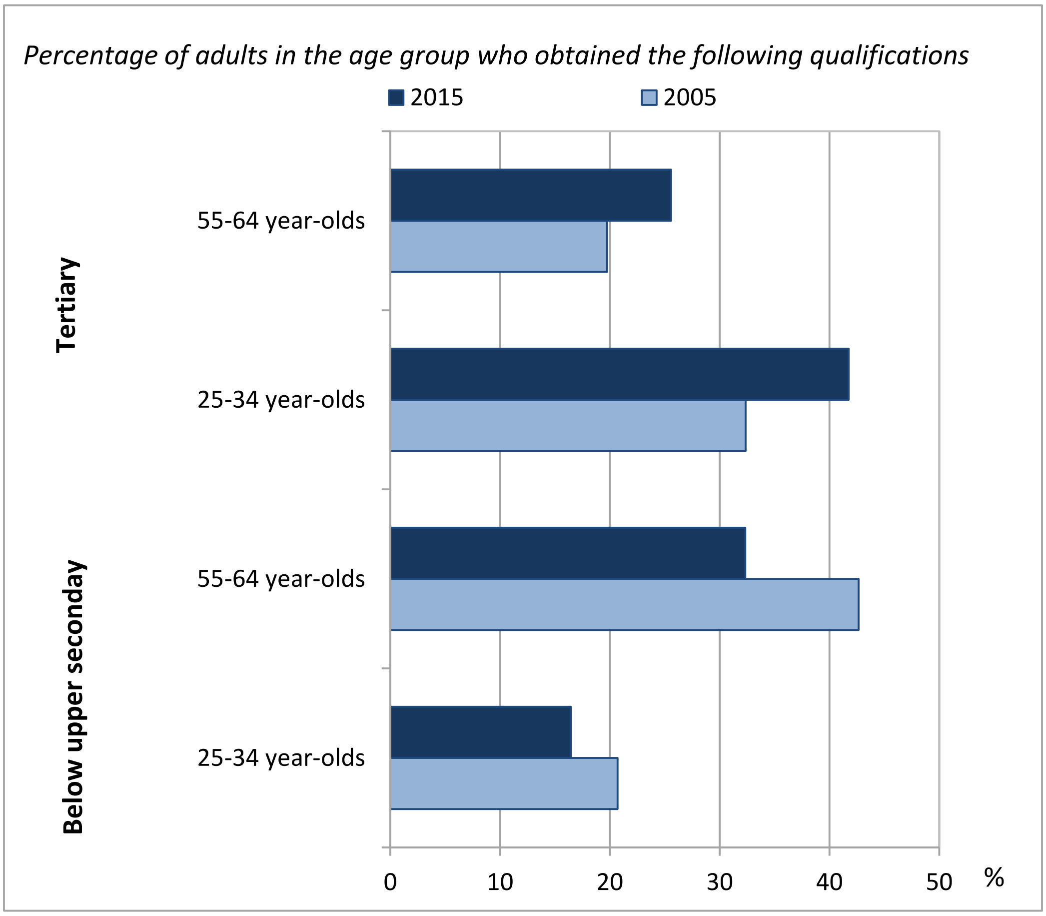
Several countries have implemented longitudinal studies that follow the participants in the PISA study as they matured. Some countries, such as Australia, Canada, the Czech Republic, Germany, and Uruguay interviewed participants who took part in PISA 2000 or PISA 2003 after a number of years. Survey instruments, attrition rates and length of follow-up differed across these countries. Other countries, such as Switzerland, linked information contained in PISA to administrative information. Denmark, administered the PIAAC questionnaire [PIAAC is the OECD Programme for the International Assessment of Adult Competencies], and assessment instruments to a sample of students who had sat the original PISA 2000 study, and matched information available in administrative data sources, to identify some of the changes individuals experienced during the 12 years that set PISA 2000 and PIAAC 2011/12 apart. Canada was the only country that re-administered the PISA assessment to a sample of participants in the PISA 2000 study, to identify achievement growth in reading performance as a function of the experiences participants had encountered over the period.

Longitudinal follow-ups of participants in PISA have generally identified a strong association between reading, mathematics and science skills at age 15 and outcomes such as school-to-work and school-to-further education and training transitions. For example, the 2009 Canadian Youth in Transition Survey followed students who were assessed by PISA in 2000. The study indicated that poorly performing 15-year-old students in PISA faced a disproportionately higher risk of not participating in post-secondary education and of poor labour-market outcomes at age 19, and even more so at age 21 (OECD, 2012d). A similar longitudinal survey in Switzerland, which followed the PISA 2000 cohort until 2010, showed that students' reading performance in PISA is strongly associated with the probability of completing upper secondary education and transitioning into tertiary education (Scharenberg et al., 2014). Similarly PISA longitudinal studies of participants in the PISA 2003 and PISA 2006 editions in Uruguay indicated a strong relationship between performance in the PISA maths test and the probability that individuals completed upper secondary education (Cardozo, 2009) or dropped out from school (Ríos González, 2014). Finally, a study conducted in Denmark indicated that poor performing students in PISA 2000 were more likely to have received income transfers for more than a year between the ages of 18 and 27 – meaning that they were unemployed or ill for long periods (Rosdahl, 2014).

At the same time, longitudinal studies of PISA participants also revealed a large degree of heterogeneity in life outcomes among individuals with similar PISA scores at age 15. Furthermore, evidence from the Australian Longitudinal Study of Australian Youth (LSAY) suggested that many of those who had poor maths scores in the PISA 2003 test were doing relatively well in terms of labour market opportunities and happiness by the time they were 19 years old but also, and crucially, that low-achieving students from a low socio-economic background had a lower likelihood of success than similar students from more affluent homes (Thomson and Hillman, 2010). These findings suggest that the learning opportunities individuals experience after the end of compulsory schooling can importantly determine individuals' outcomes and that these vary across individuals with a different socio-economic background. Moreover, post-compulsory learning opportunities are likely to play an increasingly important role in shaping individuals' life chances. In the not so recent past, the age of 15 was very close to the age at which young people completed their studies, even for those who decided to continue their education beyond the compulsory minimum. However, over the last 20-30 years, OECD countries have experienced a major expansion in educational participation and attainment (see Figure 1.1) resulting in an increase in the time spent in education and training and a consequential delay in entering the full-time labour market. In many countries today, at the age of 15, a student can expect to stay in education for another 5 to 10 years.

Figure 1.1 presents data on trends in educational attainment between 2005 and 2015 for the birth-cohorts that were 25-34 and 55-64 in the two reference years. Figure 1.1 suggests that in 2005 as many as 21% of 25-34 year-olds had not completed upper secondary education. By 2015, this percentage had dropped to 16%. At the same time, as many as 43% of 55-64 year-olds in 2005 had not completed an upper secondary degree. By 2015 this percentage had dropped to 32% in the same age group. Over the same period, an increasing proportion of individuals participated in tertiary education: in 2005 only 20% of 55-64 year-olds had obtained a tertiary degree, but by 2015 as many as 26% had done so. Among younger cohorts, in 2005 as many as 32% of 25-34 year-olds had obtained a tertiary degree and by 2015 42% had done so.

Figure 1.1 Trends in educational attainment between 2005 and 2015 in OECD countries, by age group

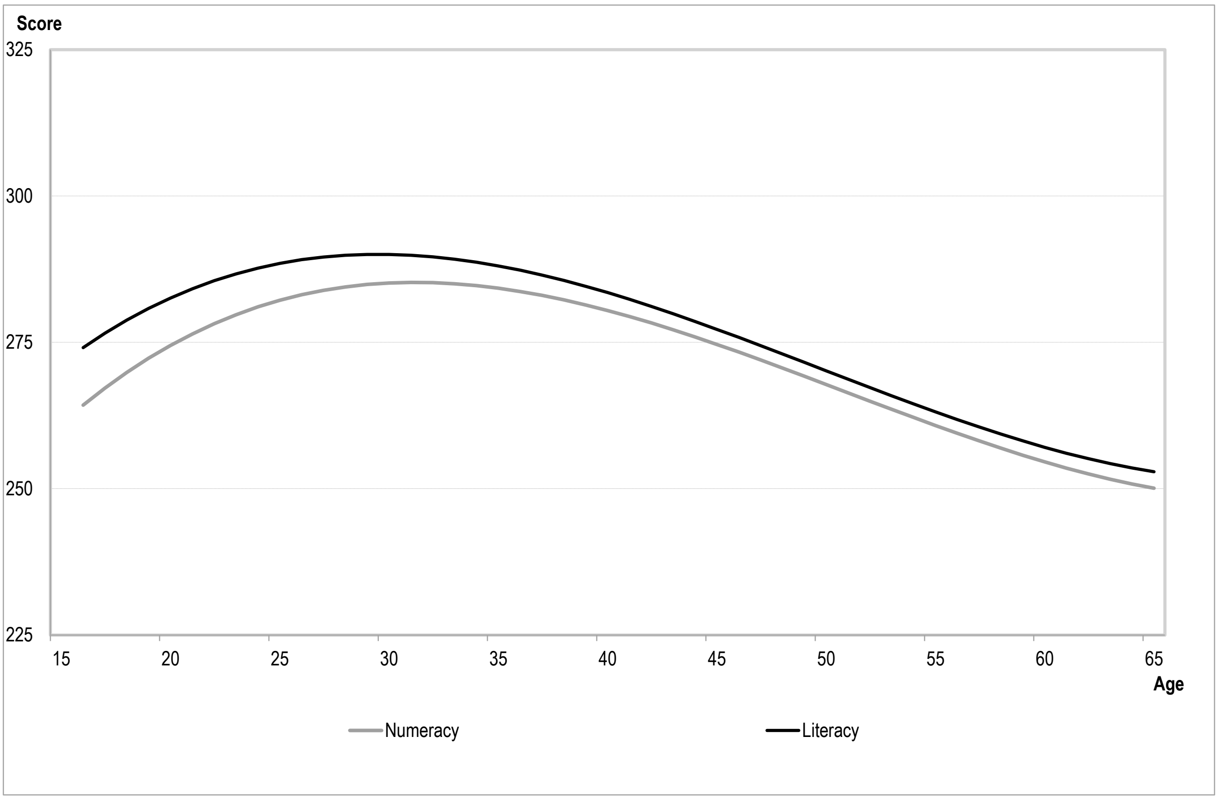


Source: OECD (2016a), *Education at a Glance 2016: OECD Indicators*, Table A1.3, OECD Publishing, Paris. http://dx.doi.org/10.1787/eag-2016-en.

The age of 15 continues to mark an important point in the educational careers of young people to the extent that, in many countries, young people (and their families) make important choices concerning school-to-school, school-to-further-education and school-to-training transitions, or are channelled through formal and informal mechanisms into different educational and training opportunities. Disparities in skill development at this age can therefore have a significant impact on life outcomes. However, the tendency to devote more and more years to the development of skills through formal schooling, further education and training, implies that the competencies measured at age 15 are now better viewed as a starting point, rather than the final level on which the effectiveness of education and training systems should be evaluated.

Furthermore, educational experiences are only one of the factors that can shape individuals' proficiency. The tasks individuals engage in at work, and the events they experience through life, have in fact the potential to reshape cognitive capacity, because brains are "plastic" and change dramatically over the life course. In particular, extensive practice and expertise are associated with profound changes in brain structures, so much so that some regions of the brain grow in size and complexity as a result of practice (Gaser and Schlaug, 2003; Draganski et al., 2006; Mechelli et al., 2004). Previous evidence from PIAAC on age-related differences in proficiency in literacy and numeracy suggests that during the years immediately following the end of compulsory schooling, proficiency in information processing skills increases. PIAAC in fact indicates that, on average, the youngest participants in the PIAAC study are generally less proficient than adults in their mid-20s (Paccagnella, 2016). Figure 1.2 reveals that the highest levels of proficiency are recorded for individuals who are around 30 years old and that, beyond this point, measured proficiency tends to gradually but steadily decline.

Figure 1.2 Age-proficiency profiles in PIAAC



Source: Paccagnella, M. (2016), "Age, Ageing and Skills: Results from the Survey of Adult Skills", *OECD Education Working Papers*, No. 132, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jm0q1n38lvc-en.

Previous analyses of data from PIAAC have identified differences in mean performance across age groups, with a particular focus on differences between prime age and older individuals (the declining part of the curve depicted in Figure 1.2) (Paccagnella, 2016). This report focuses on the years that mark the transition from the teenage years and young adulthood (the ascending part of the curve). In particular this report aims to illustrate if disparities in performance across genders and socio-economic backgrounds widen or shrink during such transition. Gender and socio-economic condition are key dimensions that identify groups that are at risk of underachievement (OECD, 2014; OECD, 2016b). The report exploits the fact that in a large number of countries, students who sat the PISA test in different editions were part of the target population of PIAAC and it is therefore possible to map some birth-cohorts in both PISA and PIAAC. Most of the paper focuses on comparisons between the birth-cohort who sat the PISA test in 2000 (year of birth 1984-1985) and the PIAAC test in 2011/2012 at the age of around 27 (for PIAAC round 2 countries this is the cohort that sat the PISA test in 2003).

At the country level, there is a reasonably close correlation between countries' performance in the different cycles of PISA and the proficiency of the corresponding age cohorts in literacy and numeracy in PIAAC. Figure 1.3 suggests that countries that perform well in PISA in a given year tend to have high performance among the relevant age cohort in PIAAC and vice versa. However, the correlations are far from perfect and these analyses do not look at the dimension of equity and the extent to which disparities are magnified or reduced by the experiences of individuals after compulsory schooling.

Figure 1.3 Mean literacy proficiency in PISA (2000 and 2003) and in the Survey of Adult Skills

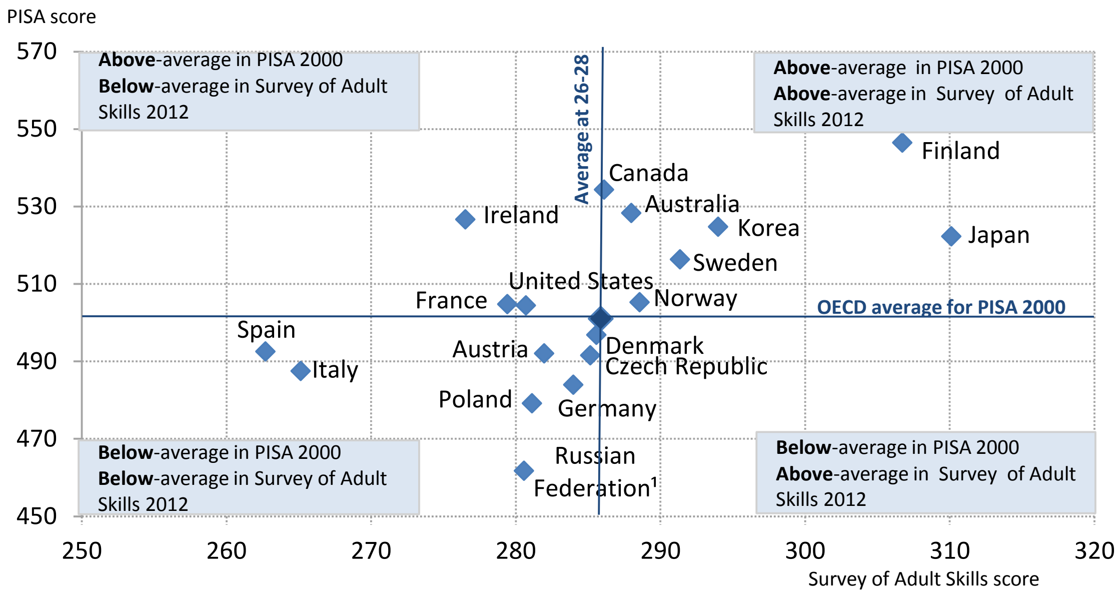
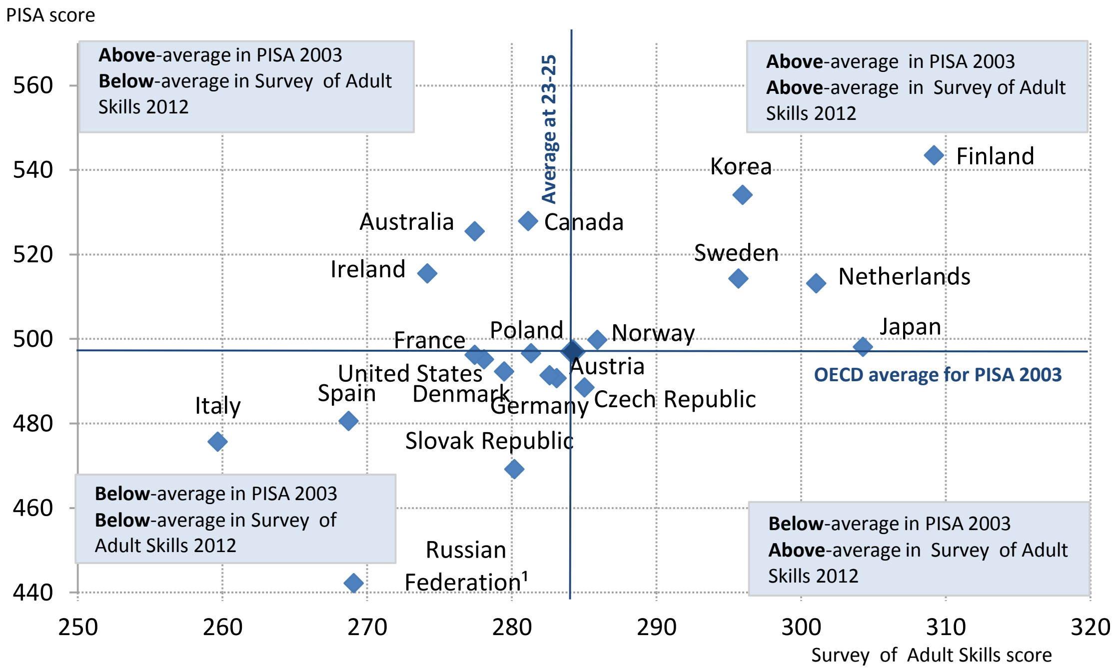


Figure 1.3 Mean literacy proficiency in PISA (2000 and 2003) and in the Survey of Adult Skills (*continued*)



1 The sample for the Russian Federation does not include the population of the Moscow municipal area. The data published, therefore, do not represent the entire resident population aged 16-65 in the Russian Federation but rather the population of the Russian Federation excluding the population residing in the Moscow municipal area.

Source: OECD (2012a), *Survey of Adult Skills (PIAAC)* Database 2012, www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Figure 1.4 Mean numeracy proficiency in PISA (2000 and 2003) and in the Survey of Adult Skills

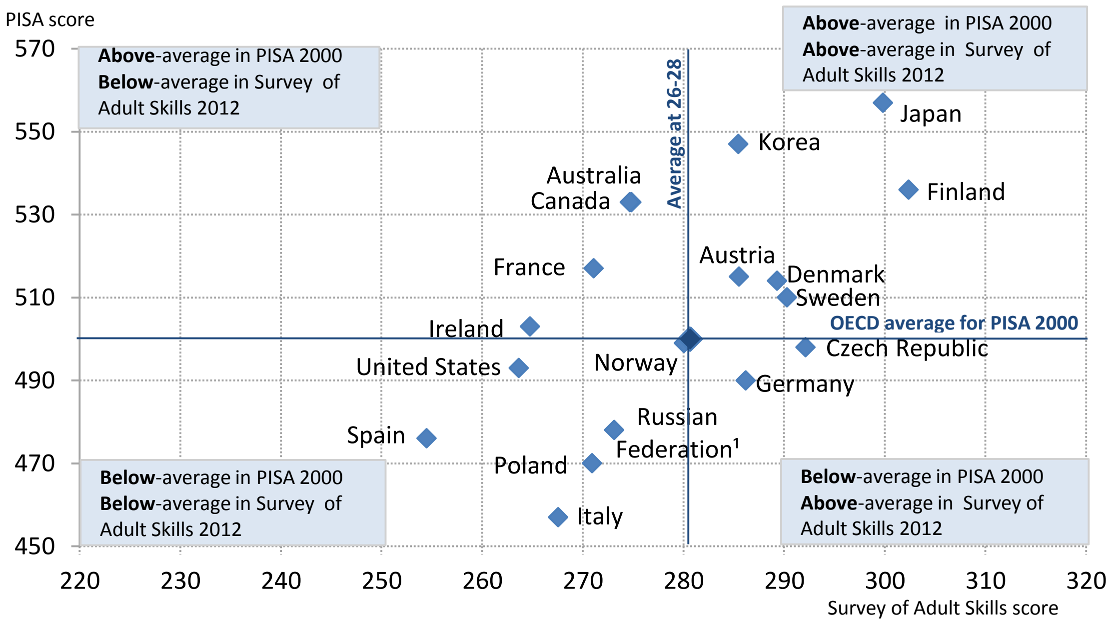
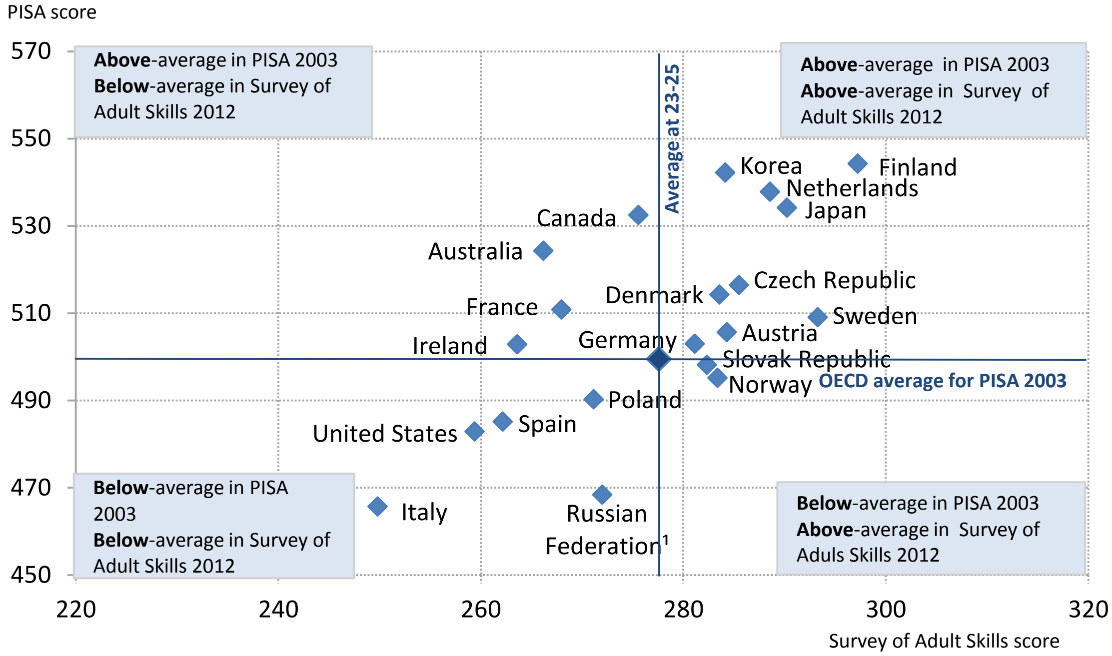


Figure 1.4 Mean numeracy proficiency in PISA (2000 and 2003) and in the Survey of Adult Skills (*continued*)



1 The sample for the Russian Federation does not include the population of the Moscow municipal area. The data published, therefore, do not represent the entire resident population aged 16-65 in the Russian Federation but rather the population of the Russian Federation excluding the population residing in the Moscow municipal area.

Source: OECD (2012a), Survey of Adult Skills (PIAAC) Database 2012, www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

This report first provides a comprehensive overview of the dimensions of the constructs measured in the PISA and the PIAAC assessments, as well as similarities and differences in target populations, administration procedures and questionnaires, and provides examples of sample items to illustrate the types of tasks PISA and PIAAC participants were required to solve. Mapping similarities and differences across the two assessments is fundamental for interpreting comparisons of proficiency scores and disparities in scores across the two assessments. Readers who are most interested in the empirical findings on the evolution of socio-economic and gender disparities in literacy and numeracy may want to skip the detailed discussion of the PISA and PIAAC assessment and questionnaire frameworks and administration procedures presented in section 1 and move directly to section 2. In summary, while the PISA and PIAAC assessment frameworks share much in common, the two studies have not been designed to be linked psychometrically and the results are presented on separate scales. Therefore, it is not possible, given the current design to estimate growth in proficiency for the age cohorts that participated in both studies by comparing results from PISA and PIAAC.

The second section of this report illustrates the evolution of disparities in proficiency in literacy and numeracy by gender, parental educational attainment and an indicator of socio-economic condition between the age of 15 and 27. These sets of analyses exploit comparability between the two assessments along key socio-economic and demographic characteristics and identify equity as a key dimension along which transition pathways should be evaluated. Section 3 develops an analysis based on a national adaptation of PISA to estimate a link between the PISA reading and mathematics and the PIAAC literacy and numeracy scales. Finally, the report reviews implications for policy and also for the design of the PISA and PIAAC studies.

SECTION 2 - A COMPARISON OF THE PISA AND PIAAC ASSESSMENT FRAMEWORKS

2.1 Assessment design

PISA and PIAAC are international comparative assessments designed to measure key competencies among 15-year-old school students and adults aged 16-65 years respectively in a wide range of countries. Both studies contain a direct assessment designed to evaluate proficiency in information processing skills and a background questionnaire designed to identify the socio-economic and demographic background of respondents, the extent to which they practice information processing skills as well as the correlates of skill development and acquisition. However, the two studies differ in their target population, sampling and administration procedures, as well as in the definition and organisation of the skill domains they assess. This section gives an overview of the commonalities and differences between PISA and PIAAC both in terms of design, assessment frameworks and questionnaires.

Design: Target population, sampling and testing time

PISA is a triennial international survey conducted since 2000. The number of participating countries has increased from 32 in 2000 to over 70 in 2015. Each PISA cycle assesses three core domains (reading, mathematics and science), one of which constitutes the major domain in a given cycle (reading in 2000 and 2009; mathematics in 2003 and 2012 and science in 2006 and 2015). The major domain receives a larger portion of testing time and questionnaire items tend to focus on students' engagement, attitudes and dispositions towards the main subject domain.

PISA is an age-based study that collects information from representative samples of students who are enrolled in 7th grade or above and are between 15 years and 3 months and 16 years and 2 months at the time of the assessment administration. This means that "15-year-old" students can be in different grades and different education levels, for example because they may have repeated grades or they may have skipped grades, but they are all in school. PISA uses a two-stage sampling procedure. The first stage is to select a sample of schools and the second stage involves the selection of students within the sampled schools. Although many countries have samples that are larger than the minimum agreed at the international level (for example to have reliable estimates at the regional/state level), sample size is generally around 150 schools and 4 500 students per country. The test takes place in a school setting and students typically take the test in a class with other students under supervision.

Response rates (for both students and schools) are high in PISA, and results of countries with response rates that fall below those detailed in the PISA technical standards are not considered comparable with those of other countries (for example the United Kingdom and the Netherlands in did not meet the target response dates in PISA 2000 and 2003). However, the PISA target population excludes all students who are not in school at the age of 15 (OECD, 2001). The selectivity of the PISA sample (population coverage) varies across countries and determines the extent to which students who were sampled in PISA can be considered to be representative of the population of adults which constituted the PIAAC target population in 2011/12 (for PIAAC round 1 countries) and 2014/2015 (for PIAAC round 2 countries). In most countries considered in this paper population coverage is high since most 15-year-olds are enrolled in school at the age of 15 because of compulsory schooling legislation. Nonetheless, at the age of 15, some socio-economically disadvantaged children may not be in school and are therefore excluded from the PISA target population and analyses presented in this report.

Until 2012, the PISA assessment was administered using printed test booklets containing a range of material from the three core assessment domains and was designed to take around 2 hours to complete. Testing material was organised around subject specific clusters designed to take around 30 minutes to complete and each booklet contained four clusters of test items. Students typically took a break after 1 hour of assessment: in PISA 2000 the booklet was broken into two groups of two clusters each and any question from the first two clusters that were not answered by the student before the break could not be attempted again after the break. In other years, the student was free to move forward and backwards in the test, to strategically solve questions depending on item difficulty, subject and response format, if they so wished.

At the end of the two hours, students were allowed to take an additional half an hour break and were then asked to complete a background questionnaire. Countries could administer additional background questions to students as national questionnaire items or administer questions from the international background questionnaires.

The PIAAC target population was defined as: "all non-institutionalised adults between the ages of 16 and 65 (inclusive) whose usual place of residence is in the country at the time of data collection. Adults are to be included regardless of citizenship, nationality or language". In other words, the target population of country X includes all persons aged 16-65 physically present in X at the date of data collection whose usual place of residence is X, including *illegal residents* (i.e. persons without a valid residence visa or other authorisation to live in the country), but excludes adults present for short stays (tourists, visitors) and adults residing in institutional settings (prisons, hospitals, aged care facilities and military bases and installations).

PIAAC is a household-based study. Trained interviewers administered the background questionnaire (BQ) and the direct assessment to representative samples of the target population. Achieved samples varied from 3 761 in Northern Ireland to 27 285 in Canada where the sample was designed to provide reliable estimates at provincial level as well as for a range of subgroups of the population such as the indigenous population and linguistic minorities The BQ took around 40 minutes to complete on average and the direct assessment, slightly less than an hour. However, no time limit was imposed so respondents could take as much time as they needed to complete the test. Response rates in PIAAC are lower than in PISA because PIAAC is a household-based study while PISA is school-based. Although response rates in PIAAC are lower than in PISA, the population coverage is very high since sample selectivity is limited to the institutionalised population, and therefore the target population reflects the population residing in the country at the time of the test administration. Response rates vary greatly across countries, but they ranged between 45% in Sweden to 75% in Korea (detailed information on response rates for individual countries can be found in Chapter 10 of OECD, forthcoming). Response bias analyses conducted to validate the quality of the PIAAC data indicate that non-respondents share common background characteristics to respondents. Furthermore, hard to reach individuals (defined as those for whom several contact attempts by the interviewers had to be made to achieve participation) do not appear to have different levels of literacy and numeracy from those individuals whose participation did not require additional effort from the interviewers (OECD, forthcoming). These may be due to compensating effects: low skilled individuals may be less likely than others to be willing to sit a test because they might fear test like situations. However, the opportunity cost of time is higher among the highly skilled, who therefore may be less willing to participate in a survey like PIAAC. The first PISA assessment in 2000 was administered in 32 countries, but to date over 80 countries and economies have participated in at least one round of the assessment. The first round of PIAAC in 2011/12 included 24 countries/subnational entities; the second round administered in 2014/15 was administered in nine additional countries, and a third round is planned for 2017/18 increasing the overall sample because of the participation of five additional countries. The PISA study is based on a 3-year cycle with an in depth assessment of one of the three domains every 9 years. PIAAC is planned to be repeated on a 10-year cycle.

Like PIAAC, the core PISA test has a strong emphasis on the notion of literacy and attempts to situate test questions in real-life contexts so that the assessment is not explicitly linked to mathematics, science or language of instruction curricula. It is designed to be valid cross-culturally and cross-nationally and to provide a benchmark measure to help countries understand the distribution of key information processing skills required for economic and social well-being.

Assessment mode

When PISA was first administered in 2000, computers and digital technologies were not as commonly used as they are today. Students in schools were not necessarily exposed to digital content and while ICT labs and courses existed in some schools, the purpose was generally to teach students how to operate digital technologies. ICT was a "study subject" rather than a means through which other subjects could be taught. Instruction material and evaluation exercises came primarily in print form and students had little experience with technology as a learning and assessment tool. A little over a decade later, in 2012, the use of computers was widespread, both in schools and society and individuals engaged with digital technology to solve a wide range of problems. Digital technologies have now become a key resource to be able to effectively solve problems in today's world and consequently ability to access, retrieve and use textual and numerical information in digital forms is necessary to be able to fully participate in society and be successful in the labour market (OECD, 2015b).

PISA started in 2000 as a paper-and- pencil assessment. A computer-based reading assessment was introduced as an optional part of the major domain of reading literacy in 2009 and, in 2012, a computer-based assessment of mathematical literacy was introduced as an optional component of the assessment. In 2009 and 2012, samples of students in a subset of countries were asked to sit an additional 40 minutes computer-based assessment designed to identify students' performance in digital reading (PISA 2009 and 2012), computer-based mathematics (2012) and complex problem solving (2012). In the majority of these countries, administration took place on the same day as the core PISA test (with the core PISA test taking place in the morning and the computer-based test in the afternoon), but in some countries the test was organised the following day or within a week of administration of the core PISA test. Administration on a different day typically occurred in countries where schools closed in the afternoon or where students did not attend school in the afternoon, or in countries where double shifts existed. Finally, in 2012, an additional sample of students was drawn for countries that implemented the optional financial literacy assessment. These students were all part of the PISA target population and completed a two hour test containing four 30 minutes clusters with materials in reading, mathematics and financial literacy. From 2015 onwards, the default mode of testing in PISA became computer-based. As a consequence of the changes in testing mode, some revisions in the organisation and the assessment of the PISA reading and mathematical literacy domain were implemented and documented in the respective frameworks.

In contrast, PIAAC was first assessed in 2012, when computers already played an important role in the work and everyday life of adults in many participating countries. Additionally, the stricter limits on testing time in PIAAC generated the need to ensure the efficiency of testing. The default option for the delivery of the assessments in PIAAC was therefore delivery via a laptop computer. At the end of the administration of the background questionnaire, individuals were handed the laptop computer and were asked to undertake the PIAAC literacy and numeracy assessments (and the Problem Solving in Technology Rich Environments assessment - PSTRE - in the majority of countries) under the supervision of the interviewer. Provisions were made for individuals with poor computer skills or little experience with the use of digital technologies. Respondents with no (or extremely limited) experience with the use of computers were given a pencil-and-paper version of the literacy and numeracy assessments.

Reporting proficiency

PISA and PIAAC conceive proficiency as a continuum of ability involving the mastering of tasks of increasing difficulty (OECD, 2013a). In order to combine multiple assessment items to derive the underlying unobservable ability of individuals, both studies make use of Item Response Theory (IRT). IRT models differentiate between observable variables, such as a correct response to an item, and non-observable variables, such as the latent ability of individuals. IRT scaling creates a common continuous scale that links the latent proficiency with item difficulty via a probabilistic function. Furthermore, IRT permits the estimation of proficiency even if test takers take only a subset of the tasks in the assessment (OECD, 2013d). To correct for measurement error in estimated proficiency levels, not only one competence value per domain is estimated for each person but based on the estimated proficiency distribution, several plausible competence values are drawn (von Davier, Gonzales and Mislevy, 2009; Wu, 2004). Until PISA 2012, five proficiency values were drawn in PISA per domain or subdomain, while ten plausible values were drawn in PIAAC and the PISA 2015 assessment.

Until 2012 PISA fitted a one parameter Rasch model to respondents when scaling item parameters (Rasch, 1960; Masters, 1982). In 2015 PISA and PIAAC a more flexible statistical model was fitted. This model, whose broadest form is the generalised partial credit model (i.e. a two-parameter item response - theory model; see Birnbaum, 1968; Muraki, 1992). The Rasch model is a special case of the model adopted in PISA 2015 and PIAAC. The main difference between the model used in PISA 2015 and PIAAC and the Rasch model is that the PISA 2015 and PIAAC model does not give equal weight to all items when constructing a score, but rather assigns optimal weights to tasks based on their capacity to distinguish between high- and low-ability students. It can therefore better accommodate the diversity of response formats included in tests like PISA and PIAAC.

In PISA and PIAAC items are translated into multiple languages. Some items in some countries may therefore function differently from how the item functions in the majority of countries. For example, terms that are harder to translate into a specific language are not always avoidable. The resulting item-by-country interactions are a potential threat to validity. Until PISA 2012, common item parameters were used for all countries, except for a very small number of items that were considered “dodgy” and therefore treated as “not administered” for some countries. In PISA 2015 and PIAAC the calibration allows for a (limited) number of country-specific deviations from the international item parameters (Glas and Jehangir, 2014; Oliveri and von Davier, 2011; Oliveri and von Davier, 2014).

Until PISA 2012 non-reached items (i.e. unanswered items at the end of test booklets) were considered as wrong answers when estimating student proficiency (i.e. in the “scoring” step) but as not administered when estimating item parameters (in the “scaling” step). In PIAAC and PISA 2015 non-reached items were treated as not administered both when estimating student proficiency and item parameters. The PIAAC and PISA 2015 approach makes the treatment of individuals’ responses consistent across the estimation of item parameters and student proficiency, and eliminates potential advantages for countries and test takers who randomly guess answers to multiple-choice questions that they could not complete in time compared to test takers who leave these non-reached items unanswered. The PIAAC and PISA 2015 treatment of non-reached items means that scores assigned to countries with many unanswered items tend to be higher compared to scores assigned using the non-reached answers as wrong answers.

Although the underlying method for estimating individual proficiency is the same in PISA and PIAAC, the measurement scale differs across the two assessments: proficiency in PISA is reported on a continuous scale set to have a mean of 500 and a standard deviation of 100 across OECD countries. In contrast, the PIAAC literacy scale was linked to IALS (International Adult Literacy Survey) and has a mean of 268 across OECD countries and a standard deviation of 47 while the PIAAC numeracy scale has a mean of 263 across OECD countries and a standard deviation of 47 (OECD, 2016b). Both, PISA and PIAAC describe levels of proficiency in reading and literacy using proficiency levels (Annex A describes the PISA and PIAAC proficiency levels).

2.2 Literacy

Definition of reading literacy in PISA and literacy in PIAAC

"Literacy" whether specified as "reading literacy" or simply as "literacy" is defined in very similar ways in PISA and PIAAC. 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, as well as the mutual influence PISA and PIAAC have had on one another.

PISA defines reading literacy in a broad sense. Reading literacy is seen as a progressive set of skills and strategies that individuals develop in various contexts and social interactions. Furthermore, reading literacy skills are not seen as skills that are only acquired during childhood and the early years of schooling but rather develop over the life cycle. In order to construct full meaning from a text, the reader need to interact with it and use previous knowledge and his/her own prior experience. The definition included in the first rounds of PISA (2000 to 2006) focuses in particular on the active interaction between the reader and the text, as one has not only to understand, but also to use and reflect on texts. Additionally, it clarifies the role reading literacy plays in supporting student involvement and engagement (OECD, 1999).

> "Reading literacy is 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."

When reading literacy became the major domain again in PISA 2009, the definition was modified slightly to put greater emphasis on the reader s "engagement with written texts". Furthermore, the 2009 framework highlights the importance of electronic texts. The notion of "written texts" in the definition implicitly refers to both printed and digital texts (OECD, 2010a).

> "Reading literacy is understanding, using, reflecting on **and engaging** with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society."

The framework proposed for the 2018 assessment of reading (at which point reading literacy will again be the major domain), includes some additional changes (OECD, 2016d). The draft reading framework states:

> "**Literacy** is 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."

In particular, the definition only refers to "literacy" rather than to "reading literacy" and the concept of "evaluating' replaces that of "reflecting on". Finally, the instrumental value literacy has for participation in society precedes that of enabling individuals to achieve their goals. The evolution of the definition of reading literacy over time has meant that the PISA and PIAAC literacy definitions have converged to very similar intended constructs. The definition of literacy used in PIAAC is (PIAAC Literacy Expert Group, 2009):

> "Literacy is 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."

Organisation of the literacy domains in PISA and PIAAC

Literacy is intended to encompass the range of cognitive strategies (including decoding) that students and adults must bring into play to respond appropriately to a variety of texts of different formats and types in the range of situations or contexts in which they read. Consequently the assessment frameworks for the literacy domain in PISA and PIAAC define several key dimensions that characterise literacy: the range of cognitive processes that are needed to comprehend texts, the range of texts to which literacy can be applied, and the type of problems and context in which text comprehension skills occur - which determine the context reflected in texts. The assessment frameworks for literacy in PISA and PIAAC organised the literacy domain along four key dimensions:

- **Content** – representing the set of artefacts, tools, knowledge, representations and cognitive challenges that constitute the corpus students and adults must respond to or be able to use.

- **Cognitive strategies** – the processes that students and adults must bring into play to respond to or use given content in an appropriate manner.

- **Context** – the range of situations and contexts that are reflected in the texts students and adults read.

- **Text type** – the range of texts students and adults read, including continuous, non-continuous and mixed texts, as well as digital and print texts.

Beyond providing an overarching definition of the domain, an assessment framework also provides a description of the main dimensions of construct assessed. Definitions of reading literacy in the PISA assessment framework and of literacy in the PIAAC assessment framework guided the development of the test items, with the aim of adequately covering several aspects of literacy both in terms of the cognitive processes individuals have to rely on to be proficient in literacy, the range of text formats and types they need to be able to work with as well as the domains and situations texts refer to. In general, given the need to limit testing time to reasonable levels (two hours in PISA and around one hour in PIAAC), PISA and PIAAC face a tension between broadening the spectrum of the range of processes, contexts, text formats and types and the number of items administered per process, context, and content. Widening coverage enables better coverage of the assessment framework but at the expense of detail and specificity at the subdomain level. Deepening coverage of specific domain facets enables the identification of the specific facets that represent areas of strength or weakness given overall levels of proficiency. The PISA and PIAAC literacy assessments were designed to ensure broad coverage of what students and adults read, and of the purposes for which they read and assessment items were developed to represent a range of difficulty in texts and tasks. Annex A illustrates some items from the PISA reading assessment and the PIAAC literacy assessment.

Processes

Table 2.1 describes the literacy processes covered in the PISA 2000, 2009, 2018 assessments and the PIAAC assessment. The PISA conceptual frameworks of 2000 and 2009 included five aspects that represent the cognitive processes that readers use in order to negotiate their way into and between texts. However, since the inclusion of too many aspects would have led to coverage problems in the assessment, the five aspects were organised in three broader categories. While the PISA reading literacy processes identified in the framework has evolved over time, the PISA and PIAAC assessments consider and describe similar processes involved in the expression of literacy.

Table 2.1 Reading and literacy processes in PISA and PIAAC

PISA 2000/2009		PISA 2018	PIAAC
Detailed categorisation	Overarching process	Process	Process
Forming broad general understanding	Access and retrieve	Locate information	Access and identify information
Retrieving information	Integrate and interpret	Understand	Integrate and interpret
Developing interpretation			
Reflecting content of a text	Reflect and evaluate	Evaluate and reflect	Evaluate and reflect
Reflecting form of a text			
		Read fluently	

The “access” aspect in PISA and PIAAC refers to selecting, locating and retrieving one or more information from a text. The “integrate and interpret” aspect involves processing what has been read and assigning a meaning to it. Additionally, it includes understanding the relations between different parts of the text. Finally, the “reflect and evaluate” aspect involves using knowledge and attitudes outside the text, relating these to text content and making judgements based on the overall information acquired.

Content

Text format

Table 2.2 illustrates that both the PISA and the PIAAC assessments of (reading) literacy contain the same text formats. In PISA, the text format definition remained unchanged until 2015. However, in 2018, the format definition will consider the unit of the text (or source dimension) and distinguish between single texts and multiple texts. PISA and PIAAC distinguish between continuous (prose) and non-continuous (document) texts (OECD, 2010a; PIAAC Literacy Expert Group, 2009). This distinction is important as each format requires the reader to apply different strategies in reading the text. Continuous texts consist of sentences formed into paragraphs. Examples for this format are novels, newspaper reports, or e-mails. Non-continuous texts rather use typographic feature to organise information. In that way tables, graphs or forms are classified as non-continuous texts. However, in reality text might not easily be distinguished into continuous and non-continuous format. Thus, two further text formats that combine continuous and non-continuous texts are used for classification in the frameworks. The first are mixed texts that describe a text that contains continuous and non-continuous elements. The second are multiple texts that include several generated texts that make sense independent of each other and might only be loosely linked or even contradictory. All of these text formats might appear in both print and electronic media. In PISA 2018 the “multiple” format will be presented as part of the new domain facet “source”, which distinguishes between single texts (same author, same publication date, and same reference title) and multiple texts (different authors, different publication dates, and different reference numbers).

Table 2.2 Reading and literacy text formats in PISA and PIAAC

PISA (2000)	PISA (2009)	PIAAC
Continuous	Continuous	Continuous
Non-continuous	Non-continuous	Non-continuous
	Mixed (continuous and non-continuous elements)	Mixed (continuous and non-continuous elements)
	Multiple (several continuous or non-continuous texts)	Multiple (several continuous or non-continuous texts)

Text type

Text type refers to the rhetorical stance of a text. PISA and PIAAC distinguish between several different text types. The rationale for introducing a classification of texts based on test type is not that task difficulty is expected to vary with rhetorical stance but, rather, that the authenticity of the assessment depends on the fact that assessment items represent the variety of texts that can occur in the real world and that individuals are expected to be able to use and process.

The early PISA rounds (from 2000 – 2006) used the rhetorical stance of a text as part of its format classification (OECD, 1999). However, since PISA 2009, the definition of text types has become a component of the reading domain (OECD, 2010a). Table 2.3 gives on overview on the different text types in PISA and PIAAC.

Table 2.3 Text types in PISA and PIAAC

PISA 2000	PISA 2009/ 2018	PIAAC
Description	Description	Description
Narration	Narration	Narration
Exposition	Exposition	Exposition
Argumentation	Argumentation	Argumentation
Instruction	Instruction	Instruction
Hypertext	Transaction	Records

The PISA and PIAAC classifications of text types are very similar and are largely overlapping. The only difference is that whereas PISA identifies *transaction* texts (in the first round these are defined as *hypertexts*, which are a slots of texts linked together in a way that single units can be read in different sequences), representing texts that have the objective of achieving a specific goal outlined in the text (i.e. the organisation of a meeting), PIAAC identifies *records* texts, which represent texts designed to standardise, present and conserve information without embedding them in other stances (i.e. an inventory such as a graph indicating oil prices, standings in a sports league table, etc.). The other categories are the same across the two studies: *description* refers to texts in which information provided refers to the properties of objects in space, *narration* refers to texts in which information refers to properties of objects in time; *exposition* refers to texts in which information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analysed; *argumentation* refers to texts which present propositions on relationships between concepts or other propositions and *instruction* refers to texts that provide directions on what to do.

Context

The information delivered through written texts varies depending on the situation in which individuals read and the purpose of their reading. For example, in working situations individuals may be required to read inventories of items that are in stock, while in their everyday life they may read reviews of restaurants to plan a night out. The contexts in which texts in the PISA and PIAAC reading/literacy assessments tasks are framed are closely related but do not exactly overlap. Because the PISA and the PIAAC target populations differ considerably in age, contexts have to differ to ensure that the tests reflect the range of everyday situations students and adults may experience in their everyday life.

The *personal* context in PISA refers to texts that are intended to satisfy students' personal interests or to maintain or develop connections with other people. *Personal use* in PIAAC has a somewhat wider frame and includes reading materials in the areas "home and family", "health and safety", "consumer economics", and "leisure and recreation". The *public* category describes texts that relate to activities and the society while *community and citizenship* refers to materials dealing with community resources and staying informed. Since the PIAAC study refers to the working-age population (although many respondents are not active in the labour market) there are a number of differences between the PISA and the PIAAC assessment within the context categories *occupational* category in PISA and the *work and occupation* category in PIAAC. PIAAC deals with various occupational but not job-specific texts. Although PIAAC assesses the use of literacy skills beyond compulsory schooling and in real-life contexts, it has to take into account that only a part of the respondents read at work. Thus, the main differences in the context categories are related to the fact that the PISA items have to accessible to 15-year-old students. The *education* context is similar in both studies, focusing on the respective educational context of the target population. In PISA the tasks are designed for the purpose of instruction and acquiring information as part of a larger learning task, while PIAAC includes texts dealing with opportunities for further learning (OECD, 2010a; PIAAC Literacy Expert Group, 2009).

Table 2.4 Reading and literacy text contexts in PISA and PIAAC

PISA	PIAAC
Reading for private use (personal)	Personal uses
Reading for public use	Community and citizenship
Reading for work (occupational)	Work and occupation
Reading for education	Education and training

Medium

Starting in 2009 PISA broadened the types of texts covered in its reading assessment to include digital texts. The PISA digital reading assessments that were administered in PISA 2009 and PISA 2012 were optional for countries and were administered in an additional optional testing session to subsamples of students who had sat the core paper-based instruments. Digital texts in the digital reading assessment differed from print texts both because of mode of administration (computer instead of paper), as well as the type of content (such as hyperlinks and non-sequential texts) and processes required to access them (navigation skills). The default mode of delivery for the PIAAC literacy assessment was computer-based, although a number of respondents used a paper version of the assessment because they lacked ICT familiarity or were unwilling to take the test using a computer. However, since the PIAAC literacy assessment was designed to enable comparisons with previous adult literacy skills surveys (IALS and ALL), the PIAAC literacy assessment contained both items that had been originally developed for paper-based administration (and therefore did not contain elements typical of digital literacy) and items that incorporated computer-based administration with the other elements that are unique to digital reading, such as the navigation of non-sequential texts.

Organisation of the literacy domain in PISA and PIAAC

Table 2.5 provides an overview of the distribution of items in PISA and PIAAC across different dimensions of the domain (OECD, 1999, 2010; PIAAC Literacy Expert Group, 2009). A smaller number of items were included in PIAAC than in PISA and a relatively low number of items were administered per facets of the domain. The distribution of items across the different dimensions of literacy varies between PISA and PIAAC.

Concerning the different strategies or processes needed to negotiate through a text, PISA and PIAAC show a markedly different focus. PISA 2000 and 2009 included the largest proportion of items (2000: 50 %; 2009: 52 %) on “integrate and interpret”. PIAAC, however, included the largest share of items on the “access and identify” category (55%). Similarly, over 60% of assessment items in PISA 2000 and 2009 involved continuous texts, but less than 25% of continuous texts were part of the PIAAC literacy item pool. By contrast, as many as 54% of texts were mixed in PIAAC, but only 4% of items were mixed in

PISA 2009 and no mixed items were present in PISA 2000. When it comes to the four different context areas, it is not surprising that PISA includes a rather high proportion on the educational context (2000: 28%; 2009: 29%) compared to a low proportion in PIAAC (10%). Instead, PIAAC includes the largest proportion of items in the personal context (50%) while PISA 2000 focused mostly on the public context (38%) and PISA 2009 gave almost equal share to the personal, public and educational context (between 27 and 29%).

Table 2.5 Overview of item distribution, by literacy dimension in PISA and PIAAC

Reading Literacy (PISA)					Literacy (PIAAC)		
	PISA 2000		PISA 2009				
	Number	%	Number	%		Number	%
Medium					**Medium**		
Print	141	100	102	78	Print	36	62
Computer	0	0	29	22	Computer	22	38
Format					**Format**		
Continuous	89	63	81	62	Continuous	13	24
Non-continuous	52	37	7	5	Non-continuous	1	2
Mixed			5	4	Mixed	29	54
Multiple			38	29	Multiple	11	20
Strategy					**Strategy**		
Access and identify	42	30	31	24	Access and identify	32	55
Integrate and interpret	70	50	67	52	Integrate and interpret	17	29
Reflect and Evaluate	29	20	33	25	Evaluate and reflect	9	16
Context					**Context**		
Occupational	22	16	21	16	Work related	10	17
Personal	26	18	37	28	Personal	29	50
Public	54	38	35	27	Society and community	13	23
Educational	39	28	38	29	Education and training	6	10
Total	**141**	**100**	**131**	**100**	**Total**	**58**	**100**

Source: PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Papers*, No. 34, OECD Publishing, Paris. http://dx.doi.org/10.1787/220348414075; OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264062658-en; OECD (1999), *Measuring Student Knowledge and Skills: A New Framework for Assessment*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264173125-en.

Item response format

The response formats in the PISA and PIAAC reading and literacy assessment distinguish between selected responses (multiple choice) and constructed responses (open and closed). Selected-response items require the choice of one or more responses from a number of response options. Closed constructed-response items are often “fill in the blanks” tasks and provide a structured setting for answering. In contrast, open-constructed-response items require respondents to communicate the answers to a task in their own words. PISA uses both item response formats in the paper-based and the electronic assessment.

However, as open-constructed response formats require manual coding by trained experts, only a limited number of these items were included in the computer-based branch and instead items implementing automated coding were used to increase efficiency. Although the PIAAC framework includes no information on the type of item response formats, a look at the task characteristics and sample items clarifies that the same item response formats as in the PISA assessments are used.

PIAAC was developed as a computer-based test with a paper-and- pencil option for adults with little or no computer experience. However, computer-based testing was first introduced as an optional part of PISA in 2009. With the inclusion of digital and electronic texts, the structure of the framework for reading literacy became somewhat different. The most important change was the introduction of the domain category "medium" and hence a general distinction between print and electronic texts. In 2012, the categories were renamed and, from that point on, "fixed texts" refer to print texts whereas "dynamic texts" refer to electronic texts. Furthermore, the new domain subcategory "environment" was introduced for the classification of electronic or dynamic texts to distinguish between authored texts that have a fixed content and message-based texts that allow the reader to participate in the text, i.e. by scrolling or clicking (OECD, 2011).

The introduction of the computer-based assessment led to the introduction of further text related characteristics that can be applied to both fixed and dynamic texts. Three main characteristics can be identified: "text objects" refer to the familiar names of texts that we use in the ever-day context, "text features" are text-based information (i.e. number of pages or linguistic complexity of a text) and "navigation tools" (which apply only to dynamic texts) describe the devices readers use to interact with a text (i.e. scrollbar, buttons, index tabs).

PIAAC also includes some categories to describe properties of dynamic texts. "Hypertexts" are texts that provide a direct link to further text and thus require clicking on a highlighted word or phrase in the text. "Interactive texts" include several texts with information that is necessary to understand the whole content. And "other navigation features" refer the devises readers use in dynamic texts and is thus similar to the description in PISA (PIAAC Literacy Expert Group, 2009).

Computer-based delivery became the main mode of administration of the PISA test in 2015. All trend items used in PISA 2015 were adapted for delivery on computer. The equivalence between the paper- and computer-based versions of trend items used to measure student proficiency in science, reading and mathematics was assessed on a diverse population of students from all countries/economies that participated in the PISA 2015 assessment as part of in an extensive field trial, conducted in all countries/economies that participated in the PISA 2015 assessment. The results of this mode-effect study, concerning the level of equivalence achieved by items ("scalar" equivalence or "metric" equivalence; see e.g. Meredith, 1993; Davidov, Schmidt, and Billiet, 2011) informed the scaling of student responses in the main study.

Factors affecting item difficulty

As the objective of both PISA and PIAAC is to provide information regarding the proficiency of populations with heterogeneous abilities, the assessments are designed to include items of varying difficulty. Ultimately item difficulty parameters for individual assessment items are estimated empirically given the number of individuals with correct responses. At the same time, to guide item development and the interpretation and presentation of results, the factors that are considered to affect item difficulty are identified on the basis of theoretical considerations. In both PISA and PIAAC, the range of factors which are considered to affect the difficulty of literacy items are organised along four dimensions: the quantity and quality of the information available in the text; the text structure and complexity; the processes and steps that are required to solve a task; and the width and depth of connections between different text parts and information outside the text.

The PISA framework describes the following factors (OECD, 2010a): the number and location of required information, as well as the amount and prominence of competing information; the length, complexity and relative abstraction of a text; the type of inferencing, interpretation, reflection or evaluation required; the familiarity of the content and the nature of the knowledge the reader needs to bring to the text; depth of understanding of the text required; the explicitness and transparency of the text structure; the relation of separate text parts to the general theme.

The PIAAC assessment focuses on two broad categories which are considered to affect item difficulty: the degree of transparency of the information present in a text; and the degree of complexity required to make inferences. Inference complexity depends on the use of paraphrasing (which require processing linguistic competences), the level at which inference occurs (whether the problem can be inferred from the text itself or is presented through a concrete question) and extra-textual inference (the need to include information from outside the text). The following issues further affect item difficulty: semantic and syntactic complexity; the amount and prominence of the information needed; the presence of competing information; and text features (degree of relation among different text parts) (PIAAC Literacy Expert Group, 2009).

Engagement in reading

Engagement in reading is seen as an important component of literacy as defined in both PISA and PIAAC. Reflecting this, PIAAC collected information on adults' literacy practices at work and in their everyday life. While PIAAC only assesses competency in reading, it nonetheless collected information on both reading *and writing* practices. It includes questions on the frequency with which individuals engage in reading (i.e., reading instructions, articles, books, …) or individuals engage in writing (i.e. writing e-mails, filling in forms, …) and distinguished between reading and writing in the work context and reading and writing at home and in everyday life.

In 2000, when reading was a major domain in PISA, the study also asked questions about students' reading practices and interests. This covered the degree of exposure to various kinds of reading material at school and at home, reading practices (i.e. frequency of reading or different text types) and students' attitude towards reading (OECD, 1999). In 2009, there was even greater focus on the behavioural aspects of reading, and the areas of perceived autonomy in reading and social interaction in reading were introduced. Additionally, metacognition, which refers to the awareness of and the ability to use various strategies when working with a text, was assessed (OECD, 2010a).

Reading components assessment

In all countries participating in PISA, some students have very poor reading skills. To give more detailed descriptive information on what students at the lower end of the literacy scale know and can do as readers, several easy items were included in the assessment in PISA 2009 which required test takers to produce simple responses. First, shorter and simpler texts were used in order to enable students to go through the whole text. Second, a closer literal match between the item and the text was ensured. Third, the items provided more direction about the location of the information that was necessary for solving the item. Fourth, instead of relating to rather abstract issues, the items content contained personal and familiar experiences. And finally, the items addressed concrete features in reflecting on and evaluating form items.

The usage of these simple items was voluntary and countries were able to choose from two alternative sets of booklets, one of which contained more items at the lower end of the continuum.

No additional changes regarding the inclusion of easier tasks were made in the reading assessment in 2012 or 2015. However, in the revised framework developed for the reading assessment in PISA 2018, where reading literacy will be the major domain, it is proposed that reading fluency be assessed. Reading fluency is defined as an individual's ability to read and process words and texts in order to understand the overall meaning of a text. Previous studies have shown that there is a correlation between reading fluency and reading comprehension (see i.e. Wagner et al., 2010). While fluent reading gives room to attention and memory processes, which can be allocated to higher-level comprehension processes, weaknesses in reading fluency divert resources towards lower level processes resulting in weaker performance in reading comprehension. Thus, the assessment of reading should be used as an indicator to describe and understand differences between students, especially those at the lower end of the proficiency scale (OECD, 2016d).

As is the case for 15-year-olds, a reasonable proportion of adults in the countries participating in PIAAC have poor reading skills. Those adults might be less proficient due to several reasons. They might never have learned reading at school properly, have learning disabilities, or might have been forced to postpone the development of reading skills (Grotlueschen and Riekmann, 2012). As PIAAC assesses skills in the language of the current country of residence but not one's home country, some adults might possess good reading skills in their mother tongue but not in the language of the PIAAC test. In order to provide more information about the strengths and weaknesses of low performing adults an additional assessment was included in PIAAC, known as the *Reading Components Assessment* (Sabatini and Bruce, 2009).

Five components were originally included in the model: alphanumeric perceptual knowledge and familiarity, word recognition, word knowledge, sentence processing, and passage fluency. However, because of difficulties in developing a measure comparable across different languages for alphanumeric perceptual knowledge and familiarity, only three components were included in the PIAAC *Reading Components Assessment*: print vocabulary, sentence processing, and passage comprehension. The reading components assessment was administered to all individuals who sat the paper-based assessment, which included individuals with no computer experience, individuals who refused to take the test via computer, individuals who failed the basic ICT assessment and who failed basic literacy and numeracy tests. By taking into account the percentage correct as well as the time needed to complete the tasks it is possible to differentiate different reading types at the lower end of the continuum and analyse their actual reading skills and potentials.

2.3 Mathematical literacy and numeracy

Gal and Tout (2014) offer a comprehensive overview of the similarities and differences between the PISA mathematical literacy and the PIAAC numeracy assessments. The discussion below draws extensively on their work. Mathematical literacy in PISA and numeracy in PIAAC are built on a closely related set of core ideas. They do not focus on skills defined by a curriculum but refer to the capacity of students and adults, respectively, to deal with mathematical or statistical problems occurring in the real world. Furthermore, mathematical skills are not defined as a minimum level but as a skill continuum ranging from very low to very high skill levels (Gal and Tout, 2014).

Despite a high degree of similarity between the two concepts, there are also several differences. This is already apparent in the designation of the PISA and PIAAC domains: while PISA uses the concept of "mathematical literacy", PIAAC refers to "numeracy". "Mathematical literacy" is widely referred to as students' capacity to apply mathematical skills to "real-life problems" (Jablonka, 2003; Tout, 2000). In contrast, "numeracy" is a wider concept encompassing real-world demand concerning mathematical skills, mathematical practices, or educational ramifications (Hoyles et al., 2002; Straesser, 2003). Furthermore numeracy does not only include skills but also dispositional components (beliefs, self-perception, motivation) concerning mathematics (Gal and Tout, 2014).

The following subsection reviews the concept definitions, how the two domains are organised, how proficiency levels are reported as well as more general issues relevant for the comparison of numeracy skills in PISA and PIAAC.

Definition of mathematical literacy and numeracy

Although PISA is an assessment of students that takes place during compulsory schooling, the assessment itself focuses on applying and transferring mathematical skills to problems outside the classroom (OECD, 2004b). As a growing proportion of situations in everyday life include mathematical reasoning or tools, mathematic literacy is a critical tool for young people to handle various situations. Thus, the basic question in an assessment would be "what is important to know for a young student to be able to handle situations that require mathematical understanding in everyday life?"

The first definition of mathematical literacy in PISA 2000 was based on this understanding as it emphasises the role of mathematics in the real world (OECD, 1999).

> "Mathematical literacy is 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."

When mathematics became the major domain for the first time in 2003, the definition was revised to put a greater emphasis not only on the engagement in mathematics but also on the use of mathematics. In that way, the definition focuses on the active component of individual behaviour, and the fact that students' willingness to use and engage with mathematics skills at age 15 is considered to be crucial to guarantee further skill development in the future (OECD, 2004a). Additionally, the reference to an "individual's current and future life" is simplified to an "individual's life" (OECD, 2004b).

> "Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to **use and engage in mathematics**, in ways that meet the needs of that **individual's life** as a constructive, concerned and reflective citizen."

A further modification of the definition took place for the assessment of mathematics in PISA 2012 when mathematical literacy became the major domain for the second time with a greater focus on the details of the mathematical processes and applications (OECD, 2013b).

> "Mathematical literacy is an individual's capacity to **formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena**. It **assists individuals** to recognise the role that mathematics plays in the world and **to make the well-founded judgments and decisions** needed by constructive, engaged and reflective citizens."

The notion of "formulate, employ, and interpret mathematics", as well as "reasoning mathematically and using mathematical concepts, procedures, facts and tools" refers explicitly to a wider range of mathematical activities. Highlighting the "variety of contexts, shows the increasing role mathematics plays not only inside classrooms but also in various situations in the real world. The phrase "assists individuals" shows clearly that mathematical literacy can be a tool that supports individuals in their decision making. And finally, "make well-founded judgements and decisions" puts emphasis on the recent changes in the flow of information (i.e. use of internet, greater communication possibilities…) that requires individuals to evaluate a great amount of information in their everyday life and the support in doing so though mathematical tools (Gal and Tout, 2014).

In contrast to the rather detailed definition of mathematical literacy in PISA 2012, PIAAC gives a broad definition of numeracy (PIAAC Numeracy Expert Group, 2009).

> "Numeracy is 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."

Nevertheless, its basic structure is similar to the PISA definition. It focuses both on mathematical processes such as "access, use, interpret, and communicate mathematical information and ideas" as well the purpose of mathematical skills which involve "engage and manage mathematical demands". A further association to PISA can be found as the numeracy definition also refers to a broad range of life situations in that mathematical or numeracy skills need to be applied.

Numeracy is further specified through the definition of "numerate behaviour," which involves managing a situation or solving a problem in a real context by responding to mathematical information and content represented in multiple ways. This distinguishes it from PISA in that the PISA framework does not include a specific definition of mathematical or numerate behaviour (PIAAC Numeracy Expert Group, 2009).

> "Numerate Behaviour involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways."

The PIAAC assessment recognises that literacy skills such as reading and writing constitute an enabling factor for numerate behaviour and that when mathematical representations involve text, performance on numeracy tasks is, in part, dependent on the ability to read and understand text. However, numeracy in PIAAC involves more than applying arithmetical skills to information embedded in text. In particular, numeracy relates to a wide range of skills and knowledge (not just arithmetic knowledge and computation), a range of responses (which may involve more than numbers), and responses to a range of representations (not just numbers in texts).

Organisation of the domain

Just as in the case of (reading) literacy, the assessment frameworks for mathematical literacy and numeracy were developed with the aim of describing and identifying key dimensions associated with numeracy proficiency. PISA and PIAAC assess numeracy by examining individuals' capacity to solve problems that mirror the set of problems that arise in the real world. The first dimensions that characterises the PISA and PIAAC numeracy frameworks refers to the set of contexts in which real world problems are embedded and whether they refer to occupational, personal, societal situations. The second dimension refers to the types of mathematical content areas that individuals are required to master and whether test questions have require individuals to identify relationships between objects and circumstances, identify patterns, make predictions and quantify the attributes of specific objects relationships and situations. These refer, broadly speaking to subject areas of algebra, geometry and probability/statistics. The final dimension refers to the type of cognitive processes that individuals need to master to be able to solve a problem, and whether that demands the application of mathematical concepts or the identification of a mathematical model to extrapolate a solution given current parameters.

Mathematical content

The competencies assessed in PISA and PIAAC are not considered to reflect the mathematical content of a curriculum. Rather they describe and assess the mathematical content occurring in various situations in the real world. In both studies, the content classification is organised in four main areas. As can be seen in Table 2.6 both use very similar descriptions describing the "big ideas" behind mathematics.

Table 2.6 Mathematics literacy and numeracy content areas in PISA and PIAAC

PISA 2003/ 2012	PIAAC
Quantity	Quantity and number
Space and shape	Dimension and shape
Change and relationships	Pattern, relationships, and change
Uncertainty and data	Data and chance

Both PISA and PIAAC build on the ideas of Steen (1990) who introduced six broad categories to identify mathematical content. In the first PISA study in 2000, when mathematical literacy was only a minor domain, only two content areas, "change and growth" and "space and shape" were assessed. Since 2003 four content areas are covered: "quantity", "space and shape", "change and relationships", and "uncertainty and data" (OECD, 2004b).

In both PISA and PIAAC, "quantity and number" describe the most pervasive and essential mathematical content area. This content area refers to individual's ability to use of numbers and number operators to quantify.

"Space and shape" and "Dimension and shape" relate to individuals' ability to describe and characterise and use objects in the world and their ability to use measurement formulas to identify their properties. In, PISA the space and shape content dimensions are directly associated with the school subject "geometry".

The "change and relationships" and the "pattern, relationships, and change" content dimensions are described in a similar way across PISA and PIAAC. Both refer to the diverse settings and patterns encountered all around us where it is necessary to understand how things change and develop. Furthermore, both frameworks highlight the importance of using basic mathematical formula to describe and understand change phenomena.

In PISA, the content area "uncertainty and change" is related to the academic field of probability theory and statistics and refers to students' ability to interpret and evaluate data in the presence of uncertainty. The area "data and change" in PIAAC is also related to probability theory and statistics and covers individuals' ability to make and interpret predictions.

Mathematical context

Both, PISA and PIAAC, focus on the assessment of mathematical skills necessary to solve problems in real world contexts. Table 2.7 shows the four context areas in PISA and PIAAC and illustrates that contexts identified in the two studies are similar both in labelling and description (OECD, 2004b; PIAAC Numeracy Expert Group, 2009).

The PISA "personal" and the PIAAC "everyday life" dimensions refer to situations pertaining to personal or family life and the interests individuals have and that guide their behaviour in daily life. Sample tasks would occur in the field of private financial situations, understanding sport statistics or using measurement in home situations. The area "societal" or "societal and community" focuses on students or adults engagement in their society as well as mathematical information presented i.e. in the media which are necessary to understand ongoing processes in the society and the world around them. The “occupational” or “work-related” dimensions could be expected to be one of the dimensions with the widest differences between PISA and PIAAC, given the respective target populations. Whereas PISA assesses skills of 15-year-old students who should be ready to make the decision to continue their study or enter the labour force upon completion of their compulsory schooling but who have little direct experience of labour markets, PIAAC focuses on the working-age population, which, by its very nature has a more direct experience of the labour market. Finally, the “scientific” and “further learning” dimension is the least similar between PISA and PIAAC. In both studies, it refers to more formal aspects of mathematics that are necessary to enable adults to participate in further educational activities.

Table 2.7 Mathematics literacy and numeracy context dimensions in PISA and PIAAC

PISA 2003	PISA 2003/ 2012	PIAAC
Personal	Personal	Everyday life
Public	Societal	Societal or community
Educational/ occupational	Occupational	Work-related
Scientific	Scientific	Further learning

Mathematical process

In order to examine the use of different cognitive processes involved in numeracy, different mathematics/numeracy items in PISA and PIAAC were developed. In the early rounds of the PISA assessment (2000-2009), the mathematical responses were called “competency clusters” and included “reproduction”, “connections” and “reflections”. However, in 2012, this classification changed to better describe the cognitive processes involved in engagement with assessment tasks and that are described in Table 2.8 (OECD, 2013b).

The first category of response is quite similar in PISA and PIAAC and refers to identifying mathematical content, recognising that mathematics can be applied to solve a problem and transform a problem into mathematical structure. The second category in both studies, employing/acting upon or using, includes the application of mathematical reasoning, using mathematical concepts and tools to solve a problem, as well as analysing information in a mathematical manner. The third category relates to reflecting upon and evaluating mathematical solutions as well as interpreting them in the context of the problem. In addition, PIAAC also includes the facet of communication which indicates that a person might have to describe their mathematical actions and interpretations to someone else (Gal and Tout, 2014).

Table 2.8 Mathematics literacy and numeracy process dimensions in PISA and PIAAC

PISA 2003	PISA 2012	PIAAC
Reproduction	Formulating	Identify, locate, or access
Connections	Employing	Act upon or use
Reflection	Interpreting	Interpret, evaluate/analyse, communicate

Item response format

PISA and PIAAC contain the same types of item response formats for the mathematical tasks and those are closely related to the response formats of (reading) literacy. They include selected responses (or multiple choice) where individuals choose one or more options from given responses and constructed responses (open and closed) where individuals communicate the answer or solution to a task in their own words. However, the need for automated scoring in the PIAAC computer-based assessment imposed limits to the use of open-constructed responses in PIAAC, which were only used in the paper-based version assessment. In order to assess at least some communication skills in the computer-based branch, in some items respondents were asked to choose predesigned texts appearing on the computer screen in order to explain a given response.

The item type of open-constructed responses are nevertheless available in PISA as the study uses expert coders to code the selected and constructed responses in both, the paper-based and the computer-based assessment. Thus, PISA has the opportunity to assess students' ability to reason and communicate mathematical skills in more detail (Gal and Tout, 2014).

Table 2.9 gives an overview of items by different facets of the mathematical literacy or numeracy domain.

As in the case of literacy, the total number of numeracy items was considerably smaller than the number of items in PISA when mathematical literacy was the main domain. Concerning the coverage of different categories of content, there was little change in the distribution between PISA 2003 and PISA 2012, with items being almost equally distributed over the four categories in both years. The distribution is similar in the PIAAC assessment with a slightly greater focus on "dimension and shape" (29%) compared to "data and change" (21%). The main difference between PISA and PIAAC concern the distribution of the items across different types of cognitive strategies. PISA 2003 had a focus on "connections" (47%), PISA 2012 focused on "employing" (47%), while PIAAC focused on "act upon or use" (61%) which was similar to PISA's category "employing". When it comes to the four different context areas, PISA 2012 has an almost equal distribution again, putting a slightly greater share of items in the "public" and "scientific" categories (26 and 29% respectively). In contrast, PIAAC contains the largest share of items in the "personal" category (45%), similar shares in the "work related" and "society" categories (around 25%), but only little emphasis on the "education" category (7%).

Table 2.9 Overview of item distribution, by numeracy dimension in PISA and PIAAC

	Mathematical literacy (PISA)					**Numeracy (PIAAC)**		
	PISA 2003			**PISA 2012**				
	Number	**%**		**Number**	**%**		**Number**	**%**
Content			**Content**			**Content**		
Change and relationships	20	24	Change and relationships	40	26	Data and change	12	21
Quantity	23	27	Quantity	38	25	Dimension and shape	16	29
Space and shape	22	26	Space and shape	39	26	Pattern, relationships, change	15	27
Uncertainty and data	20	24	Uncertainty and data	34	23	Quantity and number	13	23
Strategy			**Strategy**			**Strategy**		
Reproduction	26	31	Formulating	42	28	Identify, locate or access	3	5
Connections	40	47	Employing	71	47	Act upon, use	34	61
Reflection	19	22	Interpreting	38	25	Interpret, evaluate/analyse	19	34
Context			**Context**			**Context**		
Personal			Occupational	33	22	Work related	13	23
Public			Personal	34	23	Personal	25	45
Educational/ Occupational			Public	40	26	Society and community	14	25
Scientific			Scientific	44	29	Education and training	4	7
Total	85	100	**Total**	151	100	**Total**	56	100

Factors affecting item difficulty

Several factors influence the difficulty of the mathematical tasks in the assessments. In PISA the difficulty can be described by considering which aspects of mathematical capabilities are required in order to understand and solve a task. The frameworks distinguish the following seven factors (OECD, 2013b): mathematisation; communication; representation; reasoning and argument; devising strategies; using symbolic, formal and technical language or operations; using mathematic tools.

The PIAAC framework similarly states that differences in performance can be accounted for by differences in the following aspects: type of match/problem transparency; plausibility of distractors; complexity of mathematical information/data; type of operation/skill; and expected number of operations (Gal, 2005; PIAAC Numeracy Expert Group, 2009). While the underlying PISA and PIAAC theoretical schemes are different, there are several common features between the two assessments. For example, communication and representation in PISA might refer to type of problem transparency or plausibility of distractors that are identified in PIAAC. Similarly, mathematisation in PISA can be considered as an overarching term to identify the complexity of mathematical information or the expected number of operations identified in PIAAC. Finally, devising strategies and using symbolic, formal and technical language or operations in PISA can be considered to refer to type of operation/skill or the expected number of operations in PIAAC.

Beside mathematical content, in both PISA and PIAAC there are several other factors that can influence individual's performance on the tasks. Among the most important might be the presentation of the mathematical information, the amount of reading that is required in order to understand and solve the tasks, as well the authenticity of the task and whether students and adults have been confronted with the situation or context the task refers to.

Reporting proficiency

Reporting proficiency for mathematical literacy or numeracy is very similar to that of (reading) literacy. The only differences occur in the subscales and the reported proficiency levels for PISA. A single scale with mean 500 and a standard deviation of 100 is used when mathematical literacy is only a minor domain. In 2003 four additional subscales based on the mathematical content areas were used. In 2012 three additional subscales based on mathematical processes and four additional subscales based on content categories were included. All subscales are based on the same scaling parameters as the single proficiency scale. When the paper-based assessment of mathematics was supplemented with an optional computer-based assessment in 2012, results were also reported on a computer scale and a combined paper- and computer scale (OECD, 2014a). In contrast, PIAAC uses a single scale with a mean of 250 and a standard deviation of 50 to report numeracy proficiency (PIAAC Numeracy Expert Group, 2009). Annex B illustrates the PISA mathematics proficiency levels and the PIAAC numeracy levels.

Aids and tools

Solving mathematical problems in the real world for both students and adults often requires the use of aids and tools. Thus, in all PISA rounds, students were free to use their own calculators and other tools such as rulers to provide an authentic assessment situation which provides most informative comparison of the performance in the educational system. Until 2012, all items in the PISA assessment were designed so that the use of a calculator did not enhance students' performance. However, in 2012, some items in the paper-based assessment were constructed in a way that the use of a calculator helped to solve the problems more quickly and more easily. In the optional computer-based assessment, students were given access to a hand-held calculator, an online-calculator or a software that enabled calculations to be performed (OECD, 2013b).

PIAAC also takes into account that adults' engagement in numeracy in everyday or work situations involves the use of certain objects and tools. Thus, respondents were also allowed to use hand-held calculators as well as paper printed rulers that include metric and imperial measurements in the mathematical assessment (PIAAC Numeracy Expert Group, 2009).

Behavioural components

Beyond the measurement of mathematical competencies, it is also important to assess the attitude of individuals towards mathematics and their engagement with mathematics.

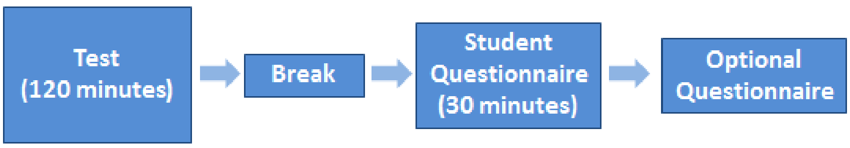
PIAAC collects information on the extent to which adults use and engage with numeracy at work and in everyday life. It includes questions on the frequency with which several mathematical tasks are performed (i.e. calculating costs, preparing charts …) or mathematical tools are used (i.e. using a calculator …). In doing so, it distinguishes between using mathematics during the current or last working life and using mathematics at home. Additionally, PIAAC assesses individuals' general willingness to learn and the extent to which they are interested in going to the bottom of complex issues (OECD, 2013d).

In 2012, when mathematical literacy was a major domain for the second time, students were asked questions regarding their attitudes, beliefs and feelings towards mathematics. The rationale for this was the hypothesis that students who feel more confident with mathematics are more likely to study mathematics or use mathematics in various situations. Two broad areas were covered: Students' willingness to do mathematics (i.e. enjoying mathematics …) and students' interest in mathematics (i.e. usefulness in real life, intentions for further studies …) (OECD, 2013b).

2.4 The PISA and PIAAC background questionnaire

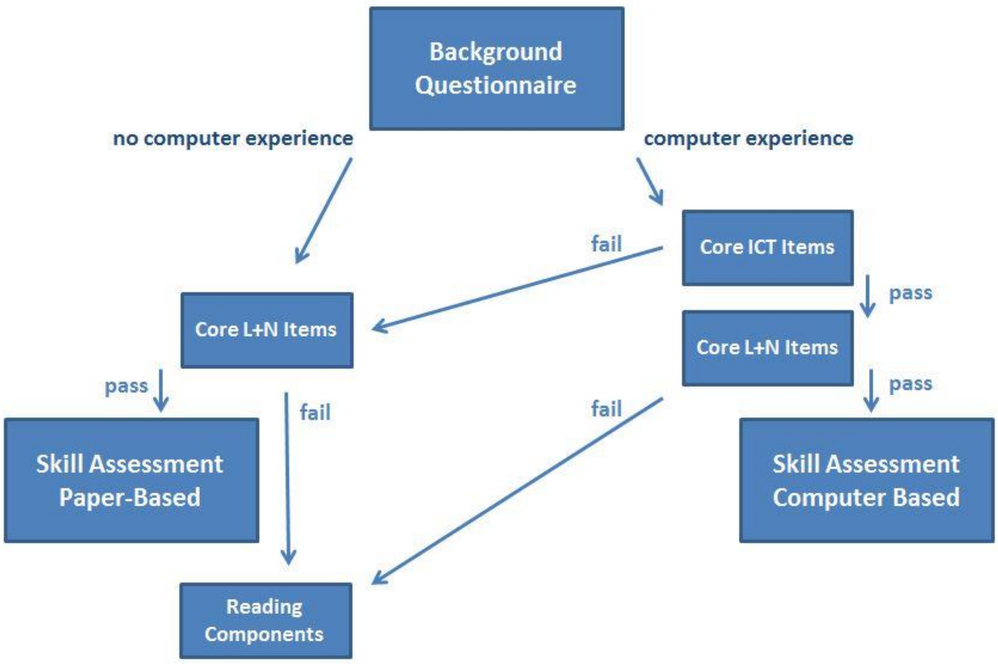
PISA and PIAAC differ not only in terms of the direct assessment design and organisation, but also in the background questionnaires. PISA administers a set of questionnaires, with the students and school principal questionnaires being the core source of information on individuals' background and learning opportunities. These questionnaires are augmented, for countries that decide to administer additional questions, by an ICT and an educational career questionnaire which are administered to participating students and a parental questionnaire that is administered to participating students' families. The sequence of the administration of the components of PISA is described in Figure 2.1 Students first sit the two hour test, then take a short break, complete the core student questionnaire and then complete any additional optional questionnaires.

Figure 2.1 PISA study design



The administration flow in PIAAC was different: participants were asked to respond to the questionnaire first, and then to complete the direct assessment. The background questionnaire was administered in a computer-assisted personal interview (CAPI) format by trained interviewers. At the end of the administration of the background questionnaire, individuals were handed the laptop computer and asked to undertake the PIAAC literacy and numeracy assessments (and PSTRE in the majority of countries) under the supervision of the interviewer. Respondents with no (or extremely limited) experience with the use of computers were given a pencil-and-paper version of the literacy and numeracy components of the assessment. Respondents with computer skills but who possessed poor literacy and numeracy skills were directed to the reading components test, which was taken in pencil-and-paper format only. Despite the fact that the reading components assessment was delivered on paper-and-pencil, information on how long individuals took to solve the assessment tasks was collected by the interviewers. Figure 2.2 illustrates the main study design in the PIAAC cycle 1 administration.

Figure 2.2 PIAAC study design



The PISA and PIAAC questionnaires differ considerably in scope as well as the position of common questions, which, in part reflects the fact that the PISA questionnaire was administered when respondents were tired (after they completed a two-hour test), and was designed for self-completion by 15-year-olds while the PIAAC questionnaire was at the very beginning of the study session, and was delivered with the aid of a trained interviewer.

The personal demographic and social background module is placed at the very beginning of the PISA questionnaire, while it is placed at the very end of the PIAAC questionnaire, except for age and sex which are defined at the start of the PIAAC interview. Common questions across the two studies include questions on gender, date of birth, parental education attainment, the number of books present in the respondent's home (at the time of the test in PISA and at age 16 in PIAAC), household composition, migration status, and if the respondent speaks at home the same language of the assessment. The PIAAC questionnaire is considerably more detailed with respect to defining household composition, language background and migration background, while the PISA questionnaire is more detailed with respect to identifying parental background (through the inclusion of questions on parental occupation and more detailed information on parental educational attainment).

In order to identify economic capital of respondents' households, the PISA background questionnaire asks participating students to report availability of a set of household resources, including durables, educational and cultural resources. PIAAC contains very detailed measures of earnt income, but no information on wealth, and household disposable income, which may include income earnt by a partner or other family member as well as benefits, capital income and pensions.

Both PISA and PIAAC contain information on literacy, numeracy practices and use of ICT devices and applications, to identify the extent to which respondents practice literacy and numeracy in everyday life and, for PISA, at school and PIAAC, in the workplace. The PISA background questionnaire generally follows the organisation of the study around a 9-year cycle with 3 editions per cycle each focusing on a specific assessment domain (PISA 2000 and 2009 focused on reading, PISA 2003 and 2012 on mathematics and PISA 2006 and 2015 on science). The background questionnaire contains detailed questions on competency related practices for the subject that is the focus of the edition. Consequently, questions on numeracy related practices are not asked when reading or science are the main domains and questions on literacy related practices are not asked when mathematics or science are the main domains.

PISA contains very detailed questions on subject specific self-beliefs (such as mathematics self-efficacy, self-concept and mathematics anxiety) as well as intrinsic motivation to learn a subject (enjoyment of reading and mathematics). In some editions (2000, 2003, 2006 and 2009) PISA also asked participating students to report their expected educational attainment and occupations. PIAAC on the other hand contains questions on participants' actual educational attainment, field of study and occupation.

A crucial difference between the PISA and PIAAC background questionnaires is that the PISA questionnaire is designed to help identify factors that are associated with skill formation and individuals' willingness and preparedness to further develop their skills whereas PIAAC attempts to identify not only factors that are associated with skill formation, but also the economic and non-economic outcomes that are associated with individual proficiency in literacy and numeracy.

PISA augmented the student questionnaire with a questionnaire administered to school principals to identify organisational characteristics of the environment in which students learn. This was made possible by the sampling design of PISA, a two-stage sampling whereby schools are the primary sampling unit and students are sampled within participating schools. In each country a minimum of 150 schools are sampled and within each school the target sample is 35 students that meet the PISA age criteria.

PIAAC on the other hand is a household-based study, in which samples of adults are surveyed in their own homes. While it was considered important to identify characteristics of the work environment and how these correlated with individuals' skill levels, the possibility of administering a questionnaire in the workplaces of PIAAC participants was not considered practical and cost effective. Therefore PIAAC participants were asked to answer a series of questions in a dedicated job requirement module. Table 2.10 illustrates the content of the PISA and PIAAC background questionnaires.

Table 2.10 A comparison of the PISA and PIAAC background questionnaires

PISA	PIAAC
General variables (for all cycles): • Student-level inputs (grade, gender, parental education and occupation, family wealth, educational resources, cultural possessions, immigration status, heritage language, age on arrival in country, family support) • School-level contexts and inputs (community size, resources, qualifications of teaching staff) • School-level processes (decision-making, admission policies, assessment and evaluation policies, professional development, teacher engagement/morale, teacher-student relations, parental involvement) • Instructional processes (learning time, disciplinary climate, teacher support) • General non-cognitive outcomes – Commitment to learning (behavioural: truancy; personal goal: educational aspirations; motivational: learning engagement, affective: sense of belonging) Domain-specific trend variables (major domain only) • Domain-specific non-cognitive outcome variables (strategies and metacognition, domain-related beliefs, self-related beliefs, motivation) • Domain-specific processes variables (Opportunity To Learn, teaching practices, teaching quality, system- and school-level support) Thematic extension variables (within individual cycles) • International options (e.g. educational career; ICT familiarity) • Context variables for additional domains (e.g. ICT-related experiences relevant for computer-based problem solving) • Descriptive and explanatory variables for specific reports (e.g. mathematics-related motivations and intentions based on the theory of planned behaviour) • Malleable variables at the school level (e.g. tracking policies, teacher certification) that are specifically selected for descriptive purposes or for causal inference	• General information: year of birth and gender • Education and training: different educational qualifications (e.g., highest qualification or current education), detailed information on continuing education • Current status and work history: respondent's occupational status (employed, unemployed, or out of the labour force), simplified employment history • Current work: respondent's occupation, sector of industry, work satisfaction, working hours, salary/earnings • Recent work: very similar to Section D but for last job; directed to respondents who were not working at the time of the interview: respondent's occupation, sector of industry, etc. • Skills used at work : activities in the current or last job (e.g. co-operation with co-workers, planning activities, time management, physical work) • Skill use literacy, numeracy, and ICT at work: reading, writing, numeracy and computer tasks carried out at work, computer experience • Skill use literacy, numeracy, and ICT in everyday life • About yourself: volunteering, social trust, political efficacy, and health • Background information: children, country of birth, citizenship, parental education

Source: OECD (2013d), *The Survey of Adult Skills: Reader's Companion*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204027-en; OECD (2014a), *PISA 2012 Technical Report*, www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

SECTION 3 - DISPARITIES IN LITERACY AND NUMERACY

3.1 Introduction

Analyses of PISA data reveal large disparities in achievement not only across countries, but also within countries across different subgroups of students. In particular, PISA data has consistently shown that, compared to other students, students from socio-economically disadvantaged households and students with an immigrant background have lower performance in the three subjects considered in PISA – reading, mathematics and science (OECD, 2013e). Gender differences also appear to be wide and did not change greatly between 2000, the first year in which PISA was administered, and 2012, the year in which the fifth wave of PISA and the first round of PIAAC were administered (OECD, 2015a).

Socio-economic disparities in PISA are estimated and monitored using the aggregate PISA index of *Economic, Social and Cultural Status* (ESCS) (OECD, 2001). The ESCS index combines information provided by students in a dedicated questionnaire on their parents' educational attainment, their parents' occupation, and if they have certain possessions in their home (including consumer durables, cultural and educational resources). PISA consistently finds a positive relationship between scores on this aggregate measure of socio-economic status and performance in reading, mathematics and science. On average in 2012, a difference of one standard deviation in the ESCS index was associated with a difference of over one-third of a standard deviation in students' mathematics scores (39 PISA points) and ESCS explained as much as 15% of the overall within-country variation in student performance in mathematics (OECD, 2013e).

Students participating in PISA are also asked questions on whether the language they speak at home differs from the language of instruction in their school, if they were born in the same country as the country of the assessment, and if their parents were born in that or another country. On average, immigration background, defined as being foreign-born or with foreign-born parents is associated with a performance disadvantaged. For example, on average, across PISA participating countries, foreign-born students scored on average 34 PISA points (around one-third of a standard deviation) lower than students with no immigration background on the mathematics scale and those who spoke a language at home that differed from the language of instruction scored 33 PISA score points (around one-third of a standard deviation) (OECD, 2013e) lower than students who spoke the language of instruction at home.

The average girl in OECD countries outperforms the average boy in reading by over one-third of a standard deviation (39 points). Gender differences in reading are even more pronounced among low-achieving students: on average across participating countries, in the bottom decile of the country-specific reading performance distribution the gender gap is as high as 52 PISA score points (or around half a standard deviation). In mathematics, the gender gap in favour of boys among average performing students corresponds to one-tenth of a standard deviation (11 points), but it increases to 20 PISA score points (or one-fifth of a standard deviation) at the top decile of the country-specific mathematics performance distribution. In science no gap can be observed among average performing students, while boys appear to be over-represented at both the top and the bottom tail of the science performance distribution (OECD, 2015a).

PISA also reveals that disparities by gender, immigration background and socio-economic status vary considerably across countries in systematic ways. Between-country differences in the size of the gender gap in academic performance, for example, have been found to be associated with the level of gender inequality in society (as measured by differential access to politics and opportunities in the labour market, as well as by attitudes towards sex roles in the family and society; see Else-Quest et al., 2010; Guiso et al., 2008; Hyde and Mertz, 2009; Reilly, 2012). However, other studies failed to identify such a relationship, because findings appeared to be sensitive to the sample of countries considered (Fryer and Levitt, 2010; Kane and Mertz, 2012; Stoet and Geary, 2013). Similarly, between-country differences in the size of socio-economic disparities in performance have been linked to the organisation of education systems, with tracking and streaming policies being associated with larger socio-economic disparities (Hanushek and Woessmann, 2006; Brunello and Checchi, 2007; Ruhose and Schwerdt, 2016).

PISA data illustrates the cumulative effect that family, social and educational factors have in shaping performance in a low-stakes standardised test at the age of 15. However important this information is, most students in OECD countries will continue their educational careers for some years after the age of 15. Most 15-year-old students can expect to stay in education or training for at least another three to four years with those who will go on to complete higher degrees looking at another 10-15 years of study. Over the past decades, educational participation has increased rapidly in many OECD countries: on average, across OECD countries, 32% of 24 to 35 year-olds had a tertiary degree in 2005. By 2015 this figure had risen to 42%, an increase of 7 percentage points (OECD, 2016a). Growing educational participation and attainment have been the result of growing demands for high levels of skills in the labour market and in everyday life. Furthermore, as the individual benefits of education are relative (and therefore depend on achieving higher levels of attainment than those obtained by peers), there is a strong incentive for individuals to acquire more education to be at the top of the education “pecking order”. PISA cannot provide information on the development of information processing skills beyond the age of 15 and the extent to which gaps in performance across genders, socio-economic and immigrant groups evolve with age.

In this context, understanding how performance gaps evolve as young people move from school to post-school education and training and through this to the labour market is extremely important for policy makers. Different types of educational and training systems and institutional arrangements in the labour market may have very different impacts in terms of the extent to which they mitigate or exacerbate the disparities that exist at the end of compulsory schooling. What is observed regarding the relationships between background characteristics and performance among a cohort of 15-year-old students might be very different from what is observed among the same cohort at later ages.

This section uses data from PISA and PIAAC to compare changes in the size of the disparities observed in PISA and PIAAC for the same birth-cohorts across countries. The goal is to identify the distribution of post-schooling learning opportunities and how these affect inequalities in skills acquisition. The focus of most analyses is the birth-cohort that sat the PISA test in 2000 and was around the age of 27 when sitting the PIAAC test in 2011/12. Some analyses also include countries who participated in the second round of PIAAC, in which case results are based on comparisons between PISA 2003 data and adults aged 26-28 in PIAAC. All analyses exclude adults sampled in PIAAC who report having been born outside of the country in which they were assessed and did not provide information on when they migrated to the country or reported having migrated to the country after the age of 10 while it includes foreign-born individuals who reported having migrated to the country prior to age 10. The age 10 threshold was chosen to ensure that individuals would have been part of the PISA sample at the age of 15 and had adequate language proficiency irrespective of their origin and destination. Previous analyses of PISA data in fact reveal large late-arrival penalties when examining academic outcomes, particularly in some countries (OECD, 2015b). Although England, the Netherlands and Northern Ireland (the United Kingdom) and the Russian Federation took part in both PISA and PIAAC, they are excluded from analyses presented in this paper. The response rate for the United Kingdom and the Netherlands in PISA 2000 and PISA 2003 was below the minimum required by the PISA technical standards. The sample for the Russian Federation in PIAAC does not include the population of the Moscow municipal area and therefore does not represent the entire resident population.

The focus of the section is on performance gaps related to gender, parental education, and family resources, three measures that are common across PISA and PIAAC. In addition to reporting their gender and the educational level of their parents, participants in PISA and PIAAC were asked to report the number of books that are available in their home (PISA) or were available in their home when they were 16 years old (PIAAC). The number of books available at home has been used in both PISA and other international large scale assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) as an indicator or socio-economic and cultural status of the respondent's family and is one of the factors that is most strongly associated with reading, mathematics and science proficiency among school-aged students (OECD, 2013e). It is strongly associated with the wealth of the students' families, with the occupational status of the respondents' parents and with other educational and cultural resources that are available for students in their homes (OECD, 2013e). Disparities related to being a non-mother tongue speaker or related to immigration background are not considered. This is because of the small sample sizes at the birth-cohort level for language minorities and foreign-born individuals in most countries in PIAAC.

The PISA and PIAAC assessments in the domains of literacy and mathematics/numeracy are not directly comparable (see section 2 of this paper). It is not possible to examine growth in achievement at the birth-cohort level, or to compare distributions to examine convergence or divergence in the distribution of skills. However, it is possible to create measures of standardised performance gaps across population subgroups that are defined in the same way across PISA and PIAAC, and to map changes in the size of these gaps between the dates at which PISA and PIAAC were administered to sample of individuals drawn from the same birth-cohort (this design is known as a "pseudo-cohort" analysis). This means that while this paper adopts a notion of learning as a dynamic process, by which individuals experience learning gains and losses as they progress through life, it can only map the extent to which initial differences in foundation skills, interacting with the experiences of individuals as they age and move through the education system and into the labour market (which are in part determined by levels of foundation skills and in part externally driven or subject to chance) widen or decrease over the time.

This does not limit the policy relevance of the analysis. Both overall performance and inequalities in performance are relevant to assessing the impact of education and transition systems. How skills are distributed across the population can have significant implications on how economic and social outcomes are distributed within society. Therefore, assessing the extent to which socio-economic background, parental educational attainment and gender determine the acquisition of information processing skills in further education, training, the labour market and everyday life is an important policy consideration. Considerable policy and academic research has considered the effects that institutional and organisational arrangements at the school level have on inequalities in skill development (van de Werfhorst and Mijs, 2010; Fruehwirth, Navarro, and Takahashi, 2016). However, much less is known at the international level about how disparities in skills evolve following the completion of compulsory education.

As people enter adulthood and age, educational and life experiences become increasingly differentiated. In particular, the educational and labour market trajectories of males and females, and of individuals coming from different socio-economic backgrounds, differ markedly. Given the evidence about skills development in the years between teenage and early adulthood (Paccagnella, 2016), it is of interest to investigate how gender and socio-economic gaps evolve in the same time frame.

In the remaining sections of this paper the term literacy will be used to indicate reading in PISA and literacy in PIAAC. Similarly, the term numeracy will be used to indicate mathematics in PISA and numeracy in PIAAC. The core set of analyses presented in this section illustrates changes in socio-economic and gender gaps in literacy and numeracy for the pseudo-cohorts that took part in PISA 2000 and PIAAC. Particular attention will be paid to the 1985 birth-cohort. People born in that year, in fact, were eligible to take part in the PISA study in 2000 and were approximately 27 years old at the time PIAAC was administered. Unfortunately, as the target PIAAC sample size in each country is about 5 000 for the entire adult population aged 16 to 65, there are often not enough observations to analyse a single birth-cohort in the PIAAC dataset. For this reason, the PIAAC age band was widened to 26-28 year-olds, thus allowing people born a year before and a year after the 2000 PISA cohort. For some analyses PIAAC round 2 countries were added and PISA 2003 was used to match the PIAAC 26-28 age range. For numeracy, given the change in framework between 2000 and 2003, analyses were developed using both PISA 2000 (matched to PIAAC 26-28 year-olds) and PISA 2003 (matched to PIAAC 23-25 year-olds) and results are presented for the PISA 2003-PIAAC 23-25 year-olds comparisons.

PISA and PIAAC data cannot be compared directly, as cognitive skills were measured in the studies with a different metrics (as shown in more detail in section 2). However, standardised gaps can be compared by computing a modified Cohen's d measure: the difference in mean scores across populations of interest (for example males and females) is divided by the pooled standard deviation across the two populations and across all countries participating in both PISA and PIAAC. While the original Cohen's d measure is a pure indication of effect size and is always positive (with larger numbers denoting a larger difference between group means), to allow for the possibility that a group may be doing better in a country but not in another, we modified the Cohen's d measure to derive a standardised measure of performance gaps which takes both positive and negative values, with larger absolute numbers indicating larger differences and the positive/negative sign indicating the group which is more advantaged. Following common practices in the literature, we consider an absolute value (positive or negative) smaller than 0.3 as a small gap, a standardised gap between 0.3 and 0.5 as a medium size gap and a standardised gap greater than 0.5 as a large gap.

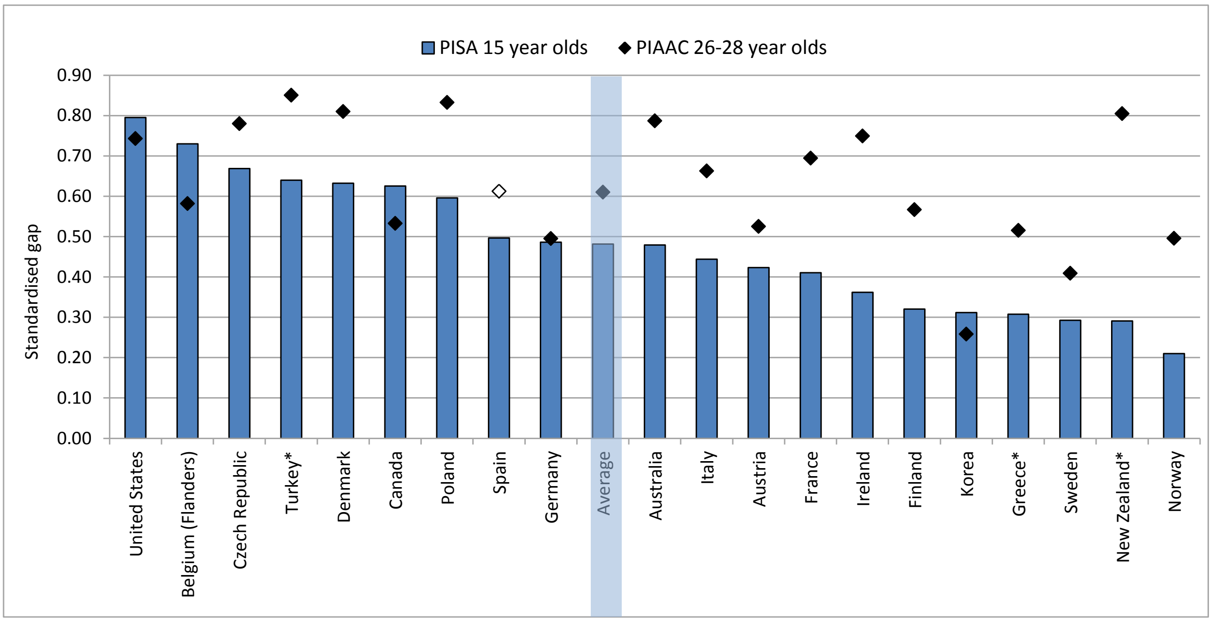
3.2 Socio-economic disparities

Socio-economic disparities in academic achievement have attracted the attention of researchers and policy makers since the 1960s (see, for example, Coleman et al., 1966; Peaker, 1971; Jencks, 1972; and comprehensive reviews such as White, 1982; McLoyd, 1998; Buchmann, 2002; Sirin, 2005). PISA has considerably advanced the understanding of between-country differences in socio-economic disparities in academic performance and helped focus evaluations of the effectiveness of education systems on questions of equity. Socio-economic disparities in performance in standardised tests in the teenage years have been considered to be the result of disparities in the quality of primary and secondary education experienced by children from different backgrounds, as well as to the differences in the out-of-school experiences of these children (Baker, Goesling and LeTendre, 2002; Bradbury et al., 2015; Downey and Condron, 2016; Downey, von Hippel and Broh, 2004; Merry, 2013).

In this section we consider disparities by parental educational attainment and the number of books present in the respondent's home (a good proxy of socio-economic and cultural capital) and examine the evolution of socio-economic disparities from age 15 to 27 in literacy and from age 15 to 24 in numeracy. While a large body of research in fact identifies the role socio-economic status has on academic performance in different countries at school, much less is known about the persistence of such disparities into young adulthood, and whether convergence or divergence in skill growth occurs when students transition from school into further education, training, the labour market and adult life.

Data presented in Figure 3.1 illustrates the difference in literacy proficiency in PISA and PIAAC for individuals whose parents did not obtain a tertiary degree and individuals with at least one parent who obtained a tertiary degree. The score gap associated with parental education is generally large at the age of 15 and that it tends to widen as the cohort of students observed in PISA move into young adulthood. In the Czech Republic, Denmark, and Poland, the standardised parental education gap is greater than 0.5 at the age of 15 and increases by age 27. By contrast, in Belgium (Flanders), Canada and the United States the gap is larger than 0.5 at age 15 but decreases by age 27. In New Zealand, Norway and Sweden the standardised parental education gap is small at age 15 (standardised gap around 0.3) but it increases 5 by young adulthood and, in New Zealand, it becomes as large as 0.8. In Korea the gap is small at age 15 and remains small by age 27. The parental education gap at the age of 15 is medium-sized in Australia, Austria, Finland, France, Germany, Ireland, Italy and Spain. The gap remains stable in Germany, while it tends to grow in the other countries.

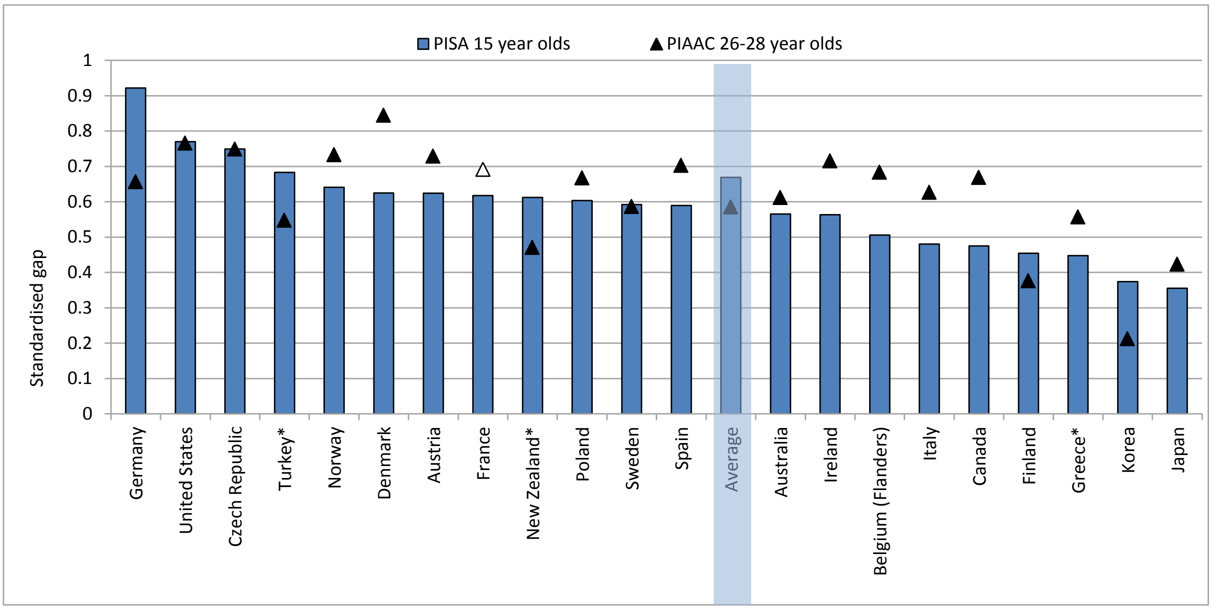
Figure 3.1 Disparities in literacy between individuals with and without tertiary educated parents in PISA 2000 and PIAAC 26-28 year-olds



Notes: The standardised gap refers to the difference in the mean scores of individuals with at least one parent educated at the tertiary level and individuals without tertiary educated level parents divided by the average standard deviation of countries participating in the study in a particular education. Countries are ranked in descending order of the gap in PISA 2000. Bars and diamonds highlighted in dark represent groups for which the gap is statistically significant at the 5% level. * Next to the country name denotes PIAAC round 2 countries for which PISA 2003 data were used to identify performance at age 15.

Sources: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Figure 3.2 Disparities in literacy between individuals who had more and individuals who had less than 100 books in their home in the teenage years in PISA 2000 and PIAAC 26-28 year-olds



Notes: The standardised gap refers to the difference in the mean scores of individuals with more than 100 books in the home at age 15 and individuals with less than 100 books divided by the pooled standard deviation of the study. Countries are ranked in descending order of the gap in PISA. Bars and diamonds highlighted in dark represent groups for which the gap is statistically significant at the 5% level. * Next to the country name denotes PIAAC round 2 countries for which PISA 2003 data were used to identify performance at age 15.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Figure 3.2 illustrates that socio-economic disparities in literacy (as measured by the number of books present in the home) are extremely large at age 15, but also that countries differ markedly in the extent to which disparities evolve as individuals transition from school into further education, training and the labour market. The figure illustrates the gap in literacy scores between individuals with less than 100 books in their homes in the teenage years (age 15 in PISA and age 16 in PIAAC) and individuals with more than 100 books in their home at the same age.

The average standardised gap observed in PISA 2000 when individuals were 15 is over 0.6. This is a large gap and it remains largely stable between the age of 15 and 27. However in Germany, the country with the largest gap at age 15, with a standardised gap of 0.88, the gap narrows considerably by the time youngsters more into young adulthood and converges with the OECD average since the standardised gap at age 27 is 0.57. On the other hand, in Canada, a country where the gap is of medium size at age 15 – corresponding to a standardised gap of 0.45, the gap increases markedly between the age of 15 and 27. By young adulthood the standardised literacy gap is 0.67. Other countries where the gap widens considerably are Belgium (Flanders), Denmark, Ireland and Italy. Korea is the only country where the gap appears to narrow from comparatively low levels and in Japan and Finland small socio-economic gaps persist as teenagers transition into adulthood.

Table 3.1 reports the standardised gap in literacy and numeracy at the age 15 (PISA 2000 and PISA 2003 respectively) and in young adulthood (PIAAC) between individuals with at least one parent who obtained a tertiary degree and those whose parents did not obtain a tertiary qualification. The standardised gap in both literacy and numeracy is relatively stable between age 15 and 27 and 15 and 24 among individuals at the 90th percentile of the performance distribution (high-achievers). The gap in literacy is 0.46 at age 15 and 0.53 at age 27. In numeracy the gap is 0.47 at age 15 and 0.45 at age 24. Table 3.1 also illustrates that, on average, the gap at age 15 is very similar among individuals at the bottom of the performance distribution (10th percentile) (0.48 in literacy and 0.33 in numeracy). The gap associated with the education level of parents grows stronger between the age of 15 and 24/27 at the bottom tail of the performance distribution. The gap increases for this group from 0.48 to 0.77 in literacy and from 0.33 to 0.72 in numeracy.

Table 3.1 Variation between PISA and PIAAC in disparities in literacy and numeracy between individuals with and without tertiary educated parents across the performance distribution, participating country average

	Literacy					
	Parental education gap 90th percentile			**Parental education gap 10th percentile**		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 26-28	22.45	(4.21)	*0.53*	32.66	(5.99)	*0.77*
PISA 2000	44.27	(1.66)	*0.46*	45.76	(2.45)	*0.48*
	Numeracy					
	Parental education gap 90th percentile			**Parental education gap 10th percentile**		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 23-25	20.17	(3.83)	*0.45*	32.21	(6.45)	*0.72*
PISA 2003	43.93	(2.03)	*0.47*	30.66	(1.56)	*0.33*

Notes: Standardised gap refers to the difference in the mean scores of individuals with at least a tertiary educated parent and individuals with no tertiary educated parent divided by the pooled standard deviation of the study. The table presents the average across the 20 countries with comparable data.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Table 3.2 presents the variation in the standardised gap in literacy and numeracy at the age 15 (PISA 2000) and in young adulthood (PIAAC) between individuals who reported having more than 100 books at home at the age of 15/16 and those who reported having less than 100 books at the same age. Results are in line with those presented for parental education and suggest that the socio-economic gap in both literacy and numeracy widens between the age of 15 and 27 and 15 and 24 at the bottom of the performance distributions. A crucial difference is that among high-achieving students in numeracy, the gap appears to be particularly wide at age 15 but shrinks by the age of 24 years.

Table 3.2 Variation between PISA and PIAAC in disparities in literacy and numeracy between who had more and individuals who had less than 100 books in their home in the teenage years across the performance distribution, participating country average

	Literacy					
	Gap 90th percentile (>100 books - <100 books)			Gap 10th percentile (>100 books - <100 books)		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 26-28	22.37	(3.37)	*0.54*	32.82	(4.98)	*0.79*
PISA 2000	49.49	(1.27)	*0.52*	60.43	(2.03)	*0.64*
	Numeracy					
	Gap 90th percentile (>100 books - <100 books)			**Gap 10th percentile (>100 books - <100 books)**		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 23-25	18.40	(3.84)	*0.41*	32.27	(4.69)	*0.72*
PISA 2003	56.49	(1.96)	*0.61*	53.87	(1.43)	*0.58*

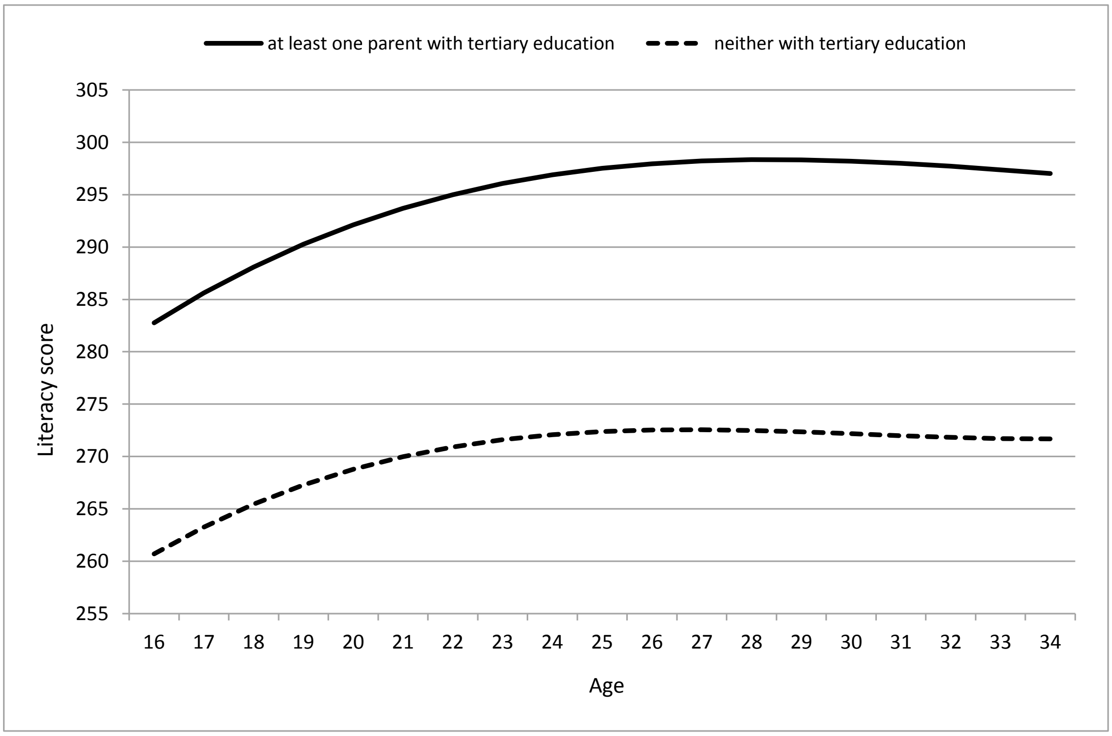
Notes: The standardised gap refers to the difference in the mean scores of individuals with more than 100 books at the age of 15 in their family home and individuals with less than 100 books divided by the pooled standard deviation of the study. The table presents the average across the 21 countries with comparable data.

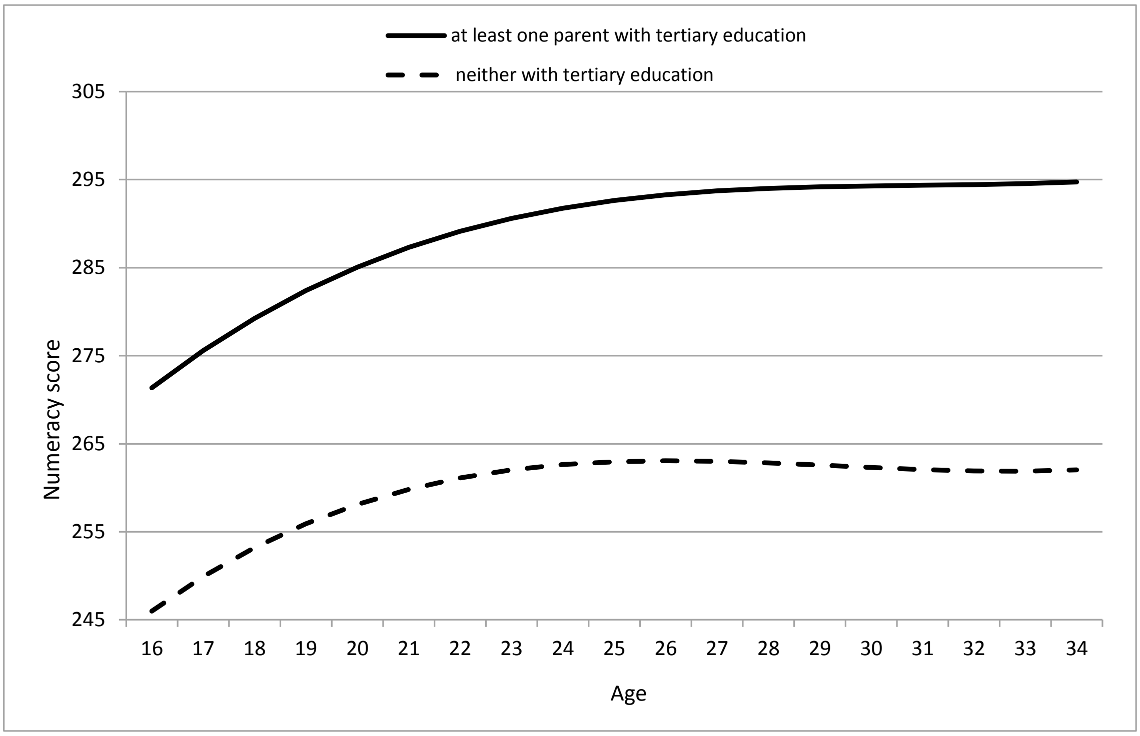
Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Relative small sample sizes in PIAAC mean that estimates of the variation at the individual country level are imprecise. However, the results suggest that socio-economic condition, as proxied by parental education and the number of books in the home, is a significant factor in shaping the opportunities that individuals with comparatively low levels of skills have.

While the comparison between PISA and PIAAC does not allow the measurement of growth in achievement, the analyses of age-proficiency profiles in literacy and numeracy by parental education and the number of books in the family home, as measured in PIAAC, provide some evidence about this. The drawback of using PIAAC data in isolation is that, rather than following the same cohort over time, we must look at different cohorts at a single point in time. We therefore have to assume that there are no cohort effects, i.e. that the age-proficiency profile is the same for people born in different years. If this assumption holds, Figures 3.3 and 3.4 suggest that, between the age of 16 and the age of 27, gaps in proficiency between individuals coming from different socio-economic backgrounds become larger, because achievement growth between the teenage years and young adulthood is steeper among socio-economically advantaged individuals. Both groups increase in skills but the growth appears to be more pronounced among high SES individuals and individuals with better educated parents. Moreover, low SES individuals and individuals with no tertiary educated parents appear to reach the age at which skills stop growing before high SES individuals and individuals with at least one tertiary educated parent.

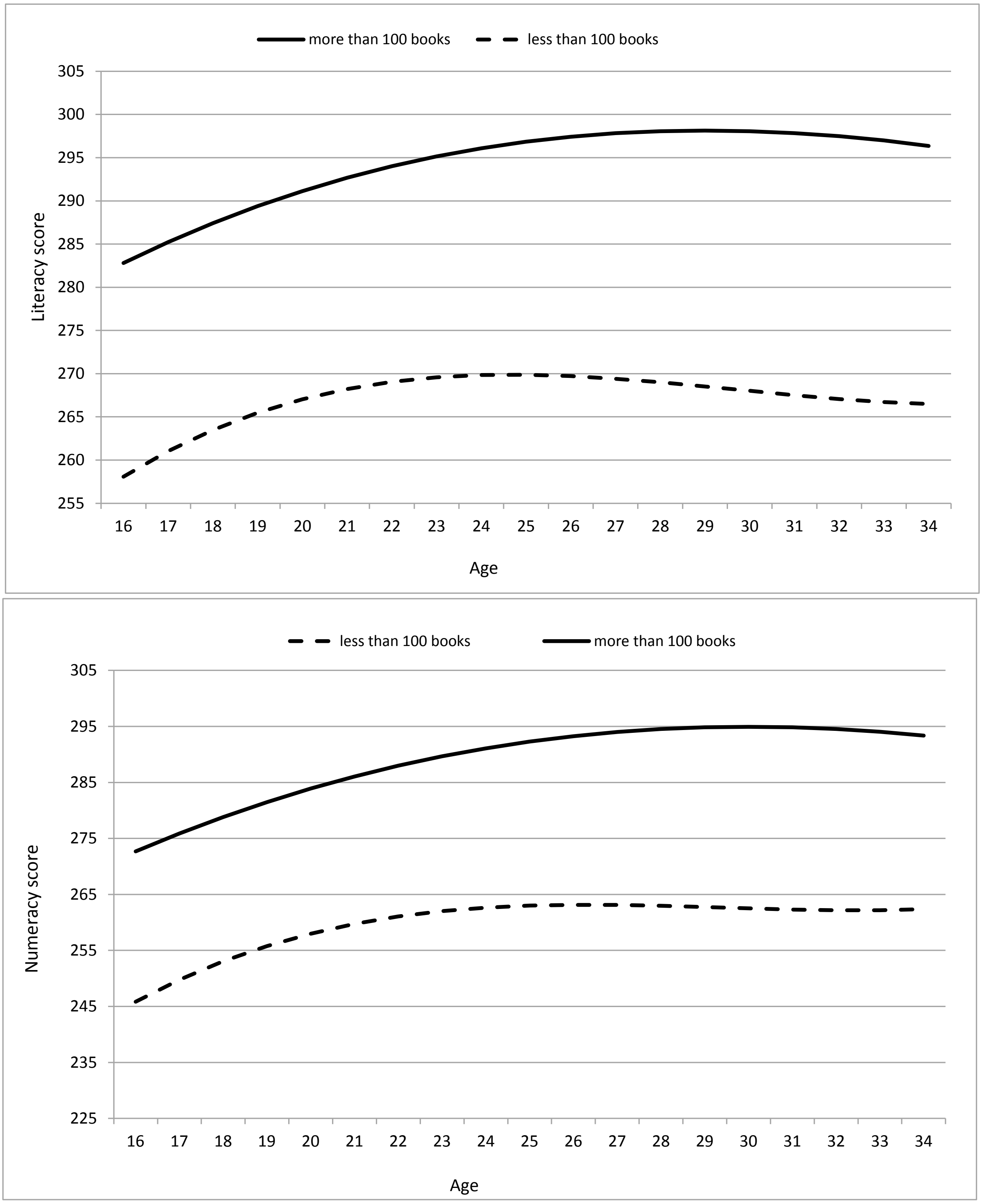
Figure 3.3 Age-by proficiency profile in literacy and numeracy in PIAAC (16-34 year-olds), by parental educational attainment





Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

Figure 3.4 Age-by proficiency profile in literacy and numeracy in PIAAC (16-34 year-olds), by number of books available at home at age 16



Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

3.3 Gender disparities

Overall, there is growing evidence of a closing of the gender gap in favour of males and of the emergence of gender gaps in favour of females in a variety of educational outcomes (Diprete and Buchman, 2013; Fortin et al., 2015). For example, in OECD countries, young women are now more likely than young men to complete secondary school, are less likely to repeat grades and now represent the majority of those who enrol into and ultimately complete tertiary degrees (OECD, 2016b; OECD, 2014c; Jacob, 2002; Goldin, Katz and Kuzienko, 2006). In fact there is evidence that females are generally awarded higher grades than males (Kimball, 1989; Snyder, Dillow, and Hoffman, 2009; OECD, 2014c; Fortin et al., 2015). However, the academic and career choices of young men and women continue to be remarkably different (OECD, 2016b; Anker, 1997; OECD, 2015a; Pope and Sydnor, 2010; Mullis et al., 2012a,b).

These choices may not only reflect the different academic strengths and weaknesses of teenage boys and girls, but may also reinforce gender disparities in these domains over time. If boys have better relative average performance in mathematics (because mathematics tends to be their strongest subject in school) and, as a result, are more likely to enrol in courses and work in jobs that require strong quantitative abilities upon completing school, they may be more likely to develop further those skills. Conversely, if girls are more likely to have a comparative advantage in reading literacy while at school, they may be more likely to make educational and career choices that will reinforce such skills, possibly at the expense of developing quantitative abilities.

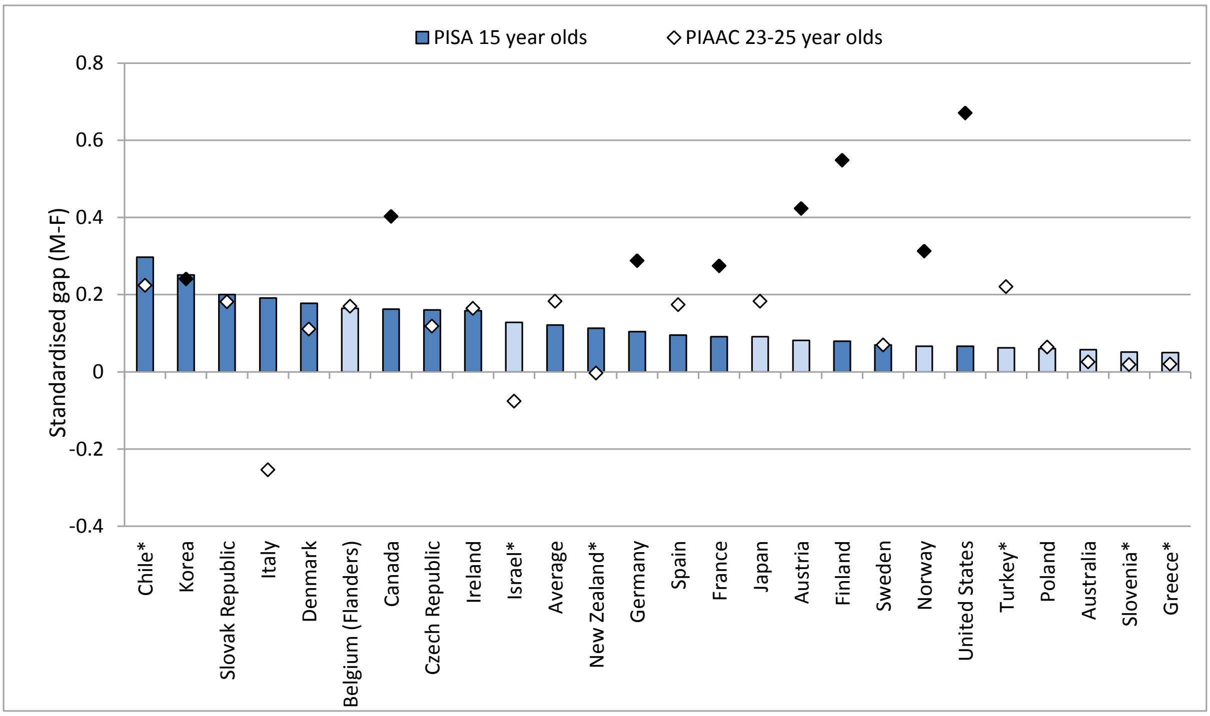
Although the existence of gender gaps in quantitative and verbal abilities has been well documented (OECD, 2015a), much less is known about how they evolve through individuals' lifetime. Better knowledge of the timing at which disparities in abilities emerge is crucial to identify the critical factors and circumstances that shape the evolution of gender gaps, and to design effective policies.

Most of the evidence on how the size of the gender gap in verbal and quantitative abilities evolves from childhood into the teenage years and then adulthood comes from Australia, Canada, the United Kingdom and the United States and focuses on the early childhood to the teenage years. This evidence suggests that gender differences in quantitative abilities widen as individuals age. For example, males and females appear to perform on a par in tests of mathematical reasoning between the ages of 4 and 10 (Spelke, 2005) but boys start to outperform girls as children complete primary school and move into secondary education (Beilstein and Wilson, 2000). Some suggest that gender differences in student's achievement in mathematics may emerge as early as kindergarten (Penner and Paret, 2008). However, the teenage years are generally identified as the point when gender gaps in quantitative abilities widen (Lindbergh et al., 2010; Machin and McNally, 2005). What is clear is that by the end of secondary school, males generally outperform females in quantitative abilities, that the difference appears to be more pronounced among high-achieving students (Machin and Pekkarinen, 2008) and that the difference tends to be more pronounced in tests that are not directly tied to a curriculum and reflect problem solving skills (Halpern, 2012). For example, young men have been obtaining higher scores in the mathematics portion of the SAT test for over two decades and the gender gap is not only statistically significant, but also quantitatively important (average standardised gap>0.3) (Halpern et al. 2007a,b).

Large scale cross-national assessments such as TIMSS, PIRLS and PISA allow the monitoring of cross-country variations in the size of the gender gap. However, they focus on specific age groups (although PIRLS and TIMSS have a grade based sampling strategy the target population is generally close to age 10) and participating students are not generally followed after they participated in the studies. There are reasons to believe that, just as countries differ in the size of the gender gap in literacy and numeracy because of social and educational factors (Guiso et al., 2008; Hyde and Mertz, 2009; Bedard and Cho, 2010), so does the evolution of such gaps.

Figure 3.5 illustrates that at age 15 the difference in proficiency of young men and women in numeracy is small in all countries considered. The average gap is 0.12 and the gap is greater than 0.20 only in Chile and Korea. On the other hand, it is smaller than 0.10 in 13 countries. However, the gender gap in numeracy is wider at age 25: on average it is 0.18, still a small gap, but it is larger than 0.30 in Austria, Canada, and Norway, and it is larger than 0.5 in Finland and the United States. Gender gaps in numeracy are small at age 15 and widen somewhat between the age of 15 and 25, with young men pulling ahead young women in quantitative abilities (Figure 3.5). However, the gender gap in numeracy remains quantitatively small in the large majority of countries.

Figure 3.5 Standardised gender gaps in numeracy in PISA 15-year-olds and PIAAC 23-25 year-olds

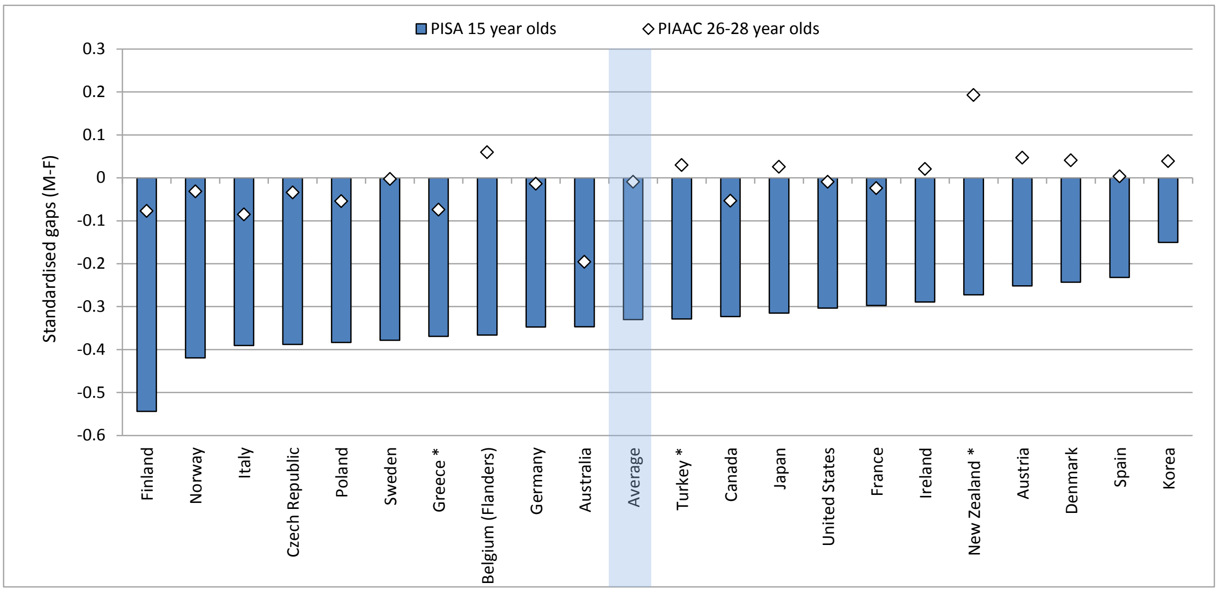


Note: The standardised gap refers to the difference in the mean scores of females – the mean scores of males divided by the pooled standard deviation. Countries are ranked in descending order of the gap in numeracy in PISA. * Next to the country name denotes PIAAC round 2 countries for which PISA 2006 data were used. Gender gaps that are statistically significant at the 5% level are marked in a darker tone.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm; OECD (2006), PISA database, www.oecd.org/pisa/pisaproducts/database-pisa2006.htm.

The gender gap in favour of females in literacy abilities follows a different trajectory with a high degree of convergence between the age of 15 and the age of 27. Figure 3.6 indicates that the gender gap is large at age 15 in the large majority of countries but it narrows and becomes small in size by age 27.

Figure 3.6 Standardised gender gaps in literacy in PISA 15-year-olds and in PIAAC 26-28 year-olds



Note: The standardised gap refers to the difference in the mean scores of females – the mean scores of males divided by the pooled standard deviation. Countries are ranked in descending order of the gap in literacy in PISA. * Next to the country name refers to PIAAC round 2 countries for which PISA 2003 data were used to identify performance at age 15. All gender gaps are statistically significant at the 5% level at age 15 and no gap is statistically significant all age 26-28.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Since it is well established that gender gaps vary across the literacy and numeracy performance distributions, we examine how the gender gap in performance varies not only for average males and females, but also for high-achieving and low-achieving males and females. We do so by calculating standardised gender gaps at the top tail of the country-specific and age-specific performance distribution (top decile of performance), at the bottom tail of the country-specific and age-specific performance distribution (bottom decile of performance).

At the age of 15, in the PISA 2000 study, the average standardised gap between boys and girls in literacy was 0.33 of a standard deviation, a moderate gap when compared to standard Cohen's d benchmarks. However the gap was as large as 0.54 in Finland and 0.42 in Norway. The gap was smaller than 0.20 in Korea. Figure 3.6 indicates that by the time 15-year-olds participating in PISA 2000 and (for PIAAC round 2 countries in PISA 2003) had transitioned into young adulthood, the gap had considerably narrowed or disappeared altogether. On average, the gender gap in literacy skills at the age of 27 was 0.01 (a negligible level) and it was no larger than 0.25 in any of the countries that were part of the analysis. New Zealand was the country with the largest estimated gap (gap=-0.19). Table 3.7 suggests that convergence occurred across the performance spectrum but was most concentrated among low performing young men: the gap at the bottom of the literacy performance distribution was as wide as 0.46 in the bottom decile but by the time the members of the cohort had reached the age of 27 the gap had effectively been closed (gap=-0.09). At the top of the distribution (90th percentile) the gap was smaller at age 15 (gap=-0.22) but had also been closed by young adulthood (gap=0.02). In numeracy, the gender gap at the top tail of the performance distribution remained stable between the age of 15 and 23-25: on average, the standardised gap at the 90th percentile of the numeracy distribution was 0.23 at age 15 and 0.23 at age 23-25. By contrast, the gap increased, but not by a large extent, at the bottom of the performance distribution, from 0.02 to 0.10. Country-specific estimates are imprecise because of the small sample size in PIAAC, but Table 3.4

suggests a high degree of heterogeneity in the extent to which the gender gap in numeracy skills evolves and whether it widens or shrinks in particular among high or low-achievers.

Table 3.3 The evolution of the gender gap in numeracy and literacy between PISA and PIAAC among high and low achievers

	Numeracy					
	Gap (M-F) 90th percentile			Gap (M-F) 10th percentile		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 23-25	10.40	(3.01)	*0.23*	4.37	(4.59)	*0.10*
PISA 2003	21.90	(1.29)	*0.23*	1.46	(1.56)	*0.02*
	Literacy					
	Gap (M-F) 90th percentile			**Gap (M-F) 10th percentile**		
	dif.	SE	Standardised gap	dif.	SE	Standardised gap
PIAAC 26-28	0.64	(3.90)	*0.02*	-3.61	(4.78)	*-0.09*
PISA 2000	-21.77	(1.41)	*-0.23*	-43.79	(2.09)	*-0.46*

Notes: Average across 21 participating countries.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Table 3.4 The evolution of the gender gap in numeracy between age 15 and 23/25 among high and low achievers

	Numeracy proficiency											
	p90 gender gap (M-F)						p10 gender gap (M-F)					
	PISA 2003			PIAAC 23-25			PISA 2003			PIAAC 23-25		
	gap	(SE)	Standardised gap	gap	(SE)	Standardised gap	gap	(SE)	Standardised gap	gap	(SE)	Standardised gap
Australia	15.73	(5.63)	0.17	4.37	(18.83)	0.10	-5.66	(5.46)	-0.06	5.83	(42.26)	0.13
Austria	21.73	(6.08)	0.23	9.50	(17.51)	0.21	-8.37	(5.67)	-0.09	20.10	(22.96)	0.45
Canada	27.17	(4.12)	0.29	20.11	(14.43)	0.45	-1.17	(3.72)	-0.01	16.81	(15.21)	0.37
Czech Republic	25.49	(4.99)	0.27	14.68	(15.27)	0.33	13.34	(7.52)	0.14	-2.95	(23.22)	-0.07
Denmark	14.83	(4.61)	0.16	13.40	(16.18)	0.30	16.19	(6.16)	0.17	-7.77	(23.12)	-0.17
Finland	17.17	(4.68)	0.18	16.50	(15.47)	0.37	-5.87	(5.46)	-0.06	27.10	(25.08)	0.60
Flanders (Belgium)	22.11	(5.78)	0.24	5.93	(13.18)	0.13	5.72	(12.77)	0.06	5.80	(21.42)	0.13
France	17.51	(6.15)	0.19	13.10	(11.54)	0.29	-7.52	(8.78)	-0.08	11.11	(16.65)	0.25
Germany	16.87	(6.48)	0.18	7.79	(15.79)	0.17	2.93	(9.21)	0.03	14.64	(21.19)	0.33
Greece*	32.79	(6.00)	0.35	4.32	(15.94)	0.10	9.91	(6.63)	0.11	-10.15	(22.74)	-0.23
Ireland	19.57	(6.82)	0.21	1.91	(26.65)	0.04	9.24	(7.36)	0.10	5.93	(25.48)	0.13
Italy	31.97	(4.80)	0.34	6.22	(13.93)	0.14	5.14	(9.49)	0.06	-25.38	(44.36)	-0.56
Japan	27.42	(8.92)	0.29	2.15	(12.63)	0.05	-8.88	(8.74)	-0.10	16.28	(14.66)	0.36
Korea	25.14	(6.60)	0.27	8.35	(13.49)	0.19	14.94	(9.57)	0.16	12.15	(16.08)	0.27
New Zealand*	24.37	(5.14)	0.26	6.61	(22.21)	0.15	2.63	(7.76)	0.03	-17.30	(23.38)	-0.38
Norway	17.14	(5.88)	0.18	23.96	(10.83)	0.53	-5.34	(7.11)	-0.06	3.68	(22.52)	0.08
Poland	21.14	(6.32)	0.23	8.22	(6.56)	0.18	-10.41	(6.68)	-0.11	-3.23	(6.72)	-0.07
Slovak Republic	23.90	(4.74)	0.26	11.65	(9.95)	0.26	9.57	(5.16)	0.10	11.45	(22.68)	0.25
Spain	19.54	(4.67)	0.21	7.34	(13.33)	0.16	-2.04	(6.62)	-0.02	14.42	(22.78)	0.32
Sweden	15.08	(5.67)	0.16	2.82	(13.83)	0.06	1.37	(5.74)	0.01	-2.00	(23.10)	-0.04
Turkey*	26.28	(11.57)	0.28	10.18	(14.27)	0.23	0.61	(7.10)	0.01	11.60	(27.86)	0.26
United States	16.25	(6.31)	0.17	39.16	(17.83)	0.87	-3.77	(6.79)	-0.04	23.26	(13.56)	0.52
Average	21.78	(1.32)	0.23	10.83	(3.31)	0.24	1.48	(1.60)	0.02	5.97	(5.12)	0.13

Note: * Denotes PIAAC round 2 countries.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

3.4 Patterns of skill development

The learning trajectories of boys and girls, of socio-economically disadvantaged and advantaged students, may either converge after they leave compulsory school and enter further education, training, the labour market and adult life more generally, or continue to diverge, possibly even more than in the school years.

Early models of skill development and acquisition focus on the childhood and teenage years and suggest that the rate at which students develop proficiency and acquire additional skills is affected by their prior skills (Dreyfus and Dreyfus, 1986). The hypothesis in these early models is that the higher the level of existing skills, the better students are able to understand new material and draw inferences from texts. According to these models, differences between high- and low-performers would tend to persist as age increases. Some even suggest that a divergence or “fan spread” pattern occurs, with the stronger getting stronger, and the weaker getting weaker over time (Catell, 1987) so that inequalities will not only persist, but widen. Learning opportunities in school may widen skills inequalities if the skill returns to educational investments is larger among those with greater initial skills endowments. Longitudinal evidence suggests that initial inequalities persist over time (Wylie and Hodgen, 2011; Bynner and Parsons, 2009).

However, little is known about the change in inequalities in skills beyond the school years. There is some evidence that information processing skills may continue to grow until the age of 25-30 (Desjardnis and Warnke, 2012; Paccagnella, 2016) and then decline. Previous empirical analyses of PIAAC data have examined in detail age-related decline in performance beyond age 30, but not enough is known about patterns of skills development and the effect they have on inequalities from the teenage years into young adulthood.

Divergence in, or a fanning out of proficiency, may occur for several reasons. Experiences of early disadvantage in skill development may lead to some groups lagging consistently behind others and, so curtail access to opportunities that are typically associated with skill acquisition. If socio-economically disadvantaged students have poorer academic results in secondary school, they may be less likely to attend selective schools, to be placed in high ability groups where learning gains are most rapid, and enter tertiary education, which is typically associated with strong development of information processing skills. They may also be less likely to work in occupations that require them to apply their information processing skills and to develop them further thanks to practice and on-the-job training opportunities.

Similarly, if girls do not develop strong quantitative abilities early on, they will be unlikely to enrol in courses and to work in jobs that require such skills, leading to few opportunities to develop quantitative abilities upon completion of school. In general, despite the fact that the schooling experience of boys and girls, and of socio-economically advantaged and disadvantaged students differ greatly – for example, because teachers tend to treat boys and girls differently in the classroom, and because socio-economically disadvantaged students tend to attend schools with more disadvantaged peers, and to have teachers who have lower expectations for them (OECD, 2015) – there is evidence that schools may offer more equal learning opportunities than do the home and social environment outside of school (Downey and Condron, 2016; Downey, von Hippel and Brih, 2004; Merry, 2013). Consequently, once individuals leave school, the differences in the opportunities to develop their skills they are offered or can access in the labour market, in further education and training, and in everyday life through non-formal and informal opportunities may lead to widening performance gaps across genders and socio-economic background.

Skills can also decline rapidly because of lack of use, a process which may start soon after the completion of compulsory schooling for young people who leave school early and have few opportunities for further development of information processing skills. Education is a powerful means through which skills are acquired, therefore failure to continue education and training beyond completion of compulsory schooling may result in overall lower levels of skill development. Socio-economically disadvantaged students are more likely to drop out of education. While there is little direct international evidence on the effect that leaving education has on the rate of skill growth/decline, there is some evidence that students' information processing skills deteriorate during the summer months when they do not attend school. The "summer slide", first identified by Heyns (1978), has been shown to vary greatly depending on the number and type of organised activities available to individuals and has been linked to inequalities in competencies between socio-economic groups that are presented with different opportunities for learning (Cooper, et al., 1996; Alexander, et al., 2007).

However, greater opportunities for low achievement cohorts relative to high achievement in post-school settings may occur if schooling environments are highly diverse in the quality of learning opportunities that they provide, for example, because of differences in the resources available to different schools, or the quality of one's peer in highly stratified systems. In these systems further education, training and labour markets may be better equipped to provide the right stimuli for those who had few opportunities at school. Convergence may also occur if groups that did not fit in the school climate and were not motivated to learn in the absence of strong extrinsic incentives, felt motivated and prone to learn in further education, training and the labour market.

In general, countries differ markedly in the extent to which the transition from schooling into further education, training, the labour market and adult life occurs. Some countries have well established orientation mechanisms that aid individuals towards the end of compulsory schooling to make choices and support them during the transition phase. Such support can take the form of additional training opportunities and mentoring schemes. Some systems offer considerable flexibility in the educational pathways that individuals can follow with the result that learners can change orientation and move between different programmes at relatively low cost (direct and/or indirect). The possibility of changing field of study and going to university after having attended vocational education and training exemplify such flexible pathways.

Explanations for socio-economic disparities

Changes in socio-economic disparities in literacy and numeracy between the age of 15 and the age of 27 may be associated with the different opportunities socio-economically advantaged and disadvantaged students have after they leave school and enter further education, training, the labour market and society. This section details factors that may lead socio-economically advantaged and disadvantaged students to have different transition pathways after compulsory school using data from PISA on educational and career expectations, as well as differences between socio-economically advantaged and disadvantaged individuals in their educational qualifications, participation in the labour market and literacy and numeracy related practices at work and in everyday life.

Studies based on PISA and other youth surveys over the past 30 years consistently find that upper secondary students tend to set ambitious educational and occupational goals and that holding higher expectations is an important drive for students to put effort in their studies and invest in skill acquisition (Marks, 2010; McDaniel, 2010; Sikora and Saha, 2007; Sikora and Saha, 2009; Croll, 2008; Goyette, 2008; Little, 1978; Reynolds et al., 2006). PISA 2006 and PISA 2009 data suggest that across OECD and partner countries and economies, a substantial share of students has ambitious education and career goals, but also that socio-economically advantaged and disadvantaged students vary greatly in their educational and occupational ambitions (OECD, 2012b).

Students' expectations about their future education and work not only reflect their academic successes and skills; they also create conditions that are conducive to academic excellence and the acquisition of skills. For example, students who expect to complete a university degree or to work in demanding jobs are more likely to choose more demanding courses and to invest greater effort in school than students who expect to complete their studies after less schooling, and with lower qualifications, or to land jobs that do not require high-level skills. Students who hold high expectations for their education and careers are more likely than those who do not to complement their school work with additional courses or related activities during their free time. Students' expectations are partly self-fulfilling prophecies, as the effort students invest to meet their expectations usually pays off. When comparing students of similar socio-economic backgrounds and academic achievement, students who expect to graduate from university are more likely to earn that degree than their peers who do not hold such expectations (Campbell, 1983; Morgan, 2005; Perna, 2000; Sewell, et al., 2003).

PISA 2000 asked students what occupation they expected to be working in by the time they were 30 years-old. Responses to this open-ended question were re-classified according to the International Standard Classification of Occupations 88 (ISCO88; International Labour Office, 1988) as well as their expectations for continuing education. Socio-economically disadvantaged students are considerably less likely than socio-economically advantaged students to expect to work as a professional or manager at the age of 30 or to work in highly status occupations.

Results presented in Table 3.5 show differences in students' career expectations between the group of students whose parents did not obtain a tertiary qualification and the group of students with at least one tertiary educated parent and between students with over 100 books in their home at the age of 15 and those with less than 100 books. On average 36% of students with no tertiary educated parent expected to work as a professional or manager by the age of 30. In contrast, 52% of students with at least a tertiary educated parent, a difference of almost 16 percentage points. Similarly, on average, across the set of countries that took part in PISA in 2000 as well as in PIAAC, only around 38% of students with less than 100 books in their home reported expecting to work as a professional or manager at the age of 30. Among students with over 100 books in their home the percentage was 52%, and the difference between the two groups corresponded to over 14 percentage points. While the socio-economic gap in expectations varies across countries, in no country did the group with less than 100 books hold more ambitious expectations than the group with more than 100 books.

Table 3.5 Disparities in students' career expectations in PISA 2000, by parental educational attainment and book availability at the age of 15

	Students' career expectations for their occupation at the age of 30											
	No parent with tertiary education				At least one parent with tertiary education				Gap (tertiary-no tertiary education)			
	Expected managerial or professional occupation		Expected social status (mean ISEI)		Expected managerial or professional occupation		Expected social status (mean ISEI)		Expected managerial or professional occupation		Expected social status (mean ISEI)	
	%	SE	Mean	SE	%	SE	Mean	SE	% point diff.	SE	mean diff.	SE
Australia	39.94	(0.02)	52.92	(0.59)	52.74	(0.01)	59.55	(0.57)	**12.80**	(0.03)	**6.63**	(0.78)
Austria	28.93	(0.01)	50.14	(0.49)	50.04	(0.02)	58.85	(0.50)	**21.11**	(0.02)	**8.70**	(0.65)
Belgium (Flanders)	35.13	(0.03)	46.23	(1.30)	55.11	(0.01)	55.03	(0.59)	**19.99**	(0.03)	**8.81**	(1.28)
Canada	45.20	(0.02)	55.64	(0.64)	56.25	(0.01)	62.21	(0.22)	**11.04**	(0.02)	**6.57**	(0.64)
Czech Republic	15.80	(0.01)	43.11	(0.48)	37.63	(0.01)	53.39	(0.50)	**21.84**	(0.01)	**10.27**	(0.54)
Germany	16.15	(0.01)	45.85	(0.50)	37.24	(0.01)	54.98	(0.43)	**21.09**	(0.02)	**9.13**	(0.66)
Denmark	30.18	(0.03)	44.95	(1.01)	50.87	(0.02)	54.94	(0.79)	**20.69**	(0.04)	**10.00**	(1.32)
Spain	40.26	(0.01)	54.92	(0.53)	59.09	(0.01)	63.30	(0.47)	**18.84**	(0.02)	**8.38**	(0.67)
Finland	38.37	(0.01)	50.15	(0.50)	56.07	(0.01)	58.88	(0.49)	**17.70**	(0.02)	**8.73**	(0.53)
France	26.64	(0.01)	51.02	(0.70)	37.91	(0.01)	57.21	(0.54)	**11.27**	(0.02)	**6.20**	(0.76)
Greece	49.40	(0.02)	55.30	(0.81)	64.13	(0.02)	62.39	(0.63)	**14.73**	(0.02)	**7.09**	(0.73)
Ireland	44.20	(0.02)	52.37	(0.67)	54.06	(0.01)	58.05	(0.43)	**9.86**	(0.02)	**5.68**	(0.70)
Italy	35.34	(0.02)	54.52	(0.51)	52.18	(0.02)	61.25	(0.55)	**16.83**	(0.02)	**6.74**	(0.66)
Korea	52.45	(0.01)	57.26	(0.58)	63.02	(0.01)	61.71	(0.46)	**10.56**	(0.02)	**4.45**	(0.63)
Norway	33.98	(0.02)	50.26	(0.66)	44.57	(0.01)	57.96	(0.47)	**10.59**	(0.02)	**7.70**	(0.79)
New Zealand	41.38	(0.02)	54.14	(0.68)	54.38	(0.01)	60.10	(0.56)	**13.00**	(0.02)	**5.96**	(0.87)
Poland	30.06	(0.03)	50.08	(1.07)	59.77	(0.02)	60.68	(0.73)	**29.71**	(0.03)	**10.60**	(1.05)
Sweden	31.58	(0.02)	51.02	(0.79)	40.49	(0.01)	56.63	(0.36)	**8.91**	(0.02)	**5.61**	(0.83)
United States	56.73	(0.04)	60.36	(1.54)	64.41	(0.01)	64.24	(0.48)	**7.67**	(0.05)	**3.89**	(1.57)
Average	36.41	(0.00)	51.59	(0.18)	52.10	(0.00)	59.02	(0.12)	**15.70**	(0.01)	**7.43**	(0.20)

Table 3.5 Disparities in students' career expectations in PISA 2000, by parental educational attainment and book availability at the age of 15 (*continued*)

	Students' career expectations for their occupation at the age of 30											
	<100 books				>100 books				Gap (>100-<100)			
	Expected		expected social		Expected		expected social		Expected managerial		expected social	
	%	SE	Mean	SE	%	SE	Mean	SE	% point diff.	SE	mean diff.	SE
Australia	42.29	(0.01)	53.84	(0.58)	51.68	(0.01)	59.34	(0.59)	**9.39**	(0.02)	**5.49**	(0.74)
Austria	27.37	(0.01)	49.34	(0.53)	44.92	(0.01)	57.03	(0.45)	**17.56**	(0.02)	**7.70**	(0.64)
Belgium (Flanders)	44.35	(0.02)	50.34	(0.72)	63.11	(0.02)	58.42	(0.68)	**18.76**	(0.02)	**8.07**	(0.87)
Canada	49.90	(0.01)	59.51	(0.29)	58.86	(0.01)	63.16	(0.25)	**8.96**	(0.01)	**3.65**	(0.29)
Czech Republic	17.64	(0.01)	43.86	(0.49)	36.98	(0.01)	53.11	(0.48)	**19.34**	(0.01)	**9.26**	(0.55)
Germany	17.24	(0.01)	46.14	(0.49)	33.99	(0.01)	54.17	(0.45)	**16.76**	(0.01)	**8.04**	(0.66)
Denmark	34.62	(0.02)	48.27	(0.78)	52.86	(0.02)	55.14	(0.72)	**18.23**	(0.03)	**6.87**	(0.89)
Spain	39.60	(0.01)	54.40	(0.56)	58.12	(0.01)	62.68	(0.47)	**18.52**	(0.02)	**8.28**	(0.70)
Finland	39.88	(0.01)	51.15	(0.53)	53.30	(0.01)	57.77	(0.53)	**13.42**	(0.02)	**6.62**	(0.66)
France	27.18	(0.01)	50.76	(0.57)	40.57	(0.01)	58.72	(0.55)	**13.39**	(0.01)	**7.97**	(0.68)
Greece	55.62	(0.02)	58.16	(0.69)	66.64	(0.02)	63.85	(0.66)	**11.01**	(0.02)	**5.69**	(0.72)
Ireland	45.33	(0.01)	53.06	(0.49)	57.39	(0.01)	59.95	(0.48)	**12.06**	(0.02)	**6.89**	(0.60)
Italy	35.39	(0.02)	54.71	(0.48)	55.17	(0.02)	62.14	(0.58)	**19.78**	(0.02)	**7.44**	(0.68)
Japan	33.76	(0.01)	53.45	(0.60)	42.10	(0.02)	56.11	(0.66)	**8.33**	(0.02)	**2.66**	(0.69)
Korea	54.26	(0.01)	57.66	(0.51)	63.37	(0.01)	62.33	(0.54)	**9.12**	(0.02)	**4.67**	(0.66)
Norway	33.66	(0.02)	51.34	(0.60)	46.24	(0.01)	58.22	(0.47)	**12.58**	(0.02)	**6.89**	(0.73)
New Zealand	43.00	(0.01)	54.88	(0.52)	52.44	(0.01)	59.15	(0.53)	**9.44**	(0.02)	**4.27**	(0.70)
Poland	37.03	(0.02)	52.58	(0.81)	64.16	(0.02)	62.37	(0.75)	**27.14**	(0.02)	**9.78**	(0.87)
Sweden	32.15	(0.01)	51.55	(0.48)	42.09	(0.01)	57.70	(0.41)	**9.94**	(0.02)	**6.14**	(0.63)
United States	58.89	(0.01)	62.03	(0.53)	67.86	(0.02)	65.36	(0.72)	**8.97**	(0.02)	**3.32**	(0.83)
Average	38.46	(0.00)	52.85	(0.13)	52.59	(0.00)	59.34	(0.13)	**14.14**	(0.00)	**6.48**	(0.16)

Notes: For each country results show the percentage of 15-year-old students in PISA 2000 who reported expecting to work in a managerial or professional occupation by the age of 30 (managerial and professional occupations were identified according to ISCO-88 major categories 1 and 2) and the average ISEI score (an index of social status for which greater numbers indicate higher social status) associated with the expected profession by whether students had at least a tertiary educated parent or not and the number of books available. Percentage point differences and mean index differences are reported in the last columns. Standard errors are reported in parentheses. Differences denoted in bold are statistically significant at the 5% level.

Source: OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm.

Table 3.6 Students' educational expectations in PISA 2003, by parental educational attainment and book availability at the age of 15

	Students' educational expectations											
	<100 books				>100 books				Gap (>100-<100)			
	Does not expect to complete secondary school		Expects to complete tertiary education		Does not expect to complete secondary school		Expects to complete tertiary education		Does not expect to complete secondary school		Expects to complete tertiary education	
	%	SE	%	SE	%	SE	%	SE	% point diff.	SE	% point diff.	SE
Australia	8.88	(0.52)	61.43	(0.97)	5.05	(0.35)	77.92	(0.74)	**-3.83**	(0.63)	**16.49**	(1.03)
Austria	41.74	(1.75)	32.33	(1.23)	17.16	(1.21)	52.18	(1.66)	**-24.58**	(2.02)	**19.85**	(1.82)
Belgium (Flanders)	13.65	(1.11)	54.96	(1.40)	5.42	(0.70)	77.48	(1.30)	**-8.23**	(0.91)	**22.52**	(1.48)
Germany	62.87	(1.56)	9.35	(0.77)	31.51	(1.76)	32.29	(1.36)	**-31.35**	(1.90)	**22.94**	(1.49)
Denmark	27.89	(1.05)	34.35	(1.22)	15.52	(0.89)	53.59	(1.16)	**-12.37**	(1.23)	**19.23**	(1.53)
Greece	11.60	(0.95)	79.19	(1.28)	3.74	(0.56)	91.54	(1.05)	**-7.86**	(0.86)	**12.35**	(1.33)
Ireland	16.08	(0.97)	57.07	(1.22)	6.00	(0.66)	81.60	(1.10)	**-10.08**	(1.15)	**24.53**	(1.40)
Italy	11.88	(1.18)	44.22	(1.66)	3.87	(0.30)	72.97	(1.14)	**-8.01**	(1.21)	**28.75**	(1.62)
Norway	38.52	(1.35)	39.44	(1.40)	20.35	(0.93)	64.49	(1.13)	**-18.17**	(1.59)	**25.05**	(1.66)
New Zealand	19.18	(0.93)	41.66	(1.16)	10.03	(0.73)	61.79	(1.22)	**-9.15**	(1.05)	**20.12**	(1.39)
Poland	41.41	(1.33)	30.40	(0.94)	13.47	(0.95)	63.80	(1.54)	**-27.94**	(1.37)	**33.40**	(1.48)
Slovak Republic	19.30	(1.53)	34.02	(1.72)	4.26	(0.56)	67.04	(1.32)	**-15.04**	(1.44)	**33.02**	(2.17)
Average	26.08	(0.35)	43.20	(0.37)	11.37	(0.26)	66.39	(0.36)	**-14.72**	(0.39)	**23.19**	(0.45)

Table 3.6 Students' educational expectations in PISA 2003, by parental educational attainment and book availability at the age of 15 (*continued*)

	Students' educational expectations											
	No tertiary educated parent				At least one parent with tertiary qualification				Gap (tertiary educated parents- non tertiary educated parents)			
	Does not expect to		Expects to		Does not expect to		Expects to		Does not expect to		Expects to complete	
	%	SE	%	SE	%	SE	%	SE	% point diff.	SE	% point diff.	SE
Australia	9.02	(0.60)	59.18	(0.89)	4.11	(0.51)	82.11	(0.68)	**-4.91**	(0.93)	**22.93**	(0.98)
Austria	38.28	(1.63)	32.40	(1.27)	21.15	(1.28)	51.80	(1.55)	**-17.13**	(1.74)	**19.40**	(1.63)
Belgium (Flanders)	12.62	(1.22)	51.01	(1.62)	5.55	(0.52)	77.72	(1.14)	**-7.08**	(1.17)	**26.71**	(1.60)
Germany	55.80	(1.58)	12.75	(0.80)	29.22	(1.82)	36.06	(1.58)	**-26.59**	(1.90)	**23.31**	(1.64)
Denmark	29.29	(1.37)	29.90	(1.41)	14.49	(0.80)	54.91	(1.19)	**-14.80**	(1.60)	**25.02**	(1.87)
Greece	11.67	(1.02)	78.82	(1.30)	5.59	(0.61)	89.75	(0.96)	**-6.08**	(0.93)	**10.93**	(1.27)
Ireland	13.57	(0.89)	60.10	(1.35)	7.51	(0.69)	80.07	(1.14)	**-6.06**	(1.08)	**19.97**	(1.61)
Italy	10.19	(0.98)	48.47	(1.41)	4.98	(0.68)	71.45	(1.36)	**-5.21**	(1.11)	**22.98**	(1.54)
Norway	39.47	(1.20)	38.02	(1.30)	18.30	(0.94)	66.68	(1.17)	**-21.16**	(1.35)	**28.66**	(1.72)
New Zealand	15.34	(1.04)	44.67	(1.42)	9.05	(0.71)	68.95	(1.25)	**-6.29**	(1.35)	**24.28**	(1.68)
Poland	36.55	(1.21)	34.80	(1.05)	7.64	(0.86)	76.58	(1.59)	**-28.91**	(1.41)	**41.78**	(1.73)
Slovak Republic	15.57	(1.41)	40.27	(1.47)	4.31	(0.65)	73.04	(1.34)	**-11.26**	(1.50)	**32.78**	(1.82)
Average	23.95	(0.35)	44.20	(0.37)	10.99	(0.26)	69.09	(0.37)	**-12.96**	(0.40)	**24.90**	(0.46)

Note: Only countries that took part in both PISA 2003 and PIAAC round 1 and administered questions on educational expectations are presented. Mean group differences that are statistically significant at the 5% level are bolded.

Source: OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

Table 3.6 illustrates socio-economic gaps in students' educational expectations. Because data are not available in PISA 2000, the data presents students responses in 2003 for the set of countries with available data. These results are in line with data presented on occupational expectations and suggest that socio-economically disadvantaged students are less likely to expect to obtain tertiary degrees and are more likely to expect not to complete secondary education than their more socio-economically advantaged peers. For example on average, across countries with available data, 26% of students who report having less than 100 books in their home at the age of 15 do not expect to complete secondary school whereas this holds true for only 11% of students who report having more than 100 books in their home (a gap of almost 15 percentage points).

Widening socio-economic disparities in literacy and numeracy skills between the age of 15 and the age of 27 may also stem from differences in participation of individuals from advantaged and disadvantaged backgrounds in post-compulsory education, further education and training and the labour market. PIAAC has in fact revealed large differences in skill acquisition related to education, training and engagement in tasks in the workplace that foster the acquisition and maintenance of literacy and numeracy. A summary measure highlighting participation (or lack of) in these opportunities for skill development is whether an individual is not in education, employment or training.

Results presented in Table 3.7 illustrate differences by parental educational attainment and by the availability of books at home at age 16 in the likelihood that youth (16-29 year-olds) in different countries are not in education, employment or training (NEET). On average, youngsters with at least one tertiary-educated parent and who reported having more than 100 books in their home at the age of 16 are more likely to be in either employment, education or training than their peers with no tertiary-educated parent or who had less than 100 books available. Across countries participating in PIAAC around 1 in 5 16-29 year-olds with at least one tertiary-educated parent is NEET, but as many as 4 in 5 16-29 year-olds without a tertiary-educated parent are in the NEET category. Similarly, 1 in 4 16-29 year-olds who reported having more than 100 books in their homes at the age of 16 is a NEET, but as many as 3 in 4 16-29 year-olds with less than 100 books were. Socio-economic disparities in NEET rates are statistically significant and quantitatively meaningful in the large majority of countries. Disparities by the number of books available to students in the teenage years are not statistically significant in Finland, Germany and Norway. Norway is the only country where disparities by parental education in NEET rates are not statistically significant.

Table 3.7 Socio-economic disparities in NEET status among 16-29 year-olds

	NEET rates among 16-29 year olds by socio-economic indicators											
	Number of books available in the home at age 16						Parental educational attainment					
	More than 100 books		Less than 100 books		Gap (Less than 100 books-More than 100 books)		At least one parent with tertiary education		Neither parent with tertiary education		Gap (Non tertiary educated parents-Tertiary educated parents)	
	%	(SE)	%	(SE)	% point dif.	(SE)	%	(SE)	%	(SE)	% point dif.	(SE)
Australia	26.35	(5.89)	73.65	(5.89)	**47.30**	(11.78)	14.63	(4.09)	85.37	(4.09)	**70.74**	(8.17)
Austria	15.51	(5.26)	84.49	(5.26)	**68.97**	(10.51)	16.08	(5.29)	83.92	(5.29)	**67.84**	(10.58)
Belgium (Flanders)	21.79	(4.78)	78.21	(4.78)	**56.43**	(9.57)	22.18	(5.35)	77.82	(5.35)	**55.64**	(10.69)
Canada	20.58	(3.15)	79.42	(3.15)	**58.85**	(6.30)	30.84	(4.13)	69.16	(4.13)	**38.32**	(8.27)
Czech Republic	36.56	(6.37)	63.44	(6.37)	**26.89**	(12.73)	2.75	(0.94)	97.25	(0.94)	**94.50**	(1.87)
Germany	46.16	(7.30)	53.84	(7.30)	7.68	(14.61)	28.50	(5.62)	71.50	(5.62)	**43.00**	(11.24)
Denmark	33.70	(7.82)	66.30	(7.82)	**32.60**	(15.64)	15.64	(5.68)	84.36	(5.68)	**68.72**	(11.36)
Spain	21.52	(3.33)	78.48	(3.33)	**56.95**	(6.66)	10.15	(2.35)	89.85	(2.35)	**79.69**	(4.70)
Finland	42.51	(6.38)	57.49	(6.38)	14.98	(12.75)	21.35	(5.03)	78.65	(5.03)	**57.30**	(10.07)
France	29.23	(3.31)	70.77	(3.31)	**41.53**	(6.62)	9.69	(1.99)	90.31	(1.99)	**80.62**	(3.98)
Greece	7.79	(2.40)	92.21	(2.40)	**84.41**	(4.80)	9.40	(2.10)	90.60	(2.10)	**81.20**	(4.20)
Ireland	17.12	(3.91)	82.88	(3.91)	**65.76**	(7.82)	12.40	(3.43)	87.60	(3.43)	**75.21**	(6.86)
Italy	15.05	(3.71)	84.95	(3.71)	**69.89**	(7.43)	3.67	(1.91)	96.33	(1.91)	**92.65**	(3.83)
Japan	14.81	(4.04)	85.19	(4.04)	**70.38**	(8.08)	36.66	(5.38)	63.34	(5.38)	**26.69**	(10.77)
Korea	31.06	(4.87)	68.94	(4.87)	**37.88**	(9.74)	16.23	(4.22)	83.77	(4.22)	**67.55**	(8.43)
Norway	56.14	(7.44)	43.86	(7.44)	-12.28	(14.88)	38.68	(8.77)	61.32	(8.77)	22.64	(17.55)
New Zealand	29.86	(4.53)	70.14	(4.53)	**40.28**	(9.06)	20.62	(4.19)	79.38	(4.19)	**58.76**	(8.39)
Poland	16.04	(2.31)	83.96	(2.31)	**67.93**	(4.61)	6.12	(1.21)	93.88	(1.21)	**87.76**	(2.43)
Slovak Republic	11.89	(2.09)	88.11	(2.09)	**76.22**	(4.18)	4.71	(1.17)	95.29	(1.17)	**90.58**	(2.35)
Sweden	33.14	(7.19)	66.86	(7.19)	**33.72**	(14.37)	26.70	(6.13)	73.30	(6.13)	**46.60**	(12.26)
United States	24.52	(5.48)	75.48	(5.48)	**50.95**	(10.96)	35.72	(6.69)	64.28	(6.69)	**28.57**	(13.39)
Average	26.25	(1.12)	73.75	(1.12)	**47.49**	(2.24)	18.22	(0.99)	81.78	(0.99)	**63.55**	(1.99)

Notes: Results illustrate the percentage of 16-29 year-olds who were not in education, employment or training (NEET), by country, by parental educational attainment and the number of books available in their homes at the age of 16. Percentage point differences that are denoted in bold are statistically significant at least at the 5% level.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

Explanations for gender gaps

Results presented in the previous section suggest that the evolution of performance differences between young men and women tends to favour young men both in literacy (where they lag behind at the age of 15 and close the gender gap by age 27) and numeracy (where they slightly outperform females at age 15 and increase their advantage by age 24). While the widening of gender gaps in numeracy is not surprising, given highly gender specific patterns of educational participation (particularly by field of study) and occupational choices, what is noteworthy is to note that the gap is small in size and remains small after students leave compulsory schooling. The fact that the gap is quantitatively small does not mean that it is not meaningful, since small gaps in skills that are valued in the labour market and, in particular, in skills that are valued in "winners take all markets" can have large consequences on gender gaps in wages. At the same time, the narrowing of the gender gap in literacy is both unexpected and quantitatively significant. Women's greater participation in tertiary education and the segregation by gender in fields of study and occupation, with women over-represented among those who choose literature and humanities courses would have been consistent with widening gaps in literacy.

This section attempts to identify possible reasons that could explain such patterns by examining if these are related to the differences between the PISA and PIAAC assessments, both in terms of test questions, structure and administration, or by the level of engagement in literacy and numeracy related practices among students at the age of 15 and young adults. Research suggests that gender differences in literacy proficiency could stem from a combination of cognitive, motivational and behavioural components (Ruble, Martin and Berenbaum, 2006). Changes in gender specific motivation during the PISA and PIAAC tests as well as cognitive and behavioural developments of young men and women from the age of 15 to the age of 27 could contribute to a relative advantage for males in PIAAC, compared to PISA.

Differences between the PISA and PIAAC assessments

The first factor affecting motivation and test-taking abilities may be that PIAAC was primarily computer-based while PISA was paper-based. The second is that PISA contained a larger pool of items in areas where women tend to perform better than men, while PIAAC contained a larger pool of items in areas where men tend to perform better. There is evidence that women may perform less well than men in problem solving tasks on digital technologies, may have poorer navigation skills and may be less interested in ICT-related skills than men (Zhou, 2014). Studies examining gender gaps in reading proficiency have in fact identified marked differences depending on the characteristics and requirements of the texts used in assessment tasks. Girls tend to do better when assessments require constructed responses and worse on tests that require greater visual-spatial ability while boys are more responsive to text content and do better in multiple-choice assessments (Castelli, Colazzo and Molinari, 1998; Lafontaine and Monseur, 2009; Oakhill and Petrides, 2007; Rosen, 2001; Schwabe, McElvany and Trendtel, 2015). Men also generally outperform women on tasks that require a greater amount of abstract information processing (Halpern and LaMay, 2000), and those that require the ability to transform a visual-spatial image in working memory, and to generate and manipulate information in a mental representation.

We examine the role differences in the two test play in shaping the gender gap through a number of robustness checks. None is perfect, but each can shed some light on the mechanisms behind observed patterns.

First, we compare the gender gap in literacy and numeracy in PIAAC among 26-28 year-olds and 16-18 year-olds. All individuals surveyed in PIAAC sat the same test in 2011/12 so the gender gap between these two groups is directly comparable. However, these two age groups belong to different cohorts and therefore we cannot control for cohort specific patterns influencing shifts in gender gaps. PISA data revealed that in the large majority of countries the gender gap in literacy and numeracy has not changed significantly between 2000 and 2012, supporting the assumption that cohort comparisons can be informative.

Figure 3.7 illustrates the estimated average literacy and numeracy skills of males and females in PIAAC from age 16 to age 34 in PIAAC countries that were also part of PISA. The gender gap in numeracy in favour of males appears to widen between age 15 and age 27. More specifically, Figure 3.7 indicates that both males and females appear to become more proficient as they age, but achievement growth appears to be steeper among males than among females and to grow for a longer time. Given evidence from PISA that gender gaps in numeracy did not change much between 2003 and 2012 (OECD, 2014c), the results present in Figure 3.7 strengthen evidence derived from comparisons of PISA and PIAAC on widening gaps in numeracy and suggest that such effect may be driven by a more pronounced growth in proficiency among males than among females.

However, Figure 3.7 indicates that, contrary to analyses based on comparisons between PISA and PIAAC, the gender gap in literacy, as measured in PIAAC, is already small at age 16 and remains relatively constant between the age of 16 and 27. If anything, women's literacy skills appear to start declining at an earlier age than men's and the decline appears to be steeper.

Figure 3.7 Average age-proficiency profiles in PIAAC (16-34 age group), by gender

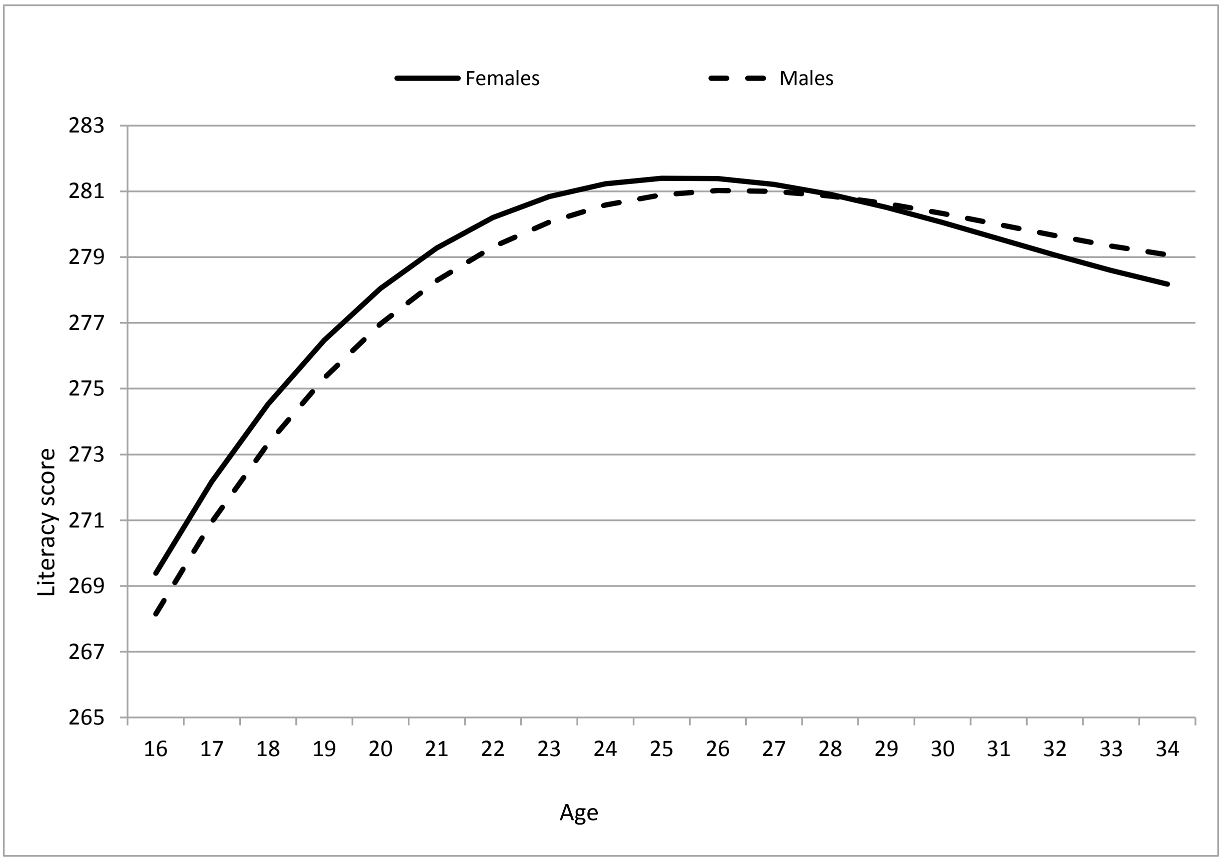
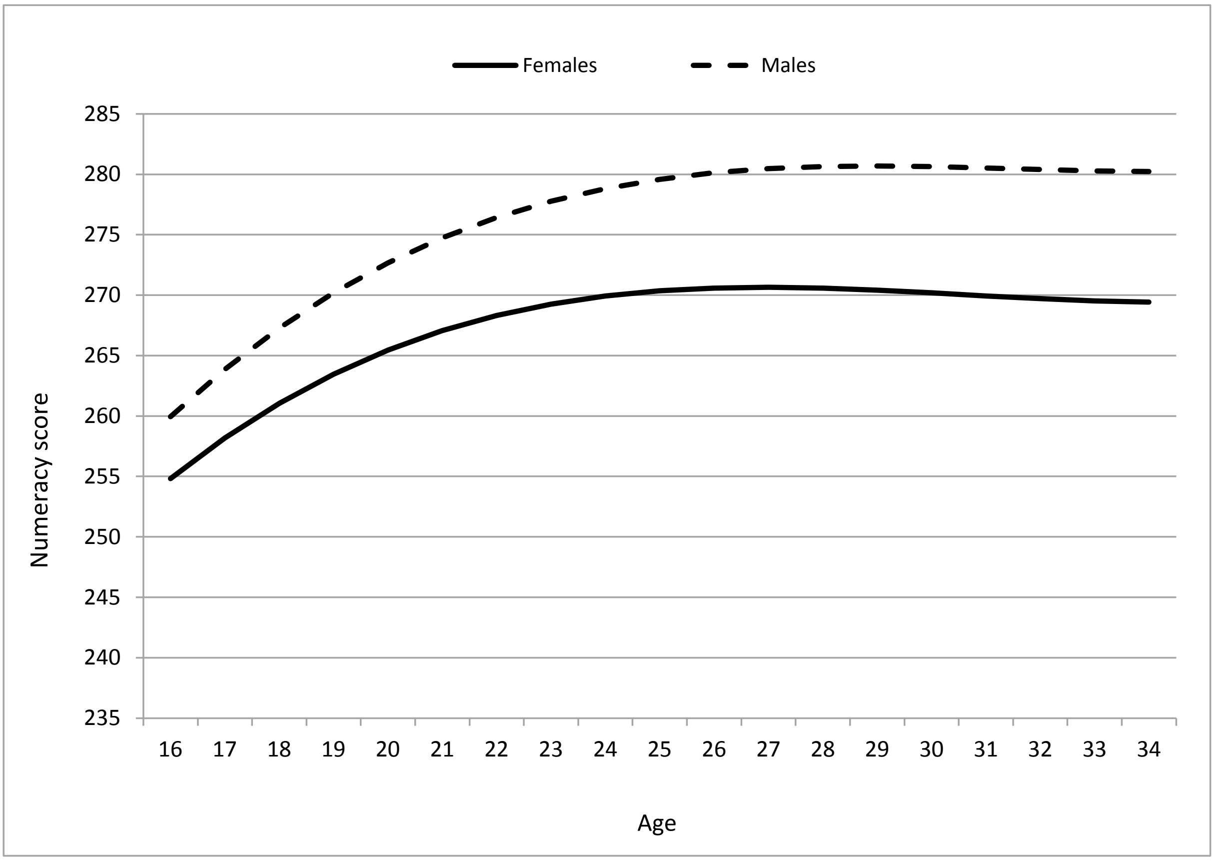


Figure 3.7 Average age-proficiency profiles in PIAAC (16-34 age group), by gender (*continued*)



Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

In the following sections we analyse two potential reasons for the observed differences in the gender gap in literacy between PISA and PIAAC: administration mode (computer vs. paper) and item coverage.

The role of computer-based delivery

A possible reason for the observed difference between the results highlighted in Figures 3.5, 3.6 and 3.7 lies in the fact that PISA was delivered as a paper-and-pencil assessment in 2000, while PIAAC was computer-based. We compare the gender gap in literacy and numeracy for the 16-18 year-old age group and the 26-28 age group in the six countries that took part in ALL in 2003/2006, PIAAC in 2011/12 and PISA in 2003 and 2012. Under the assumption of absence of strong trends over the period considered in gender gaps in proficiency in literacy, this comparison allows us to identify potential differences in the size of the gender gap across paper-based assessments (PISA and ALL) and computer-based assessments (PIAAC).

Figure 3.8 compares standardised gender gaps in literacy and numeracy for the four countries with comparable information in PISA 2000, ALL and PIAAC: Canada, Italy, Norway and the United States. Results suggest that computer-based delivery is unlikely to be an explanation for the differences in gender gaps observed PISA and PIAAC. Although results have to be considered with care given the small sample size in ALL and PIAAC for the relevant age group, estimates suggest that the gap in PISA is of medium size and is estimated precisely, while the gender gap in literacy in ALL and PIAAC is quantitatively small (point estimates) and, because it cannot be precisely estimated, it cannot be considered to be different from

0 (although it is statistically different from the estimated gap in PISA). In numeracy, results appear to indicate that the gender gap widens between the age of 15 and 26-28.

Figure 3.8 Standardised gender gaps in literacy and numeracy in PISA 2000, ALL and PIAAC, average across Canada, Italy, Norway and the United States

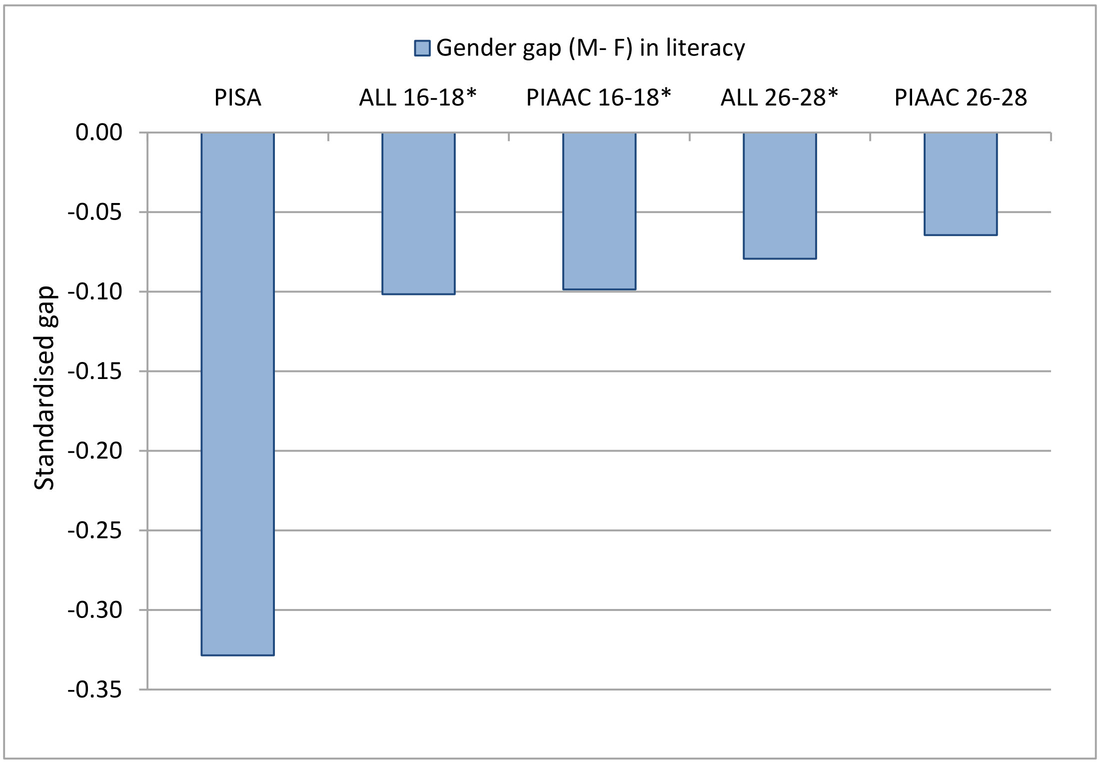
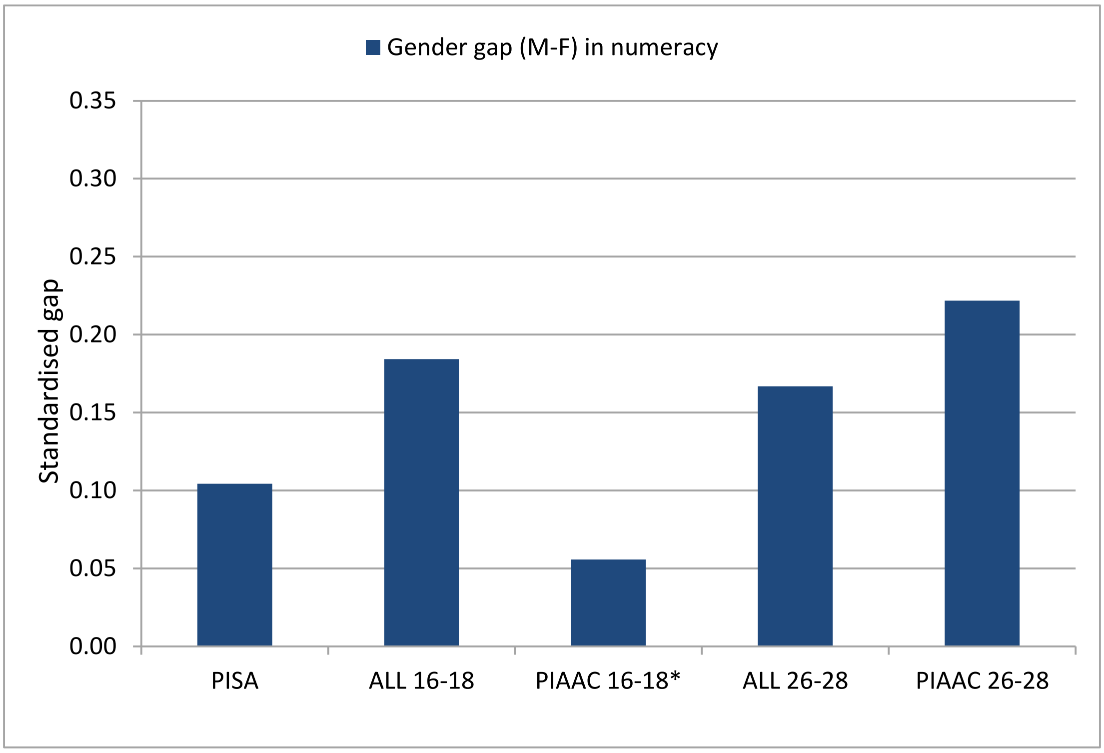


Figure 3.8 Standardised gender gaps in literacy and numeracy in PISA 2000, ALL and PIAAC, average across Canada, Italy, Norway and the United States (*continued*)



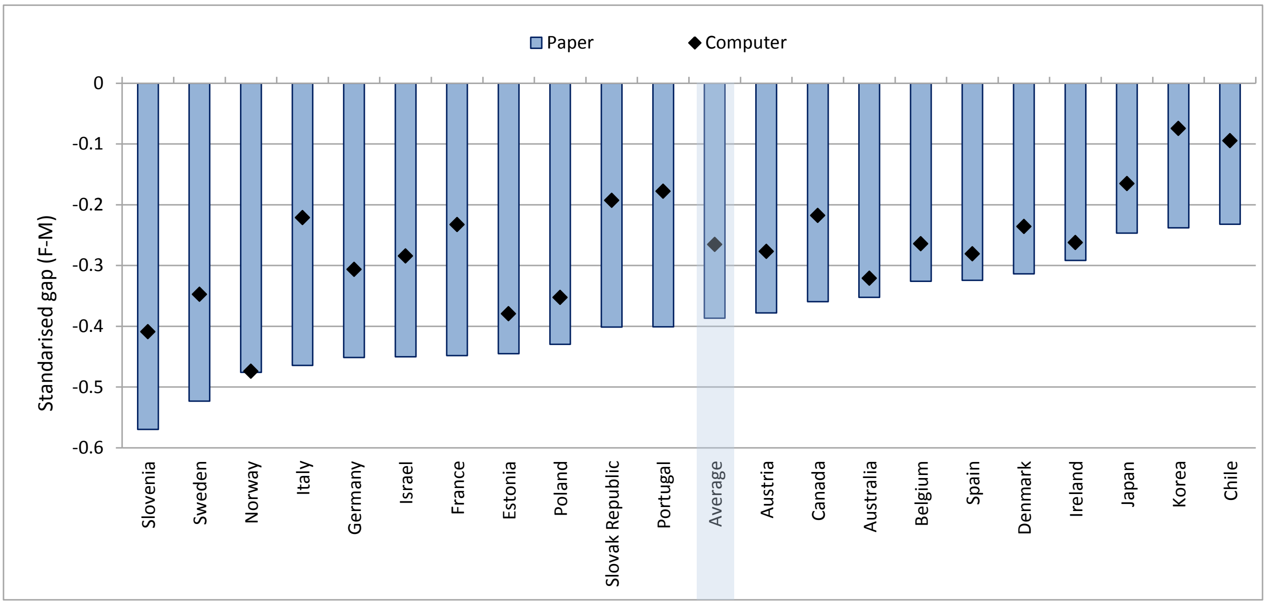
Notes: Estimates represent the average standardised gap in literacy (top panel) and numeracy (bottom panel) across the following four countries: Canada, Italy, Norway and the United States. Age groups considered in ALL and PIAAC are 16-18 year-olds and 26-28 year-olds. * Next to the survey denotes an estimated gap that is *not* statistically different from 0 (at the 5% level).

Source: Statistics Canada (2003), International Adult Literacy and Skills Survey (ALL) Canada, www5.statcan.gc.ca/olc-cel/olc.action?ObjId=89M0022X&ObjType=2&lang=en&limit=0 (accessed 1 March 2017); OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/; OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm.

In order to examine in depth the role that computer delivery may play in shaping the gender gap in literacy and numeracy, we also use data from PISA 2012 to identify gender gaps in literacy in a computer-based and a paper-based assessment. In 2012 a subset of countries participating in PISA administered, on top of the core instruments, an additional optional assessment in computer-based reading. Since the same students sat both the paper-based core instruments and the optional computer-based tests, we can examine within student differences in performance in the two tests to identify the role played by technology in shaping the gender gap in the two domains. It is important to note that the PISA computer-based and paper-based literacy assessments differ along two dimensions: mode of administration and skills tested. Computer delivery constitutes a tool that students have to use to solve test questions. Gender differences in familiarity with computers, and motivation to use computers may therefore lead to different test results of boys and girls when sitting a computer-based and a paper-based assessment. However, computer delivery also allowed examination of students' proficiency in digital reading, which requires a unique set of skills when compared to print reading. Gender differences in such skills may contribute to different test results of boys and girls in digital and print reading performance. A mode-effect study was conducted in preparation of the transition of PISA from a paper-and-pencil test to a computer-delivered test. Results from the study revealed that, at the international level, mode of delivery was not associated with performance differences across boys and girls (OECD, 2016c).

Results presented in Figure 3.9 illustrate that at age 15 compared to girls, boys tend to underperform in both the paper-based and in the computer-based reading assessments. However, their underachievement is considerably smaller in the computer-based reading assessment. In the large majority of countries, boys' underachievement relative to girls in the PISA computer-based assessment remains considerably wider than the estimated gender gap observed in the same set of countries in the youngest age group tested in PIAAC. This suggests that while differences between computer-based and paper-based testing may play a role in explaining the large convergence in literacy observed when comparing the birth-cohort that sat the PISA test in 2000 and the PIAAC test in 2011/12, other factors are most likely to contribute to the vanishing performance gap in literacy between males and females.

Figure 3.9 The gender gap in literacy in the computer and paper administered test, PISA 2012



Notes: Only countries that took part in PISA 2012 and the first round of PIAAC were considered in the analysis. Countries are sorted in descending order to the standardised gap in literacy (females-males) in the paper-based version of the assessment. All gender gaps are statistically significant at the 5% level.

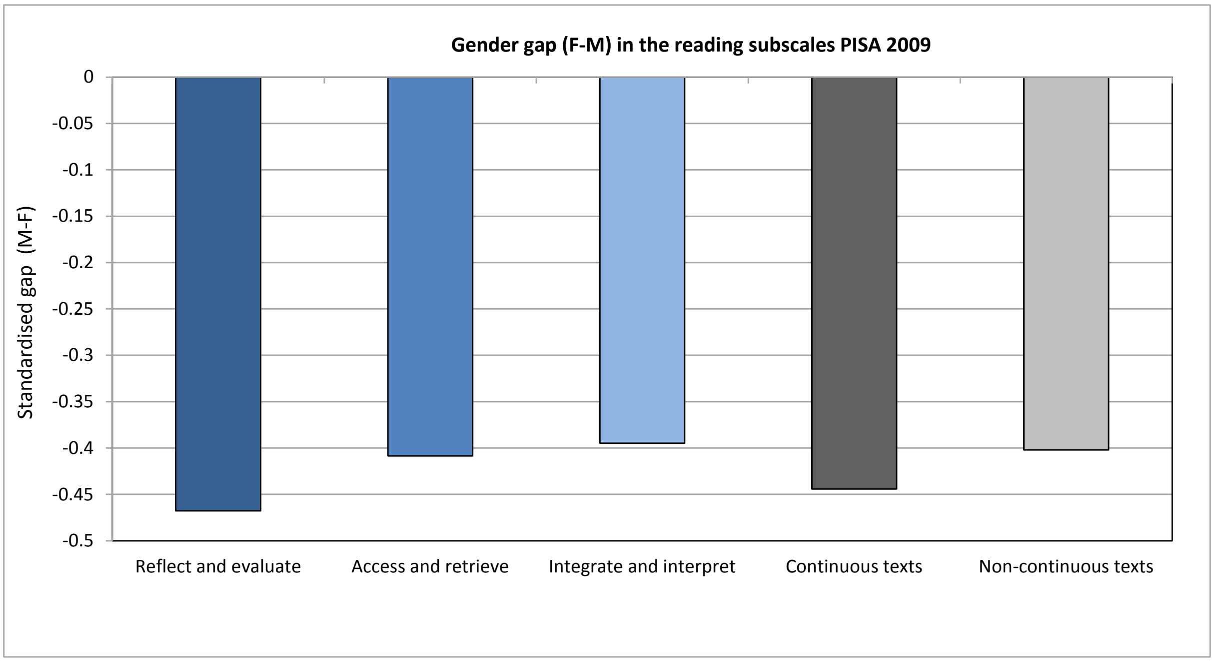
Source: OECD (2012c), PISA database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm.

The role of assessment coverage

In order to examine the role played by assessment coverage, we compare the gender gap in the PIAAC literacy assessment with the gender gap in the different reading subscales in PISA 2000. This comparison allows us to evaluate whether gender gaps observed among reading subscales in PISA are similar to gender gaps observed for PIAAC literacy items that share similar item characteristics. As discussed in section 2 the PIAAC item pool contains a considerably larger proportion of items that require individuals to access and identify information (55% of all the items) than of items that require individuals to integrate and interpret information (29% of all the items). By contrast, in PISA 2000, only around 30% of the literacy items required individuals to access and identify information and, as many as 50%, required individuals to integrate and interpret information. Similarly, as little as 24% of literacy items in PIAAC involved continuous texts compared to 63% of literacy items in PISA 2000 whereas only 2% of items in PIAAC involved non-continuous items compared to 37% of literacy items in PISA 2000.

Figure 3.10 illustrates how the gender gap in reading differs depending on the specific cognitive processes assessed and across the different text formats used. As expected, the gender gap in favour of girls tends to be smaller in non-continuous texts and wider in continuous texts, but at the same time the gap is medium-sized across both text types. Similarly, while the gender gap is considerably wider when boys and girls are required to reflect upon and evaluate written texts than when they are required to integrate and interpret texts, the standardised gap is above 0.3 for all subscales examined.

Figure 3.10 Average standardised gender gaps, by reading subscales in PISA 2009



Note: Average across 23 countries with common data in PISA 2009 and PIAAC cycle 1.

Source: OECD (2009), PISA database, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

No literacy subscales were developed for PIAAC, however, it is possible to compare the performance of males and females in PIAAC on items that differ with respect to the cognitive processes required to solve the item and the type of text format involved. Table 3.8 illustrates the average percentage of correct responses across the set of PIAAC round 1 countries that took part in PISA 2000 of young men and women between 25 and 29 (to have a broad enough sample size to obtain reliable estimates) by item characteristic. Table 3.8 reveals that there is no difference between men and women in their performance across any of the cognitive processes. However, across the different text formats, men were more likely than women to correctly answer the literacy item using a non-continuous text while they were less likely than women to correctly answer literacy items using a continuous text format. These results suggest that differences in assessment focus and in the item pool are very unlikely to be responsible for the convergence in literacy gaps between PISA 2000 and PIAAC.

Table 3.8 Gender differences in the percentage of literacy items answered correctly among 25-29 year-olds in PIAAC, by item characteristics

	Men		**Women**		**Gender Gap (Men - Women)**
	%	S.E	%	S.E	% point Dif.
Percentage of literacy items answered correctly	56.70%	(0.005)	56.97%	(0.004)	-0.26%
Strategy					
Evaluate and reflect	46.54%	(0.008)	46.96%	(0.007)	-0.42%
Integrate and interpret	51.28%	(0.006)	51.04%	(0.005)	0.24%
Access and identify	61.66%	(0.005)	61.81%	(0.005)	-0.16%
Format					
Multiple	52.05%	(0.008)	50.86%	(0.007)	1.19%
Continuous*	59.10%	(0.006)	61.14%	(0.005)	**-2.03%**
Mixed	57.81%	(0.005)	57.93%	(0.005)	-0.12%
Non-continuous*	58.23%	(0.012)	46.41%	(0.011)	**11.83%**

Note: Average across countries participating in both PISA 2000 and PIAAC cycle 1. Sample restricted to PIAAC participants between the ages of 25 and 29. * Next to the country round indicates items for which the gender gap is statistically significant.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

Overall, the analyses presented in this section suggest that while test design and characteristics play a role in shaping the size of the gender gap in literacy, this effect does not fully account for the strong convergence in the gender gap in literacy detailed in Figure 3.5. Unfortunately, it is not possible to identify the role played by administration procedures on the specific test-taking motivation males and females have: it is possible that a school-based assessment may be less conducive to males' motivation to engage with the test than a household-based assessment. Gender differences in competency related practices, educational and career pathways.

The previous section identifies a range of test characteristics and administration procedures which may explain why males appear to catch up with females in literacy after the end of compulsory schooling. In this section we examine a range of factors which may promote differences in skill development among males and females between the age of 15 and the age of 27 in both literacy and numeracy. The first factor which may contribute to the observed patterns in the evolution of gender gaps is the different opportunity to engage in literacy and numeracy related practices at the age of 15 and in the years following the end of compulsory schooling.

Formal education is one of the main mechanisms through which individuals learn. In primary and lower secondary schools boys and girls tend to attend the same schools and, even in the case when they attend single-sex schools they are generally exposed to similar curricula (OECD, 2015). However, by the time students move into upper secondary schooling, and certainly once they enrol in post-secondary education, the likelihood that boys and girls will be differentially exposed to course content designed to foster the acquisition of quantitative abilities increases. At the secondary school level, tracking and streaming practices lead to the over-representation of girls in academic tracks and, conversely, the over-representation of boys in vocational tracks that have a technical orientation because girls tend to have higher grades overall (OECD, 2016b). Moreover, at the upper secondary level, many education systems offer students the possibility of enrolling in programmes that focus on particular subjects or they allow students attending the same school to choose courses based on preferences and prior ability. Enrolment in technical and vocational tracks and teacher and parental expectations that teenagers read materials that hold little appeal among teenage boys may unintentionally have a negative effect on boys' literacy skills at this age group.

Beyond compulsory schooling, girls appear to be less likely than boys of similar mathematics ability to opt for educational pathways that contain a lot of mathematics rather than a lot of reading. Possible reasons include: girls' lower levels of mathematics self-efficacy and self-concept, orientation and guidance on the part of teachers and parents that follows gendered notions of female and male fields of studies and the fact that when boys and girls perform on a par in mathematics, girls generally outperform their male peers in reading and humanities subjects (OECD, 2015). Similarly, in most countries, university courses and pathways are differentiated along academic fields of study or disciplines (Clark, 1986). Four broad academic areas can be identified: natural sciences, humanities and social sciences, science-based professions and social professions. These, in turn, are aligned to the reading/mathematics opposition (Becher, 1994). The differentiation between mathematics and language/humanities continues in the labour market: sub-major units in the International Standard Classification of Occupations of professionals are drawn along the mathematics/humanities dichotomy (ILO, 2012). Girls are considerably more likely to enrol in courses at the post-secondary level in the humanities and social sciences, while boys are over-represented in the physical sciences, engineering and mathematics, the so called STEM fields (science, technology, engineering and mathematics) (OECD, 2016b).

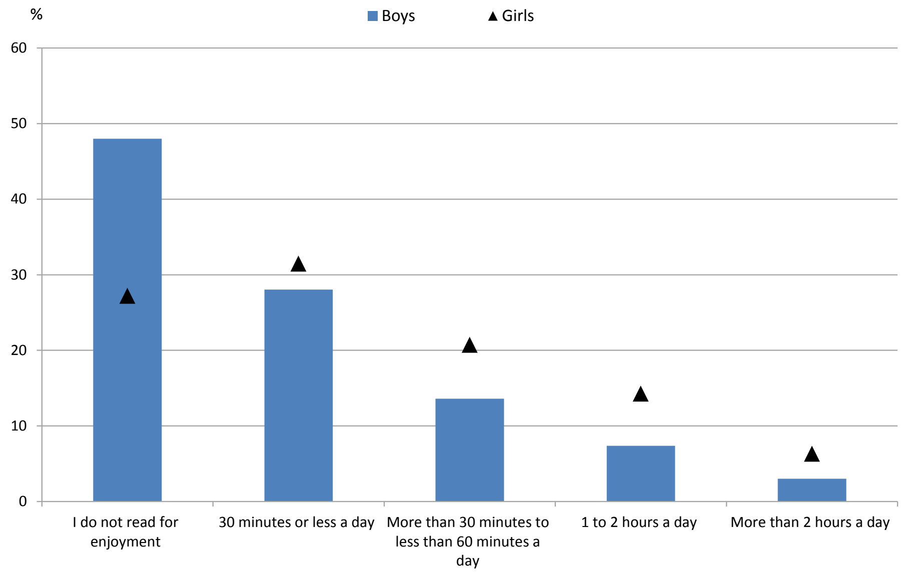
Boys' literacy skills may however benefit from the transition into further education, training, the labour market and adult life more generally. In young adulthood, given the possibility of being able to choose their own reading materials, men are more likely to engage in reading activities in everyday life and to do so while at work. Differences in the occupations men and women work in, may also determine that women work in jobs that do not stimulate skill acquisition while men do. Moreover, differences in the distribution of labour within the household across genders may mean that women will find little time and motivation to engage extensively in competency enhancing practices (Miranda, 2011).

PISA and PIAAC data cannot be used to identify to what extent differences in engagement between men and women in literacy and numeracy related practices, as well as differences in their educational and labour market trajectories shape the evolution of gender gaps in literacy and numeracy from age 15 to age 27. However, these data provide information on engagement in literacy and numeracy practices by gender, making it possible to examine whether a change in men and women's engagement in literacy and numeracy practices between the age of 15 and the age of 27 is consistent with changes in the gender gap observed, given the strong positive association between practice and proficiency.

Figures 3.11, 3.12 and 3.13 and Tables 3.9 and 3.10 map gender differences observed in PISA in reading related practices at the age of 15 (note that since no such data were available in PISA 2000 and reading for practices are based on PISA 2009 data). Figures 3.14, 3.15 and 3.16 map gender differences observed in PIAAC in literacy and numeracy practices for young adults between the age of 25 and 29.

At the age of 15 girls appear to be significantly more likely to read for enjoyment than boys. Across countries that took part in PISA 2000 and PIAAC, almost one in two boys does not read for enjoyment, while only a third of girls does not. Girls are also more likely to read more complex materials at the age of 15 than boys, such as fiction books. However, boys are more likely than girls to read comic books and newspapers than girls. Few differences on the other hand emerge with respect to gender differences at age 15 in online reading behaviours.

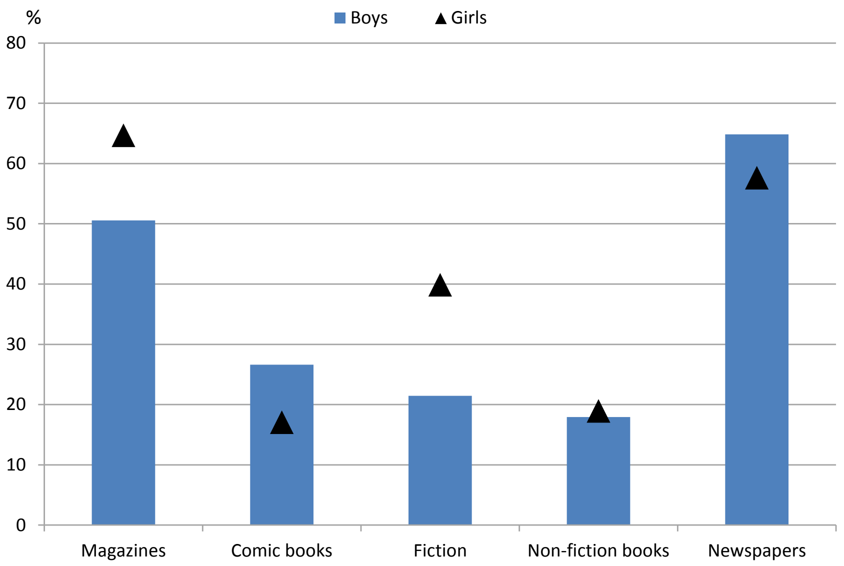
Figure 3.11 Percentage of students, by time spent reading for enjoyment in PISA 2009



Notes: All gender gaps are statistically significant. The OECD average is calculated on the 28 countries that took part in PISA 2009 and were part of either PIAAC round 1 or PIAAC round 2.

Source: OECD (2009), *PISA database*, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

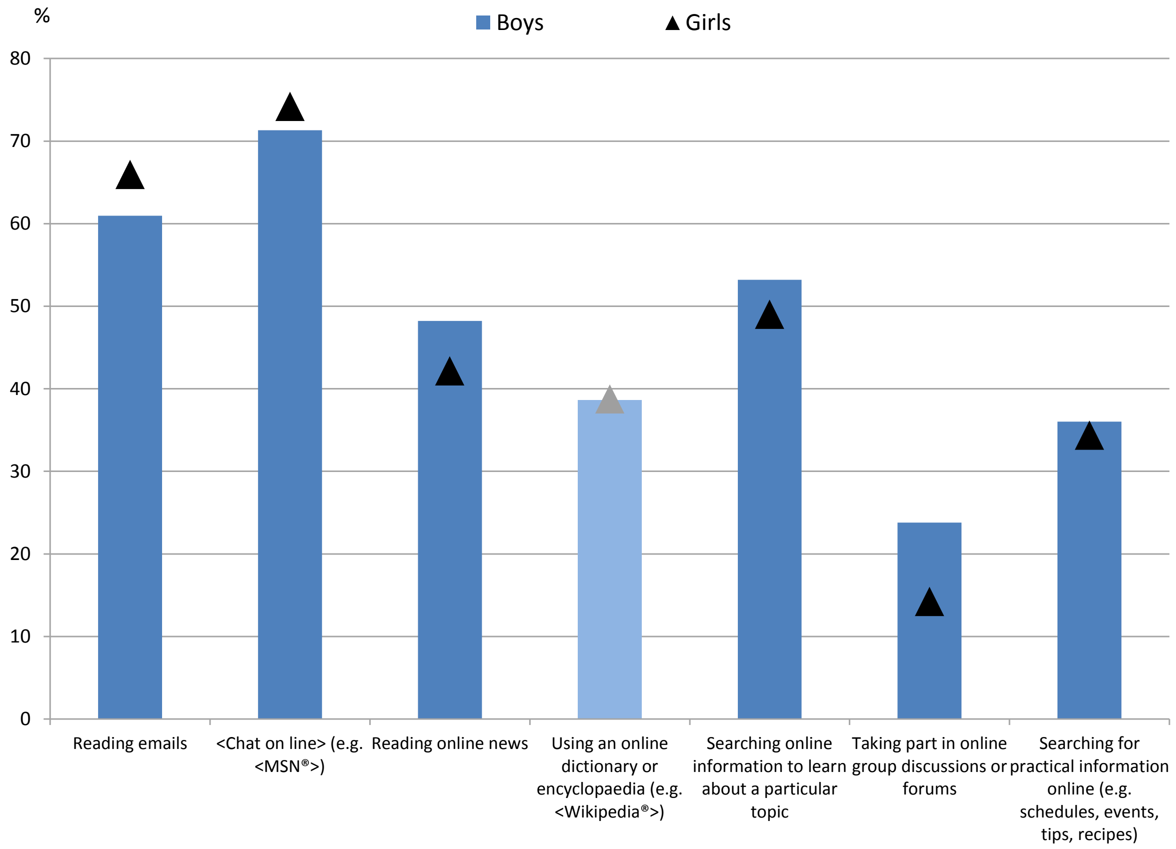
Figure 3.12 Percentage of boys and girls who reported that they read the following materials because they want to "several times a month" or "several times a week" in PISA 2009



Notes: All gender differences are statistically significant at the 5% level. The OECD average is calculated on the 28 countries that took part in PISA 2009 and were part of either PIAAC round 1 or PIAAC round 2.

Source: OECD (2009), *PISA database*, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

Figure 3.13 Percentage of boys and girls who reported that they were involved in online reading activities "several times a week" or "several times a day" in PISA 2009



Notes: Statistically significant gender gaps (5% level) are denoted in a dark shade. The OECD average is calculated on the 28 countries that took part in PISA 2009 and were part of either PIAAC round 1 or PIAAC round 2.

Source: OECD (2009), *PISA database*, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

Table 3.9 Gender disparities in reading tasks for school in PISA 2009

Texts	Percentage of boys and girls who report reading for school (either in the classroom or as homework) the following texts more than twice a week					
	Girls		Boys		Gender gap	
	%	SE	%	SE	Dif.	SE
Information	55.9	(0.21)	52.8	(0.21)	**-3.1**	(0.27)
Poetry	46.8	(0.25)	40.1	(0.22)	**-6.6**	(0.28)
Diagrams	55.2	(0.21)	54.0	(0.20)	**-1.1**	(0.27)
Fiction	67.1	(0.20)	57.4	(0.21)	**-9.7**	(0.26)
Newspapers	50.5	(0.22)	47.8	(0.20)	**-2.7**	(0.27)
Manuals	28.4	(0.19)	36.0	(0.21)	**7.7**	(0.27)
Tables	59.1	(0.21)	59.8	(0.20)	**0.6**	(0.26)
Advertising	42.0	(0.21)	41.1	(0.20)	**-0.9**	(0.26)

Notes: Average across the 33 OECD countries that took part in PISA in 2009 and in PIAAC cycles 1 and 2. Bold signals statistical significance at the 5% level. Country-specific results are available from the author upon request.

Source: OECD (2009), *PISA database*, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

Table 3.10 Gender disparities in reading tasks for school in PISA 2009

Tasks	Percentage of boys and girls who report engaging for school (either in the classroom or as homework) in the following reading tasks more than twice a week					
	Girls		Boys		Gender gap	
	%	SE	%	SE	Dif.	SE
Find information	60.9	(0.21)	60.5	(0.19)	-0.4	(0.26)
Explain cause	65.7	(0.20)	60.7	(0.19)	**-5.0**	(0.26)
Explain behaviour	64.8	(0.21)	57.4	(0.20)	**-7.5**	(0.27)
Learn about writer	41.7	(0.23)	38.4	(0.21)	**-3.2**	(0.26)
Explain purpose	65.9	(0.21)	60.1	(0.20)	**-5.8**	(0.27)
Memorise text	27.3	(0.19)	27.4	(0.19)	0.1	(0.23)
Learn history	34.9	(0.20)	35.0	(0.19)	0.0	(0.26)
Describe table	34.8	(0.20)	39.0	(0.20)	**4.2**	(0.26)
Explain connection	39.0	(0.20)	41.2	(0.19)	**2.2**	(0.26)

Notes: Average across the 33 OECD countries that took part in PISA in 2009 and in PIAAC cycles 1 and 2. Bold signals statistical significance at the 5% level. Country-specific results are available from the author upon request.

Source: OECD (2009), *PISA database*, www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm.

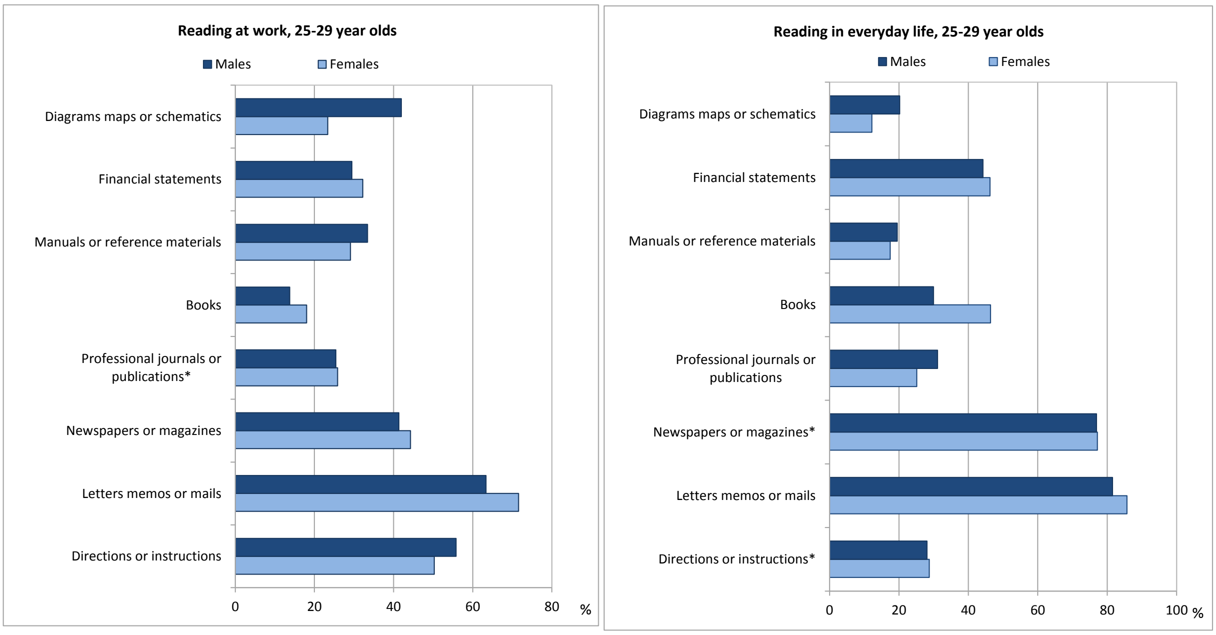
Tables 3.9 and 3.10 suggest that boys and girls not only differ in the extent to which they read for enjoyment and the types of reading material they access in their free time, but also in the extent to which they read for school purposes (whether at school or doing homework). Girls are considerably more likely than boys to read poetry and fiction, while boys are more likely to read manuals. Differences also emerge when examining differences between boys and girls in the types of tasks boys and girls do for school: girls are more likely than boys to engage in reading activities that require them to explain behaviour, to explain causes, and explain purpose. On the other hand boys are more likely than girls to engage in reading to describe tables.

Some of these differences reflect differences in motivation to meet curricular requirements and complete school assignments: girls in fact spend more time doing homework and, on average, they tend to value what is taught in class more than boys (OECD, 2015). These differences however also reflect gender differences in academic selection and the fact that girls are more likely to be in academic tracks while boys are over-represented in vocational tracks. It may also reflect course selection at school, with boys being more likely to choose courses with a strong applied and technical content. These differences tend to be particularly pronounced among low-achievers, a category where boys are over-represented (boys are more likely to repeat grades than girls and to be identified as all-round low-achievers in PISA) (OECD, 2015).

Participants in PIAAC were asked to report the frequency with which individuals participate in a range of literacy and numeracy related practices in the context of their job and in their private life. Reading practices at work and in everyday life considered in PIAAC include reading documents such as directions instructions, letters, memos, e-mails, articles, books, manuals, bills, invoices, diagrams and maps. Writing practices at work considered in PIAAC include writing letters, memos, e-mails, articles, reports, and forms. Numeracy related practices at work include calculating prices, costs or budgets; using fractions, decimals or percentages; using calculators; preparing graphs or tables; using algebra or formulas; using advanced mathematics or statistics (calculus, trigonometry, regressions). Respondents could indicate that they engage in a specific practice "never", "less than once a month", "less than once a week but at least once a month", "at least once a week but not every day" or "every day". Figures 3.14, 3.15 and 3.16 illustrate gender differences in the percentage of 25-29 year-olds across PIAAC participating countries who report engaging at least weekly in specific reading, writing and numeracy related practices at work and in everyday life.

Gender gaps in reading and writing practices tend to narrow considerably by the time students mature and are part of the young adult sample in PIAAC. Rates of engagement in reading and writing practices tend to converge and, in fact, young men in some countries engage in more reading activities than young women. On the other hand, women lag behind in the use of numeracy practices at work and in everyday life. In particular, women are less likely to use advanced mathematics and statistics, to use algebra or formulae, and to use or calculate fractions and percentages. Gender gaps are either not statistically significant or in favour of females for using calculators both in everyday life and at work, and for calculating costs and budgets, reflecting the greater prevalence of women in administrative rather than managerial positions.

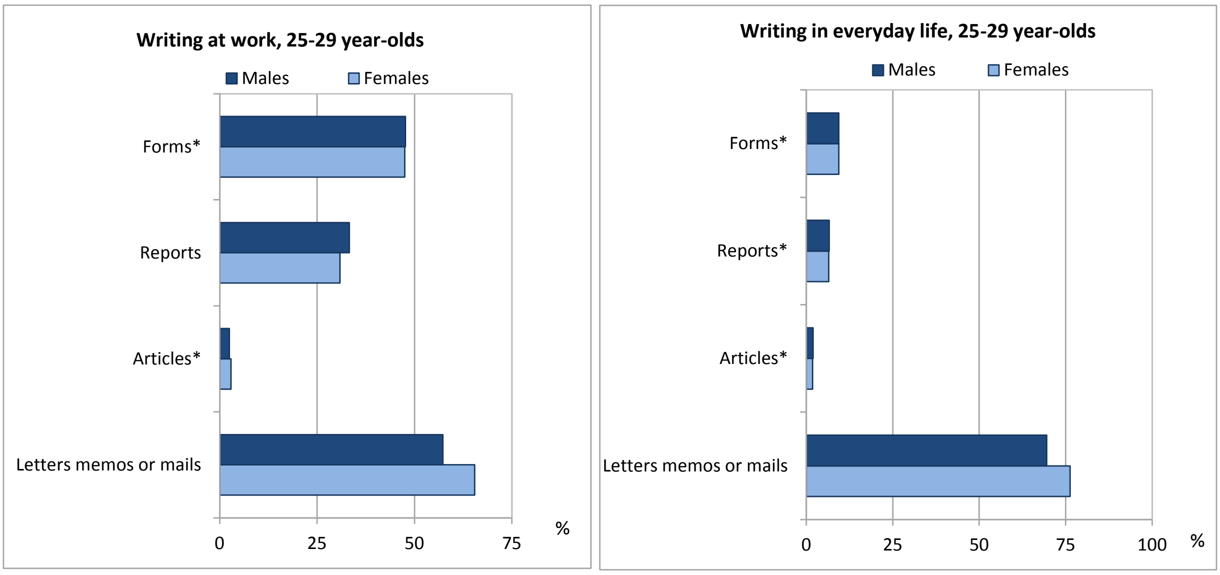
Figure 3.14 Gender differences in reading at work and in everyday life among 25-29 year-olds in PIAAC



Notes: The figure illustrates the average percentage of males and females who report engaging in such practice at least once a week across countries in PIAAC round 1 that were also part of PISA 2000. * Next to the reading practice indicates a difference between the percentage of males and females that is not statistically significant at the 5% level.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

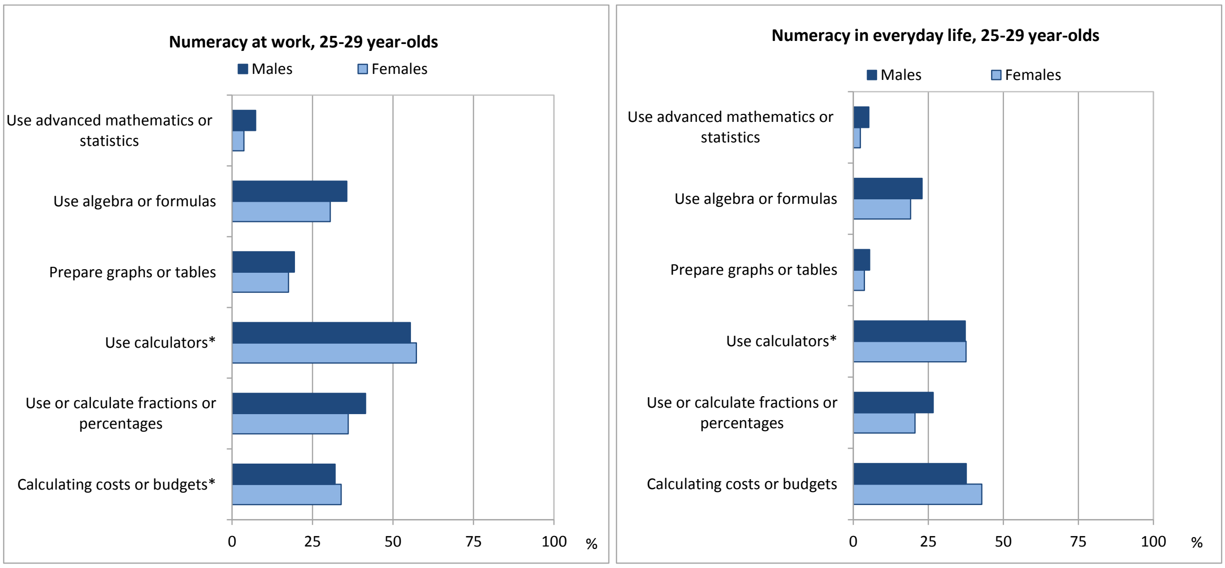
Figure 3.15 Gender differences in writing practices at home and at work among 25-29 year-olds in PIAAC



Notes: The figure illustrates the average percentage of males and females who report engaging in such practice at least once a week across countries in PIAAC round 1 that were also part of PISA 2000. * Next to the reading practice indicates a difference between the percentage of males and females that is not statistically significant at the 5% level.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

Figure 3.16 Gender differences in numeracy related practices at home and at work among 25-29 year-olds in PIAAC

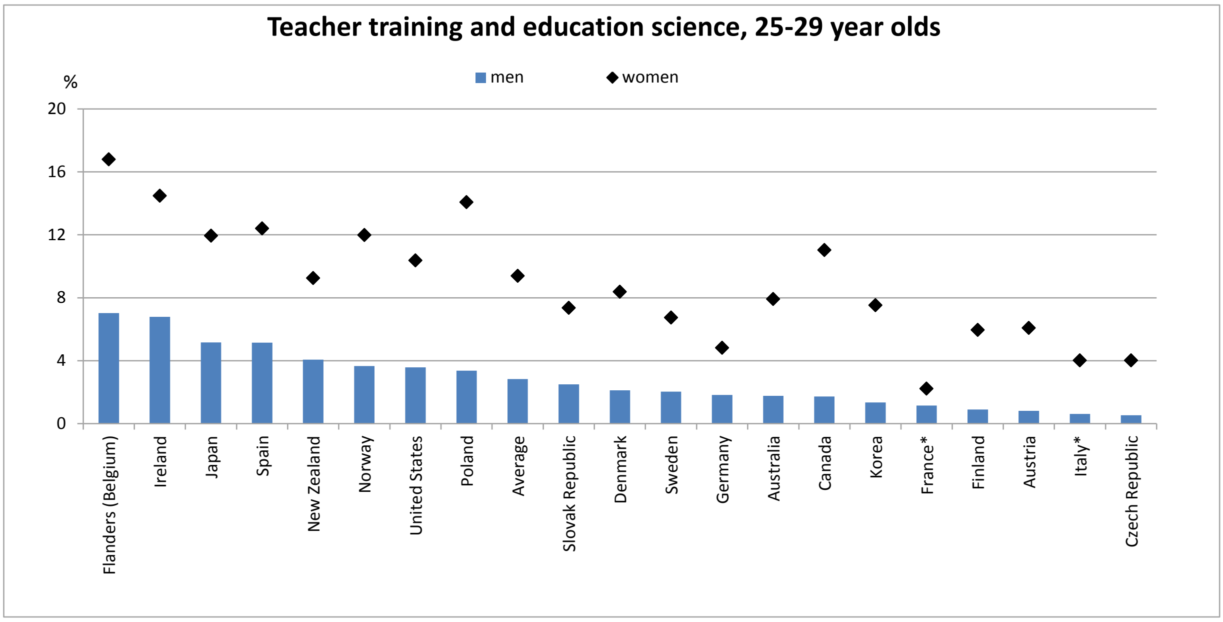


Notes: The figure illustrates the average percentage of males and females who report engaging in such practice at least once a week across countries in PIAAC round 1 that were also part of PISA 2000. * Next to the reading practice indicates a difference between the percentage of males and females that is not statistically significant at the 5% level.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

In order to identify potential differences by gender in the choice of field of study, which may lead to differential exposure to numeracy. Figure 3.17 illustrates the percentage of young men and women in PIAAC (the age range is 25-29 to have reasonably precise estimates by increasing the sample size) whose highest qualification is in teacher training and education science; humanities, languages and arts; engineering, manufacturing and construction. Results reveal that young men are significantly less likely than young women to have obtained their highest qualification in teaching and education science, and are significantly more likely to have obtained their highest qualification in engineering, manufacturing and construction. Gender differences in the humanities, languages and arts are smaller and, because of the small sample size at the individual country level, generally not statistically significant at the 5% level.

Figure 3.17 Gender differences in field of study in PIAAC, 25-29 year-olds



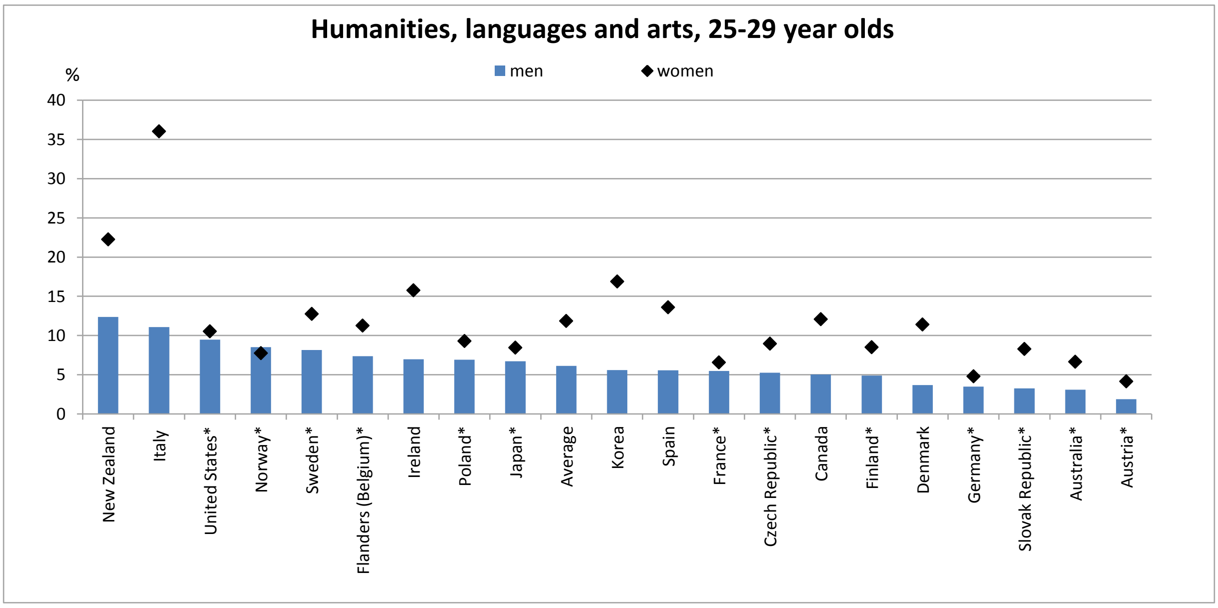
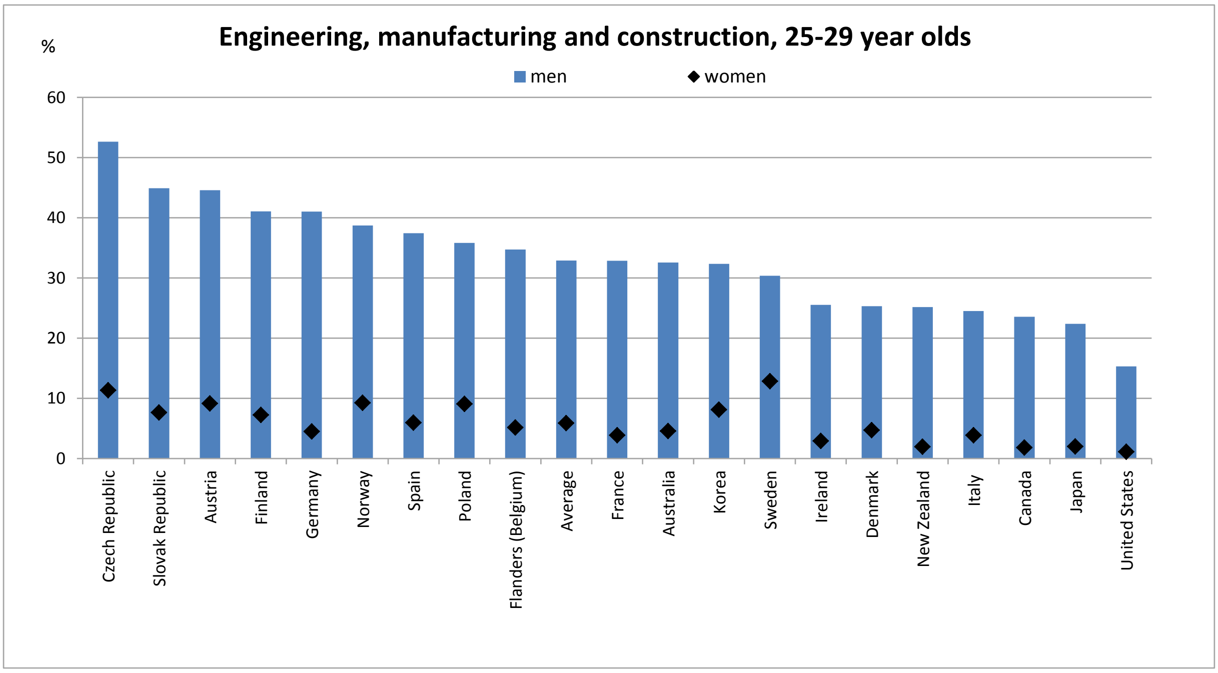


Figure 3.17 Gender differences in field of study in PIAAC, 25-29 year-olds (*continued*)



Note:* Next to the country name denotes countries in which the gender gap is not statistically significant. In each panel countries are ranked in descending order of the percentage of men who obtained their highest qualification in the specific field under consideration.

Source: OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

Differences between men and women not only exist in educational attainment, skill levels and skill use but also in occupational choices. At the age of 15 girls are less likely to expect to work in STEM occupations and, in young adulthood, they appear to be less likely to have enrolled in fields of study that are maths intensive, such as engineering, manufacturing and construction (OECD, 2015). Differences in the mathematics content of the courses young men and women access upon completing compulsory schooling may explain some of the observed differences in the growth in the gender gap between PISA and PIAAC. Results presented in this section indicate that young men and women tend to make educational choices that reflect gender stereotypes on masculine and feminine occupations, with men being over-represented in engineering and mathematics and females being over-represented in teaching, education, the arts and humanities.

However, the fact that the growing gap in numeracy is concentrated at the bottom end of the performance distribution and that this group is the least likely to participate in education and training opportunities upon completing compulsory schooling, suggests that the use of numeracy at home and in the workplace rather than education opportunities plays a larger role in explaining divergence in the gender gap in numeracy.

SECTION 4 - PISA AND PIAAC IN POLAND

4.1 Introduction

As discussed in section 2 of this report, while the PISA and PIAAC assessments of reading/literacy and mathematics/numeracy were not designed to be directly comparable and the results are, therefore, on different scales, the assessment framework in these domains share many characteristics. The absence of a psychometric link between the two studies in these domains mean that, as has been previously discussed in this paper, it is not possible to establish whether information processing skills develop after the age of 15, whether the rate of achievement growth differs across countries, or to identify institutional and organisational factors that may differentially shape growth in achievement.

Longitudinal studies based on PISA samples offer one potential avenue for looking at the achievement growth of young people from the age of 15 years old. A number of countries have such studies: Australia (Thomson and Hilman, 2010; NCVER, 2013), Canada (OECD, 2012b); others were established in Poland (Domański et al., 2012), Germany (Blossfeld, Schneider, Doll, 2009), and Denmark (Danish Ministry of Education, 2014). In Denmark a sample of participants in the PISA 2000 study were administered the PIAAC assessment. However, longitudinal studies at the individual level do not necessarily allow answering questions on achievement growth, unless the same assessment is used (as was the case in the Canadian context) or a linkage study is developed to link the assessments at different time points.

In typical linking studies, a linking design is applied to control for variations in measurement instruments. Statistical linking can be achieved either by administering a range of items to the same test takers or administering common test items to different test takers. A *single group design* involves administering the two tests for which researchers would like to establish a link to the same examinees. A variation of the single group design is an *equivalent group design*, which consists in administering the two tests to equivalent groups of examinees, rather than to the exact same group of examinees. Finally, a *non-equivalent group with anchor test design* consists in administration of two tests to different populations, in which the two tests contain a set of common test items (for details see Dorans et. al., 2011: 25-31).

To date, there is no known study that has attempted to establish a statistical link between PISA and PIAAC using any of these designs. This section compares performance across the PISA and PIAAC studies using *statistical matching* (Rässler, 2004; 2012). It is based on Polish data. In 2012, both PISA and PIAAC were administered in Poland and the PISA study contained a sample of students that was selected based on attendance to grade 10, and who therefore were older than 15 and could have been eligible to sit the PIAAC assessment. Scale concordance is used to map the PISA and PIAAC to one another on the basis of the partial sample overlap in Poland and the existence of comparable background information in the Polish PISA and PIAAC questionnaires.

The aim of the section is to establish a "*proof of concept*" of the feasibility of comparing data from PISA and PIAAC in the absence of a design that allows more traditional forms of linkages between the two studies and stimulate discussions about the possibility of establishing formal links between the two studies in the future.

According to the classification by Holland and Dorans (2006) linking PISA and PIAAC is an exercise that falls between two categories of alignment: vertical scaling and concordance analysis. Concordance occurs when tests have similar constructs, similar reliability and similar difficulty and population, while in vertical scaling, situations with dissimilar difficulty and population might be considered. To some extent, the concordance analysis described in the Holland and Dorans (2006) framework could be considered as a special case of moderation analysis. Therefore the attempt to link PISA and PIAAC could be more precisely named concordance/moderation analysis.

The analyses developed in this section could serve as a test case for national studies of PISA and PIAAC which considered the possibility of identifying analytical tools to enable examining issues related to achievement growth in the absence of international efforts in this sense. This section illustrates that the presence of background variables that are common across PISA and PIAAC allows identifying respondents that share common characteristics which, in turn, can be used to develop pseudo-equivalent groups.

4.2 Past evidence on linking and statistical matching

Linking tests which measure similar constructs but do not share a common framework and test conditions is not a new idea (Linn, 1975; Feuer et al., 1998). International large scale assessments have been frequently linked to national assessments. Beaton and Gonzalez (1993) applied the distribution-matching procedure using the linear equating method in a single group design to link the US National Assessment of Educational Progress (NAEP) and the International Assessment of Educational Progress (IAEP). Pashley and Phillips (1993) used a projection technique based on linear regression applied to a sub-sample of participants that sat both NAEP and IAEP to establish a link between the two studies. The results of these two alternative approaches allow the placement of IAEP results on the NAEP scale. The two approaches yielded similar results for average performing countries but were less consistent in their findings for countries at the top and bottom of the performance distribution.

Four studies have been conducted to link NAEP and TIMSS with the aim of predicting TIMSS scores for states in the United States (for which NAEP scores were available but not TIMSS scores). The first study linked the 1996 NAEP results for grade 4 and grade 8 and the 1995 TIMSS results for grade 4 and grade 8 (Johnson, 1998). The second used the 2000 NAEP grade 8 results and the 1999 TIMSS grade 8 results (Johnson et al., 2003). The third used results from the 2007 NAEP grade 8 maths results and the 2007 TIMSS grade 8 maths results (Phillips, 2009). The most recent study linked 2011 NAEP and 2011 TIMSS assessments in mathematics and science (grade 8 for both studies) (Jia et al., 2011). The first three linking studies used an equivalent group design, assuming that the student samples that took the NAEP and the TIMSS assessments were randomly equivalent. These papers each use different statistical methods to match the characteristics of the score distributions. The most recent study (Jia et al., 2011) was based on a single group design. To take advantage of this, the statistical approaches used were: calibration, statistical projection and statistical moderation for robustness checks. The study design involved two additional samples of students on top of the core NAEP and TIMMS samples. In the first sample, students were assessed during the 2011 NAEP testing window (following NAEP administration procedures) with booklets containing items from both the NAEP and the TIMSS assessment. In the second sample students were assessed during the 2011 TIMSS testing window (following TIMSS administration procedures) with items from both the TIMSS and NAEP assessment. The last approach (statistical projection), used an equivalent group design ignoring the fact that some students responded to questions from both assessments. Findings revealed that similar predicted TIMSS results could be predicted using NAEP scores irrespective of the approach taken and statistical projection was used to report final results because of its simplicity (Jia et al., 2014).

Cartwright et al. (2003) linked reading literacy assessment tests administered to grade 10 students in 2000 as part of the British Columbia’s Foundation Skills Assessment (FSA) and the 2000 PISA reading assessment. A single group design was used, with 2 800 grade 10 Canadian students participating in both FSA and PISA both of which were administered in the period April to May. Several methods of linking were tried with the equipercentile method being chosen as the most appropriate. Radwan and Xu (2012) attempted to link the Ontario Secondary School Literacy Test (OSSLT) with the PISA 2009 reading assessment. A single group design was used, in which a group of students sat both tests. The equipercentile procedure and the fixed item parameter (FIP) procedure were used, both providing similar results. PISA was also linked to the ACT’s College Readiness PLAN assessment in the United States (ACT, 2009). A special linking study was conducted involving the administration of one PLAN battery and one PISA booklet within a 4-month window to the same sample of students.

Several studies have been conducted to link different international assessments. For example, in 2000 PISA was linked to the International Adult Literacy Survey (IALS). This linking study used a common item and single group design using IRT methodology (Yamamoto, 2002). More specifically, fifteen prose items from IALS were administered as part of the PISA 2000 assessment. The results indicated that PISA students could be reliably placed on the IALS scale (Yamamoto, 2002).

The idea of using additional background information available about students, such as gender, ethnicity, and socio-economic background, to link different test forms has been investigated from the late 1980s. However, this has been proposed only in the context of classical linking designs where a link based on common item or common persons was possible and the additional information was only used to facilitate linking based on classical designs. Lawrence and Dorans, (1990) used matched samples in equating with set common items for the Scholastic Aptitude Test (SAT) forms. Yu et al. (2004) used a propensity score stratification technique on the Pre-Professional Skills Test (PPST) when comparing a paper-based to computerised administration. In this study, variables such as gender, ethnicity, educational background, job related information and teaching experience were used to ensure conditional independence. Paek, Liu, and Oh (2006) exploited a propensity score stratification technique to link the Preliminary Scholastic Aptitude Test (PSAT) and SAT. Gender, ethnicity, high school grade level and common item scores were used for appropriate matching.

Only one study has been found that explores the possibility of linking two assessments in the absence of common items and where the assessments cover non-equivalent or not fully equivalent groups using statistical matching (Haberman, 2015). Haberman (2015) proposes a linking method for non-equivalent groups of examinees where the test forms either lack common linking items or have unsatisfactory linking items. Haberman’s approach uses background information about examinees to construct sample weightings via minimum discriminant information (MDIA; Haberman, 1984). He shows that a pseudo-equivalent groups approach that uses background questions to link different test forms yields very similar results to classical approaches.

4.3 The Polish experiment

As has been indicated in section 2, PIAAC and PISA were not designed to be linked and no formal linking studies have been undertaken at the international level. However, participating countries are allowed some flexibility in the administration of the tests to maximise the policy value of the studies given their context and policy concerns. Countries can, for example, oversample some students or schools, add questions at the end of the core PISA instruments and administer the PISA instruments to additional samples of students. The main limiting factor to country optional study operations is that these should not have a bearing on the main study administration, thereby invalidating common test conditions and procedures across the different countries and economies that participate in PISA.

In 2012, Poland implemented a national option which involved administering PISA to a sample of students who attended grade 10 (and were, therefore, aged 16 years or more when they took the assessment). The Polish grade based sample participated in PISA following the same administration protocols and procedures that were applied to the core international target population of 15-year-olds. Results for the additional national sample were scaled together with the international sample and put on the same scale, but were not part of the international PISA database. The difference between the two samples was that the international sample was based on age (with the target population being

15-year-olds in grade 7 or above, irrespective of grade attended), while the Polish national sample was based on the grade students attended and thus included age groups outside the PISA age range, primarily 16 and 17-year-old students but also some 18 and 19-year-old students.

Students aged 18 years or more could be attending grade 10 for different reasons. For example 36% of students aged 18 and 19 (39 students in the sample) had repeated a year or more. In addition, the older students could be in grade 10 if they started school at a later than the usual age in Poland (age 7). It is thus reasonable to assume that 18- and 19-year-olds in the Polish grade 10 PISA sample are a highly selected group (for example, repeaters are generally at the bottom end of the academic performance distribution, except for students who were forced to repeat a grade because of accidents and illnesses information that we do not have), and it is not reasonable to assume that background variables would account for the selection. Therefore, 18- and 19-year-old students in PISA were excluded from the analysis. The age distribution of the Polish sample for PISA 2012 (which combines the national and international samples) and the Polish sample for PIAAC is presented in Table 4.1.

Table 4.1 Distribution of age in Polish samples of PISA and PIAAC

	Age						
	<16	16	17	18	19	>19	Total
PISA	3 526	3 642	1 088	79*	28*	0	8 152
PIAAC	0	93	123	157	702	8 291	9 989
Total	3 526	3 735	1 211	236	730	8 291	18 141

Note: * Excluded from linking.

Table 4.1 shows that the samples of the two assessments slightly overlap and are highly unbalanced. A large number of students aged 17 took PISA but the number of 17 year-olds who took PIAAC is small. No student aged 20 years or more took the PISA test. The age differences between the two groups mean that a classical equivalent group design cannot be used to link the studies. However, other methods such as pseudo-equivalent group design can be used to derive concordance between the two assessments. Since the age of respondents in both studies is observed, the concordance design could be treated as a missing data design on the assumption of random missing cases, as in Rubin’s (1974) work. In this framework, groups can be considered to be equivalent conditional on age and other background variables. Central to the idea of the potential outcome framework by Rubin (1974) is the assumption that respondents can be characterised by two potential outcomes regardless of whether or not they received the treatment.

Rubin’s framework indicates that conditional independence – which is fulfilled when assignment to a test condition is independent of outcomes given a set of background characteristics – is achieved in the Polish context if the performance of participants in the two tests was not determined by whether they sat the PISA or PIAAC. In a random group design, conditional independence holds because randomisation typically balances data between groups. However, assignment to either PISA or PIAAC was not random in Poland since it was highly dependent on age. One way to achieve conditional independence and estimate a link between the PISA and PIAAC scales is to use propensity scores as defined by Rosenbaum and Rubin (1983). The propensity score is the conditional probability that an individual received a particular test (PISA or PIAAC), given a set of observable characteristics.

Because the PIAAC target population includes individuals between the age of 16 and 65, school-aged students who were 15 years old and who participated in the Polish PISA sample were excluded from the analysis (the PISA sample was reduced from the original 8 152 participants to 4 837 participants). As there was no clear cut off for the upper age bound participating in PISA (since participation was linked to attendance of grade 10), the PIAAC population was not restricted. Estimates of propensity scores were obtained using a logistic model, in which participation in PISA was the dependent variable. In terms of covariates, age was the most important predictor, since the probability of being sampled for either PISA or PIAAC was a function of age. Gender and the number of books at home at the age of 15 (16 in PIAAC) were included in both studies and were included as covariates since they might be associated with the selection process.

Propensity score weighting was used to reweight the PIAAC sample to match PISA data. The reweighting procedure allows the estimation of ex post distribution parameters for the two tests in two pseudo-equivalent groups. Once pseudo-equivalent groups are defined and score distributions for the two tests are estimated, concordance analysis is equivalent to estimating a scale transformation function. In order to transform PISA scores to the PIAAC metric, the following linear transformation was used (for a detailed discussion see Kolen and Brennan, 2004:31-32).

$$l_{PIAAC}\left(PISA\right)=\frac{\sigma_{PIAAC}}{\sigma_{PISA}}\left(PISA\right)+\left[\mu_{PIAAC}-\frac{\sigma_{PIAAC}}{\sigma_{PISA}}\mu_{PISA}\right]$$

While the transformation of the PIAAC score to the PISA metric was based on the following transformation:

$$l_{PISA}\left(PIAAC\right)=\frac{\sigma_{PISA}}{\sigma_{PIAAC}}\left(PIAAC\right)+\left[\mu_{PISA}-\frac{\sigma_{PISA}}{\sigma_{PIAAC}}\mu_{PIAAC}\right]$$

Where $PISA$ denotes the PISA scores and $PIAAC$ denotes scores from the PIAAC assessment, μ_{PISA} and σ_{PISA} are the mean and standard deviation of PISA scores; μ_{PIAAC} and σ_{PIAAC} denote mean and standard deviation of PIAAC scores after reweighting. Both equations express a simple linear function where $\sigma_X / \sigma_Y \left(X\right)$ represents the slope and $\mu_X - \sigma_X / \sigma_Y \, \mu_Y$ represents the intercept.

The analyses presented are based on the full set of five plausible values from PISA and the first five plausible values in PIAAC (results should be equivalent when using any combination of five plausible values from PIAAC). Bootstrapping was used to reflect estimation uncertainty (Efron, 1982: 29-35). In order to apply the transformations developed using Polish data to data from other countries and then performing statistical analysis with the linked scales, an additional linking error must be introduced. This error should reflect the additional uncertainty that derives from making inferences about countries other than Poland based on a concordance analysis based on Polish data.

Figure 4.1 shows the predicted probabilities and associated weights (lower part) of being classified as PISA students for the PIAAC sample. Both probabilities and weights approach zero at the age of 20. The left side of the figure shows results for the full age range in PIAAC, while right side of the figure shows results for individuals in the 16-20 year-old age range. For respondents aged 19 years, average weights equal 0.11, for 19.5 year-olds, average weights are 0.02 and for 20 year-olds, they are 0.002. This means than effectively only the 1 000 of the youngest respondents from PIAAC were used for linking. The very small weights associated with the older PIAAC participants mean that, in practical terms, they do not contribute any information in the estimation.

Figure 4.1 Predicted probablities and weights from Model 2 against age in PIAAC sample

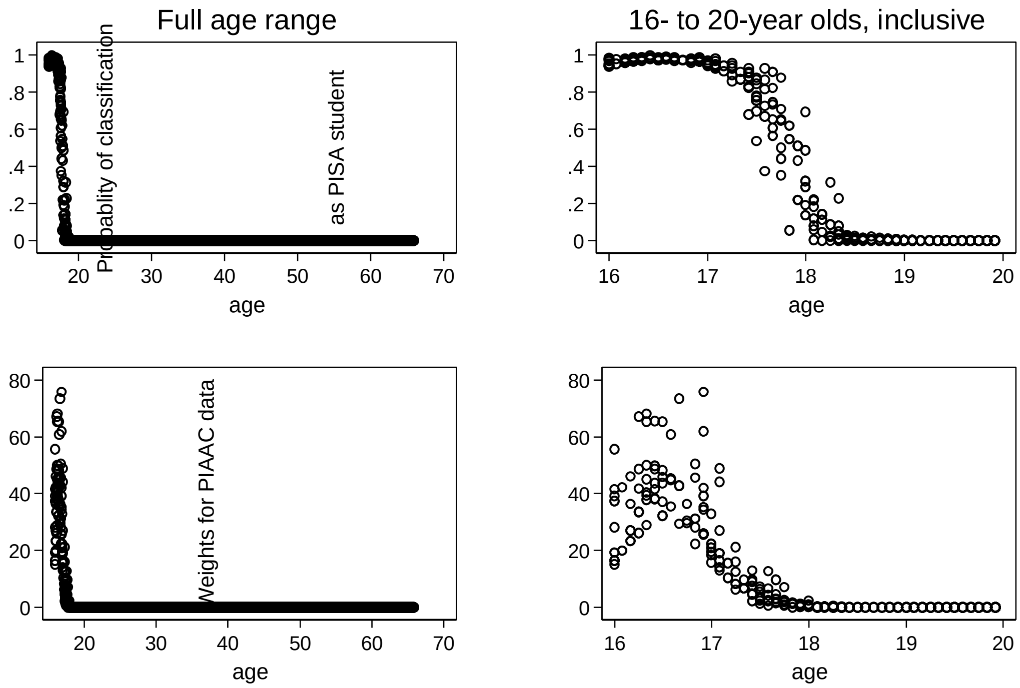


Table 4.2 presents descriptive statistics with mean values for age, number of books in the family home, the percentage of females, the percentage of respondents that reports being still in education (this variable was not used in the propensity score model and is presented here for validation purposes only). For PIAAC, results are presented before and after weighting.

Table 4.2 Mean values for age, number of books, % of females, % respondents in school in PISA sample and PIAAC sample before and after weighting

	PISA sample	PIAAC	
		Unweighted	*Weighted*
Age	16.61	31.67	16.62
Number of books (at age 15)	3.24	3.20	3.26
% of females	52%	49%	47%
% of respondent in school	100.00%	34.45%	97.41%

Table 4.3 presents the linking constants that can be used to transform the PIAAC scales into PISA scales and vice versa. Those constants together with equations 4 and 5 can be applied to scale the transformation. Constants are presented for each plausible value (the first five) and can be applied to respondent level data. To transform scales from aggregates, e.g., to transform the PISA mean result to PIAAC scale transformation, constants should be derived from average slopes and intercepts from the five plausible values.

Table 4.3 Linking constants for linear linking

Assessment	Weights based on:		From PISA to PIAAC		From PIAAC to PISA	
		Plausible values	Slope	Intercept	Slope	Intercept
Numeracy/mathematics	age, gender, number of books	1	0.4641	11.7342	2.1549	-25.2856
		2	0.4647	10.5375	2.1519	-22.6751
		3	0.4875	5.1472	2.0511	-10.5575
		4	0.4822	-0.4105	2.0739	0.8512
		5	0.4778	5.2215	2.0931	-10.9292
		Average	0.4753	6.4460	2.1050	-13.7192
Literacy/reading	age, gender, number of books	1	0.4592	37.3293	2.1779	-81.2992
		2	0.4918	18.9556	2.0332	-38.5406
		3	0.4833	25.7384	2.069	-53.2537
		4	0.4855	21.5606	2.0598	-44.4111
		5	0.4584	40.7725	2.1815	-88.9459
		Average	0.4756	28.8713	2.1043	-61.2901

To illustrate the precision of the linking conversion, tables for scores defining student ability levels in PIAAC and PISA are shown (Tables 4.4 and 4.5), together with the transformed scores, and standard errors associated with each transformed score. Linking errors for mid-distribution scores are much smaller than at the extremes of the distribution, by as much as a third.

Table 4.4 Concordance of PISA to PIAAC scores for numeracy and literacy scales

PISA maths score	PIAAC numeracy		PISA reading score	PIAAC literacy	
	Score	(error)		Score	(error)
358	179.08	(9.92)	262	155.64	(14.24)
420	207.45	(7.81)	335	189.42	(11.33)
482	235.82	(6.22)	407	222.74	(8.63)
545	264.66	(5.61)	480	256.52	(6.30)
607	293.03	(6.27)	553	290.31	(4.98)
669	321.40	(7.89)	626	324.09	(5.48)
			698	357.41	(7.39)

Table 4.5 Concordance of PIAAC to PISA scores for mathematics and reading scales

PIAAC numeracy score	PISA maths		PIAAC literacy score	PISA reading	
	Score	(error)		Score	(error)
176	351.24	(21.44)	176	305.53	(26.69)
226	460.56	(14.05)	226	413.72	(17.86)
276	569.87	(12.59)	276	521.92	(11.52)
326	679.18	(18.53)	326	630.12	(12.39)
376	788.49	(27.41)	376	738.32	(19.54)

Tables 4.6 and 4.7 illustrate the PISA and PIAAC proficiency level descriptions and the threshold scores that subject matter experts responsible for the development of PISA and PIAAC assigned to the PISA reading/mathematics proficiency levels and the PIAAC literacy/numeracy proficiency levels. Furthermore, tables 4.6 and 4.7 identify the threshold that would have been estimated through trough the analyses presented in this section. This was done to identify whether the results from the concordance analyses are in line with the characterisation of the level of difficulty conducted by subject matter experts for PISA and PIAAC. Comparisons between definitions of proficiency levels in PISA and PIAAC using linked scores are presented in Table 4.6 (reading and literacy) and Table 4.7 (mathematics and numeracy).

Table 4.6 Proficiency levels of reading and literacy on common scales

		linked to			
PIAAC description for level of proficiency		(PISA)	(PIAAC)		**PISA description for level of proficiency**
At Level 5, tasks may require to construct syntheses of similar and contrasting ideas or points of view; or evaluate evidence based arguments. Tasks often require respondents to be aware of subtle, rhetorical cues and to make high-level inferences or use specialised background knowledge.	376	(738)			
			(357)	698	*At Level 5, tasks require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. Reflect and evaluate tasks may require the reader to hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text.*
At Level 4, tasks require respondents to perform multiple-step operations to integrate, interpret, or synthesise information from complex or lengthy continuous, non-continuous, mixed, or multiple type texts.	326	(630)			
			(324)	626	*At Level 5, tasks involve working with deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge.*
			(290)	553	*At Level 4, tasks involve retrieving information require the reader to locate and organise several pieces of embedded information. Some tasks require interpreting the meaning of nuances of language by taking into account the text as a whole.*
At Level 3, texts are often dense or lengthy, and include continuous, non-continuous, mixed, or multiple pages of text. Tasks require the respondent to identify, interpret, or evaluate one or more pieces of information, and often require varying levels of inference.	276	(522)			
			(257)	480	*At Level 3, tasks require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase.*
At Level 2, tasks require respondents to make matches between the text and information, and may require paraphrasing or low-level inferences.	226	(413)			
			(223)	407	*At Level 2, some tasks require recognising the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low-level inferences.*
			(189)	335	*At Level 1a, tasks require to locate one or more independent pieces of explicitly stated information; to recognise the main theme or author's purpose in a text about a familiar topic; or to make a simple connection between information in the text and common, everyday knowledge.*
At Level 1, tasks require the respondent to read relatively short digital or print continuous, non-continuous, or mixed texts to locate a single piece of information that is identical to or synonymous with the information given in the question or directive.	176	(306)			
Below Level 1,The tasks require the respondent to read brief texts on familiar topics to locate a single piece of information.			(156)	262	*At Level 1b,tasks require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type.*

Note: description of PIAAC ability levels are shortened from OECD 2013f: 64; PISA ability levels are shortened from OECD 2010b: 47.

Table 4.7 Proficiency levels for mathematics and numeracy on common scales

PIAAC description for level of proficiency		linked to (PISA)	linked to (PIAAC)		**PISA description for level of proficiency**
At Level 5, tasks require the respondent to understand complex representations and abstract and formal mathematical and statistical ideas, possibly embedded in complex texts.	376	(788)			
At Level 4, tasks require the respondent to understand a broad range of mathematical information that may be complex, abstract or embedded in unfamiliar contexts.	326	(679)			
			(321)	669	*At Level 6, students can conceptualise, generalise and utilise information based on their investigations and modelling of complex problem situations, and can use their knowledge in relatively non-standard contexts.*
			(293)	607	*At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions.*
At Level 3, tasks require the respondent to understand mathematical information that may be less explicit, embedded in contexts that are not always familiar and represented in more complex ways.	276	(570)			
			(265)	545	*At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions.*
			(236)	482	*At Level 3, students can execute clearly described procedures, including those that require sequential decisions.*
At Level 2, tasks require the respondent to identify and act on mathematical information and ideas embedded in a range of common contexts where the mathematical content is fairly explicit or visual with relatively few distractors.	226	(461)			
			(207)	420	*At Level 2, students can interpret and recognise situations in contexts that require no more than direct inference.*
			(179)	358	*At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined.*
At Level 1, tasks require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit with little text and minimal distractors.	176	(351)			

Note: descriptions of PIAAC ability levels are shortened versions of those obtained from OECD, 2013f: 76; description of PISA ability levels are shortened from OECD, 2012c: 267.

The lower threshold for levels 1 on the PIAAC numeracy scale is a score of 176. Roughly, this shows the upper bound for those lacking basic numeracy skills. A similar threshold was set by the PISA experts group to a PISA score of 358. After linking, a PIAAC score of 176 is estimated to correspond to 351 on the PISA scale and likewise a PISA score of 358 was estimated to correspond to 179 on the PIAAC scale. This means that using linking, without knowing the threshold for one scale, the threshold on the other scale can be predicted reasonably precisely.

The Level 5 threshold on the PIAAC numeracy scale was linked to a score of 788 on the PISA mathematics scale, which is well above the highest PISA threshold for the PISA mathematics assessment (which is 669). The PISA scale across OECD countries has a mean of 500 with a standard deviation of 100. From this it can be inferred that the PIAAC level 5 threshold for numeracy is 2.7 standard deviations above the mean on the PISA scale. If we assume that numeracy has a standard normal distribution, then only about 0.5% would be at this level. Results were similar for literacy. The PIAAC level 5 in literacy was linked to a PISA score of 738 in reading, well above the threshold defining the PISA proficiency level 6. What these results suggest is that the Level 5 threshold for PIAAC may be more demanding than the highest threshold identified in PISA.

These comparisons constitute a data driven exercise to identify if major differences emerge, rather than a proper validation of results. They constitute a baseline level of confirmation that results obtained in the concordance analyses presented do not yield unreasonable results and are broadly in line with experts' judgement on the scores identifying levels of proficiency in the two assessments and content areas. A proper validation study would require additional data which would help to develop other statistical approaches and linking plans, in line with the work conducted on NAEP and TIMSS by Jia and colleagues (Jia et al., 2014).

SECTION 5 - CONCLUSIONS AND IMPLICATIONS

The OECD PISA and PIAAC studies are key instruments that help participating countries develop solid evidence to map, monitor and promote high quality learning opportunities for all. The fact that developmental activities are currently underway to prepare the 2021/22 PIAAC and the 2021 PISA assessments makes this a key moment to: i) evaluate the benefits of the two studies for policy makers and researchers; ii) articulate the existing links and relationships between the two studies; iii) identify the scope of the analyses that can be developed given current synergies; and iv) illustrate the benefits that would arise as a result of further integration.

PISA has consistently identified equity as a key dimension of success of education systems, and has argued that reducing the influence of socio-economic background on skill development should be a crucial objective of education policy. Since the first wave in 2000, the PISA study has in fact revealed that 15-year-old students with a disadvantaged socio-economic background are less likely than their more advantaged peers to develop strong foundation skills, that boys and girls are not equally likely to perform on par in reading and mathematics but also, and crucially, that socio-economic and gender disparities in education outcomes vary markedly across education systems. The first cycle of PIAAC has revealed that, on average, information processing skills tend to improve after the age of 15 and peak in the late twenties and such skills reflect the learning opportunities young adults have in post-compulsory education training and the early years upon entering the labour force. However, little is known so far about not only the extent to which disparities in skills evolved after the end of compulsory schooling, but also in whether disparities in learning opportunities after post-compulsory schooling magnify or reduce disparities observed at the age of 15.

The analyses presented in this report illustrate that, in most countries, socio-economic disparities in literacy and numeracy at the age of 15 not only persist in young adulthood, but tend to widen. Further education and participation in the labour market are the crucial mechanisms that are associated with skill acquisition beyond compulsory schooling. Socio-economically disadvantaged youngsters are considerably less likely than their more advantaged counterparts to attend post-secondary education and training, and are more likely to drop out of education without a secondary degree. They are also more likely to be unemployed or out of the labour force and to work in jobs that require little advanced on-the-job training or practice of higher order thinking skills. Although it is not possible to establish causality given the nature of the data examined, this report has revealed divergence in the opportunities for skill development between socio-economically advantaged and disadvantaged youngsters upon completion of compulsory schooling and an overall increase in socio-economic gaps in proficiency in the large majority of countries.

On average, socio-economic disparities in skills appear to grow larger in particular among individuals at the bottom tail of the performance distribution. For example, between the age of 15 and 27 the standardised achievement gap at the 9th decile of the literacy performance distribution between individuals with and without tertiary educated parents remained stable at around 44% of a standard deviation. However, the gap between individuals with and without tertiary educated parents grew considerably at the 1st decile of performance (from 0.46% of a standard deviation to 0.71% of a standard deviation). Results on gaps in numeracy and between individuals of different socio-economic background (according to a measure of the number of books available in the respondents' home) are very similar.

Germany and Korea represent important outliers. In Germany gaps in literacy and numeracy by socio-economic condition were the highest among students who sat the PISA test in 2000, but were in line with the OECD average when considering the same birth-cohort at the age of 27. Korea also bucked the trend of increasing socio-economic gaps in skills after the end of compulsory schooling: gaps were already among the lowest at the age of 15 and grew smaller by the age of 27.

These findings suggest that schooling represents an equalising force in terms of skill development. In most countries, whatever their social background, students follow a broadly similar programme of study, at least until the completion of lower secondary level education. Even if participation in a common educational experience does not eliminate socio-economic disparities in learning outcomes, it moderates the effects of differences in social background and the differences in access to resources as well as in the associated dispositions and attitudes to learning on performance. With the completion of compulsory schooling, the experience of young people from different backgrounds diverges both in terms of the types of education programmes as well as in the timing of entry to the labour market and occupational destinations. Unsurprisingly, inequalities in information processing skills appear to widen after the end of compulsory schooling, in particular, among low-achieving individuals.

The striking cross-national variation in socio-economic disparities in skills gaps at age 15 and the evolution of such gaps between the age of 15 and the age of 27 raises the question of what policies and institutional arrangements may explain such variability. Extensive policy analysis and research has been devoted to the analysis of the features of education systems that are most strongly associated with socio-economic gradients (or lack of) in literacy and numeracy. However, much less is known about which factors contribute to convergence or divergence in socio-economic gaps after the end of compulsory schooling. Results on the widening socio-economic gap at the bottom of the performance distribution identify a target group of policy interventions – students without tertiary educated parents and who are low-achievers at the age of 15. They also help to formulate hypotheses as to why gaps widen in several countries after schools ceases to be able to exert their equalising effect since this is the group which is less likely to enjoy opportunities for further skill development through education and training.

Over the years, PISA results have helped the education policy community develop a more nuanced understanding of gender disparities in education, revealing, for example, that boys are severely over-represented among low-achievers in literacy while girls remain under-represented among the highest performers in numeracy. Results presented in this report highlight that in the vast majority of countries gender differences in literacy observed in PISA narrow considerably and, in many countries disappear altogether by age 27 in PIAAC. The analyses developed in section 3 suggest that the high rate of convergence across males and females in literacy from the age of 15 in PISA to the age of 27 in PIAAC could be related, at least in part, by some of the differences in test administration between PISA and PIAAC. Analyses of PIAAC data for the young age group of 16-30 year-olds in fact reveal that the gender gap in literacy is already considerably narrower at the age of 16 in PIAAC than it is in PISA. These results are further corroborated by a comparison of analyses of data from ALL, covering similar cohorts and age groups. Differences in assessment coverage appear to play a less important role. At the same time, we identify marked differences in the reading practices of males and females as they age: while at age 15 boys are considerably less likely than girls to read at school and for enjoyment, young men and women’s reading practices converge once they leave compulsory schooling and have greater opportunities to choose their own reading materials or are required to read at the workplace.

It has been argued that the evidence that large scale international assessments provide is bounded by their specificities – so much so that even studies covering similar populations and subject areas can yield different rankings (Rindermann, 2007; Wu, 2009). Differences in gender gaps in PISA and PIAAC estimated in this report further support the need to carefully considered test characteristics when drawing implications and policy messages from quantitative analyses of large scale assessments. In particular, this study suggests that motivational factors may play an important role in explaining convergence in performance in literacy between males and females between PISA and PIAAC and, that males generally display better relative performance in PIAAC compared to PISA.

According to expectancy value theory, the motivation individuals exert when engaging in a task is a function of their expected performance on the task and the degree to which they value the task (Eccles and Wigfield, 2002). Test-taking motivation, defined as the willingness to work with effort and persistence on a test (Baumert and Demmrich, 2001) can therefore be considered to depend on perceived success on a given test, beliefs about the amount of effort the test will consume, perceived importance of the test, and affective reactions to the individual test items (Wise and DeMars, 2005). In high-stakes assessments it is generally assumed that test takers will invest their maximum effort in the test. In low-stakes assessments such as PISA and PIAAC however, test-taking motivation is likely to be more variable, and the association between motivation and test performance to be stronger (Barry et al., 2010; Chan et al., 1997; Cole, Bergin and Whittaker, 2008; Eklöf, 2010; Wise and DeMars, 2010). As a result, specificities in test conditions are likely to have an important bearing on overall estimates of performance, as well as estimates of differences across population subgroups that differ in expected performance and task value (Braun, Kirsch and Yamamoto, 2011; Wolf and Smith, 1995). Experimental evidence based on the reading performance of 12th graders on the National Assessment of Educational Progress (NAEP) reading assessment in the United States suggest that the gender gap in reading is narrower when monetary incentives are introduced. Both boys and girls appeared to exert more effort when monetary incentives were introduced than in the absence of monetary incentives, but boys' reaction to monetary incentives was stronger than girls' (Braun, Kirsch and Yamamoto, 2011).

Computer-based administration and, more crucially, the fact that the PIAAC test was conducted as a one-to-one study in respondents' homes in the presence of a trained interviewer rather than as a group exercise in a school setting using paper booklets like in PISA may entice greater engagement with the test material, greater motivation and greater effort among males. Although we cannot directly test for the effect administration procedures and settings have on differences across in the gender gap in literacy across PISA and PIAAC previous research suggests that boys' performance in PISA may be particularly affected by test-taking motivation and ability to maintain engagement with the test throughout the test session (Borgonovi and Biecek, 2016). It is possible that, in the teenage years, boys may need stronger motivational incentives to display their knowledge and skills and that their performance in standardised, low-stakes assessments may be more dependent on motivation and engagement than girls' performance. The presence of a trained interviewer supervising the test-taking session on a one-to-one basis may provide such incentive. Furthermore, in the teenage years many boys may display lack of engagement with school, and with reading in particular as a way to build their social identity and status among peers (Paechter, 1998; Francis, 2000; Warrington, Younger and Williams, 2000; Smith and Wilhelm, 2002; Smith and Wilhelm, 2006). As a consequence, they have an incentive to display lack of engagement and motivation to perform in school tests.

Another important difference between PISA and PIAAC relates to the length of the test session: PISA is designed to take two hours to complete and a strict two hour limit is enforced. The PIAAC study is designed to be completed within around one hour, although respondents can engage with the test for as long as they want to. Boys who participate in the PISA assessment appear to be particularly prone to fatigue and engagement effects in particular in the reading part of the test (Borgonovi and Biecek, 2016): the percentage of correct responses boys provide on a set of reading items is higher when these items are at the beginning rather than at the end of the test. While this effect is also visible among girls, the difference in the percentage of correct answers that girls provide at the beginning compared to the end of the test is smaller than that observed among boys (Borgonovi and Biecek, 2016). These results suggest that introducing measures of conscientiousness in the PISA and PIAAC assessment, together with standardised measures aimed at capturing test-taking engagement and fatigue may be necessary to more accurately compare the performance of individuals on the two tests. Furthermore, these results suggest that the influence of test engagement and motivation on the PISA test may be larger than in PIAAC, given the structure of administration (group session in a school under the overall supervision of a teacher or other member of the school staff vs. one-to-one session with a trained interview, with the test taking place after a lengthy one-to-one interaction between the test taker and the interviewee during the CAPI delivered questionnaire).

Recent evidence on the opening of a gender gap in favour of women in participation in tertiary level education suggests that the underachievement of boys in school may have important consequences for their future life chances. Further research and analyses should aim to investigate the extent to which boys' underachievement is magnified by testing conditions, or other factors, to ensure that boys are sufficiently motivated in school and, in particular, during testing sessions so that they do not drop out of their studies and their potential is adequately recognised in formal evaluations.

Results presented in section 3 also suggest that males may outperform females in digital literacy tasks. While the gender gap in digital reading is smaller than that in paper-based reading, girls still outperform boys in digital reading in PISA and in PIAAC based on Figure 3.8. Digital literacy requires test takers to form a mental representation of digital spaces and to use navigation tools. It requires individuals to acquire a unique set of skills that is not necessary when reading printed texts (Rasmusson and Åberg-Bengtsson, 2014), for example requiring much greater use of visual-spatial skills (Castelli et al., 1998; Lee, 2007; Zhang and Salvendy, 2001). The increasing digitisation of everyday life may ultimately represent an opportunity to engage boys with reading, but may also represent a challenge for girls in the future.

Most research on gender gaps in quantitative abilities illustrates that gender gaps are small in the teenage years, being one-tenth of a standard deviation, on average. However, it is well-recognised that such differences tend to be wider among high-achievers and that differences across genders in perceived self-efficacy in mathematics, mathematics self-concept, and mathematics anxiety are large. Even small differences in performance, especially when compounded by large differences in attitudes, dispositions and self-beliefs, may lead to an important under-representation of women in fields of study and occupations that require strong quantitative skills (OECD, 2014c).

We observe that gender gaps in numeracy widen between the age of 15 and 27. On average, the standardised gap grows from one-tenth of a standard deviation to around one-fifth of a standard deviation. Among high-achievers (defined as individuals in the top decile of the country-specific numeracy distribution), the gender gap remains relatively unchanged from the age of 15 to the age of 27, at around one-fifth of a standard deviation. The nature of the data used cannot shed light on the mechanism that lead to a widening of such gaps, but it is possible that the under-representation of women in fields of study and occupation that does not require strong quantitative abilities may further curtail the development of numeracy skills among women. Practice makes one perfect, and lack of practice of numeracy skills in post-secondary, tertiary education, training and the labour market among women may contribute to the observed widening of the gender gap.

All analyses developed in this report focused on examining the evolution of disparities in literacy and numeracy from the age of 15 to young adulthood because of the absence of an established link between PISA and PIAAC. In section 4 analyses of the Polish PISA 2012 grade based data were used to estimate a concordance between the PISA reading and the PIAAC literacy scales and the PISA mathematics and the PIAAC numeracy scales. The approach developed in section 4 to identify a link between PISA and PIAAC could be further expanded to other countries that implemented a grade based sample in PISA 2012 to identify a more robust and precise link between the PISA and PIAAC scales.

This report identified clear similarities in the assessment frameworks of PISA and PIAAC in the two key areas of literacy/reading and numeracy/mathematics. Despite important similarities between the two studies which results from a high degree of cross-fertilisation, no explicit effort has been made so far to establish a closer and explicit integration in the development of PISA and PIAAC. This report has clearly illustrated that while at the moment PISA and PIAAC can be examined in conjunction to produce evidence on important policy questions, an overarching assessment strategy that explicitly considered interrelationships between the two studies might considerably enhance their respective policy value. For example, although at the conceptual phase of the first cycle of PIAAC the possibility of oversampling the birth-cohorts that had taken part in the early cycles of PISA was considered, no oversampling was implemented. The small age-specific sample at the country level means that estimates that attempt to combine data from PISA and PIAAC are inevitably very imprecise. Differences in administration also leave questions unanswered on the role test setting and administration has on observed results for particular population subgroups.

The creation of explicit and stronger synergies between PISA and PIAAC could take several forms and ultimately enable researchers and policy makers to identify patterns of achievement growth in the critical years that mark the transition from age 15 to age 30, how these differ across countries and population subgroups, and which institutional and organisational arrangements are associated with positive/negative transition pathways. In particular, being able to locate PISA respondents on the PIAAC literacy scale would aid investigation of the learning growth occurring after the ages of 15 and understanding some of the structural (e.g. features of different education and training systems) and cyclical (e.g. youth unemployment) factors that that affect learning gain at the cohort level.

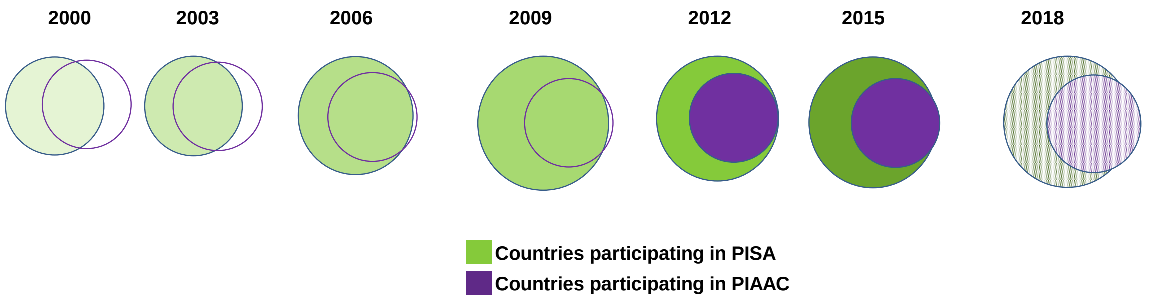
The following synergies would significantly enhance the possibility of combining information from the two studies. The first would be to oversample PISA cohorts in PIAAC to have sample sizes that enable analyses of specific age groups. However, in the absence of greater integration between the assessment frameworks, oversampling would not enable to examine patterns of achievement growth. The second would be to establish a formal link between independently scaled PISA and PIAAC tests. The existence of a psychometric link between the PISA and PIAAC scales would provide invaluable information relevant to several areas of research and policy interest and would not require modifications to the current governing structure of the two studies.[1] Several possibilities exist to create a psychometric link between PISA and PIAAC and proposals have been developed to conduct a PISA-PIAAC link study during the Main Study administration of PISA 2018. The timing of such development made the proposal unfeasible, insights from such proposal could be used to develop a link study in the future.

Finally, greater synergies could be established by creating explicit reporting mechanisms between the PISA and PIAAC experts groups and boards to ensure that developmental activities undertaken by the two studies respond to an overall, broader assessment strategy. While the creation of formal links between the bodies responsible for the development of PISA and PIAAC could bring significant policy value, and would be facilitated by the fact that both studies are steered by the OECD, the governance structures and decision-making of the two studies is distinct and the development of the two studies has not been subject to close and formal co-ordination before. Moreover, although the creation of formal links between PISA and PIAAC would help to answer a set of questions and benefit directly countries that participate in both programmes, not all countries do. Each programme also has a unique role to play in promoting evidence based policy making in countries, given international comparisons and there may be opportunity costs in creating stronger synergies between PISA and PIAAC for each study constituency.

1. The current arrangement means that the assessment frameworks for the PISA and PIAAC literacy/reading and numeracy/mathematics are developed independently, subject to a degree of information and mutual learning.

The overlap in country coverage has shifted dramatically over the years, mostly because of a rapid expansion of coverage in PISA. In 2000, students from 43 countries and economies took part in the first edition of PISA. By 2015 this number rose to 72 countries and economies and is projected to increase to 77 countries and economies by 2018. Additional countries are taking part in the PISA for Development Initiative. The first round of the first cycle of PIAAC administered in 2011/12 included 23 countries and subnational entities. An additional 9 countries took part in the second round of the first cycle of PIAAC in 2015 and an additional 5 countries are expected to take part in the third round in 2018. Figure 5.1 illustrates the overlap between PISA and PIAAC country coverage.

Figure 5.1 Changes in the overlap in country coverage between PISA and PIAAC 2000-2018



Even though country overlap is increasing in absolute terms (more countries participated in both PISA and PIAAC by 2018 than participated in PISA 2000 and the first round of the first cycle of PIAAC in 2012), relative overlap is decreasing (around 50% of the countries participating in PISA 2018 are expected to be represented in PIAAC by 2018. By contrast around 90% of the countries who participated in PISA 2000 will be represented in PIAAC by 2018). This means that an increasingly smaller proportion of countries might see value in closer integration. The middle ground of creating a psychometric link bears the promise of the most effective solution to enhance the policy value of PISA and PIAAC in the medium term.

REFERENCES

ACT (2011), *Technical Manual for the PLAN-PISA International Benchmarking Linking Study*, ACT.

Alexander, K. L., D.R. Entwisle and L.S. Olson (2007), "Lasting consequences of the summer learning gap", *American Sociological Review*, Vol. 72/2, pp.167–180.

Anker, R. (1997), "Occupational segregation by sex", *International Labour Review*, Vol. 13, Geneva.

Baker, D. P., B. Goesling and G.K. LeTendre (2002), "Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley Effect" on mathematics and science achievement", *Comparative Education Review*, Vol. 46/3, pp. 291-312.

Barry, C. et al. (2010), "Do examinees have similar test taking effort? A high-stakes question for low-stakes testing", *International Journal of Testing*, Vol. 10/4, pp. 342–363.

Baumert, J., and A. Demmrich (2001), "Test motivation in the assessment of student skills: The effects of incentives on motivation and performance", *European Journal of Psychology of Education*, Vol. 16/3, pp. 441–462.

Beaton, A. E., and E.J. Gonzales (1993), *Comparing the NAEP Trial State Assessment Results with the IAEP International Results, National Academy of Education Panel on the NAEP Trial State Assessment (Research Report)*.

Becher, T. (1994), "The significance of disciplinary differences", *Studies in Higher Education*, Vol. 19/2, pp. 151-161.

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

Beilstein, C. D. and J.F. Wilson (2000), "Landmarks in route learning by girls and boys", *Perceptual and Motor Skills*, Vol.91/3, pp. 877-882.

Blossfeld, H.-P., T. Schneider and J. Doll (2009), "Methodological advantages of panel studies: Designing the new National Educational Panel Study (NEPS) in Germany", *Journal for Educational Research Online*, Vol. 1/1, pp. 10-32.

Borgonovi, F. and P. Biecek (2016), "An international comparison of students' ability to endure fatigue and maintain motivation during a low-stakes test", *Learning and Individual Differences*, Vol. 49, pp. 128–137.

Bradbury, B. et al (2015), *Too Many Children Left Behind: The US Achievement Gap in Comparative Perspective*, Russell Sage Foundation.

Braun, H., I. Kirsch and K. Yamamoto (2011), "An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment", *Teachers College Record*, Vol. 113/11, pp. 2309–2344.

Brunello, G. and D. Checchi (2007), “Does school tracking affect equality of opportunity? New international evidence”, *Economic Policy*, Vol. 22/52, pp. 782–861, http://dx.doi.org/doi:10.1111/j.1468-0327.2007.00189.x.

Buchmann, C. (2002), “Measuring family background in international studies of education: Conceptual issues and methodological challenges”, in Porter, A.C. and Gamoran, A. (Eds.) *Methodological Advances in Cross-national Surveys of Educational Achievement.*, pp. 150-197, National Academy Press, Washington, DC.

Bynner, J. and S. Parsons (2009), “Insights into basic skills from a UK longitudinal study”, in Reder, S., and Bynner J. (Eds.), *Tracking Adult Literacy and Numeracy Skills–Findings from Longitudinal Research*, pp. 27-58. Routledge, New York.

Campbell, R. (1983), “Status attainment research: End of the beginning or beginning of the end?”, *Sociology of Education*, Vol. 56/1, pp. 47-62.

Cardozo, S. (2009), “Experiencias laborales y deserción en la cohorte de estudiantes evaluados por pisa 2003 en Uruguay: nuevas evidencias”, *REICE- Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Vol. 7/4, pp. 198-218.

Cartwright, F. et al. (2003), *Linking Provincial Student Assessments with National and International Assessments*, Statistics Canada.

Castelli, C., L. Colazzo and A. Molinari (1998), “Cognitive variables and patterns of hypertext performances: Lessons learned for educational hypermedia construction”, *Journal of Educational Multimedia and Hypermedia*, Vol. 7, pp. 177-206.

Chan, D. et al. (1997), “Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation”, *Journal of Applied Psychology*, Vol. 82/2, pp. 300–310.

Clark, B. R. (1986), *The Higher Education System: Academic Organization in Cross-national Perspective*, University of California Press.

Cole, J. S., D.A. Bergin and T.A. Whittaker (2008), “Predicting student achievement for low-stakes tests with effort and task value”, *Contemporary Educational Psychology*, Vol. 33, pp. 609–624.

Coleman, J.S. et al. (1966), *Equality of Educational Opportunity*, U. S. Government Printing Office, Washington, DC.

Cooper, H. et al. (1996), “The effects of summer vacation on achievement test scores: A narrative and meta-analytic review”, *Review of Educational Research*, Vol. 66/3, pp. 227–268.

Croll, P. (2008), “Occupational choice, socio-economic status and educational attainment: A study of the occupational choices and destinations of young people in the British household panel survey”, *Research Papers in Education*, Vol. 23, pp. 243-268.

Cronbach, L. J. (1970), *Essentials of Psychological Testing* (3rd ed.), Harper and Row, New York.

Danish Ministry of Education (2014), *Summary of the Danish PISA-PIAAC Survey*, *Research Report*, http://eng.uvm.dk/-/media/UVM/Filer/English/PDF/140626-PISA-PIAAC-engelsk-sammenfatning.ashx.

Desjardins, R. and A. Warnke (2012), "Ageing and skills: A review and analysis of skill gain and skill loss over the lifespan and over time", *OECD Education Working Papers*, No. 72, OECD Publishing, Paris. http://dx.doi.org/10.1787/5k9csvw87ckh-en.

DiPrete, T. A., and C. Buchmann (2013), *The Rise of Women: The growing Gender Gap in Education and What it Means for American schools*, Russell Sage Foundation.

Domański, H. et al. (2012), "From school to work: Individual and institutional determinants of educational and occupational career trajectories of young poles" *ASK, Research&Methods*, Vol. 21, pp. 123-144.

Dorans, N. J., T. P. Moses and D.R. Eignor (2011), "Equating test scores: Toward best practices, Statistical models for test equating, scaling, and linking" (pp. 21-42), in Von Davier, M. and A. Alina (Eds.) *Statistical Models for Test Equating, Scaling, and Linking*, Springer, New York.

Downey, D. B. and.D. J. Condron (2016), "Fifty years since the Coleman Report rethinking the relationship between schools and inequality", *Sociology of Education*.

Downey, D. B., P.T. von Hippel, and B. Broh (2004), "Are schools the great equalizer? School and non-school sources of inequality in cognitive skills*", American Sociological Review*, Vol. 69/5, pp. 613-635.

Dreyfus, H. and S. Dreyfus (1986), *Mind over Machines*, The Free Press, New York.

Eccles, J. S. and A. Wigfield (2002), "Motivational beliefs, values, and goals", *Annual Review of Psychology*, Vol. 53, pp. 109–132.

Efron, B. (1982), "The jackknife, the bootstrap and other resampling plans", *Society for Industrial and Applied Mathematics Series*, Volume 38, Stanford University, Stanford. http://dx.doi.org/10.1137/1.9781611970319.

Eklöf, H. (2010), "Skill and will: Test-taking motivation and assessment quality", *Assessment in Education: Principles, Policy and Practice*, Vol. 17/4, pp. 345–356.

Else-Quest, N.M., J.S. Hyde and M.C. Linn (2010), "Cross-national patterns of gender differences in mathematics: A meta-analysis", *Psychological Bulletin*, Vol. 136/1, pp.103–127.

Feuer, M. J. et al. (1998), *Equivalence and Linkage of Educational Tests*, National Academies Press, Washington D.C.

Fortin, N.M., P. Oreopoulos, S. Phipps, S. (2015), "Leaving boys behind gender disparities in high academic achievement", *Journal of Human Resources*, Vol. 50/3, pp. 549-579.

Francis, B. (2000), *Boys, Girls, and Achievement: Addressing the Classroom Issue*, Routledge/Falmer Press, London.

Fruehwirth, J.C., S. Navarro and Y. Takahashi (2016), "How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects", *Journal of Labor Economics*, Vol. 34/4, pp. 979–1021, http://doi.org/10.1086/686262.

Fryer, R.G., Jr., and S.D. Levitt (2010), "An empirical analysis of the gender gap in mathematics", *American Economic Journal: Applied Economics*, Vol. 2, pp. 210–240.

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

Gal, I. and D. Tout (2014), “Comparison of PIAAC and PISA frameworks for numeracy and mathematical literacy”, *OECD Education Working Papers*, No. 102, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jz3wl63cs6f-en.

Goyette, K. A. (2008), “College for some to college for all: Social background, occupational expectations, and educational expectations over time”, *Social Science Research*, Vol. 37/2, pp. 461-484.

Grotlueschen, A. and W. Riekmann (2012), *Funktionaler Analphabetismus in Deutschland*, Ergebnisse der ersten leo. - Level-One Studie, Muenster: Waxmann Verlag GmbH.

Guiso, L. et al. (2008), “Culture, gender, and math”, *Science*, Vol. 320/5880, pp. 1164–1165.

Haberman, S. J. (2015), “Pseudo-equivalent groups and linking”, *Journal of Educational and Behavioral Statistics*, Vol. 40/3, pp. 254-273.

Haberman, S. J. (1984), “Adjustment by minimum discriminant information”, *The Annals of Statistics*, Vol. 12/3, pp. 971-988.

Halpern, D.F. (2012), *Sex Differences in Cognitive Abilities*, 4th edition, Psychology Press, Taylor and Francis Group, New York, NY.

Halpern, D. F. et al. (2007a), “The science of sex differences in science and mathematics”, *Psychological Science in the Public Interest*, Vol. 8/1, pp. 1-51.

Halpern, D. F., et al. (2007b), “Sex, math and scientific achievement”, *Scientific American Mind*, Vol. 18/6, pp. 44-51.

Halpern, D. F., and M. L. LaMay (2000), “The smarter sex: A critical review of sex differences in intelligence”, *Educational Psychology Review*, Vol. 12/2, pp. 229-246.

Hanushek, E.A. and L. Woessmann (2006), “Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries”, *The Economic Journal*, Vol. 116/510, pp. 63-76, http://dx.doi.org/doi:10.1111/j.1468-0297.2006.01076.x.

Heyns, B. (1978), *Summer Learning and the Effects of Schooling*, Academic Press, New York.

Hogan, J. and Holland, B. (2003), “Using theory to evaluate personality and job–performance relations: A socioanalytic perspective”, *Journal of Applied Psychology*, Vol. 88, pp. 100 –112.

Holland, P. W. and N.J. Dorans (2006), “Linking and equating”, in R. L. Brennan (Ed.), *Educational Measurement* (pp. 189–220), Westport, Praeger.

Hoyles, C. et al. (2002), *Mathematical Skills in the Workplace, Science*, Technology and Mathematics Council, London, http://eprints.ioe.ac.uk/1565/1/Hoyles2002MathematicalSkills.pdf.

Hyde, J.S. and J.E. Mertz (2009), “Gender, culture, and mathematics performance”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106/22, pp. 8801–8807.

International Labour Office (2012), International Standard Classification of Occupations (ISCO-08), Geneva.

International Labour Office (1990), International Standard Classification of Occupations (ISCO-88), Geneva.

Jablonka, E. (2003), “Mathematical literacy”, in *Second International Handbook of Mathematics Education* (pp. 75–102), Dordecht, Kluwer, Netherlands.

Jacobs, J. et al. (2002), “Ontogeny of children’s self-beliefs: Gender and domain differences across grades one through 12”, *Child Development*, Vol. 73, pp. 509-527.

Jencks, C. et al. (1972), *Inequality: A Reassessment of the Effect of Family and Schooling in America*, Basic Books, New York.

Jia, Y. et al. (2014), *2011 NAEP-TIMSS Linking Study: Technical Report on the Linking Methodologies and Their Evaluations* (NCES 2014-461), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Johnson, E. G. (1998), *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): A Technical Report, Research and Development Report*, ETS.

Johnson, E.G. et al. (2003), *2000 NAEP–1999 TIMSS Linking Report. (NCES 2005-01)*, National Center for Education Statistics, U.S. Department of Education, Washington, DC.

Kane, J.M. and J.E. Mertz (2012), “Debunking myths about gender and mathematics performance”, *Notices of the American Mathematical Society*, Vol. 59/1, pp.10–21.

Kaplan, D. and A. Turner (2013), “Data fusion with international large scale assessments: A case study using the OECD PISA and TALIS surveys”, *Large-scale Assessments in Education*, Vol. 1/1, pp. 1-26.

Kimball, M. M. (1989), “A new perspective on women's math achievement”, *Psychological Bulletin*, Vol. 105/2, pp. 198.

Kolen, M. J. and R.L. Brennan (2004), *Test Equating, Scaling, and Linking*, Springer, New York.

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*, Vol. 8/1, pp. 69-79.

Lawrence, I. M., and N.J. Dorans (1990), “Effect on equating results of matching samples on an anchor test”, *Applied Measurement in Education*, Vol. 3/1, pp. 19-36.

Lee, H. (2007), “Instructional design of web-based simulations for learners with different levels of spatial ability”, *Instructional Science*, Vol. 35, pp. 467-479.

Lindberg, S. M. et al. (2010), “New trends in gender and mathematics performance: A meta-analysis”, *Psychological Bulletin*, Vol. 136/6, pp. 1123.

Linn, R. L. (1975), “Anchor test study: The long and the short of it”, *Journal of Educational Measurement*, Vol. 12/3, pp. 201-2014.

Little, A. (1978), *The Occupational and Educational Expectations of Students in Developed and Less-Developed Countries*, Sussex University, Institute of Development Studies, Sussex.

Machin, S. and S. McNally (2005), “Gender and student achievement in English schools”, *Oxford Review of Economic Policy*, Vol. 21/3, pp. 357-372.

Machin, S., and Pekkarinen, T. (2008), “Global sex differences in test score variability”, *Science*, 322(5906), pp. 1331-1332.

Marks, G.N. (2010), “Meritocracy, modernization and students‘ occupational expectations: Cross-national evidence”, *Research in Social Stratification and Mobility*, Vol. 28, pp. 275-289.

McDaniel, A. (2010), “Cross-national gender gaps in educational expectations: The influence of national-level gender ideology and educational systems”, *Comparative Education Review*, Vol. 54, pp. 27-50.

McLoyd, V. (1998), “Socioeconomic disadvantage and child development”, *American Psychologist*, Vol 53, pp. 185–204.

Merry, J. J. (2013), “Tracing the US deficit in PISA reading skills to early childhood evidence from the United States and Canada”, *Sociology of Education*.

Messick, S. (1989), “Validity”, in Linn, R.L. (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). American Council on Education/Macmillan, New York.

Miranda, V. (2011), “Cooking, caring and volunteering: unpaid work around the world”, *OECD Social, Employment and Migration Working Papers*, No. 116, OECD Publishing. http://dx.doi.org/10.1787/5kghrjm8s142-en.

Morgan, S. (2005), *On the Edge of Commitment: Educational Attainment and Race in the United States*, Stanford University Press, Stanford, CA.

Mullis, I. V. et al. (2012a), *TIMSS 2011 International Results in Mathematics*, International Association for the Evaluation of Educational Achievement, Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Mullis, I. V. et al. (2012b), *PIRLS 2011 International Results in Reading*, International Association for the Evaluation of Educational Achievement, Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Oakhill, J. V. and A. Petrides (2007), “Sex differences in the effects of interest on boys' and girls' reading comprehension”, *British Journal of Psychology*, Vol. 98/2, pp. 223-235.

OECD (forthcoming), *Technical Report of the Survey of Adult Skills*, Second Edition.

OECD (2016a), *Education at a Glance 2016: OECD Indicators*, OECD Publishing, Paris. http://dx.doi.org/10.1787/eag-2016-en.

OECD (2016b), *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264258051-en.

OECD (2016c), *Survey of Adult Skills (PIAAC)* (Database 2012, 2015), www.oecd.org/skills/piaac/publicdataandanalysis/.

OECD (2016d), *PISA 2018 - Draft Analytical Frameworks* , www.oecd.org/pisa/pisaproducts/PISA-2018-draft-frameworks.pdf.

OECD (2015a), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229945-en.

OECD (2015b), *Immigrant Students at School: Easing the Journey towards Integration*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264249509-en.

OECD. (2014a), PISA 2012 Technical Report, www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OECD (2014b), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208780-en.

OECD (2013a), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

OECD (2013b), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264190511-en.

OECD (2013c), *PISA 2012 Released Mathematics Items*, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

OECD (2013d), *The Survey of Adult Skills: Reader's Companion*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204027-en.

OECD (2013e), *PISA 2012 Results: Excellence through Equity (Volume II): Giving Every Student the Chance to Succeed*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264201132-en.

OECD (2012a), Survey of Adult Skills (PIAAC) Database 2012, www.oecd.org/skills/piaac/publicdataandanalysis/

OECD (2012b), *Grade Expectations: How Marks and Education Policies Shape Students' Ambitions*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264187528-en.

OECD (2012c), PISA database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm

OECD (2012d), *Learning beyond Fifteen: Ten Years after PISA*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264172104-en.

OECD (2011), *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI)*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264112995-en.

OECD (2010a), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264062658-en.

OECD (2010b), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264091450-en

OECD (2004a), "Improving Skills for More and Better Jobs: Does Training Make a Difference?", in *OECD Employment Outlook 2004*, OECD Publishing, Paris. http://dx.doi.org/10.1787/empl_outlook-2004-6-en

OECD (2004b), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264101739-en.

OECD (2003), PISA database, www.oecd.org/pisa/data/database-pisa2003.htm.

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264195905-en.

OECD (2000), PISA database, www.oecd.org/pisa/data/database-pisa2000.htm.

OECD (1999), *Measuring Student Knowledge and Skills: A New Framework for Assessment*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264173125-en.

Paccagnella, M. (2016), "Age, Ageing and Skills: Results from the Survey of Adult Skills", *OECD Education Working Papers*, No. 132, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jm0q1n38lvc-en.

Paechter, C. (1998), Educating the Other: Gender, Power, and Schooling, Falmer, London.

Paek, I., J. Liu and H. Oh (2006), *Investigation of Propensity Score Matching on Linear/nonlinear Equating Method for the PSAT/NMSQT (ETS Statistical Report SR-2006-55)*, Educational Testing Service, Princeton, NJ.

Pashley, P. J. and G.W. Phillips (1993), *Toward World-class Standards: A Research Study Linking International and National Assessments*, Educational Testing Service, Princeton, NJ.

PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Papers*, No. 34, OECD Publishing, Paris. http://dx.doi.org/10.1787/220348414075.

PIAAC Numeracy Expert Group (2009), "PIAAC Numeracy: A Conceptual Framework", *OECD Education Working Papers*, No. 35, OECD Publishing, Paris. http://dx.doi.org/10.1787/220337421165.

Peaker, G. (1971), *The Plowden Children Four Years Later*, National Foundation for Educational Research in England and Wales, London.

Penner, A. M. and M. Paret (2008), "Gender differences in mathematics achievement: Exploring the early grades and the extremes", *Social Science Research*, Vol. 37/1, pp. 239-253.

Perna, L. W. (2000), “Differences in the decision to attend college among african americans, hispanics, and whites”, *Journal of Higher Education*, pp. 117-141.

Phillips, G.W. (2009), *The Second Derivative: International Benchmarks in Mathematics for American States and School Districts*, American Institutes for Research, Washington, DC.

Pope, D.G. and J.R. Sydnor (2010), "Geographic variation in the gender differences in test scores", *The Journal of Economic Perspectives*, Vol. 24/2. pp. 95-108.

Radwan, N. and Y. Xu (2012), *Comparison of the Performance of Ontario Students on the OSSLT/TPCL and the PISA 2009 Reading Assessment*, Northwest Evaluation Association.

Rässler, S. (2012), "Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches", *Lecture Notes in Statistics* Vol. 168. Springer Science and Business Media.

Rässler, S. (2004), "Data fusion: identification problems, validity, and multiple imputation", *Austrian Journal of Statistics*, Vol. 33/1-2, pp. 153-171.

Rasmusson, M. and L. Åberg-Bengtsson (2014), "Does performance in digital reading relate to computer game playing? A study of factor structure and gender patterns in 15-year-olds' reading literacy performance", *Scandinavian Journal of Educational Research*, Vol. 59/6, pp. 1-19.

Reilly, D. (2012), *Gender, Culture, and Sex-typed Cognitive Abilities*, PLoS ONE, 7, e39904.

Reynolds, J. et al. (2006), "Have adolescents become too ambitious? High school seniors' educational and occupational plans, 1976 to 2000", *Social Problems*, Vol. 53, pp. 186-206.

Rindermann, H. (2007), "The g-factor in international cognitive ability comparisons: The homogeneity of results in PISA, TIMMS PIRLS and IQ tests across nations", *European Journal of Personality*, Vol. 21/5, pp. 667–706.

Ríos González A. (2014), "Calendario y determinantes de riesgo educativo: la cohorte PISA 2006-2011 en Uruguay" [Timing and determinants of the fall in educational risk: PISA 2006-2011 cohort in Uruguay], *Revista de Ciencias Sociales*, Vol. 35, pp. 109-136.

Rosdahl, A. (2014), Fra 15 til 27 år. PISA 2000-eleverne I 2011/12 (From 15 to 27 years. The PISA 2000-students in 2011/12), SFI-Rapport 14:13, SFI – Det Nationale Forskningscenter for Velfærd, Copenhagen.

Rosen, M. (2001), "Gender differences in reading performance on documents across countries", *Reading and Writing: An Interdisciplinary Journal*, Vol. 14, pp. 1-38.

Rosenbaum, P. R. and D.B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, Vol. 70/1, pp. 41-55.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies", *Journal of educational Psychology*, Vol. 66/5, pp. 688-701.

Ruble, D. N., C. L. Martin and S.A. Berenbaum (2006), "Gender development", in W. Damon and N. Eisenberg, *Handbook of Child Psychology*.

Ruhose, J. and G. Schwerdt (2016), "Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries", *Economics of Education Review*, Vol. 52, pp. 134–54. https://dx.doi.org/10.1016/j.econedurev.2016.02.004.

Sabatini, J. and K. Bruce (2009), "PIAAC Reading Component: A Conceptual Framework", *OECD Education Working Papers*, No. 33, OECD Publishing, Paris. http://dx.doi.org/10.1787/220367414132.

Scharenberg et al. (2014), *Education Pathways from Compulsory School to Young Adulthood: The First Ten Years*, TREE, Basel.

Schwabe, F., N. McElvany and M. Trendtel (2015), "The school age gender gap in reading achievement: Examining the influences of item format and intrinsic reading motivation", *Reading Research Quarterly*, Vol. 50/2, pp. 219-232.

Sewell, W. et al. (2003), "As we age: A review of the Wisconsin longitudinal study, 1957-2001", *Research in Social Stratification and Mobility*, Vol. 20, pp. 3-111.

Sikora, J. and L. J. Saha (2009), "Gender and professional career plans of high school students in comparative perspective", *Educational Research and Evaluation*, Vol. 15, pp. 387-405.

Sikora, J. and L. J. Saha (2007), "Corrosive inequality? Structural determinants of educational and occupational expectations in comparative perspective", *International Education Journal: Comparative Perspectives*, Vol. 8, pp. 57-78.

Sirin, S. R. (2005), "Socioeconomic status and academic achievement: A meta-analytic review of research", *Review of Educational Research*, Vol. 75/3, pp. 417-453.

Smith, M. W. and J. Wilhelm (2006), *Going with the Flow: How to Encourage Boys (and Girls) in their Literacy Learning*, Heinemann, Portsmouth, NH.

Smith, M.W. and J. Wilhelm (2002), *Reading don't Fix no Chevys: Literacy in the Lives of Young Men*, Heinemann, Portsmouth, NH.

Snyder, T. D., S.A. Dillow and C.M. Hoffman (2009), *Digest of Education Statistics 2008, (NCES 2009-020)*, National Center for Education Statistics, Institute of Education Sciences, US Department of Education, Washington DC.

Spelke, E. S. (2005), "Sex differences in intrinsic aptitude for mathematics and science?: A critical review", *American Psychologist*, Vol. 60/9, pp. 950.

Statistics Canada (2003), International Adult Literacy and Skills Survey (ALL) Canada, www5.statcan.gc.ca/olc-cel/olc.action?ObjId=89M0022X&ObjType=2&lang=en&limit=0 (accessed 1 March 2017).

Steen, L. A. (1990), *On the Shoulders of Giants: New Approaches to Numeracy*, National Research Council, Washington, D.C.

Stoet, G. and D.C. Geary (2013), *Sex Differences in Mathematics and Reading Achievement are Inversely Related: Within- and Across-nation Assessment of 10 years of PISA Data*, PLoS ONE, 8, e57988.

Straesser, R. (2003), *Mathematics at Work: Adults and Artefacts. In Learning Mathematics to Live and Work in our World: Proceedings of the 10th International Conference on Adults Learning Mathematics*, Johannes Kepler Universitaet Linz, Linz, Austria.

Thomson, S. and K. Hillman (2010), *Against the Odds. Influences on the Post-school Success of 'Low Performers'*, NCVER Publishing.

Tout, D. (2000), *What is Numeracy? What is Mathematics?,* Presented at the International Congress on Mathematics Education 2000, Tokyo, Japan.

van de Werfhorst, H.G. and J.J. Mijs (2010), "Achievement inequality and the institutional structure of educational systems: A comparative perspective", *Annual Review of Sociology*, Vol. 36, pp. 407-428, http://dx.doi.org/10.1146/annurev.soc.012809.102538.

von Davier, M. E. J. von, Gonzales and R. J. Mislevy (2009), "What are plausible values and why are they useful?", *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, Vol. 2, pp. 9–36.

Wagner, R. K. et al. (2010), *Test of Sentence Reading Efficiency and Comprehension*, Pro-Ed, Austin, TX.

Warrington, M., M. Younger and J. Williams (2000), "Students' attitudes, image, and the gender gap", *British Educational Research Journal*, Vol. 26, No. 3, pp. 393-407.

White, H. (1982), "Regularity conditions for Cox's test of non-nested hypotheses", *Journal of Econometrics*, Vol. 19/2, pp. 301-318.

Wise, S. L. and X. Kong (2005), "Response time effort: A new measure of examinee motivation in computer-based tests", *Applied Measurement in Education*, Vol. 18/2, pp. 163–183.

Wolf, L. F. and J.K Smith (1995), "The consequence of consequence: motivation, anxiety, and test performance", *Applied Measurement in Education*, Vol. 8/3, pp. 227–242.

Wu, M. (2009), "A comparison of PISA and TIMMS 2003 achievement results in mathematics", *Prospects*, Vol. 39/1, pp. 33–46.

Wu, M. (2004), "Plausible values", *Rasch Measurement Transaction*, Vol. 18/2, pp. 976–978.

Wylie, C. and E. Hodgen (2011), *Forming Adulthood. Past, Present and Future in the Experiences and Views of the Competent Learners*, New Zealand Council for Educational Research (NZCER), Wellington.

Yamamoto, K. (2002), *Estimating PISA Student Scores on the IALS Prose Literacy Scale. Technical Report*, Educational Testing Service, Princeton, NJ. www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33680659.pdf.

Yu, L. et al. (2004), *Investigating Differences in Examinee Performance Between Computer-based and Handwritten Essays, (ETS Research Report RR-04-18)*, Educational Testing Service, Princeton, NJ.

Zhang, H. and G. Salvendy (2001), "The implications of visualisation ability and structure preview design for web information search tasks", *International Journal of Human-Computer Interaction*, Vol. 13/1, pp. 75-95.

Zhou, M. (2014), "Gender difference in web search perceptions and behavior: Does it vary by task performance?", *Computers and Education*, Vol. 78, pp. 174-184.

ANNEX A – MEASURING READING IN PISA AND LITERACY IN PIAAC

Although proficiency levels are used to report results in both PISA and PIAAC, in the first PISA rounds from 2000 till 2009, the literacy scale was divided into five levels. Following 2009, the lowest level was renamed as Level 1a while a new Level 1b for those who scored below Level 1 (< 335) was introduced. Additionally, the scale was expanded to Level 6 for those scoring at the top of the scale.

By dividing the continuous measurement scales into different proficiency levels, it is possible to get a description of the characteristics of the tasks that adults estimated to perform at a certain level can solve. Table A.1 provides an overview of the proficiency levels in literacy used in PISA and PIAAC (OECD, 2002; PIAAC Literacy Expert Group, 2009).

Table A.1 Comparison of the PISA 2000 and PIAAC literacy proficiency levels

PISA 2000 reading literacy levels (print reading)			PIAAC literacy levels		
			Below Level 1	< 176	Below Level 1 tasks require the respondent to read brief texts on familiar topics to locate a single piece of information.
Level 1	335 - 407	The reader must: locate one or more independent pieces of explicitly stated information according to a single criterion; identify the main theme or author's purpose in a text about a familiar topic; or make a simple connection between information in the text and common, everyday knowledge. Typically, the requisite information is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.	Level 1	176 - 225	At Level 1, tasks require the respondent to read relatively short digital or print continuous, non-continuous, or mixed texts to locate a single piece of information that is identical to or synonymous with the information given in the question or directive.
Level 2	408 - 480	The reader must: locate one or more pieces of information that may be needed to meet multiple criteria; identify the main idea, understand relationships or construe meaning within a limited part of the text by making low-level inferences; form or apply simple categories to explain something in a text by drawing on personal experience and attitudes; or make connections or comparisons between the text and everyday outside knowledge. The reader must often deal with competing information.	Level 2	226 - 275	At Level 2, tasks require respondents to make matches between the text and information, and may require paraphrasing or low-level inferences.

Table A.1 Comparison of the PISA 2000 and PIAAC literacy proficiency levels (*continued*)

PISA 2000 reading literacy levels (print reading)			PIAAC literacy levels		
Level 3	481 - 552	The reader must: recognise the links between pieces of information that have to meet multiple criteria; integrate several parts of a text to identify a main idea, understand a relationship or construe the meaning of a word or phrase; make connections and comparisons; or explain or evaluate a textual feature. The reader must take into account many features when comparing, contrasting or categorising. Often the required information is not prominent but implicit in the text or obscured by similar information.	Level 3	276 – 325	At Level 3, texts are often dense or lengthy, and include continuous, non-continuous, mixed, or multiple pages of text. Tasks require the respondent to identify, interpret, or evaluate one or more pieces of information, and often require varying levels of inference.
Level 4	553 - 625	The reader must: locate sequence or combine several pieces of embedded information; infer the meaning of a section of text by considering the text as a whole; understand and apply categories in an unfamiliar context; or hypothesise about or critically evaluate a text, using formal or public knowledge. The reader must draw on an accurate understanding of long or complex texts in which competing information may take the form of ideas that are ambiguous, contrary to expectation, or negatively worded.	Level 4	326 - 375	At Level 4, tasks require respondents to perform multiple-step operations to integrate, interpret, or synthesise information from complex or lengthy continuous, non-continuous, mixed, or multiple type texts.
Level 5	> 625	The reader must: sequence or combine several pieces of deeply embedded information, possibly drawing on information from outside the main body of the text; construe the meaning of linguistic nuances in a section of text; or make evaluative judgements or hypotheses, drawing on specialised knowledge. The reader is generally required to demonstrate a full, detailed understanding of a dense, complex or unfamiliar text, in content or form, or one that involves concepts that are contrary to expectations. The reader will often have to make inferences to determine which information in the text is relevant, and to deal with prominent or extensive competing information.	Level 5	> 375	At Level 5, tasks may require to construct syntheses of similar and contrasting ideas or points of view; or evaluate evidence based arguments. Tasks often require respondents to be aware of subtle, rhetorical cues and to make high-level inferences or use specialised background knowledge.

Source: OECD (2002), *PISA 2000 Technical Report*, www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33688233.pdf; PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Papers*, No. 34, OECD Publishing, Paris. http://dx.doi.org/10.1787/220348414075.

For countries that implemented the optional digital reading assessment in 2009 an additional single scale based only on digital reading tasks was constructed. The metric of that scale was set that the mean and the standard deviation of the equally weighted countries participating in this assessment are the same as these countries' mean and standard deviation in the print assessment (OECD, 2013b). As there were only a small number of items in the digital reading assessment, only four levels were used for reporting proficiency, as can be seen in Table A.2.

Table A.2 The PISA 2009 digital reading proficiency levels

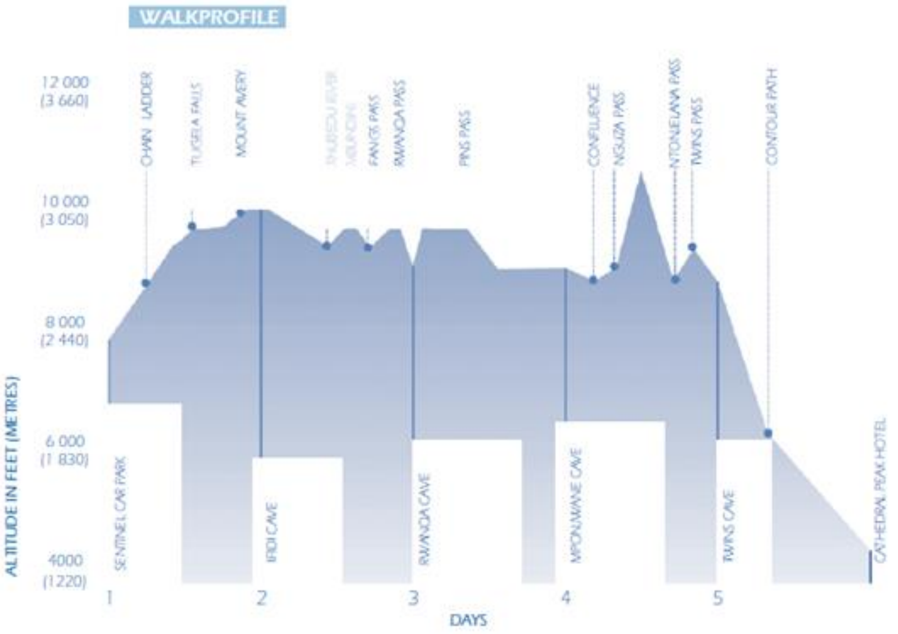
PISA 2009 reading literacy levels (only digital reading)		
Level 2	407 – 480	Tasks at this level typically require the reader to locate and interpret information that is well-defined, usually relating to familiar contexts. They may require navigation across a limited number of sites and the application of web-based navigation tools such as drop-down menus, where explicit directions are provided or only low-level inference is ced for. Tasks may require integrating information presented in different formats, recognising examples that fit clearly defined categories.
Level 3	481 - 552	Tasks at this level require that the reader integrate information, either by navigating across several sites to find well-defined target information, or by generating simple categories when the task is not explicitly stated. Where evaluation is called for, only the information that is most directly accessible or only part of the available information is required.
Level 4	553 - 625	Tasks at this level may require the reader to evaluate information from several sources, navigating across several sites comprising texts in a variety of formats, and generating criteria for evaluation in relation to a familiar, personal or practical context. other tasks at this level demand that the reader interpret complex information according to well-defined criteria in a scientific or technical context.
Level 5	> 625	Tasks at this level typically require the reader to locate, analyse and critically evaluate information, related to an unfamiliar context, in the presence of ambiguity. They require generating criteria to evaluate the text. Tasks may require navigation across multiple sites without explicit direction, and detailed interrogation of texts in a variety of formats.

Source: OECD (2012a), *PISA 2009 Technical Report*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264167872-en.

Examples of literacy items in PISA and PIAAC

Previous sections of this paper described the PISA and PIAAC (reading) literacy assessment frameworks. This subsection shows sample items from the PISA reading literacy and the PIAAC literacy assessments. In order to illustrate commonalities across the two assessments, the sample items chosen can be considered to be "matching" items along the context, medium, reading process, and item response format dimension and therefore can illustrate some of the similarities and differences between the PISA and PIAAC problem tasks items (OECD, 2010, 2013). Chosen sample items displayed in Figures A.1 and A.2 were in the "personal" context category, were initially developed for "paper-and-pencil administration", were part of the item pool reflecting "access and retrieve" and "access and identify" reading process and were part of the pool of items that required a form of "constructed response". Chosen sample items displayed in Figures A.1 and A.2 were in the "personal" context category, were initially developed for "paper-and-pencil administration", were part of the item pool reflecting "access and retrieve" and "access and identify" reading process, and were part of the pool of items that required a form of "selected responses".

Figure A.1 PISA 2009 literacy sample item I



Question 2: AFRICAN TREK

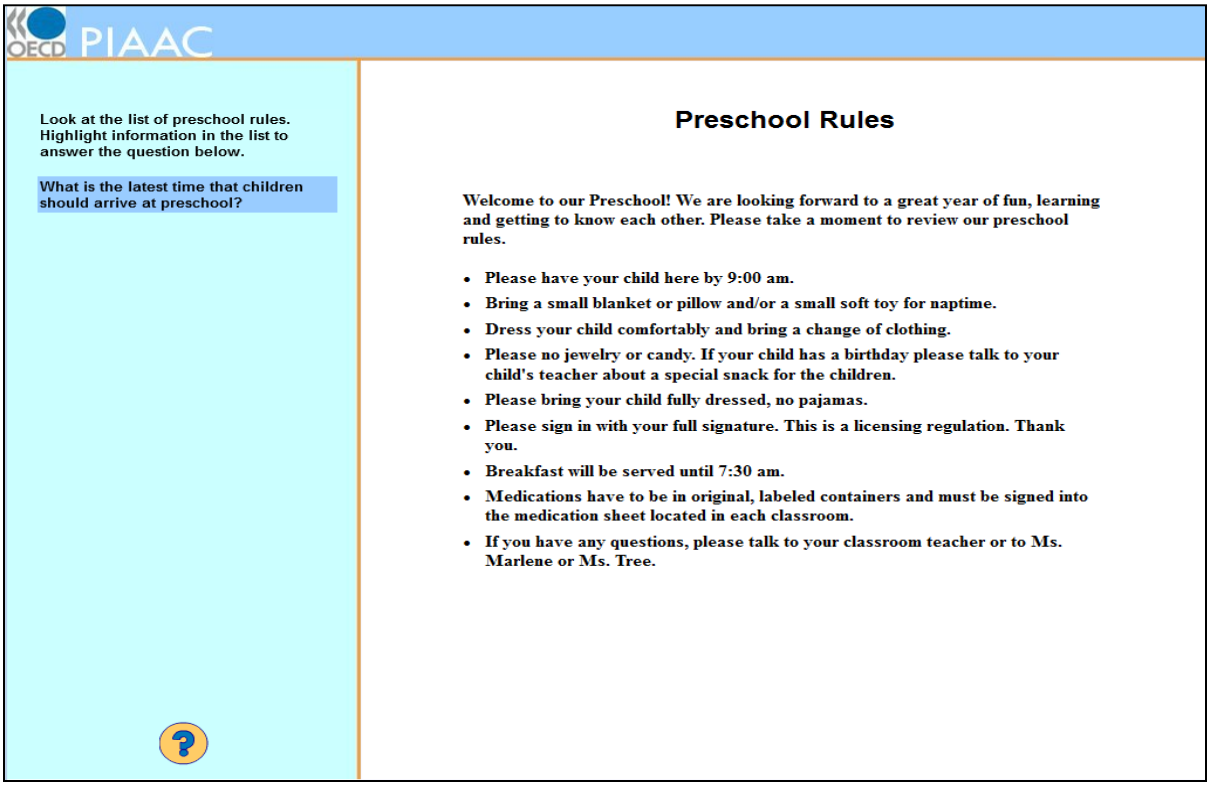
At what altitude is the Sentinel Car Park? Give your answer in feet and metres.

.... feet

.... metres

OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264062658-en.

Figure A.2 PIAAC 2012 literacy sample item I



OECD (2012b), *PIAAC Literacy Sample Items*, www.oecd.org/skills/piaac/Literacy%20Sample%20Items.pdf.

Figure A.3 PISA 2009 literacy sample item II

African Trek

The Northern Drakensberg Trek: South Africa / Lesotho

Fact file

OVERVIEW

- The Northern Drakensberg Trek involves crossing the northern Drakensberg escarpment at high altitudes. The route, which is approximately 40 miles (65km) long, straddles the border between South Africa and Lesotho, taking 5 strenuous days to complete. The Trek is filled with highlights, including breathtaking views over the Amphitheatre to the Devil's Tooth as you make your way to the Chain Ladder, and sunrise seen from Mponjwane, which is well worth setting your alarm for.
- Start: The Sentinel car park, Royal Natal National Park.
- Finish: The Cathedral Peak Hotel.
- Difficulty and Altitude: This is a high-mountain walk in one of the more remote areas of the Drakensberg Range. The going can be quite tough and the days long. Good navigation skills are essential for safe crossing.

Question 9: AFRICAN TREK

On the morning of which day of the trek would you see the sunrise mentioned in the overview?

A. Day 1.

B. Day 2.

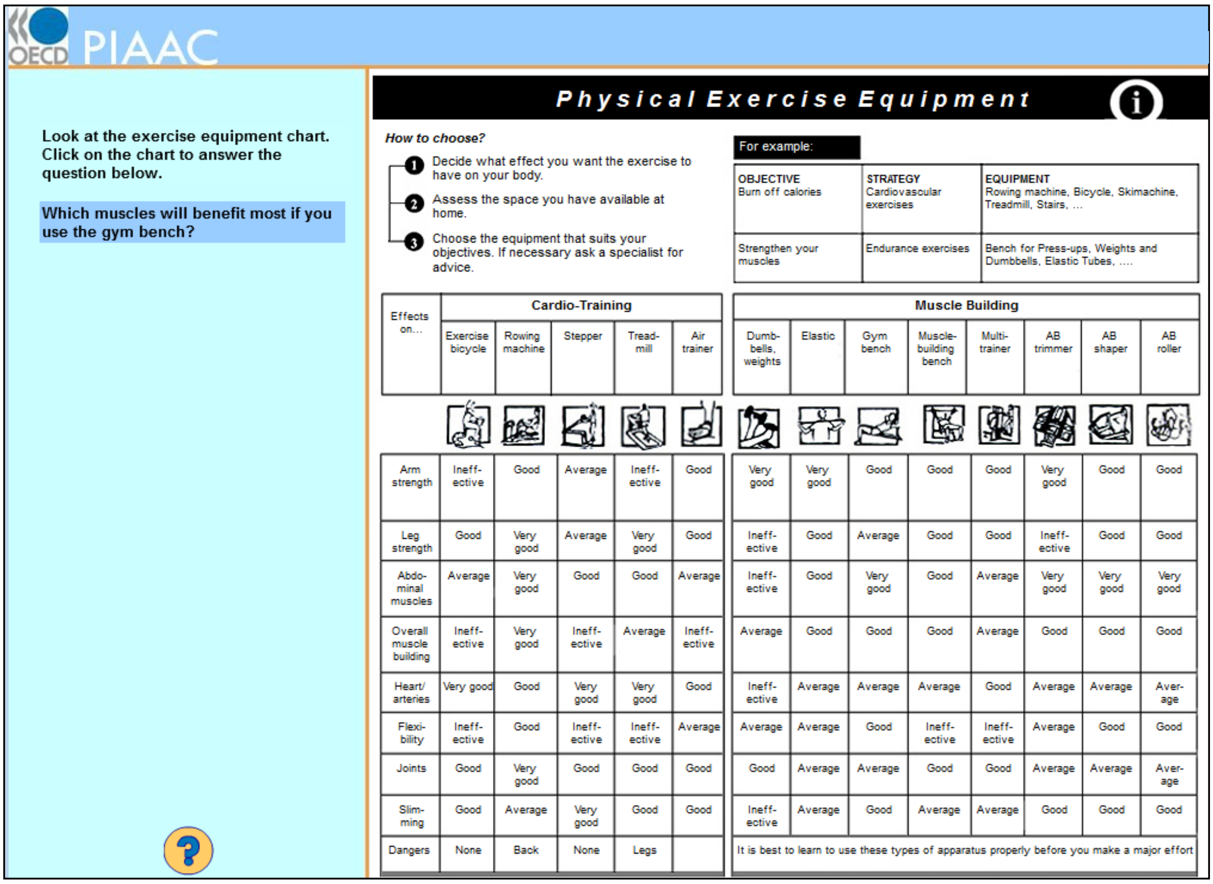
C. Day 3.

D. Day 4.

E. Day 5.

OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264062658-en.

Figure A.4 PIAAC 2012 literacy sample item II



OBJECTIVE	STRATEGY	EQUIPMENT
Burn off calories	Cardiovascular exercises	Rowing machine, Bicycle, Skimachine, Treadmill, Stairs, …
Strengthen your muscles	Endurance exercises	Bench for Press-ups, Weights and Dumbbells, Elastic Tubes, ….

Effects on…	Cardio-Training					Muscle Building							
	Exercise bicycle	Rowing machine	Stepper	Tread-mill	Air trainer	Dumb-bells, weights	Elastic	Gym bench	Muscle-building bench	Multi-trainer	AB trimmer	AB shaper	AB roller
Arm strength	Ineff-ective	Good	Average	Ineff-ective	Good	Very good	Very good	Good	Good	Good	Very good	Good	Good
Leg strength	Good	Very good	Average	Very good	Good	Ineff-ective	Good	Average	Good	Good	Ineff-ective	Good	Good
Abdo-minal muscles	Average	Very good	Good	Good	Average	Ineff-ective	Good	Very good	Good	Average	Very good	Very good	Very good
Overall muscle building	Ineff-ective	Very good	Ineff-ective	Average	Ineff-ective	Average	Good	Good	Good	Average	Good	Good	Good
Heart/ arteries	Very good	Good	Very good	Very good	Good	Ineff-ective	Average	Average	Average	Good	Average	Average	Aver-age
Flexi-bility	Ineff-ective	Good	Ineff-ective	Ineff-ective	Average	Average	Average	Good	Ineff-ective	Ineff-ective	Average	Good	Good
Joints	Good	Very good	Good	Good	Good	Good	Average	Average	Good	Good	Average	Average	Aver-age
Slim-ming	Good	Average	Very good	Good	Good	Ineff-ective	Average	Good	Average	Average	Good	Good	Good
Dangers	None	Back	None	Legs		It is best to learn to use these types of apparatus properly before you make a major effort							

OECD (2012b), *PIAAC Literacy Sample Items*, www.oecd.org/skills/piaac/Literacy%20Sample%20Items.pdf.

REFERENCES

OECD (2013), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264204256-en.

OECD (2012a), *PISA 2009 Technical Report*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264167872-en.

OECD (2012b), *PIAAC Literacy Sample Items*, www.oecd.org/skills/piaac/Literacy%20Sample%20Items.pdf.

OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264062658-en.

OECD (2002), *PISA 2000 Technical Report*, www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33688233.pdf.

PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Papers*, No. 34, OECD Publishing, Paris. http://dx.doi.org/10.1787/220348414075.

ANNEX B – MEASURING MATHEMATICS IN PISA AND NUMERACY IN PIAAC

The PISA scale for mathematical literacy also consists of six levels. However it is not differentiated at the lower but at the higher end and thus, includes level 6 (> 670) instead of level 1b. Table B.1 gives an overview on the levels used in PISA and PIAAC.

Table B.1 PISA 2012 and PIAAC proficiency levels in mathematics and numeracy

PISA mathematical literacy levels			PIAAC numeracy levels		
Level 1	> 357.8	At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined	Below Level 1	< 176	Tasks at this level require the respondents to carry out simple processes such as counting, sorting, performing basic arithmetic operations with whole numbers or money, or recognising common spatial representations in concrete, familiar contexts where the mathematical content is explicit with little or no text or distractors.
Level 2	> 420.1	At Level 2, students can interpret and recognise situations in contexts that require no more than direct inference.	Level 1	176-225	At Level 1, tasks require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit with little text and minimal distractors.
Level 3	> 482.4	At Level 3, students can execute clearly described procedures, including those that require sequential decisions.	Level 2	226-275	At Level 2, tasks require the respondent to identify and act on mathematical information and ideas embedded in a range of common contexts where the mathematical content is fairly explicit or visual with relatively few distractors.
Level 4	> 544.7	At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions	Level 3	276-325	At Level 3, tasks require the respondent to understand mathematical information that may be less explicit, embedded in contexts that are not always familiar and represented in more complex ways.
Level 5	> 607.0	At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions.	Level 4	326-375	At Level 4, tasks require the respondent to understand a broad range of mathematical information that may be complex, abstract or embedded in unfamiliar contexts

Table B.1 PISA 2012 and PIAAC proficiency levels in mathematics and numeracy (*continued*)

PISA mathematical literacy levels			PIAAC numeracy levels		
Level 6	> 669.3	At Level 6, students can conceptualise, generalise and utilise information based on their investigations and modelling of complex problem situations, and can use their knowledge in relatively non-standard contexts.	Level 5	> 375	At Level 5, tasks require the respondent to understand complex representations and abstract and formal mathematical and statistical ideas, possibly embedded in complex texts.

Source: OECD (2016), *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264258051-en; OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264190511-en.

Sample items mathematical literacy/ numeracy

As for literacy, this subsection provides examples of sample items for mathematical literacy or numeracy. These items were chosen because they are the same across the context, mathematical content, mathematical process, and item response format dimensions in both PISA and PIAAC (OECD, 2013b). Figures B.1 and B.2 refer to the following dimensions: “societal” context, “uncertainty and data / data and chance” content, “interpreting” mathematical process and “selected response” response format. Figures B.3 and B.4 refer to the following dimensions: “occupational” context, “space and shape/dimension and space” content, “employing/act upon or use” mathematical process and “constructed response” response format.

Figure B.1 PISA 2012 mathematical literacy sample item I

CABLE TELEVISION

The table below shows data about household ownership of televisions (TVs) for five countries.

It also shows the percentage of those households that own TVs and also subscribe to cable TV.

Country	Number of households that **own TVs**	Percentage of households that **own TVs** compared to **all households**	Percentage of households that **subscribe to cable television** compared to households that **own TVs**
Japan	48.0 million	99.8%	51.4%
France	24.5 million	97.0%	15.4%
Belgium	4.4 million	99.0%	91.7%
Switzerland	2.8 million	85.8%	98.0%
Norway	2.0 million	97.2%	42.7%

Question 1: CABLE TELEVISION

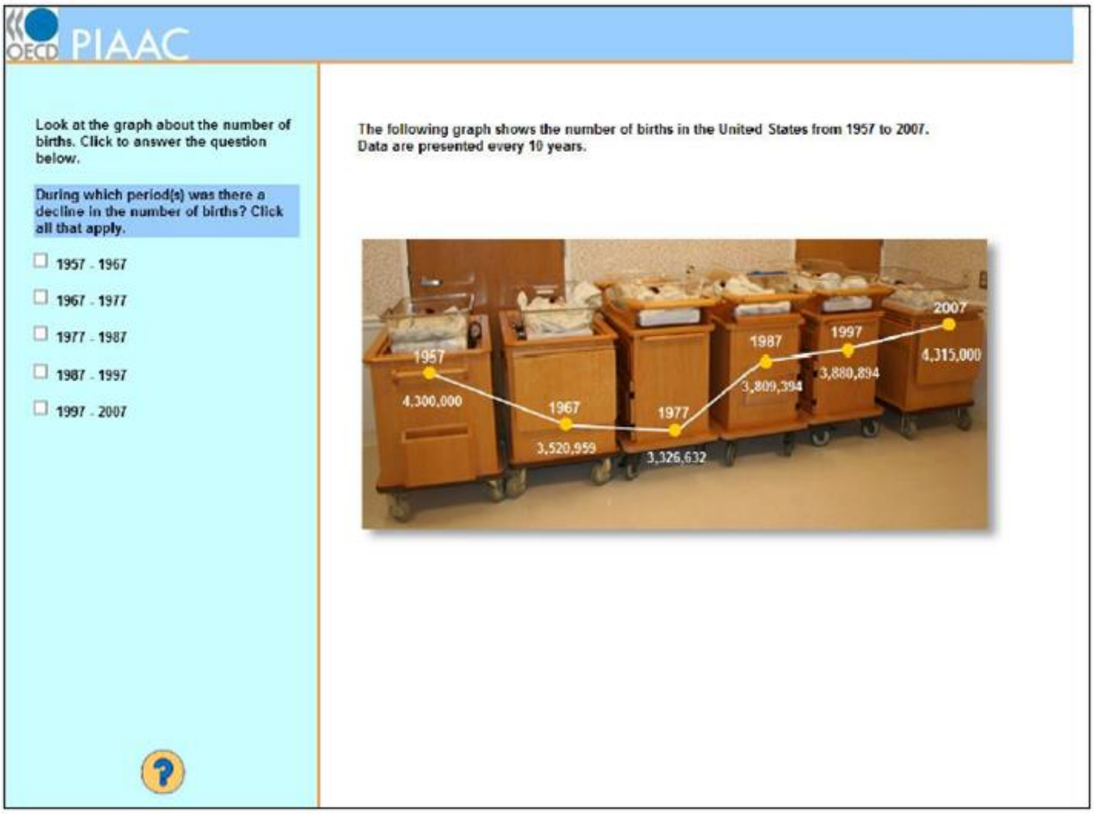
The table shows that in Switzerland 85.8% of all households own TVs.

Based on the information in the table, what is the closest estimate of the total number of households in Switzerland?

A 2.4 million
B 2.9 million
C 3.3 million
D 3.8 million

Source: OECD (2013b), *PISA 2012 Released Mathematics Items*, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Figure B.2 PIAAC 2012 mathematical literacy sample item I

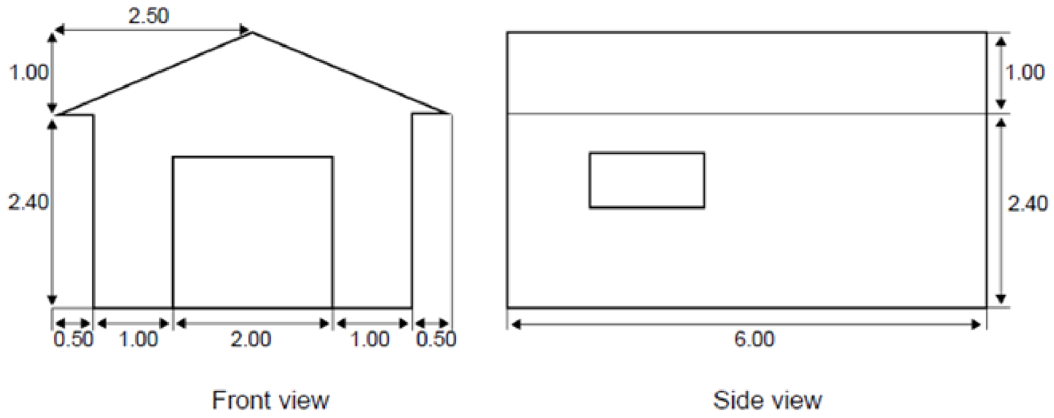


Source: OECD (2012), *PIAAC Numeracy Sample Items,* www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf.

Figure B.3 PISA 2012 mathematical literacy sample item II

Question 2: GARAGE

The two plans below show the dimensions, in metres, of the garage George chose.



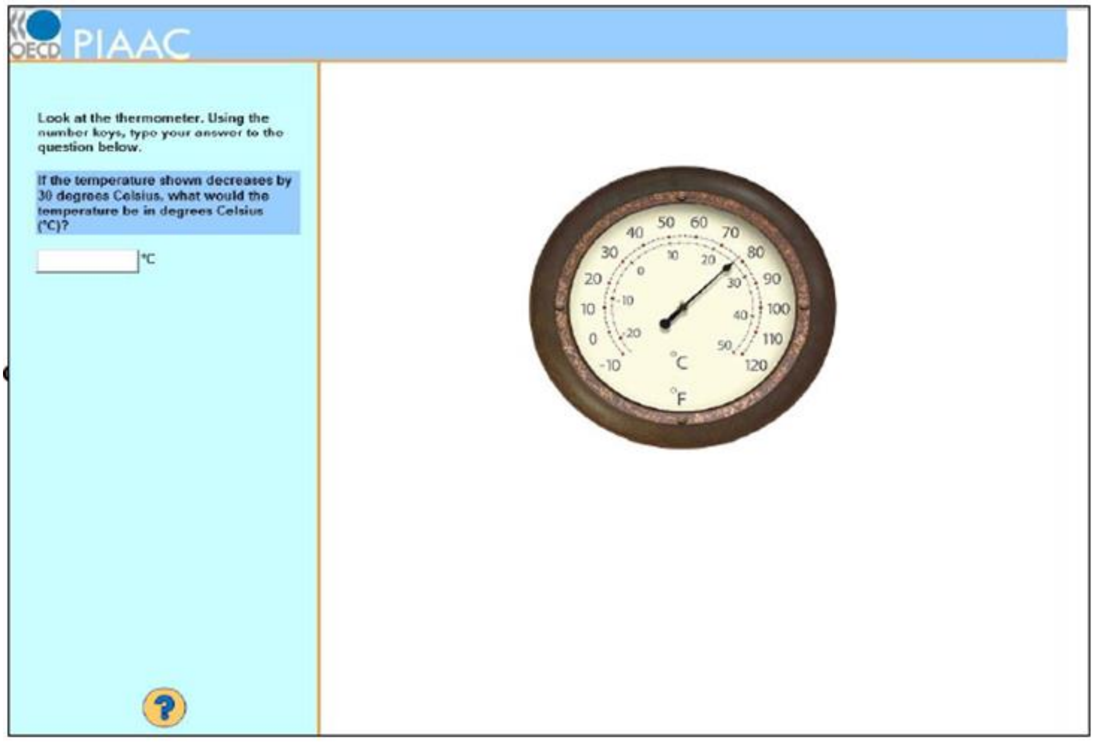
Note: Drawing not to scale.

The roof is made up of two identical rectangular sections.

Calculate the **total** area of the roof. Show your work.

Source: OECD (2013b), *PISA 2012 Released Mathematics Items*, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Figure B.4 PIAAC 2012 mathematical literacy sample item II



Source: OECD (2012), *PIAAC Numeracy Sample Items,* www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf.

REFERENCES

OECD (2016), *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264258051-en.

OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264190511-en.

OECD (2013b), *PISA 2012 Released Mathematics Items*, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

OECD (2012), *PIAAC Numeracy Sample Items*, www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf.