Self-assessments or tests?

Studies in Higher Education, 2015
Vol. 40, No. 3, 482–504, http://dx.doi.org/10.1080/03075079.2015.1004237

Self-assessments or tests? Comparing cross-national differences in patterns and outcomes of graduates' skills based on international large-scale surveys

Martin Humburg* and Rolf van der Velden

The Research Centre for Education and the Labour Market (ROA), Maastricht University, Maastricht, The Netherlands

In this paper an analysis is carried out whether objective tests and subjective self-assessments in international large-scale studies yield similar results when looking at cross-national differences in the effects of skills on earnings, and skills patterns across countries, fields of study and gender. The findings indicate that subjective skills measures do not correlate well with objective measures of similar constructs when looking at cross-national differences. Countrywise associations between subjective skills measures and earnings do not correlate well with those found using objective skills measures. Moreover, cross-national differences in the level of subjective skills measures do not correlate well with cross-national differences in skill levels based on objective tests. Nor do gender differences found using subjective skills measures correlate with those found using objective skills measures. This does not mean that self-assessments cannot be used, but they need to be restricted to analysing within-country differences. Within countries, self-assessments do a good job in predicting skills differences across fields of study and also in predicting the effect of skills on earnings. When comparing gender differences in skills levels within countries, however, one needs to be aware that females tend to overestimate their skills levels in typical 'female' domains like literacy.

Keywords: skills; graduates; self-assessment; test; outcomes; cross-national differences

1. Introduction

Higher education graduates are often seen as the drivers of innovation and technological progress in developed countries (Hanushek and Woessmann 2008; Humburg and van der Velden, forthcoming). Against this background, international large-scale skills surveys are an important instrument to assess the stock of skills of graduates and to investigate how skills are associated with individuals' labour market outcomes.

To date, only very few international large-scale studies on graduates' skills have been conducted. For example, graduate surveys such as the Research into Employment and Professional Flexibility (REFLEX) and its successor the Higher Education as a Generator of Strategic Competences (HEGESCO) survey have assessed graduates' skills in various fields of study and in comparable ways across (mostly) European countries. In the absence of inexpensive and easy-to-administer objective skills

*Corresponding author. Email: martin.humburg@icfi.com

measures, these studies deployed a framework in which graduates were asked to self-assess their level of skills in several domains. These studies document important differences between the skills of graduates from different fields of study and find significant relationships between the level of skills and individuals' labour market outcomes (Allen and van der Velden 2011; Allen, Pavlin, and van der Velden 2011).

The data recently collected in the frame of the Programme for the International Assessment of Adult Competencies (PIAAC; OECD 2013) offer a new opportunity to investigate the stock of graduates' skills and to shed further light on the association between objective and subjective measures of skills at various aggregate levels. In contrast to the REFLEX and HEGESCO surveys which relied on individual self-assessment, the PIAAC survey objectively tested individuals' skills in three domains: numeracy, literacy and problem solving in technology-rich environments.

Focusing on higher education graduates, this paper for the first time compares subjective and objective skills measures obtained through international large-scale studies in relation to:

- cross-national differences in the effect of skills on hourly earnings (separate for males and females);
- skills level patterns across countries;
- skills level patterns across fields of study;
- skills level patterns across gender.

This is done separately for numeracy, literacy and problem solving in technology-rich environments.

The remainder of this paper proceeds as follows. Section 2 presents the conceptual framework and the hypotheses. Section 3 describes the data and the methods used. In Section 4 the results are presented and Section 5 concludes.

2. Conceptual framework

Before elaborating on the different ways one can measure graduates' skills, there is a need to define more accurately what is meant by the term 'skills' and how that differs from the term 'competence' as often used in this type of research. The term 'competence' has been defined as 'the ability to successfully meet complex demands in a particular context through the mobilisation of psychosocial prerequisites (including both cognitive and non-cognitive aspects)' (Rychen and Salganik 2003, 43). A basic characteristic of this definition is its holistic nature: it refers to using different cognitive and non-cognitive components at the same time (knowledge, attitudes, motivation, etc.). Moreover it involves the notion of 'orchestration' (Rychen and Salganik 2003, 45), that is, the ability to use these constituent elements in a deliberate and meaningful way, very much like an orchestra conductor makes different musicians play together. And finally, it refers to the successful completion of an authentic task: competence has to be shown in practice. This last aspect is what most definitions of competence have in common (Van der Klink and Boon 2003; Le Deist and Winterton 2005).

Given this definition, most assessments in large-scale surveys measure skills rather than competencies, that is, they measure one of the constituent elements to successfully meet a complex task. Or at least one can say that large-scale assessments are largely restricted to cognitive components of competence (Rychen and Salganik 2003), which would be an undesirable restriction of the competence concept (Weinert

2001). This is one of the reasons why this paper resorts to using the term 'skills'. Skills are defined as one of the individual components of competencies, not only motoric skills, but also knowledge, attitudes, motivations, etc.[1]

There is probably little doubt that the assessment of skills is best done in a situation where people can really demonstrate their skills levels in different situations that involve carrying out certain tasks, such as an assessment centre (Allen and van der Velden 2005). This usually involves carrying out authentic tasks, the successful completion of which indicates a certain competence level. Although this is often regarded as a 'gold standard' (Ward, Gruppen, and Regehr 2002, 67), the expenses involved preclude any large-scale execution. Testing is therefore often regarded as a good alternative and has gained wide-spread recognition through the initiatives of organisations such as the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Cooperation and Development (OECD). Especially the success of the Programme for International Student Assessment (PISA) has contributed to the recognition that large-scale assessments can give an accurate idea of the stock of skills in a population. However, this is by and large confined to the cognitive domain, such as literacy, science or problem solving. Attempts trying to broaden the scope of such large-scale assessments to soft skills such as team working, have bitterly failed (Murray 2003). Moreover, large-scale assessments involve large amounts of resources both for the development and for the data collection. Average tests involve at least one hour of testing time, usually at school or at home.

This is one of the reasons why researchers have turned to other sources of assessing graduates' skills such as using self-assessments, i.e. asking respondents to report on their own level of skills. Richter and Johnson (2001) discuss some of the advantages of using self-assessments when measuring skills. Clear advantages are that self-assessments are relatively inexpensive, can be applied to a wide range of skills (including hard-to-measure 'soft' skills), require less response burden, and potentially offer access to information that is usually unobserved by a test or an assessor (for example, the respondent's own mind, see Allen and van der Velden 2005).

However, these advantages come at a price. The main problem with self-assessments is that they introduce all kinds of biases, both intentional and unintentional (Allen and van der Velden 2005). A first indication is that objective and subjective assessment of the same performance do not correlate very high. In a meta-analysis of 44 self-assessment studies in higher education, Falchikov and Boud (1989) reported correlations between self-assessed and external measures of performance ranged from −0.05 to 0.82, with a mean correlation of 0.39. Moreover, reports from students and peers or teachers may not coincide (Falchikov and Goldfinch 2000). Others report that the self-bias is relatively stable (Bouffard et al. 2011). Respondents may have an upward-biased idea about their own level or give socially desirable answers, the famous example being that most drivers think that their driving skills is above average. But the assessment may also be unintentionally biased because the concept is unclear or the scale has no clear anchors. This problem arises when the values in the answering scale have no clear intuitive meaning that is the same for all respondents. For example assessing your skill level on a scale ranging from 'low' to 'high' has a different meaning for people coming from different levels of education or fields of study. Someone coming from a mathematics programme may rate his or her math skills as moderate compared to fellow-students but as high compared to someone coming from humanities. Although some surveys avoid this problem by explicitly labelling the anchors in the scale (O*NET 2014) or by the use of vignettes (King et al. 2004), this anchoring problem is largely ignored in most self-assessment studies.

Still, this has not prevented researchers in the past (the authors included) to use self-assessments as a way to measure the skills of graduates in different kinds of large-scale surveys and one can expect this to continue in the future as well. The question is to what extent does one get a distorted picture if one relies on such data? Does one get valid information in terms of predictive validity? In other words, do test results and self-assessed skills predict the same outcomes? And what about biases when comparing across relevant groups such as gender, fields of study or countries? Are self-assessments more prone to such cross-cultural biases and if so which are the types of comparisons that one should be particularly careful about?

Ideally, one would like to compare test results and self-assessments directly at the individual level. However, the existing international large-scale surveys usually have either test results or self-assessments. Instead this paper compares different data sets and looks at the correlation structure and skills patterns across groups. The following hypotheses are formulated:

> *Hypothesis 1*: Subjective skills measures contain more measurement error and therefore show lower returns to skills in an earnings regression due to attenuation bias.

> *Hypothesis 2*: Aggregate measures of subjective skills are an unreliable measure of the relative stock of skills in a country as country- and culture-specific response styles exist and individuals do not share a common reference frame.

> *Hypothesis 3*: Within countries, subjective skills measures capture parts of the skills level differences between groups, such as graduates from different fields of study, as individuals share a common reference frame and have obtained information on their relative skills level through educational and occupational tracking and sorting.

3. Data and method

This paper uses data from the PIAAC survey, the REFLEX survey and its successor the HEGESCO survey. The REFLEX and HEGESCO surveys were designed as representative international graduate surveys that would provide accurate information on the transition from higher education to work. They were conducted in 16 and 5 countries[2] (years 2005 and 2008), respectively. Sample sizes varied between 5000 and 10,000 graduates per country. The total response was over 45,000 graduates and the overall response rate was 31%.[3] The sampling frame consisted of people graduating from higher education (ISCED 5A) in these countries in 2000 and 2003, respectively. The questionnaire was thus sent to graduates five years after their graduation. The data collection was via a combined Internet/postal survey. In the survey, information on various inputs and outcomes of education were collected. Of particular interest for this study is the information collected on individuals' self-assessed level of skills, field of study and hourly earnings. The question on the self-assessment was formulated as follows: 'Here is a list of competencies. Please provide the following information: (1) How do you rate your own level of competence? (2) What is the required level of competence in your current work?'[4] This was followed by a list of 19 skills ranging from 'Mastery of your own field or discipline' to 'Ability to write and speak in a foreign language'. Both own level and the required level needed to be answered on the Likert scale ranging from 1 'very low' to 7 'very high'. Only the endpoints of the scale were defined.

The PIAAC survey was conducted in 2012 in 24 countries[5] to measure the skills of the adult population between 16 and 65 years of age. The survey is designed to provide valid and reliable estimates of adults' competences in key information-processing skills, and to understand the antecedents as well as the outcomes of key skills on life chances (OECD 2013). National samples contain over 5000 adults between the age of 16 and 65 years (minimum response rates were set at 50%). Respondents were interviewed using computer-assisted personal interviews, although for the testing pencil-and-paper data-collection strategies were also used. PIAAC measures respondents' proficiency in literacy and numeracy as well their capacity to solve problems in technology-rich environments. All of these key information-processing skills are considered essential for the development of higher order cognitive skills as well as for gaining access to and understanding knowledge domains (OECD 2013). The OECD (2013, 59) defines literacy as 'the ability to understand, evaluate, use, and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential'. Numeracy is defined as 'the ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life'. And finally, problem solving in technology-rich environments is defined as 'the ability to use digital technology, communication tools, and networks to acquire and evaluate information, communicate with others and perform practical tasks'.

Given time restrictions, respondents took a subset of all items (adaptive testing) and item response theory and multiple imputation techniques were used to calculate 10 plausible values for literacy, numeracy and problem solving. Each of these plausible values provides an unbiased estimate of the person's true score of their proficiency level (OECD 2013). The scales have a range of 0 to 500, with an average of around 270 for OECD countries. Similar to the REFLEX and HEGESCO data, the PIAAC data also contain information on fields of study and hourly earnings.

In order to be able to compare the patterns of subjective and objective measures of skills across countries and fields of study, and their association with hourly earnings, the analysis is based on European countries[6] contained in both the REFLEX/HEGESCO data and the PIAAC data. These countries are Austria, Belgium (Flanders), the Czech Republic, Estonia, Finland, France, Germany, Italy, the Netherlands, Norway, Poland, Spain and the UK.

The PIAAC sample is restricted to individuals who resemble the individuals surveyed in the REFLEX and HEGESCO surveys. Consequently, the PIAAC final sample is obtained by only keeping individuals with a higher education degree of ISCED 5A between the ages of 20 and 35 years. Despite the fact that the PIAAC study was not explicitly designed as a graduate survey the number of observations is large enough to do our analysis. Table A1 provides the main descriptive statistics of both data sets.

The REFLEX/HEGESCO sample is restricted to individuals not older than 35 years who did not live abroad at the time of the survey and who did not follow any other education programmes after the higher education programme they were surveyed for (so that they all have ISCED 5A as the highest degree attained).

As indicated in Section 1, there is no direct comparison possible between the skills measured in PIAAC and in REFLEX/HEGESCO at the individual level. But even at a conceptual level, the assessments are hard to compare. The domains in PIAAC have been carefully defined and measured with a large battery of items whose psychometric quality has been well established and which give a reliable, cross-cultural valid indication of the concepts that are measured. In REFLEX/HEGESCO no attempt was made to use self-assessment to measure comparable constructs as in PIAAC. Instead the items were used to measure concepts such as professional expertise, functional flexibility, innovation and knowledge management, mobilisation of human resources and international orientation (Allen and van der Velden 2011). Table 1 gives an overview of the items and their related constructs.

The reliability (Cronbach's alpha) of the scales varies from a low 0.52 for the three items comprising the scale of professional expertise to 0.83 for the six items making up the scale of mobilisation of human resources. No indications were found that the reliability differed across countries. As the concepts measured in REFLEX/HEGESCO are different from the concepts measured in PIAAC, it was decided to look for single items in the REFLEX/HEGESCO questionnaire that coincide best with the concepts measured in PIAAC. The item 'analytical thinking' in the REFLEX/HEGESCO questionnaire is the item that best matches with the concept of numeracy as defined in the PIAAC project. Further, the item 'ability to write reports, memos or documents' comes quite close to the literacy concept in PIAAC, although a major difference is that the literacy scale is more related to reading and understanding information from written text and not to writing. However, one can argue that reading and writing skills are highly correlated as was also shown in the correlation between the use of these type of skills in the PIAAC data ($r = 0.56$; $p < .01$). Finally, the 'ability to

Table 1. Items in REFLEX/HEGESCO and related demands.

Items per demand	Cronbach's alpha
Professional expertise	*0.52*
Mastery of your own field or discipline	
Analytical thinking	
Ability to assert your authority	
Functional flexibility	*0.59*
Knowledge of other fields or disciplines	
Ability to rapidly acquire new knowledge	
Ability to negotiate effectively	
Innovation and knowledge management	*0.76*
Alertness to new opportunities	
Ability to use computers and the Internet	
Ability to come up with new ideas and solutions	
Willingness to question your own and others' ideas	
Mobilisation of human resources	*0.83*
Ability to perform well under pressure	
Ability to coordinate activities	
Ability to use time efficiently	
Ability to work productively with others	
Ability to mobilise the capacities of others	
Ability to make your meaning clear to others	
Other items	
Ability to present products ideas or reports to an audience	
Ability to write reports, memos or documents	
Ability to write and speak in a foreign language	

Source: Allen and van der Velden (2011).

use computers and the Internet' is rather close to the concept of problem solving in technology-rich environments as measured in PIAAC.

Two types of analyses are performed. On the one hand, skills estimates are compared countries, across fields of study and gender. This is done for all selected respondents, whether working or not. To make the skills measures comparable, standardised measures are used in which standardisation is based on all selected respondents in the data set.

For the analysis on the skills effect on earnings, the analysis will be restricted to individuals with a substantial labour market participation of 30 or more working hours per week. Self-employed were excluded, and so were respondents with extreme earnings: the 1st and the 100th earnings percentile were excluded to limit the influence of outliers on the results. The returns to skills are then estimated using simple Ordinary Least Squares (OLS) regression with the log of hourly earnings as the dependent variable and the skills measure standardised by country as the independent variable of interest. In the estimations using REFLEX/HEGESCO data, controls include field of study dummies. For the pooled regression, country dummies were included. In the estimations using PIAAC data, additional controls are added for potential labour market experience since not all individuals in the sample graduated in the same year. The potential labour market experience variable is constructed as the number of years since the highest educational degree was obtained. The first plausible value is used for each of the skill domains (standardised by country). All earning regressions are performed separately for males and females.

In all analyses, estimates derived from both data sets are compared. As the estimates are based on two different data sets, each data point will have two confidence intervals. The inverses of the product of these intervals are used as weights in the scatter plots and the correlation analyses to give more weight to those data points which are more reliable. These inverses are rescaled so that the sum of all weights equals one.

4. Analysis

The analysis starts with cross-national differences in the external validity of the skills measures. This is done by looking at the cross-national differences in the effect of skills on earnings. Figure 1 provides the estimates for the effect of numeracy on earnings for females (1(a)) and males (1(b)). Figure 2(a) and 2(b) provides similar estimates for literacy and Figure 3(a) and 3(b) for problem solving in technology-rich environments. Coefficients and standard errors are presented in Tables A2 and A3.

The average effect of numeracy skills on earnings is, as hypothesised, higher for the objective PIAAC numeracy measure than for the subjective REFLEX/HEGESCO measures. In the PIAAC data, a one standard deviation increase in numeracy skills is associated with a 6.8% ($p < .01$) increase in hourly earnings for females, and a 6.6% ($p < .01$) increase in hourly earnings for males (see Table A2). In the REFLEX/HEGESCO data, a one standard deviation increase in self-reported analytical skills is associated with a 4.3% ($p < .01$) increase in hourly earnings for females and a 5.8% ($p < .01$) increase in hourly earnings for males (see Table A3). The differences between the objective and subjective measures are surprisingly low if one takes into account that the objective measure is based on a full test and the subjective measure is based on a single item. There is less than a percentage point difference for males and about a 2.5 percentage point difference for females. Note that these estimates are obtained after controlling for country and fields of study fixed effects.

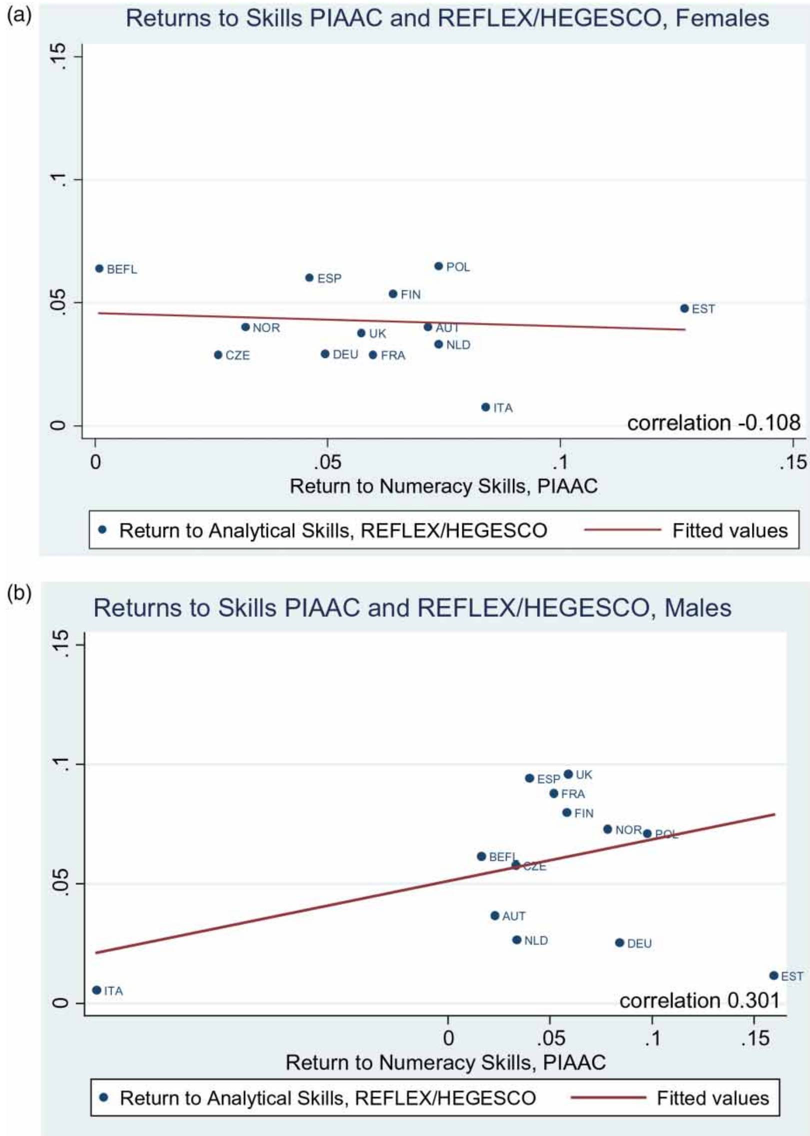


Figure 1. (a) Scatter plot of the returns to numeracy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), females. (b) Scatter plot of the returns to numeracy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), males.

While these average returns to skills seem to suggest that both measures capture the same constructs, the scatter plots presented in Figure 1(a) and 1(b) illustrate that the returns to skills estimated from the two data sources for particular countries may differ. For females, the estimated returns to skills from PIAAC and REFLEX/HEGESCO data are fairly similar for some countries such as Finland, France, the Netherlands and Norway, but very different for other countries like Austria, Estonia and Italy. As displayed in Figure 1(a), the correlation between the returns to skills in the different countries for both measures for females is close to zero

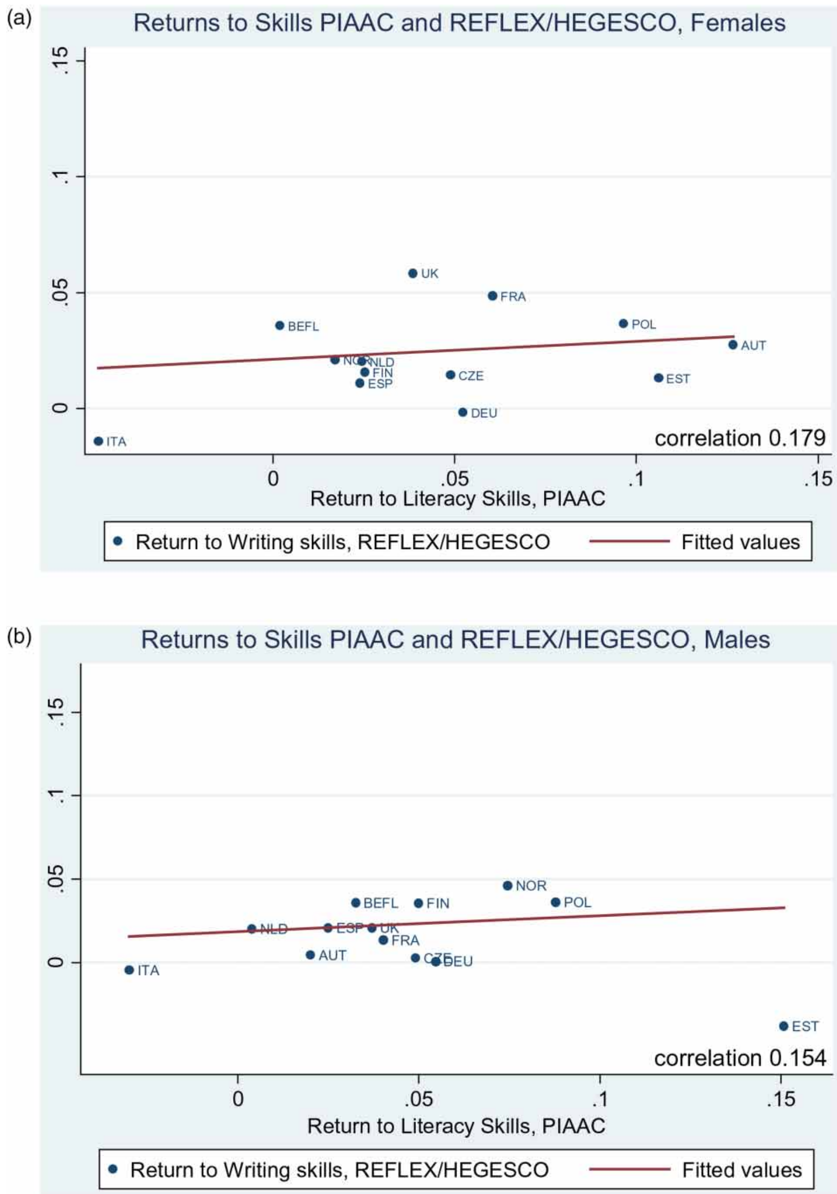
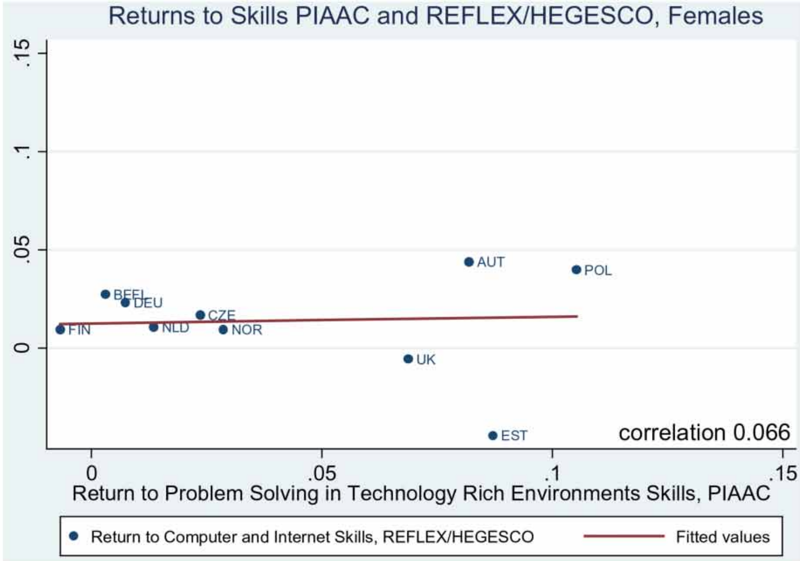


Figure 2. (a) Scatter plot of the returns to literacy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), females. (b) Scatter plot of the returns to literacy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), males.

($r = -0.11$.; $p = .73$). For males, the correlation seems to be a bit larger ($r = 0.30$; $p = .32$), but this is mainly driven by the negative but statistically insignificant effect of numeracy skills on earnings in Italy. Leaving Italy out, the correlation is close to zero ($r = 0.07$; $p = .83$).

Figure 2(a) and 2(b) displays similar results for literacy skills. Again the average effect of literacy skills on earnings is higher for the objective PIAAC literacy

(a)



(b)

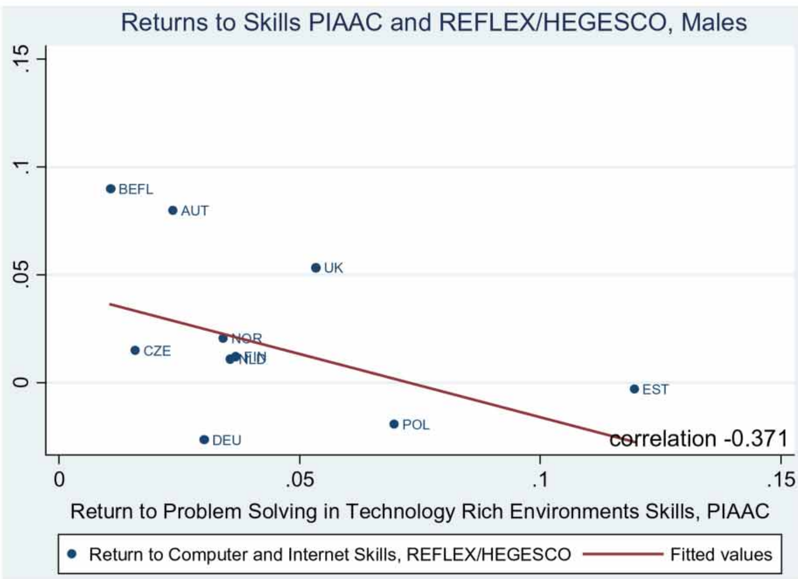


Figure 3. (a) Scatter plot of the returns to problem-solving skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), females. (b) Scatter plot of the returns to problem-solving skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO), males.

measure than for the subjective REFLEX/HEGESCO measures. In the PIAAC data, a one standard deviation increase in literacy skills is associated with a 6.1% ($p < .01$) increase in hourly earnings for females, and a 5.4% ($p < .01$) increase in hourly earnings for males (Table A2). In the REFLEX/HEGESCO data, a one standard deviation increase in self-reported ability to write reports, memos and documents is associated with a 2.1% ($p < .01$) increase in hourly earnings for females and a 1.6% ($p < .05$) increase in hourly earnings for males (Table A3). This difference between objective and subjective measures is much larger than that in the case of numeracy skills. Again one can see that the returns to skills estimated from the two data sources may differ, for both females and males (Figure 2(a) and 2(b)).

Figure 3(a) and 3(b) displays the results for problem solving in technology-rich environments. The average effect of this type of skill on earnings is lower than that for numeracy and literacy. In the PIAAC data, a one standard deviation increase in problem-solving skills is associated with a 5.3% ($p < .01$) increase in hourly earnings for females, and a 4.5% ($p < .01$) increase in hourly earnings for males (Table A2). Again the average effect of the subjective measures is lower. In the REFLEX/HEGESCO data, a one standard deviation increase in self-reported ability to use computers and the Internet is associated with a 1.7% ($p < .01$) increase in hourly earnings for females and a 2.1% ($p < .05$) increase in hourly earnings for males (Table A3). Importantly, one can note again that the returns to skills estimated from the two data sources may differ, for both females and males (Figure 3(a) and 3(b)).

Figures 4–6 plot the coefficients of country dummies (the reference country is Austria) from a regression of numeracy (and literacy and problem-solving skills) on country dummies, field of study dummies and a dummy which equals one if the respondent is female, performed with each of the two data sets. The figures assess whether subjective skills measures in REFLEX/HEGESCO can capture the same cross-country differences as the objective skills measures in PIAAC do. Unlike the former analysis, skills measures are used which are standardised on the whole data set and not country-wise (otherwise the comparison across countries would not have been meaningful).

The correlation between the country differences of the aggregated objective and subjective skills measures is very weak in the case of numeracy skills ($r = 0.12$; $p = .70$) and even negative in the case of literacy ($r = -0.32$; $p = .32$) and especially problem solving in technology-rich environments ($r = -0.57$; $p = .11$). In consequence,

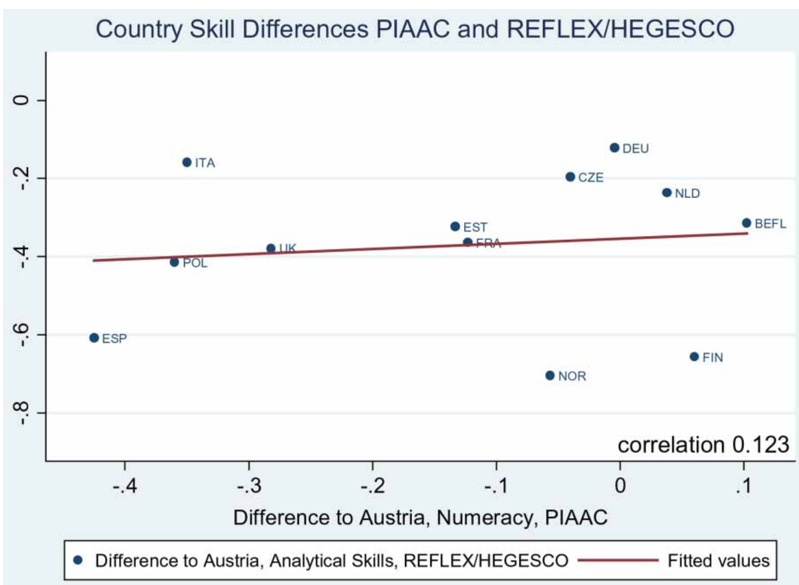


Figure 4. Scatter plot of the country differences in numeracy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

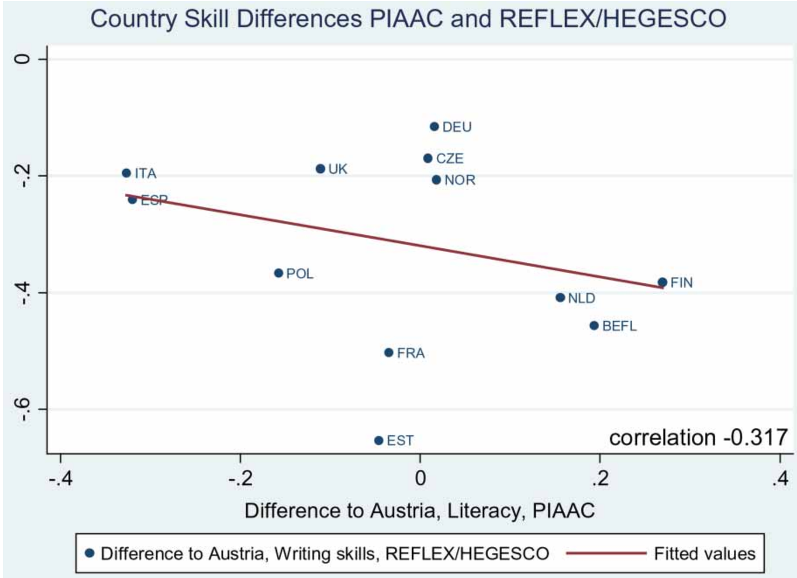


Figure 5. Scatter plot of the country differences in literacy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

if the subjective REFLEX/HEGESCO skills measure was used to make inferences about the relative stock of skills at the country level (after controlling for country differences in the relative shares of fields of study or gender), this would substantially underestimate the relative stock of skills in some countries, for example, Finland and

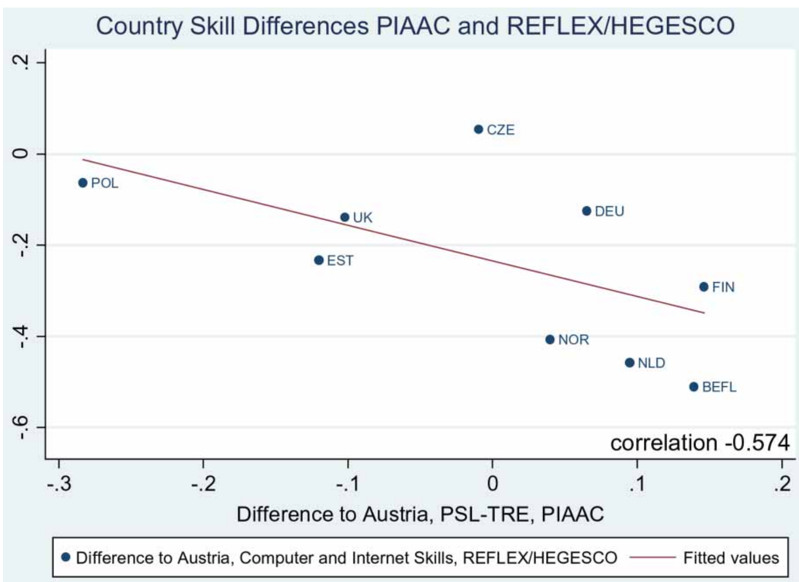


Figure 6. Scatter plot of the country differences in problem-solving skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

Belgium-Flanders (all three skill domains), and considerably overestimate the relative stock of skills in, for example, Italy (note that in Italy the PIAAC problem-solving test was not taken).

Figures 7–9 present scatter plots of the differences across fields of study for numeracy, literacy and problem solving, respectively. This gives an idea whether subjective skills measures can capture within-country differences between fields of study as measured in PIAAC. The scatter plots present the coefficients of the field of study dummies (reference category is health and welfare) from a regression of numeracy (and literacy and problem-solving skills) on country dummies, field of study dummies and a dummy which equals one if the respondent is female, performed with each of the two data sets. Coefficients and standard errors are reported in Tables A4 and A5.

In contrast to the country comparison, there is a very strong correlation between the mean scores of objective and subjective skills at the field of study level. In the case of numeracy skills (Figure 7), this correlation amounts to 0.91 ($p < .01$) and in the case of problem-solving skills (Figure 9) to 0.81 ($p < .05$). Only in the case of literacy skills (Figure 8) there is no correlation between the objective and subjective skills measures at the field of study level ($r = 0.06$; $p = .90$). This is mainly driven by the results for science, mathematics and computing that combines a relatively low score on 'writing reports, memos or documents' with a relatively high score on literacy skills. If one takes into account the high correlation between the test scores for numeracy and literacy ($r = 0.86$; $p < .01$), this should come as no surprise. The literacy test scores also capture some general cognitive skills and not just literacy skills. These general cognitive skills

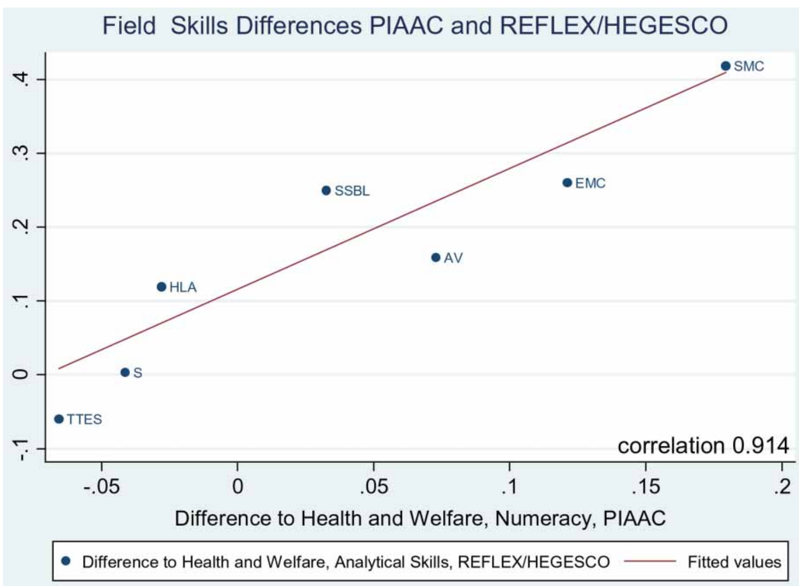


Figure 7. Scatter plot of the field of study differences in numeracy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).
Notes: There are no observations for 'General Programmes' in the REFLEX/HEGESCO data so this category was not taken up in the graph. The reference category is health and welfare. AV – agriculture and veterinary, EMC – engineering, manufacturing and construction, HLA – humanities, languages and arts, SMC – science, mathematics and computing, S – services, SSBL – social sciences, business and law, TTES – teacher training and educational sciences.

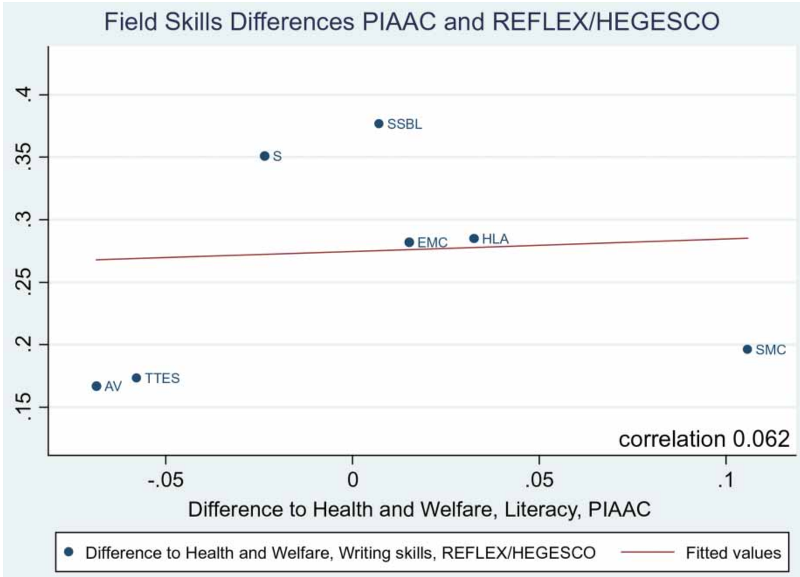


Figure 8. Scatter plot of the field of study differences in literacy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

are clearly higher among graduates of the science programmes. The high correlations suggest that subjective skills measures are a strong indicator for the relative position of individuals with a particular field of study background in the objectively measured skills distribution.

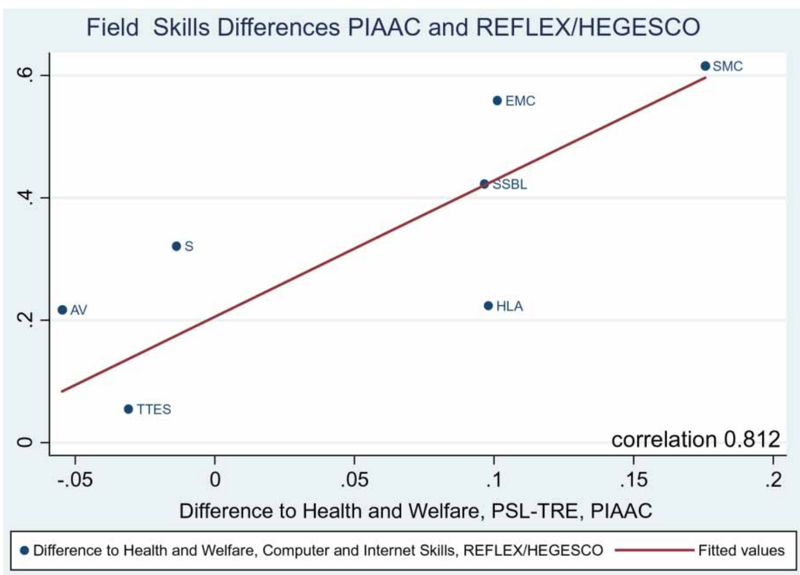


Figure 9. Scatter plot of the field of study differences in problem-solving skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

Finally, Figures 10–12 present the gender differences in skills countrywise for numeracy, literacy and problem solving, respectively. Similar to the previous figures, the scatter plots present the coefficients of the female dummy from a regression of numeracy (and literacy and problem-solving skills) on country dummies, field of study dummies and a dummy which equals one if the respondent is female, performed with each of the two data sets. This analysis gives an idea of whether subjective skills measures are biased by the fact that females tend to have a different response style. Coefficients and standard errors are again reported in Tables A4 and A5.

First one can look at the overall effect of gender on skills (see Tables A4 and A5). The effect of being female on the objective numeracy score is −0.305 ($p < .01$; Table A4), while the corresponding effect for subjective analytical skills is −0.244 (Table A5). These coefficients are similar, indicating that females' subjective assessment of their numeracy skills corresponds well with their actual level of these skills. This is similar for problem-solving skills. Here the effect is −0.221 ($p < .01$; Table A4), while the corresponding effect for the 'ability to use computers and the Internet' is −0.160 ($p < .01$; Table A5). For literacy the situation is different; here, the scores of females on the subjective assessment are larger than the scores on the objective assessment, namely 0.068 ($p < .01$; Table A5) versus −0.145 ($p < .01$; Table A4), suggesting that females tend to overestimate their literacy skills.

In line with these average results, Figures 10–12 show that for numeracy and problem solving in technology-rich environments most of the estimates are negative, indicating that in most countries females have lower scores than males both on the objective test scores and on the self-assessed measures. But the gender effects are not similar across countries. An interesting example is the strong positive self-assessment of females in Poland as well as the slight positive self-assessment of females in the UK. Both scores are remarkable given the fact that the objective scores of the

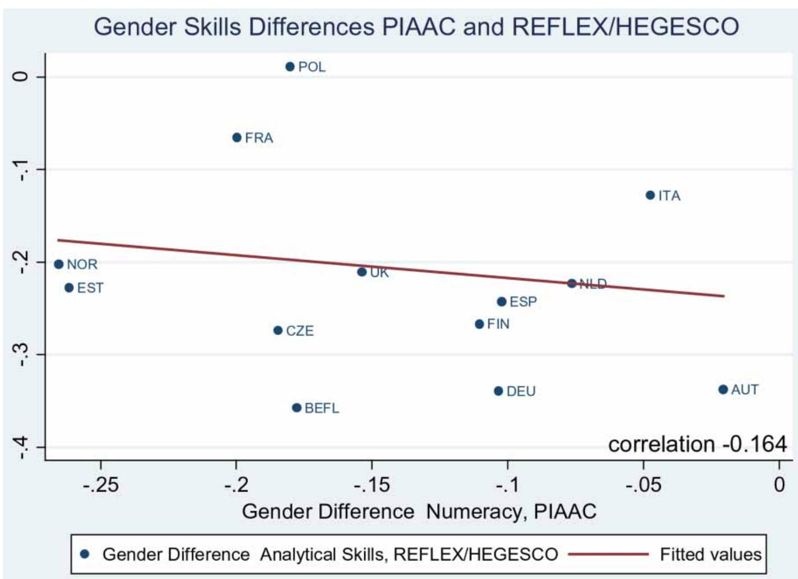


Figure 10. Scatter plot of the gender differences per country in numeracy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

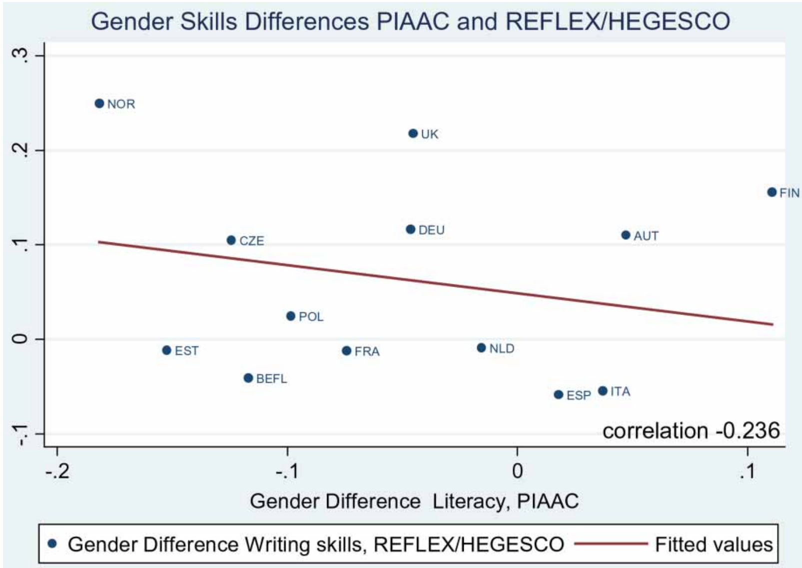


Figure 11. Scatter plot of the gender differences per country in literacy skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

females in these two countries are actually much lower than that of their male counterparts. These anomalies also occur in some other cases, which is the main reason why the correlations between the gender differences across countries for both measures are not only low but even negative.

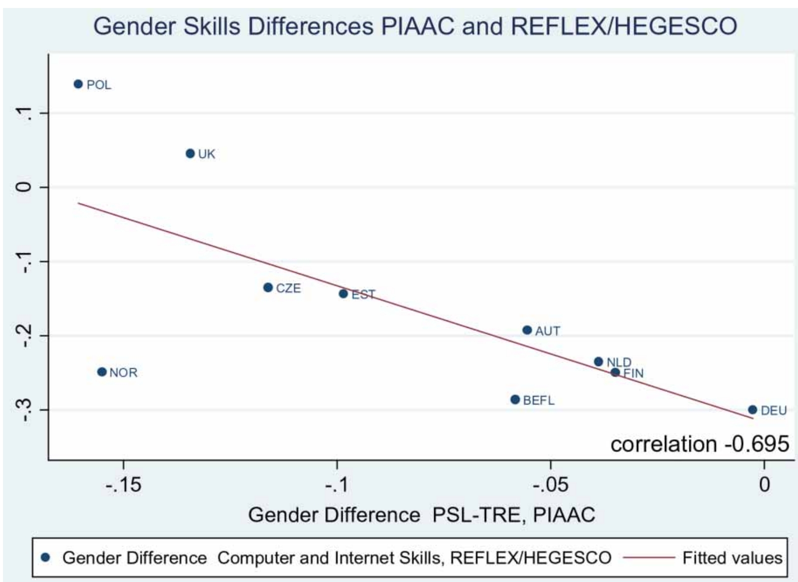


Figure 12. Scatter plot of the gender differences per country in problem-solving skills: test (PIAAC) versus self-assessed measures (REFLEX/HEGESCO).

5. Conclusion

In this paper an analysis is carried out whether objective tests and subjective self-assessments in international large-scale studies yield similar results when looking at cross-national differences in the effects of skills on earnings, and skills patterns across countries, fields of study and gender. As a caveat, it is important to point out that the evidence presented in this paper can only be understood as indicative, given that the data sets that are used contain different individuals, measure different skills and were conducted in different years. However, as skills are to a large extent the outcomes of education and labour market systems, and as these systems are fairly stable over time, some interesting conclusions can be drawn from this analysis.

In line with the first hypothesis, the results show that the returns to skills using a subjective skills measure are lower than the returns to earnings using an objective skills measure. But more importantly, the results also show that aggregate measures of subjective skills are an unreliable measure of the relative stock of skills in a country compared to objective skills measures, which confirms *Hypothesis 2*. This is due to the fact that response styles differ both across countries and across gender. However, in line with *Hypothesis 3*, subjective skills measures can be used within countries, to capture skills level differences among graduates from different fields of study.

The overall conclusion is that self-assessments, at least of the types that were used in the REFLEX/HEGESCO study are not adequate to capture the 'true' cross-national variation in skills levels, nor do they predict adequately the cross-national relations between skills and earnings or skills and gender. This is an important conclusion and indicates that one should be very careful when using such measures in a cross-national perspective. Probably, the main reason is that there are cross-cultural biases in answering scales for self-assessment, especially when these scales are not properly anchored (King et al. 2004; Allen and van der Velden 2005). This leads to biases when one tries to use such scales to estimate the relative stock of skills in a particular country, compared to other countries. Exactly for this reason one should be very careful using self-assessments in any cross-national comparison. Country differences in the effect of skills on earnings based on self-assessment are quite different from country differences in the effect of skills on earnings based on objective test scores. And the same applies for country differences in the effect of gender on skills.

This does not mean that these self-assessments cannot be used at all. Self-assessments very accurately predict within-country differences in skills levels between fields of study. And they also have a good within-country predictive validity if one looks at the effect of skills on earnings. Even within-country gender differences point in the same direction, but here one has to take into account that in typical 'female' domains like literacy females tend to overestimate their skills level.

As indicated above, a caveat for this analysis is that different data sources need to be used to compare the effects of objective and subjective skills measures. In future research, it is important to include subjective skills measures in skills surveys such as PIAAC or PISA to arrive at firmer conclusions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. Note that we follow here the Anglo-Saxon definition of 'skills' as used for example by the OECD in their reports on PIAAC and PISA, rather than the narrow concept of skills as used in the German language.
2. The REFLEX project involved partners from Austria, Belgium (Flanders), the Czech Republic, Estonia, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden (here no data on self-assessment were collected), Switzerland and the UK. The HEGESCO survey was conducted in Hungary, Lithuania, Poland, Slovenia and Turkey.
3. For more information, see Allen and van der Velden (2011) and Allen, Pavlin, and van der Velden (2011) or the project's website www.reflexproject.org.
4. The REFLEX questionnaire is available from the authors upon request.
5. PIAAC data are currently available for Australia, Austria, Belgium (Flanders), Canada, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, the UK and the USA.
6. The analysis is constrained to European countries to ensure more cross-cultural comparability. Japan is in many ways very different from the European countries, both in labour market structure and in skills levels. To prevent distortion Japan was therefore excluded. Also Sweden is excluded as Sweden did not collect data on the self-assessment of skills in the REFLEX project.

References

Allen, J., S. Pavlin, and R. van der Velden, eds. 2011. *Competencies and Early Labour Market Careers of Higher Education Graduates in Europe*. Ljubljana: University of Ljubljana.

Allen, J., and R. van der Velden. 2005. *The Role of Self-assessment in Measuring Skills*. REFLEX Working paper 2. Maastricht: ROA.

Allen, J., and R. van der Velden, eds. 2011. *The Flexible Professional in the Knowledge Society: New Challenges for Higher Education. Higher Education Dynamics, 35*. Dordrecht: Springer.

Bouffard, T., C. Vezeau, M. Roy, and A. Lengelé. 2011. "Stability of Bias in Self-evaluation and Relations to Well-being Among Elementary School Children." *International Journal of Educational Research* 50 (4): 221–29.

Falchikov, N., and D. Boud. 1989. "Student Self-assessment in Higher Education: A Meta-analysis." *Review of Educational Research* 59 (4): 395–430.

Falchikov, N., and J. Goldfinch. 2000. "Student Peer Assessment in Higher Education: A Meta-analysis Comparing Peer and Teacher Marks." *Review of Educational Research* 70 (3): 287–322.

Hanushek, E., and L. Woessmann. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46 (3): 607–68.

Humburg, M., and R. van der Velden. Forthcoming. "What Is Expected of Higher Education Graduates in the 21st Century?" In *Oxford Handbook of Skills and Training*, edited by J. Buchanan, D. Finegold, K. Mayhew, and C. Warhurst. Oxford: Oxford University Press.

King, G., C. J. L. Murray, J. A. Salomon, and A. Tandon. 2004. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207.

Le Deist, F., and J. Winterton. 2005. "What Is Competence?" *Human Resource Development International* 8 (1): 27–46.

Murray, T. S. 2003. "Reflections on International Competence Assessments." In *Key Competencies for a Successful Life and a Well-functioning Society*, edited by D. S. Rychen and L. H. Salganik, 135–59. Göttingen: Hogrefe and Huber.

OECD. 2013. *Skills Outlook: First Results from the OECD Survey of Adult Skills*. Vol. 1. Paris: OECD Publishing.

O*NET. 2014. Accessed July 1, 2014. http://www.onetcenter.org/overview.html.

Richter, L., and P. B. Johnson. 2001. "Current Methods of Assessing Substance Use: A Review of Strengths, Problems, and Developments." *Journal of Drug Issues* 31 (4): 809–32.

Rychen, D. S., and L. H. Salganik. 2003. "A Holistic Model of Competence." In *Key Competencies for a Successful Life and a Well-functioning Society*, edited by D. S. Rychen and L. H. Salganik, 41–62. Göttingen: Hogrefe and Huber.
Van der Klink, M. R., and J. Boon. 2003. "Competencies: The Triumph of a Fuzzy Concept." *Journal of Human Resources Development and Management* 3 (2): 125–37.
Ward, M., L. Gruppen, and G. Regehr. 2002. "Measuring Self-assessment: Current State of the Art." *Advances in Health Sciences Education* 7: 63–80.
Weinert, F. E. 2001. "Concept of Competence: A Conceptual Clarification." In *Defining and Selecting Key Competencies*, edited by D. S. Rychen and L. H. Salganik, 45–65. Göttingen: Hogrefe & Huber.

Appendix

Table A1. Descriptive statistics.

	PIAAC				REFLEX/HEGESCO			
	Female		Male		Female		Male	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	29.1	3.7	29.5	3.7	29.5	2.2	30.0	2.2
Potential labour market experience	4.6	3.7	4.6	3.5	5 years by design			
Field of study								
AV	0.02		0.02		0.03		0.04	
EMC	0.07		0.28		0.09		0.37	
HLA	0.16		0.09		0.11		0.05	
HW	0.15		0.05		0.18		0.06	
S	0.03		0.03		0.03		0.04	
SMC	0.09		0.16		0.06		0.11	
SSBL	0.34		0.30		0.34		0.29	
TTES	0.14		0.06		0.17		0.05	
N	2944 (60%)		1953 (40%)		8563 (58%)		6123 (42%)	

Notes: In both data sets, skills measures and earnings have been standardised to have mean 0 and standard deviation 1. Descriptive statistics of the PIAAC data are based on individuals for which information on gender, country, the field of study and literacy skills are available. Descriptive statistics of the REFLEX/HEGESCO data are based on individuals for which information on gender, country, the field of study and the ability to write memos, documents and reports are available. AV – agriculture and veterinary, EMC – engineering, manufacturing and construction, HLA – humanities, languages and arts, HW – health and welfare, SMC – science, mathematics and computing, S – services, SSBL – social sciences, business and law, TTES – teacher training and educational sciences.

Table A2. Returns to skills across countries, PIAAC.

	Average		Austria		Belgium		Czech Republic		Estonia		Finland		France	
Skill	F	M	F	M	F	M	F	M	F	M	F	M	F	M
Numeracy	0.068***	0.066***	0.072	0.023	0.001	0.017	0.026	0.033	0.127***	0.160***	0.064***	0.058**	0.060**	0.052**
s.e.	0.008	0.009	0.043	0.052	0.024	0.045	0.047	0.035	0.041	0.048	0.02	0.026	0.028	0.024
Literacy	0.061***	0.054***	0.127***	0.020	0.002	0.033	0.049	0.049	0.106**	0.151***	0.025	0.050**	0.060	0.040*
s.e.	0.011	0.008	0.036	0.052	0.022	0.043	0.048	0.033	0.044	0.049	0.022	0.024	0.027	0.024
Problem solving	0.053***	0.045***	0.082**	0.024	0.003	0.011	0.024	0.016	0.087**	0.120**	−0.007	0.037*		
s.e.	0.016	0.008	0.036	0.049	0.026	0.040	0.031	0.037	0.042	0.058	0.024	0.021		
N	1806	1348	64	65	89	64	95	77	148	77	170	122	145	121

	Germany		Italy		Netherlands		Norway		Poland		Spain		UK	
	F	M	F	M	F	M	F	M	F	M	F	M	F	M
Numeracy	0.049	0.084**	−0.048	−0.030	0.074***	0.034	0.032*	0.078***	0.074***	0.098***	0.046	0.04	0.057**	0.032**
s.e.	0.036	0.038	0.446	0.094	0.026	0.027	0.018	0.019	0.022	0.028	0.054	0.058	0.024	0.018
Literacy	0.052	0.055	0.096	−0.008	0.024	0.004	0.017	0.075***	0.097***	0.088***	0.024	0.025	0.038	0.037
s.e.	0.034	0.037	0.065	0.071	0.02	0.03	0.018	0.017	0.021	0.024	0.051	0.051	0.027	0.029
Problem solving	0.007	0.030			0.014	0.036	0.029*	0.034*	0.105***	0.070**	.	.	0.069***	0.053***
s.e.	0.049	0.050			0.020	0.030	0.016	0.020	0.022	0.027	.	.	0.023	0.027
N	81	77	33	32	99	90	183	148	326	218	88	64	272	193

Notes: Coefficients from countrywise OLS regressions of log hourly earnings on skills standardised by country (first plausible value, one skill at a time), controlling for potential labour market experience (years since leaving education), and field of study. In the pooled OLS regressions to produce the average results, additional controls for country are included. Standard errors are clustered by country. Problem-solving skills measures are not available for France, Italy and Spain. The R^2 of the countrywise log hourly earnings regressions for females is 0.23 for numeracy and problem solving and 0.22 for literacy. The R^2 of the countrywise log hourly earnings regressions for males is 0.29 for numeracy, 0.30 for literacy and 0.28 for problem solving.

*$p < .10$.

**$p < .05$.

***$p < .01$.

Table A3. Returns to skills across countries, REFLEX/HEGESCO.

	Average		Austria		Belgium		Czech Republic		Estonia		Finland		France	
Skill	F	M	F	M	F	M	F	M	F	M	F	M	F	M
Analytical skills	0.043***	0.058***	0.040**	0.037	0.064***	0.061***	0.029***	0.058***	0.048*	0.012	0.054***	0.080***	0.029	0.088**
s.e.	0.006	0.008	0.018	0.023	0.018	0.019	0.008	0.01	0.027	0.051	0.011	0.014	0.02	0.039
Ability to write reports, memos or documents	0.021***	0.016**	0.028	0.005	0.036*	0.036**	0.014*	0.003	0.013	−0.038	0.016	0.036**	0.049***	0.014
s.e.	0.004	0.005	0.020	0.019	0.019	0.018	0.008	0.010	0.030	0.037	0.012	0.013	0.018	0.034
Ability to use computers and the Internet	0.017***	0.021**	0.044**	0.080***	0.027	0.090***	0.017*	0.015	−0.045	−0.003	0.009	0.012	0.072***	0.045
s.e.	0.005	0.007	0.022	0.028	0.019	0.020	0.009	0.012	0.029	0.049	0.013	0.017	0.020	0.037
N	5215	4348	167	220	181	174	1202	1176	190	90	504	362	252	128

	Germany		Italy		Netherlands		Norway		Poland		Spain		UK	
	F	M	F	M	F	M	F	M	F	M	F	M	F	M
Analytical Skills	0.029	0.025	0.008	0.005	0.033***	0.027**	0.040***	0.073***	0.065***	0.071**	0.060***	0.094***	0.038**	0.096***
s.e.	0.019	0.019	0.017	0.018	0.009	0.012	0.009	0.014	0.024	0.031	0.014	0.019	0.018	0.027
Ability to write reports, memos or documents	−0.002	0.001	−0.014	−0.004	0.020**	0.020*	0.021	0.046***	0.037	0.036	0.011	0.021	0.058***	0.021
s.e.	0.023	0.016	0.015	0.019	0.009	0.010	0.010	0.012	0.025	0.032	0.014	0.018	0.022	0.026
Ability to use computers and the Internet	0.023	−0.027	0.025	0.041*	0.011	0.011	0.009	0.020	0.040	−0.019	0.003	0.028	−0.006	0.053*
s.e.	0.018	0.020	0.017	0.021	0.010	0.013	0.012	0.016	0.027	0.041	0.016	0.020	0.022	0.028
N	211	273	378	360	633	492	381	322	259	218	584	326	273	208

Notes: Coefficients from countrywise OLS regressions of log hourly earnings on skills standardised by country (one skill at a time), controlling for field of study. In the pooled OLS regressions to produce the average results, additional controls for country are included. Standard errors are clustered by country. The R^2 of the countrywise log hourly earnings regressions for females is 0.12 for all three types of skills. The R^2 of the countrywise log hourly earnings regressions for males is 0.15 for analytical skills, 0.12 for the ability to write reports, memos or documents, and 0.13 for the ability to use computers and the Internet.

*$p < .10$.

**$p < .05$.

***$p < .01$.

Table A4. OLS regression of skills, PIAAC.

Dependent variable	Numeracy	Literacy	Problem solving
Gender	−0.305***	−0.145***	−0.221***
	0.029	0.030	0.033
Country			
Austria	Ref.	Ref.	Ref.
Belgium	0.184*	0.369***	0.307***
	0.099	0.097	0.096
Czech Republic	−0.069	0.018	−0.011***
	0.089	0.090	0.089
Estonia	−0.234**	−0.088	−0.253***
	0.088	0.089	0.086
Finland	0.111	0.517***	0.320***
	0.090	0.091	0.082
France	−0.218**	−0.065	–
	0.090	0.089	
Germany	−0.007	0.031	0.144
	0.093	0.092	0.088
Italy	−0.617***	−0.626***	–
	0.103	0.107	
Netherlands	0.070	0.298***	0.203**
	0.089	0.089	0.085
Norway	−0.097	0.037	0.093
	0.096	0.093	0.082
Poland	−0.637***	−0.301***	−0.591***
	0.082	0.082	0.080
Spain	−0.751***	−0.611***	–
	0.088	0.091	
UK	−0.500***	−0.212**	−0.205***
	0.085	0.084	0.078
Field of study			
AV	0.103	−0.039	−0.110
	0.124	0.125	0.147
EMC	0.232***	0.017	0.200***
	0.055	0.057	0.062
HLA	−0.047	0.077	0.187***
	0.052	0.054	0.061
HW	Ref.	Ref.	Ref.
S	−0.071	−0.048	0.016
	0.079	0.082	0.111
SMC	0.331***	0.212***	0.376***
	0.058	0.060	0.067
SSBL	0.082*	0.021	0.254***
	0.045	0.047	0.051
TTES	−0.110**	−0.106*	−0.084
	0.054	0.056	0.063
N	4897	4897	3776

Notes: AV – agriculture and veterinary, EMC – engineering, manufacturing and construction, HLA – humanities, languages and arts, HW – health and welfare, SMC – science, mathematics and computing, S – services, SSBL – social sciences, business and law, TTES – teacher training and educational sciences.
*$p < .10$.
**$p < .05$.
***$p < .01$.

Table A5. OLS regression of skills, REFLEX/HEGESCO.

Dependent variable	Analytical skills	Ability to write reports, memos or documents	Ability to use computers and the internet
Gender	−0.244***	0.068***	−0.160***
	0.017	0.018	0.017
Country			
Austria	Ref.	Ref.	Ref.
Belgium	−0.319***	−0.460***	−0.532***
	0.056	0.057	0.056
Czech Republic	−0.196***	−0.168***	0.060
	0.043	0.040	0.038
Estonia	−0.326***	−0.665***	−0.237***
	0.058	0.064	0.054
Finland	−0.661***	−0.384***	−0.298***
	0.048	0.046	0.042
France	−0.367***	−0.513***	−0.662***
	0.052	0.057	0.056
Germany	−0.121**	−0.114**	−0.129***
	0.051	0.050	0.046
Italy	−0.162***	−0.198***	−0.333***
	0.047	0.046	0.043
Netherlands	−0.235***	−0.412***	−0.471***
	0.045	0.044	0.041
Norway	−0.706***	−0.204***	−0.415***
	0.053	0.047	0.047
Poland	−0.420***	−0.372***	−0.065
	0.055	0.054	0.048
Spain	−0.613***	−0.243***	−0.611***
	0.048	0.045	0.045
UK	−0.379***	−0.193***	−0.140***
	0.055	0.052	0.050
Field of study		0.191***	
AV	0.176***	0.055	0.235***
	0.049		0.053
EMC	0.300***	0.295***	0.598***
	0.030	0.033	0.030
HLA	0.121***	0.321***	0.221***
	0.039	0.041	0.041
HW	Ref.	Ref.	Ref.
SMC	0.463***	0.225***	0.664***
	0.036	0.039	0.036
S	0.030	0.376***	0.348***
	0.052	0.048	0.049
SSBL	0.298***	0.411***	0.476***
	0.028	0.029	0.029
TTES	−0.043	0.195***	0.053
	0.035	0.036	0.036
N	14,670	14,686	14,714

Notes: AV – agriculture and veterinary, EMC – engineering, manufacturing and construction, HLA – humanities, languages and arts, HW – health and welfare, SMC – science, mathematics and computing, S – services, SSBL – social sciences, business and law, TTES – teacher training and educational sciences.
*$p < .10$.
**$p < .05$.
***$p < .01$.