Cabus, Sofie; Napierala, Joanna; Carretero, Stephanie

Working Paper

The returns to non-cognitive skills: A meta-analysis

JRC Working Papers Series on Labour, Education and Technology, No. 2021/06

Provided in Cooperation with:
Joint Research Centre (JRC), European Commission

Suggested Citation: Cabus, Sofie; Napierala, Joanna; Carretero, Stephanie (2021) : The returns to non-cognitive skills: A meta-analysis, JRC Working Papers Series on Labour, Education and Technology, No. 2021/06, European Commission, Joint Research Centre (JRC), Seville

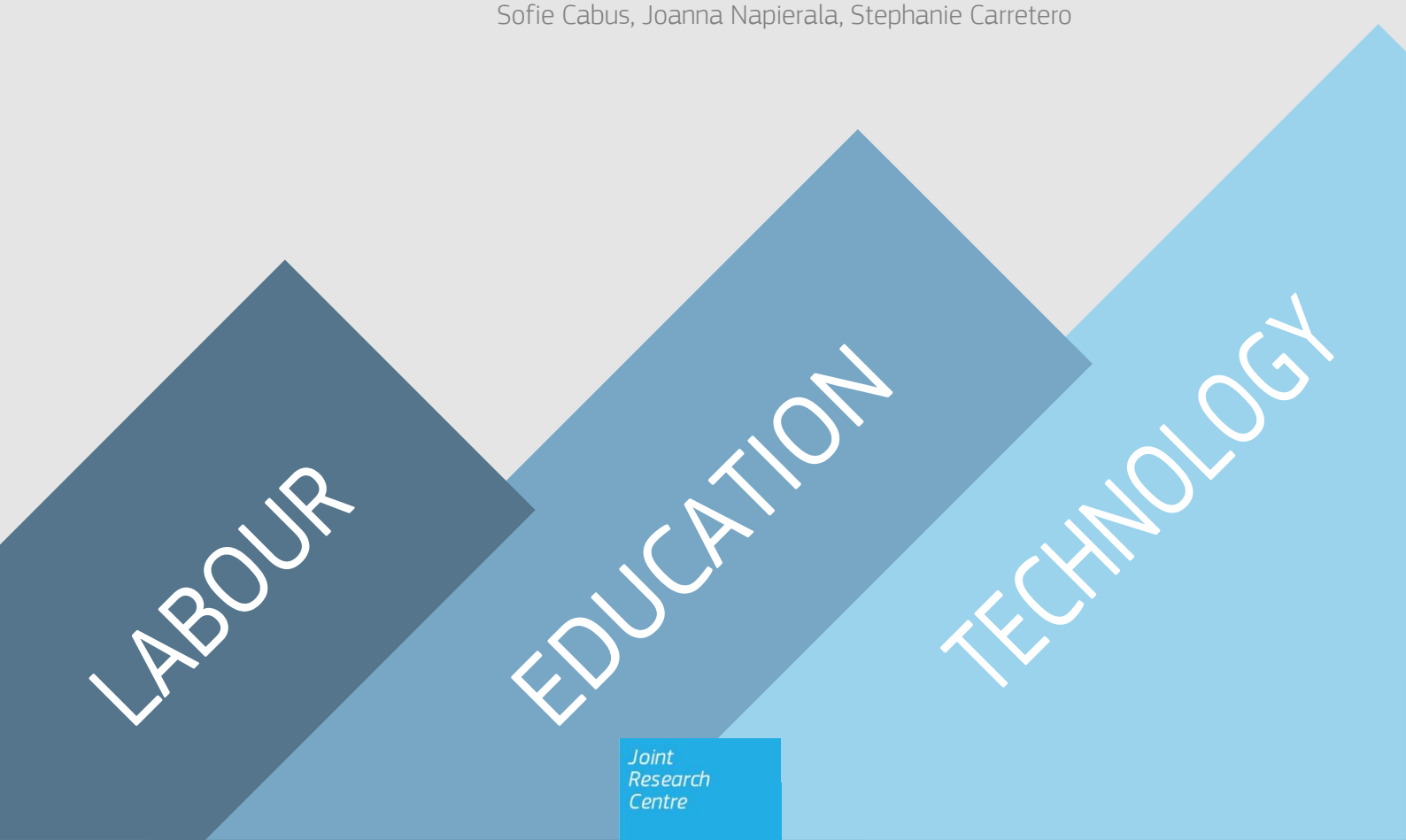
This Version is available at:
https://hdl.handle.net/10419/233211

JRC Technical Report

The Returns to Non-Cognitive Skills: A Meta-Analysis

JRC Working Papers Series on
Labour, education and Technology
2021/06

Sofie Cabus, Joanna Napierala, Stephanie Carretero



Joint
Research
Centre

Contact information
Name: Sofie Cabus
Email: Sofie Cabus <sofie.cabus@kuleuven.be>

EU Science Hub
https://ec.europa.eu/jrc

JRC123308

Seville: European Commission, 2021

How to cite this report: Cabus, S., Napierala, J., Carretero, S. *The Returns to Non-Cognitive Skills: a Meta-Analysis*, Seville: European Commission, 2021, JRC123308.

The Returns to Non-Cognitive Skills: A Meta-Analysis

Sofie Cabus (University of Leuven, Belgium), Joanna Napierala (Joint Research Centre), Stephanie Carretero (Joint Research Centre)

Abstract

This paper discusses the returns to non-cognitive skills based on results of a meta-analysis. The systematic literature review of articles published in the last decade and analysing labour market outcomes and non-cognitive skills allowed us to extract more than 300 estimates linking earnings and non-cognitive skills, most often measured by the Big Five inventory. The results of meta-analysis point to heterogeneity in the estimated signs and significance of a particular non-cognitive skill. We observe that conscientiousness and openness are two personality traits that bring higher earnings, while agreeableness and neuroticism (low emotional stability) are associated with receiving lower earnings. Some gender differences are also observed. Older and female participants seemed to benefit more from programmes targeted at developing non-cognitive skills than younger participants and men. However, there is a positive selection of female participants to enrol to programmes with better prospects (e.g. longer in duration).

Keywords: Big Five; Meta-analysis; Non-cognitive skills; Earnings, Programme effectiveness; Returns

Authors: Sofie Cabus (University of Leuven, Belgium), Joanna Napierala (Joint Research Centre), Stephanie Carretero-Gomez (Joint Research Centre)

Acknowledgements: We would like to thank Milos Kankaras and Giorgio di Pietro for valuable comments to this paper.

Joint Research Centre reference number: JRC123308

Contents

1. Introduction

The debate on what skills people need for the future, particularly in the context of the changing world of work, is very lively (e.g. Gonzalez-Vazquez et al. 2019). There is a common understanding that new advanced technologies (e.g. robotics, artificial intelligence, internet of everything) are going to affect the world of work. However, there is a disagreement among scholars about the share of jobs being at high risk of automation (Frey and Osborne, 2013; Arntz et al. 2016; Nedelkoska & Quintini, 2018; Lordan, 2018). While some jobs might undergo transformation due to a change in tasks content (Eurofound, 2020), others are expected to be created as it has happened in previous waves of technological change (e.g. Gonzalez-Vazquez et al. 2019). However, the bottom line is that the nature of jobs is transforming and, consequently, the demand for workers' skills is changing too. This implies, in turn, that the returns to skills are also expected to change. The World Economic Forum (2015) emphasises that "*to thrive in today's innovation-driven economy, workers need a different mix of skills than in the past*" (p.2).

Some changes in the demand for skills are already observed, for example Jaimovich and Siu (2018) show that between 1980 and 2000 there was a positive change in the importance of social skills in the occupations, which translated into an increase in the demand for high-skilled female workers. Many ongoing discussions on what skills will be needed for the future are pointing at growing importance of socioemotional skills (Puerta, Valerio, & Bernal, 2016). A recent analysis of online job advertisement indicates that teamwork and adapting to change were the two most frequently mentioned skills by hiring employers (CEDEFOP, 2019). Some forward-looking studies indicate that workers equipped with cognitive and meta-cognitive skills (e.g. critical thinking), non-cognitive skills (e.g. empathy, work readiness and collaboration), and digital skills, are expected to better fit into future work environments (OECD, 2019). CEDEFOP survey data (2016) shows that most of the jobs, which are anticipated to expand until 2025, require at least a moderate level of digital skills and a high level of non-cognitive skills. These forecasts find reflection in the situation of workers, who are expected to earn more in occupations that require a combination of non-cognitive skills with moderate or advanced use of ICT skills (Gonzalez-Vazquez et al. 2019). In addition, since the 80s, the increasing importance for social skills on the labour market, the skills in which humans have advantage over machines, translates into observed higher wage premiums (Deming, 2017).

Indeed, several studies point to the growing importance of non-cognitive skills for employability and earnings. In this paper, we aim at enhancing knowledge and add some more recent evidence on the relationship between non-cognitive skills and earnings. In particular, we discuss non-cognitive skills in the context of people's labour market positions, operationalised by levels of earnings. We have systematically collected articles discussing non-cognitive skills and labour market performance that were published in the last decade (2009-2019). We have narrowed down our research to this period as the prior studies in this area had been reported to have several limitations related to e.g. reverse causality and measurement error (Heckman et al., 2006). We have retained 29 empirical studies, delivering 333 estimates, from which we can draw conclusions, with regard to what kind of non-cognitive skills are rewarded on the labour market.

Second, in order to draw conclusions at a much more disaggregated level than ever done before, we have constructed two databases from the research findings. While the first database focuses on observational data (see Table 1), collected from estimates obtained

from Mincer equations or similar single-equation regression models (following Montenegro & Patrinos, 2014 and Patrinos, 2016), the second database delivers causal evidence on programme effectiveness (following Kluve, 2016). All these programmes aimed at increasing participants' non-cognitive skills (see Table 2).

We address the three following research questions:

1. What is the state of the art of knowledge on relationship between non-cognitive skills and earnings
2. What is the relationship between non-cognitive skills and earnings based only on quantitative and robust studies.
3. Are training programmes, targeted at non-cognitive skills and improvement of participants' employability, also effective in increasing participants' earnings?

Consequently, we will start with the systematic literature review to address the first question. Then the second question will be addressed by applying meta-analysis to studies with observational data; and the third question by using the database with causal evidence from training programmes targeted at non-cognitive skills.

Additionally, we have clustered the non-cognitive skills retrieved from the 29 studies by using a homogenous definition. For example, with regard to the first database, most studies have constructed standardised scales for non-cognitive skills using surveys that underlie the Big Five inventory. For the second database we could extract programmes solely focusing on non-cognitive skills from those programmes that combine them with academic (often vocational-oriented) skills. Consequently, these two databases together offer a unique disaggregated perspective on the heterogeneous returns to non-cognitive skills, from which conclusions could be drawn for policymakers, school leaders and programme designers.

Finally, the collected study and programme characteristics allow us to make statistical inference for: (1) men and women separately; (2) the returns to non-cognitive skills controlling and not controlling for educational attainment; (3) programme effectiveness targeted at non-cognitive skills (un)conditional on employment; and (4) the effectiveness of trainings in non-cognitive skills by programme duration, population characteristics, and the timing of the data collection after programme ending. Statistical inference from the two databases is facilitated using different meta-analysis techniques described in the section 4.

This paper proceeds as follows. In section 2, we discuss the issue with the measurement on non-cognitive skills. In section 3 we introduce the findings on the relationship between non-cognitive skills and earnings based on previous studies. In section 4 we explain our data and the process of building databases. In section 5, we present different techniques (and steps followed) in the meta-analysis. The main results with regard to the returns to personality traits are presented in Section 6. Sections 7 and 8 deal with the effectiveness of programmes targeted at interpersonal and soft skills and personality traits. Section 9 includes a discussion on robustness of the results. Section 10 concludes.

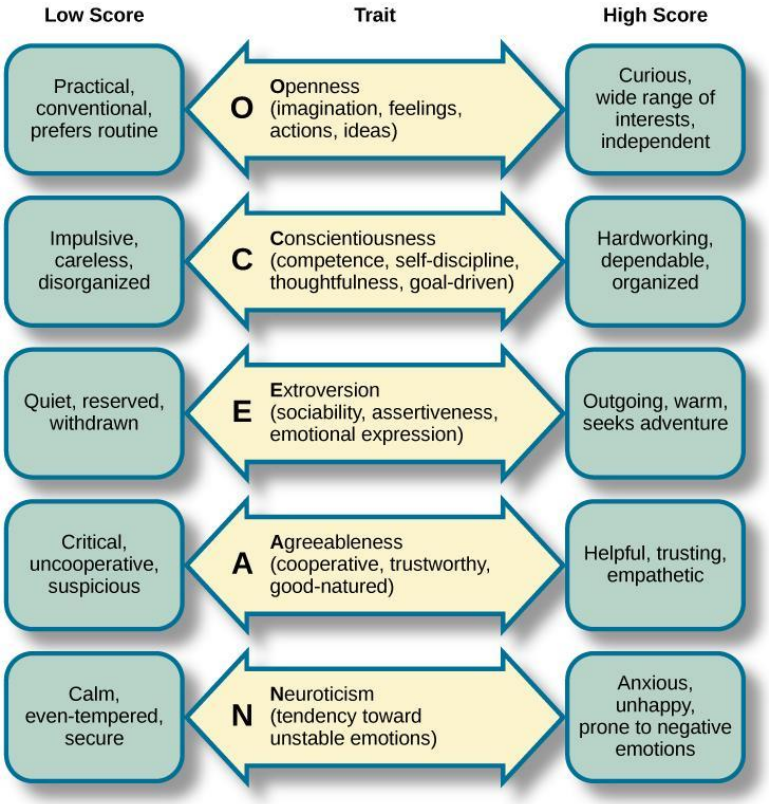
2. Addressing measurement of non-cognitive skills

While there is a common understanding of what cognitive skills are, the attempts of defining what constitute the remaining important skills seems far more complex. Probably, defining these skills by exclusion and labelling them 'non-cognitive' as Messick did (1978, p.2) added more complexity rather than simplifying the understanding of what those skills are. Non-cognitive skills can be defined as patterns of thoughts, feelings and behaviour that affect social interactions with others (Borghans et al., 2008). Farrington et al. (2012) highlight that the term 'non-cognitive' is unfortunate as it reinforces a false dichotomy. Also other authors indicate potential shortcomings in the use of the term 'non-cognitive'. For example, Duckworth and Yeager (2015) point at its lack of precision, encompassing heterogeneity of categories and difficulty in measurement. Moreover, Kankaras (2017) points at the inaccuracy of the term itself, which in his opinion is "too broad to be useful" and "incomplete". In particular, it implies the absence of cognitive activities: this is also misleading, as the use of those skills in many instances fundamentally depends on cognition (Ibidem). Despite these deficiencies in the term non-cognitive, economists have a preference to use it to refer to about behavioural characteristics and personality traits (Sánchez Puerta et al., 2016).
"Soft skills" is another label used by Heckman and Kautz (2012) who tried to summarize what important skills, apart from cognitive, are still missed or mismeasured in achievement tests. Yet, in the literature we can find plenty of other definitions (e.g. non-cognitive skills, social and emotional skills, 21st Century skills, character skills, life skills) used by researchers referring to very similar and overlapping concepts of skills (e.g. Sánchez Puerta et al., 2016). Sometimes, researchers may refer to the same concepts using different labels. For example, Deming (2017) deemed social skills increasingly important for the job market, which he defined as sociability and participation in sports or clubs, or extraversion. He showed that the returns to these social skills in the U.S. increased from 2.0% in the late 1980s and 1990s to 3.7% between 2004 and 2012. He additionally argued that these findings were robust to controlling for non-cognitive skills, defined as locus of control and self-esteem. By contrast, Heckman and Kautz (2013) classified locus of control and self-esteem as "character skills". Furthermore, social skills are also labelled interpersonal skills (Acemoglu & Autor, 2010; Autor & Handel, 2013), soft skills (Heckman & Kautz, 2012), or people skills (Borghans et al., 2014). As a result, given the variety in the definitions underlying those skills, it is not surprising that "*hard evidence on soft skills*" (Heckman & Kautz, 2012) is difficult to deliver.

One may then rightly wonder whether (and how) we can draw unambiguous conclusions from the previous literature on non-cognitive skills, as this confusion among what those skills are may seriously hamper statistical inference. Yet, despite this lack of agreement on defining those skills in theoretical discussions, in empirical analysis researchers most often referred to the Personality model called "Big Five" to assess non-cognitive skills of individuals. This model emerged in the field of psychology as a meaningful taxonomy for studying individual differences (Barrick & Mount, 1991). It recognises five broad dimensions of personality traits: conscientiousness, openness, extraversion, agreeableness, neuroticism. The personality traits of the Big Five model are often presented with the acronym CANOE or OCEAN. In this model, each person has five traits (See chart 1), each scored on a continuum from high to low. The main reason of the popularity of the Big Five model among researchers is that its validity of the measurement has been demonstrated through self-, observer-, and peer-ratings (Fiske, 1949; Goldberg, 1990; Norman, 1963; Smith,1967). This model, although criticised by some researchers (e.g. Block 2010), is considered as the most accurate

approximation of the basic personality trait dimensions (Funder, 2001) and the Big Five personality traits tend to be relatively stable over our lifespan (Spielman et al., 2014), with conscientiousness and agreeableness having some tendency to increase, and neuroticism and extroversion tending to decline slightly, with age (Ibidem).

Figure 1 Big Five model – continuum of traits from low to high



Source: Spielman et al. (2014) p. 399

There are also other alternative methods to measure personality traits. For example, the HEXACO model differs from the Big Five through its inclusion of a sixth factor, Honesty-Humility (also known as the H-factor) (Ashton & Lee 2007). Nevertheless, the main advantage of Big Five model, as compared with other competing models of human personality differences, is the highest amount of accumulated empirical evidence.

3. What do we know about relationship between non-cognitive skills and earnings?

Individuals may treat accumulation of human capital – in terms of skills, knowledge, and abilities – as an investment, expecting that the incurred costs of education and training may bring them higher earnings in the future. The relevance of skills in explaining differences in earnings was first analysed by Becker (1964) who though focused only on cognitive skills. While a series of meta-analyses studies summarizing the effectiveness of personality traits in predicting job performance were carried out in the 1990s (Barrick & Mount, 1991; Tett et al., 1991, 1999), few of them focused on relation of personality traits and earnings. The conceptual framework in which personal traits were included in the explanation of the

residual variance in the individual level of earnings not captured by the usual measures of schooling and work experience was proposed by Bowles and colleagues (2001). Yet, the research on returns to skills started to gain on importance only when the skills became acknowledged as the key ingredient in modern knowledge-based economies (Hanushek & Woessmann, 2008; Broecke, 2015; Hanushek et al., 2015). However, difficulty in measurement of skills and omitted variables (e.g. skills not taken into account) are pointed by Hampf et al. (2017), next to reverse causation, as potential threats to causal identification of the relationship between skills and labour market outcomes.

Individuals' personal traits may have impact on wages by directly affecting individual's productivity, or indirectly by impacting their level of schooling or work experience (Heckman et al., 2006). The direct effect stems from personality being considered part of a worker's endowment, which is rewarded directly by the employer. Looking more closely at indirect impact, we see that the non-cognitive skills explain the variance in educational attainment (see Cunha et al., 2010). For example, whilst conscientiousness has the most predictive power to explain the educational attainment (Almlund et al. 2011), the correlation between neuroticism, extraversion, and agreeableness and educational attainment is negative (Goldberg et al., 1998; van Eijck & de Graaf, 2004). Additionally, Roberts et al. (2007) enlist five reasons why personality traits matter for the occupational achievements. Firstly, because people's characteristics matters for individuals' choices of education paths and also choices of occupations. For example, more conscientious people may prefer conventional jobs, such as accounting, whilst extraverts may prefer to become teachers or work in business management. Secondly, people with certain personal characteristics may be treated in a preferential way. For instance, more extraverted, conscientious, and less neurotic applicants may be perceived in a better way by recruiters and receive job offers more frequently (Cook et al., 2000). Third, personality traits may positively affect individuals' adaptability to work environments and this translates into their better productivity. Fourth, people behaviour may be determined by their personality traits. For example, some people may drop schools or jobs as consequences of "attrition" or "deselection pressures" (Cairns & Cairns, 1994). Finally, personality traits may promote certain kinds of task effectiveness. For instance, neurotic - emotionally unstable - individuals may experience either too much or too little external stimulation, and consequently this may lead them to perform their tasks poorly (Gardner & Cummings 1988). Extraversion is a strong predictor of workers performance, especially those in the upper tail of income distribution. Therefore, when looking at the relationship between non-cognitive skills and wages, it is important to control for occupational level and occupational status. Using occupation as a control variable is also important in the context of bidirectional relations between personality traits and job performance, as for example, extraversion may be required in some jobs whilst introversion in others (Tett at el., 1991).

Existing research results suggest that the relationship between social and emotional skills and income varies also in terms of gender. For example, Mueller & Plug (2006), in a model in which they control for occupation, industry and job characteristics, find that male earned more when being less agreeable, more emotionally stable, and more open to experience, whilst for women, conscientiousness and being open to experience were associated with higher earnings. On average, men are less agreeable than women and they are alone getting wage premium for that trait (Mueller and Plug, 2006). Similar results were observed in a study of O'Connell & Sheikh (2011): men obtained 1.3% lower earnings per each unit increase on the agreeableness scale, while this effect was not observed for women. In this study, all personality factors, except extraversion, had significant effects on earnings with

openness being the most significant predictor of the variance in earnings, followed by emotional stability and conscientiousness (O'Connell and Sheikh, 2011). Apart from the same gender association between agreeableness and earnings, Nyhus & Pons (2005) argue that when looking at impact of personality traits on variation in earnings, it is crucial to control for educational level. In this line of research, Gensowski (2014) is showing that men with post-graduate education benefit more from conscientiousness or extraversion when compared to men with a bachelor's degree or less. The relationship between non-cognitive skills and income is also moderated by cultural context, which is the key environmental determinant of a person's personality. For example, Rammstedt et al. (2017) explain that the unexpected negative association between openness and income observed in several countries, like Germany or Latvia, could be attributed to cultural differences. Also the analysis of a large Dutch sample conducted by Gelissen & de Graaf (2006) reported that openness was negatively correlated to earnings for men (but uncorrelated for women).

4. Data

In order to bring more understanding on the relationship between non-cognitive skills and labour market outcomes, in the first step, we performed the systematic literature review. We have screened relevant literature published in English between 2009 and 2019. We departed from a list of key words written in Boolean terms (see Box 1). In our search, we have used extensive list of skills. We understood labour market performance broadly e.g. employability, level of wages, job quality and length of job search. In particular, we have used a set of two separate Boolean terms that allow us to distinguish between literature on non-cognitive skills and outcomes on the labour market, on the one hand, and literature on personality traits and outcomes on the labour market, on the other hand.

Box 1: Boolean terms with regard to a systematic literature on skills and the labour market

Economic literature on non-cognitive skills and the labour market

((non-cognitive OR cognitive) AND (education OR skills OR competences OR IQ)) AND ((labour OR labor) OR (work) OR (employed OR employment OR earnings))

Psychological and educational literature on non-cognitive skills and the labour market

(personality traits OR soft skills OR character skills OR human literacy OR 21st century skills OR life skills OR key competences OR social skills OR emotional skills) AND ((labour OR labor) OR (work) OR (employed OR employment OR earnings))

Notes In the first phrase of Boolean terms we have mentioned non-cognitive and cognitive together whereas many articles discuss both terms together. This expands the list of search results, after which we have assessed the relevance of the articles including 'cognitive skills'. Articles were only kept in case they mention both non-cognitive and cognitive skills.

With regard to the key words on personality traits, we have additionally searched on key words, which we have found in the literature and had overlapping meaning, for example: socio-emotional skills; social and emotional skills; transversal competences; transversal skills; open-mindedness; creativity; entrepreneurship; digital competence; digital skills; resilience; adaptability; critical thinking; problem solving; teamwork; flexibility; innovation; sociability; empathy; collaboration; emotional control; self-efficacy; openness to learn and to

change; curiosity; creativity; planning/organisation; responsibility; persistence; communication; initiative; positivity.
We used the search engines: LIMO, Web of Sciences and Scopus, ERIC and Google Scholar. To conclude with, we have added articles to the database using other methods, namely cross-citations of heavily cited papers in the past 10 years and snowball methods. Eventually, we have found 117 relevant articles (see Figure 1).

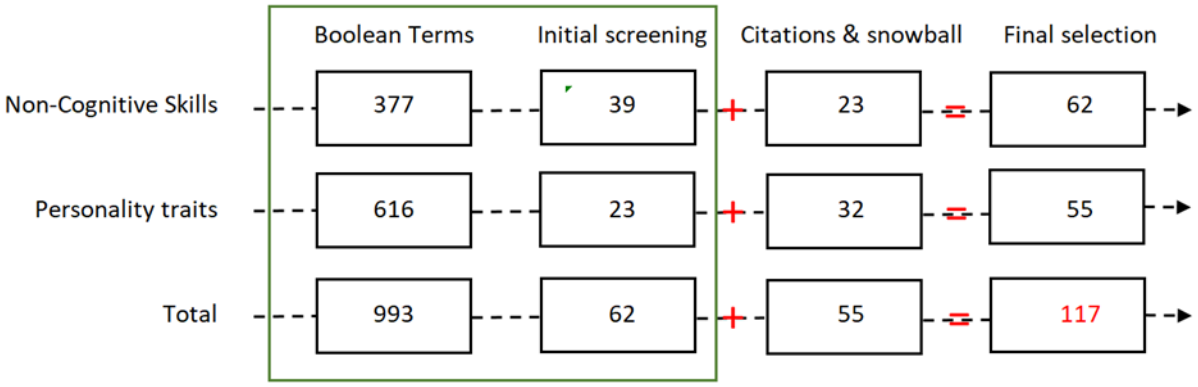


Figure 2: Selection of articles based on systematic literature review.

These 117 studies fit the additional following inclusion criteria:

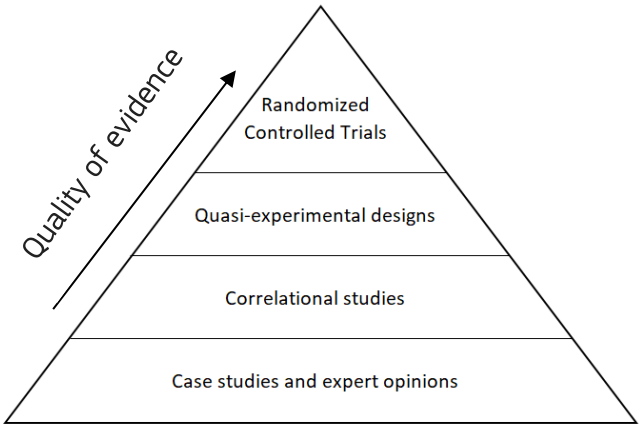
- Published in peer reviewed journals, or accessible online in case of reports, between 2009 and 2019;
- English-written articles, working papers, or reports, which are downloadable by KU Leuven login[1] or by using open access;
- Books are excluded[2];
- Geographical coverage: all countries are included

In the following step, the articles were coded with values from 1 to 4 using the classification from the evidence-based pyramid (Paul & Leibovici, 2014; Berlin & Golub, 2014) (see Figure 3). This was done to distinguish between papers with the highest level of evidence (e.g. randomized controlled trials (RCT) or experimental studies), which were coded with the value of 1, from the quasi-experimental studies, coded as 2 (e.g. difference-in-differences, regression discontinuity design or matching estimators). The correlational studies were coded with a value of 3 because they lack of (random) assignment to a treatment, and the authors of such studies face endogeneity problems that may bias the presented estimates (e.g. omitted variable bias, error-in-variables bias, and reversed causality). The interpretation of results of such studies needs to be treated with caution. In the fourth category we retained studies that are qualitative in nature (e.g. case studies), that consist of expert opinions or

[1] KU Leuven has access to a wide variety of journal articles and reports. It additionally supports the search engine LIMO that contains over 1.5 billion items (e-journals, books and e-books, databases and other media). LIMO and Web of Science yield similar results. As such, the KU Leuven login was not an issue in downloading most of the items for this study.
[2] We agreed upon meeting with JRC that books were excluded from the search for articles. Books are not easy accessible for download and costly, and not easy to summarize in the database as they often cover many other topics than relevant for this study. If dissertations were considered useful, we have searched for the publication of the relevant chapter in a journal, accordingly.

that relate to conceptual (theoretical) models without (causal) empirical evidence. Articles presenting systematic literature reviews can be classified according to the studies that the review discusses. For example, a meta-analysis relying on quasi-experimental studies can be classified in category 2. A literature review of correlational studies can be placed in category 3, etc. However, the articles presenting literature reviews that use narrative approaches will be classified in category 4.

Figure 3: The evidence-based pyramid



Source: A modified version of the evidence-based pyramid (originally produced by Jan Glover, David Izzo, Karen Odato and Lei Wang, EBM Pyramid and EBM Page Generator, copyright 2006 Trustees of Dartmouth College and Yale University). https://guides.lib.uci.edu/ebm/pyramid, accessed on 18 February 2021.

The final version of database containing 117 articles was used as a basis to extract articles for further meta-analysis. As the evidence on other topics linking non-cognitive skills and labour market performance is still scarce, we decided to narrow down the meta-analysis to analytical work focusing on relationship between non-cognitive skills and level of earnings. These articles were divided into two subsets: the first set includes 15 studies (see Table 1) that cover the analysis of relationship between non-cognitive skills and earnings in which estimates were presented by using Mincer equations or correlations (we called it Database I) and the second set of articles consists of studies measuring the impact on earnings of training programmes targeting at non-cognitive skills (we called it Database II, see Table 2). As the Boolean algorithm used for the systematic literature review did not include terms like "intervention"; and "training programmes" an additional search was done to update the list of already identified research articles. The article written by Kluve et al. (2019) presenting the systematic review on this topic was also used for that purpose. This additional effort to select other relevant articles, identified 10 additional studies, which were added together with other articles, and estimates extracted from these studies are included in the database II. In terms of quality it is worth mentioning that all studies but one in database II are randomized controlled trials and all studies except two are covering a developing country.

In general, we have noticed that the quality of evidence with regard to the labour market outcomes ascribed to non-cognitive skills improved considerably over the last decade. For example, of the 29 studies from database II, 13 studies are based on randomised controlled trials, and 9 studies use a quasi-experimental method. As such, only a minority of studies

(7) is correlational in nature. Kluve et al. (2019) argue that there is a general ongoing trend post-2010 of assessing impact of training programmes in developing countries. The training programmes mostly involve training of interpersonal or soft skills, or work readiness skills, and are sometimes provided in combination with personality traits and/or cognitive skills (e.g. vocational training through job placement or basic skills training, numeracy, literacy). The training programmes run for 6 to 12 months, however, a small number of programmes are considerably shorter (e.g. Groh et al., 2016) or longer (e.g. Rodriguez-Planas, 2012).

Table 1: Selected studies on the returns to non-cognitive skills in Database I

ID	First author	Year	Geographical coverage	Meth.[1]	Skills[2]	Definitions of skills
1	Acosta	2015	Colombia	3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
2	Adhitya	2019	Indonesia	3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
3	Albandea	2018	France	3	Other	perseverance; self-esteem; risk-taking
4	Balcar	2016	Czech Republic	3	Other	scale of 15 soft skills
5	Chowdhury	2017	Bangladesh	3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
6	Cunningham	2016	Peru	3	Big Five & other	consistency of interest; cooperation; kindness; perseverance; conscientiousness; emotional stability; openness
7	Deming	2017	U.S.	3	Other	social skills; locus of control; self-esteem
8	Diaz	2013	Peru	2,3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
9	Fletcher	2013	U.S.	2,3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
10	Gensowski	2014	U.S.	2	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
11	Girtz	2012	U.S.	2,3	Other	locus of control; self-esteem
12	Heineck	2010	Germany	2,3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
13	Heineck	2011	United Kingdom	2,3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
14	Hilger	2018	Bangladesh	2,3	Big Five	agreeableness; conscientiousness; extraversion; neuroticism; openness
15	Lindqvist	2011	Sweden	2	Other	scale for non-cognitive skills

Table 2: Selected studies on the returns to non-cognitive skills in Database II

ID	First author	Year	Geographical coverage	Meth.[1]	Training programme targeting at…	Programme duration
16	Adhvaryu	2018	India	1	Workplace based training of soft skills.	12 months
17	Adoho	2014	Liberia	1	Livelihood and life skills training in combination with vocational skills.	12 months
18	Alvares de Azevedo	2013	Kenya	1	A comprehensive employability skills program including life skills training.	24 months
19	Blattman	2014	Uganda	1	Business skills training with focus on improved decision-making, psychological health and self-esteem.	12 months
20	Calero	2016	Brazil	1	Socio-emotional skills in combination with vocational skills and basic skills	6 months
21	Calero	2017	Brazil	1	Socio-emotional skills and work-readiness skills in combination with vocational skills and basic skills	6 months
22	Card	2011	Dominican Republic	1	Self-esteem and work-readiness in combination with vocational training.	Variable duration
23	Cho	2013	Malawi	1	Psycho-social well-being in combination with entrepreneurial and vocational skills	3 months
24	De Coulon	2010	England	2	Remedial intervention aiming to improve non-cognitive skills	12 to 24 months
25	Gertler	2013	Jamaica	1	Psycho-social stimulation.	24 moths
26	Groh	2016	Jordan	1	Soft skills (mainly work readiness).	0.5 months
27	Ibarraran	2014	Dominican Republic	1	Personality traits and work readiness in combination with vocational training.	2 months
28	Premand	2016	Tunisia	1	A track providing entrepreneurship education in Tunisian universities, including a module for personality traits.	12 months
29	Rodriguez-Planas	2012	U.S.	1	Social skills and work readiness, sense of community membership, in combination with educational services to increase academic performance.	60 months

5. Description of the meta-analysis approach

There are three different meta-techniques used throughout the analytical sections.

Firstly, with regard to Section 6, we present an estimate of the weighted mean of the estimates ($\bar{E}$), which we define as follows:

$$\bar{E} = \frac{\sum \omega_i E_i}{\sum \omega_i}, \quad (1)$$

where ω_i is the inverse variance weight for effect size $i \in \{1,2,\dots,k\}$ with k being the total number of estimates. Owing to ω_i, we give studies with a higher precision a higher weight in the average effect size. This model corresponds to a "fixed effects model" or a "common-effects model" in meta-analysis (Borenstein et al. 2009), and we estimate this model using a weighted least squares (WLS) regression.

$$E_i = \alpha_0 + \varepsilon_i, \quad (2)$$

where $\hat{\alpha}_0$ is the mean weighted return to non-cognitive skills that only deviates from the 'true' estimate on the returns by a sample error ε_i.

As such, in Section 6 we prefer a fixed or common-effects model estimation above other meta-analysis techniques. There are several reasons for this choice. First, many coefficients (81.2% of 245 coefficients) on the personality traits included in database I are coming from a standardised questionnaire measuring the 'Big Five'. Second, whereas the coefficients on personality traits are coming from correlational (59.1%) or, at most, quasi-experimental studies (40.9%), we do not aim to present 'causal evidence' on the returns to personality traits but rather make descriptive observations. A common-effects model that gives studies with higher precision (i.e. those studies with a larger sample size and a smaller standard error) a greater weight is then deemed sufficient. A final reason for choosing the fixed effects model, is that it introduces less bias than, for example, the random effects model, in case when the average sample size of the studies is large (Poole and Greenland 1999; Furukawa et al., 2003). This is the case for database I with an average sample size of ($\bar{N}_r = 9223$).

In Section 7, we additionally discuss the random effects model in order to estimate 'the genuine' effect size (ES), or impact, of the training programmes targeting at non-cognitive skills (Schwarzer et al., 2015). As such, the outcome variable in Section **Error! Reference source not found.** is an effect size (ES), and not an estimate (%) like in Section **Error! Reference source not found.**, whereas the underlying data is coming from different type of studies (for a discussion, see Section **Error! Reference source not found.**).

On the other hand, there is a substantial between-study heterogeneity in database II that pushes us into the direction of using a random effects model. The between-study heterogeneity in database II comes from the fact that: (1) almost always authors have used their own survey instruments to measure non-cognitive skills and labour market outcomes;

(2) the sample size ($\overline{N}_r = 1103$)[3] is lower in database II than in database I (that could rely more on administrative data); (3) there is more heterogeneity in the model specifications that authors used to estimate the effect size (e.g. the way the outcome variable earnings got transformed using the logarithm or not); and, importantly, from database II we expect to retrieve a 'genuine effect size', or causal effect, as all studies are randomized controlled trials. For all of these reasons we prefer the random effects model compared to the fixed effects model (Schwarzer et al., 2015).

A random effects model introduces a second source of error, denoted with parameter θ_r, that reflects the heterogeneity of the underlying populations between the studies included in the meta-analysis. It is a matter of fact that different studies stem from different populations around the world. Consequently, there is not one single effect size that measures the true impact of the training programmes in non-cognitive skills, but rather a distribution of effect sizes. We rewrite equation (2) as follows:

$$ES_i = \beta_0 + \theta_r + \xi_i \quad . \tag{3}$$

In practical terms, we add the variance of the distribution of the effect sizes to the regression by estimating θ_r in a parameter called $\hat{\tau}$. The subscript r denotes a study $r \in \{1,2,\dots,R\}$. The variance that cannot be explained by the between-study heterogeneity and sample error $(\theta_r + \xi_{ir})$ is then captured by β_0, the weighted mean effect size.

In Section 7, we also present a multivariate analysis. For this purpose, we prefer the residual maximum likelihood (REML) random effects model that supresses the constant β_0. Several studies have shown that ML-estimators have better properties in estimating the between-study variance (Sidik and Jonkman 2007; Viechtbauer 2005). Furthermore, random effects models allow including a set of control variables X_j.

$$ES_i = \beta_0 + \sum \gamma_j X_j + SE^2 + \theta_r + \xi_{ir} \tag{4}$$

Equation (4) explores where the variation in effect sizes comes from by gradually including control variables like programme duration, demographic characteristics of study participants (age, gender, education), and several model specifications that authors use to estimate the effect size (log-linear specifications, standard deviation reported or not, and when the data got collected after the end of the training programme).

A standard error squared is included in equation (4) as a control variable. This model refers to a Precision Effect Estimate with Standard Error Meta-Regression Analysis (PEESE-MRA) (Stanley and Doucouliagos, 2014), and has been estimated by several other authors (among

[3] There is one study (De Coulon, 2010) with a sample size of 99,260 that yields two effect sizes. We did not include this study in this average as that would falsely lead to the conclusion that the average is equal to 4609 instead of 1103.

others, Vooren et al., 2019). The coefficients can then be interpreted in absolute levels of effectiveness, captured into the control variable, instead of relative differences.

6. The returns to non-cognitive skills

In this section, we present the main results of meta-analysis applied to the estimates extracted from articles belonging to database I. The presented coefficients in Table 3 can be read as returns (in %) to non-cognitive skills. For example, a one standard deviation increase in non-cognitive skills increases the returns to earnings with 1.3% (see column of the unweighted mean and variable 'non-cognitive skills' in Table 3). 95%-confidence intervals are presented in Table 3 below the coefficients between brackets.

Table 3: The returns to non-cognitive skills based on database I

		Weighted mean	
	Unweighted mean	**Not Controlled for Education**	**Controlled for Education**
Cognitive skills (29/7/22)	0.112 ***	0.168 ***	0.088 ***
	[0.07;0.15]	[0.12;0.21]	[0.05;0.12]
Non-cognitive skills (245/55/190)	0.013 *	0.025 **	0.007 *
	[-0.001;0.02]	[0.004;0.05]	[-0.001;0.02]
Big five (199/45/154)	0.013	0.006	0.004 *
	[0-.005;0.02]	[-0.03;0.04]	[-0.001;0.01]
Agreeableness (39/9/30)	-0.026 *	-0.006	-0.021 ***
	[0-.05;0.001]	[-0.02;0.01]	[-0.03;-0.01]
Conscientiousness (40/9/31)	0.043 **	0.067 ***	0.014 ***
	[0.01;0.08]	[0.03;0.11]	[0.01;0.02]
Extraversion (40/9/31)	0.051 **	0.078 ***	0.002 *
	[0.01;0.09]	[0.05;0.11]	[-0.0003;0.003]
Neuroticism (40/9/31)	-0.026	-0.070 **	-0.016 ***
	[0-.03;0.01]	[-0.13;-0.01]	[-0.02;-0.01]
Openness (40/9/31)	0.004	-0.055	0.022 ***
	[0-.02;0.03]	[-0.14;0.04]	[0.01;0.04]
Other definitions (46/10/36)	0.016	0.044 **	0.030 ***
	[-0.01;0.04]	[0.01;0.08]	[0.01;0.05]

Notes: (a) Number of coefficients between brackets (mean/unconditional/conditional).
(b) Null hypothesis that the mean returns to personality traits is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***). 95%-confidence intervals between brackets.
(c) Coefficients weighted by using the inverse of the variance in a Weighted Least Squares (WLS) regression. This corresponds to the estimation of a fixed effects model.
(d) Other definitions include 'consistency of interest'; 'cooperation'; 'kindness'; 'locus of control'; 'perseverance'; 'risk-taking'; 'scale of 15 social skills'; 'scale of non-cognitive skills'; and 'self-esteem'. Although some of these results correspond to Big Five dimensions, we have excluded them and we built a separate category to assure measurement invariance.
(e) The 'not controlled weighted mean coefficients' are an indication that the coefficients were not controlled for (hard measures of) educational attainment. On the contrary, 'conditional weighted mean coefficients' are controlled for educational attainment.

The results presented in columns come from two models: (1) the *unweighted* mean average returns to non-cognitive skills and its standard deviation; and (2) the *weighted* mean average returns to non-cognitive skills. The latter model corresponds to the estimation of what is

called in meta-analysis a “fixed effects model” or a “common-effects model” as explained in the previous chapter. Further, we distinguish between coefficients *not* controlled, and controlled for level of education. Hereby, we wish to account for the fact that educational attainment, or intellectual ability, move together with non-cognitive skills as considered ‘favourable’ by employers (Patrinos, 2016; Deming, 2017).

Next to coefficients describing results of the overall returns to non-cognitive skills, we present also the more detailed ones that point to the personal traits measured by the Big Five model. In a few cases we have reversed a dimension of the Big Five Inventory (corresponding to the third variable in row three in Table 4) as to deal with the heterogeneity of the impact of its dimensions on earnings. For example, the opposite of the dimension ‘Neuroticism’ is called ‘Emotional stability’. In case a particular study would measure ‘Emotional stability’ instead of ‘Neuroticism’, with both dimensions lying on one continuum, we reversed the dimension so as to indicate ‘Neuroticism’. Doing so, the sign of the estimated coefficients in all studies including estimates on Neuroticism move into the same direction. Notwithstanding, we find that non-cognitive skills measured by the Big Five model as a whole yield only very small returns. Controlling for educational attainment in the included studies, the returns to an increase of one standard deviation on the dimensions of the Big Five inventory as a whole are equal to 0.4%. This holds true, too, for the scale of non-cognitive skills (0.7%) including the Big Five and other definitions of non-cognitive skills.

The picture is different when looking at the specific skills measured by the Big Five model. If agreeableness increases with one standard deviation, and when controlling for educational attainment, the returns are equal to -2.1%. We also find a negative return to neuroticism (-1.6%). A positive return is found for conscientiousness (1.4%), openness (2.2%) and for the non-cognitive skills measured differently than with the Big Five model (3.0%).

The remaining impact of personal traits, namely extraversion, is very small (0.2%). However, when not controlling for educational attainment, the estimate of extraversion increases to 7.8%. This means that people, who yield high scores on extraversion, in general have a higher level of education, that, moreover, can be correlated with occupational status. Education is then likely acting as a proxy for occupational status. For example, managers will have higher returns for extraversion than regular employees, and they are more likely to be higher educated, too. Controlling for educational attainment when looking at the returns to non-cognitive skills is then important, because otherwise we would falsely attribute the positive and significant estimate to the personality trait of extraversion, while, in fact, educational attainment (then again, associated with occupational status) is the driving factor of this significance. Further, we argue that the reversed reasoning applies to neuroticism. Without controlling for educational attainment, the returns are significantly negative and equal to -7.0%. Taking into account the level of education, however, this estimate drops to -1.6%. There is an indication that people, who yield high scores on neuroticism, are in general lower educated. However, caution is again in place with interpretation of the coefficients, as indicated in the example of managers above, whereas the data do not allow us to draw conclusions on the relationship between educational level and occupational status. In sum, our findings suggest that personality traits are compensated (in a good or in a bad way) on the labour market by one’s educational attainment, but we are not able to position this variable as a causal mechanism due to omitted variable bias.

In Table 4, we present a summary of results based on models for the general population and broken down by gender where educational attainment is controlled for .

Table 4: The returns to non-cognitive skills by study population (database I)

	Weighted mean controlled for education		
	Female	**Male**	**Female & Male**
Cognitive skills (4/11/10)	0.013	0.044 *	0.110 ***
	[-0.01;0.03]	[-0.01;0.09]	[0.06;0.16]
Non-cognitive skills (54/79/57)	0.003	0.006	0.032 ***
	[-0.001;0.01]	[-0.002;0.01]	[0.02;0.04]
Big five (50/70/34)	0.004	0.006 *	0.014
	[-0.002;0.01]	[-0.0008;0.01]	[-0.008;0.04]
Agreeableness (10/14/7)	-0.029 ***	-0.014 ***	-0.021 *
	[-0.04;-0.02]	[-0.02;-0.01]	[-.04;0.003]
Conscientiousness (10/14/8)	0.012 **	0.014 ***	0.021 **
	[0.003;0.02]	[0.01;0.02]	[0.004;0.04]
Extraversion (10/14/8)	0.001 ***	0.003	0.059 ***
	[.0003;0.001]	[-0.006;0.01]	[0.05;0.07]
Neuroticism (10/14/8)	-0.027 ***	-0.002	-0.029 **
	[-0.04;-0.02]	[-0.008;0.004]	[-0.05;-0.004]
Openness (10/14/8)	0.028 **	0.026 **	-0.019
	[0.004;0.05]	[0.007;0.05]	[-0.05;-0.004]
Other definitions (4/9/23)	-0.070 ***	-0.001	0.040 ***
	[-0.10;-0.03]	[-0.08;0.08]	[.02;0.05]

Notes: (a) Number of coefficients between brackets (mean/unconditional/conditional). (b) Null hypothesis that the mean returns to personality traits is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***). 95%-confidence intervals between brackets.
(c) Coefficients weighted by using the inverse of the variance in a Weighted Least Squares (WLS) regression. This corresponds to the estimation of a fixed effects model.
(d) Other definitions include 'consistency of interest'; 'cooperation'; 'kindness'; 'locus of control'; 'perseverance'; 'risk-taking'; 'scale of 15 social skills'; 'scale of non-cognitive skills'; and 'self-esteem'. Although some of these results correspond to Big Five dimensions, we have excluded them and we built a separate category to assure measurement invariance.
(e) The 'not controlled weighted mean coefficients' are an indication that the coefficients were not controlled for (hard measures of) educational attainment. On the contrary, 'conditional weighted mean coefficients' are controlled for educational attainment.

An observation from Table 3 is that the negative returns to agreeableness stems from the studies based on female populations only (-2.9%) and to lower extent from studies based on male populations only (-1.4%). Similarly, neuroticism plays an important role for women with a negative return equal to -2.7%. This estimate is equal to -0.2% and is not significant for male only populations. Overall, what we observe based on the analysis of estimates for individual gender groups is that non-cognitive skills matter more for women, for example, extraversion and neuroticism does not matter for men earnings at all while it matters for women. At the same time, cognitive skills seem to matter more for men (4.4%) than for

women (1.3% and not significant).[4] Of course, men and women are sorted differently across occupations and economic sectors. Although in most of the studies these variables were control for in the models still these unobserved determinants of the returns to non-cognitive skills may explain, at least, in part, the statistical inference made. Therefore, conclusions about the male and female population should not be interpreted in a deterministic manner.

The highest returns to cognitive skills are found, however, for the mixed gender populations (11.0%) (See table 4). These higher returns (than for male and female only) can likely be explained by the fact that mixed gender populations were included in, for example, larger and more comprehensive studies. This may indicate between-study heterogeneity in the reason why (and for whom) data were collected on particular study populations.

Cognitive skills are often used in studies as an equivalent of years of schooling, for example in Mincer equations. There is indeed a lot of overlap between these two variables. Nonetheless, cognitive skills are still capturing variation unmeasured by educational attainment, such as the level of one's intellectual ability as compared to others. The fact that about half of the estimate of 16.8% can be attributed to a year of schooling (see Table 3) is in line with the global average private return to a year of schooling of 10% (Patrinos, 2016). This also strengthens the credibility of our findings for the returns to non-cognitive skills.

At the same time, presented findings illustrate an inconsistent set of results, with coefficients for cognitive skills varying between 1.3% (female only; Table 4) to 16.8% (full populations; Table 3) depending on underlying populations in the studies. These results illustrate that there is a relatively large degree of disagreements among studies on the returns to cognitive and non-cognitive skills. The fact that we observe this can be attributed, at least in part, to the correlational nature of the findings. Omitted variables bias is certainly present in many (if not all) of the studies, and its impact on statistical inference has been demonstrated above by using the example of occupational status. As follows, we specifically focus on studies using other, more credible research designs.

7. Returns to non-cognitive skills measured by mean effect sizes

In this section, we present the results of the analysis from the database II that includes information on programme effectiveness in equipping people with non-cognitive skills. Therefore, we will no longer refer to returns to non-cognitive skills in percentages, but from now on, we will talk about *mean effect sizes*. To be more precise, an effect size lies between 0 and 1 with 0.2 standard deviations (SD) being considered as a small effect; 0.4-0.7 SD a moderate effect; and a large effect for values above 0.8 SD.

Table 5 presents the main results of our meta-analysis. The table below reports the unweighted and the weighted mean effect sizes. With regard to the weighted mean effect sizes, we have estimated two models. The first model A corresponds to a fixed effects model, and the second model B to a random effects model. We follow a traditional empirical

[4] The fact that this estimate is not significant should be interpreted with caution, whereas we only have four estimates to draw conclusions from for female only populations.

approach for conducting fixed and random effects models, as discussed in Section 5. With regard to the random effects models, we additionally include the tau² statistic. This value indicates that for all models estimated the hypothesis of homogeneity is rejected; which indicates that we should control for between-study heterogeneity. As a result of this conclusion we prefer the random effects over the fixed effects model. An elaborated discussion on accounting for between-study heterogeneity is given in Section 5.

Further, we distinguish between estimates unconditional and conditional on employment status. Whereas the latter model only produces results for employed people only, the former also takes unemployed (with an income equal to zero) into account.

Table 5: Impact of training programmes targeted at development of non-cognitive skills on earnings (mean Effect Sizes; Database II)

		Fixed Effects (Model A)		**Random Effects (Model B)**			
	Unweighted mean (SE)	**Unconditional**	**Conditional**	**Unconditional**	**Tau²**	**Conditional**	**Tau²**
All (59/42/17)	0.143 ***	0.085 ***	0.009 **	0.118 ***	0.014	0.041 ***	0.001
	[0.10;0.19]	[0.07;0.10]	[0.002;0.02]	[0.08;0.16]		[0.01;0.07]	
Non-cognitive skills only (28/18/10)	0.183 ***	0.138 ***	0.026 **	0.192 ***	0.016	0.049 **	0.003
	[0.12;0.24]	[0.12;0.16]	[0.01;0.05]	[0.13;0.26]		[0.001;0.10]	
+ cognitive skills (31/24/7)	0.111 ***	0.025 **	0.007 *	0.039 ***	0.006	0.049 **	0.001
	[0.04;0.18]	[0.00;0.05]	[-0.001;0.01]	[-0.01;0.08]		[0.01;0.09]	

Notes: (a) Based on 59 effect sizes from database II.
(b) Number of effect sizes between brackets (mean/conditional/unconditional).
(c) Null hypothesis that the weighted mean effect size is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***). 95%-confidence intervals between brackets.
(d) Effect sizes weighted by using the inverse of the variance in a fixed effects model. Weighted mean effect size from Model A additionally controlled for between-study heterogeneity in a random effect model.

We present all mean effect sizes from database II together in the first row. Looking at Model B, unconditional on employment, which is the preferred model for all results below, we conclude that the weighted mean effect size is equal to 0.118 SD, significant at 1% level. Looking at the second row in Table 5, i.e. non-cognitive skills only, we actually present the effect size of training programmes without a cognitive component in the training programme. This unconditional effect size in Model B is equal to 0.192 SD. This is higher than the overall result from all studies (in the first row) – but still considered a small effect size (supra). Programmes that, apart from non-cognitive skills, also put emphasis on training cognitive skills, are considered *the least effective*. The effect size is equal to 0.039 SD.

In Table 6, we additionally present the weighted mean effect sizes by programme characteristics. We consider: (a) study population; (b) programme duration; and (c) the point in time at which post-treatment data were collected (i.e. a short-term or long-term follow-up).

First, let us consider the study population. Largest effect sizes (0.187 SD) are found for the female only population. The effect sizes for the male only and mixed populations is equal to

0.081 SD (not significant) and 0.049 SD, respectively. When looking at the type of the skills training programme (not reported in Table 6), then we conclude that these effect sizes for all study populations are again driven by training programmes in non-cognitive skills only.

Next, let us look at results presented starting from row (b) in Table 6. The effect sizes for shorter programmes (6 months or less) are equal to 0.047 SD. These estimates increase to 0.154 SD for programmes with a longer duration (more than 6 months).

Starting from row (c) we additionally show results of models in which we look at the point in time when the data got collected by the authors. From these results we can conclude that there are in general larger effects of the training on earnings in the short-run than in the long-term.

Table 6: Unconditional weighted mean effect sizes by programme characteristics (Database II)

	Model 1		**Model 2**		**Model 3**	
(a) study population						
Female (17)	0.187	***				
	[0.11;0.26]					
Male (5)	0.081					
	[-0.06;0.23]					
Female & Male (20)	0.049	**				
	[0.01;0.09]					
(b) programme duration						
≤6 months (16)			0.047	**		
			[0.01;0.08]			
> 6 months (26)			0.154	***		
			[0.09;0.22]			
(c) follow-up						
short-term (17)					0.159	**
					[0.09;0.23]	
long-term (23)					0.060	***
					[0.02;0.10]	

Notes. (a) Number of effect sizes between brackets.
(b) Null hypothesis that the weighted mean effect size is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***). 95%-confidence intervals between brackets.
(c) Weighted mean effect size controlled for between-study heterogeneity in a random effect model.

8. Study characteristics versus effect sizes – multivariate analysis

In this section, we explore the extent to which study characteristics determine the estimated effect sizes in a multivariate meta-regression.

Table 7 presents the results of a multivariate meta-regression using a set of control variables that may determine the magnitude of the effect size (details on this regression are given in Section 5). We explore: (1) programme duration; (2) the demographic characteristics as age, gender and education; and (3) the variable 'follow-up' (how much time after

programme ending were the data collected). The dependent variable is the unconditional weighted mean effect size.

In Model 1, we only include programme duration as an explanatory variable. Programmes with a longer duration (more than 6 months) are more effective than programmes with a very short duration (6 months or less). This is in line with what to expect a priori: the longer a programme takes, the more time trainer (coach, or mentor) has to work with participants and try to increase their level of non-cognitive skills. Controlling for age and gender in Model 2 does not alter these conclusions. In Model 3, we additionally include educational attainment (in years) of the study participants. The variable education is not significant. The number of effect sizes in the regression drop from 42 to 36, so we do not include educational attainment in Model 4.

In Model 4, we additionally account for the variable 'follow-up'. We do not estimate a significant contribution of this variable to the overall model.

Table 7: Programmes characteristics that explain the unconditional weighted mean effect sizes (Database II)

	Model 1		**Model 2**		**Model 3**		**Model 4**	
SE^2	0.017		0.010		0.010		0.010	
	(0.02)		(0.02)		(0.02)		(0.02)	
Programme duration (Ref. ≤6 Months)								
>6 Months	0.147	***	0.167	***	0.145	***	0.109	**
	(0.03)		(0.03)		(0.03)		(0.05)	
Control variables								
Age (std)			0.089	***	0.080	***	0.090	***
			(0.02)		(0.02)		(0.02)	
Gender (std)			0.000		0.018		0.001	
			(0.02)		(0.03)		(0.03)	
Education (std)					-0.004			
					(0.02)			
Model specifications								
Follow-up (Ref. Short-term)								
Long-term							0.058	
							(0.05)	
Number of ES	42		42		36		40	

Notes. (a) Null hypothesis that the weighted mean effect size is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***).
(b) Standard errors between brackets.
(c) Effect sizes regressed on the variance (SE^2) and using a Residual Maximum Likelihood (REML) random effects model accounting for the heterogeneity in distribution of the effect sizes and allowing for the inclusion of control variables. We additionally suppress the constant.
(d) 'Follow-up' indicates that we compare effect sizes at long-term follow up after programme ending with effect sizes at short-term follow-up.

Age (years standardised) is positive and significant across all models estimated. This indicates that older participants gain more from the training programmes in non-cognitive skills than younger participants. Nonetheless, we need to make this interpretation with

caution as it may be the case that age picks up features of programme design in the multivariate regression analysis. For example, programmes reaching out to older persons, included in this meta-analysis, could have been better designed than those programmes targeting at young participants. Or else, younger persons were lacking motivation to engage in the trainings due to having bad experience from the period of formal schooling. Another aspect to take into account is that some characteristics (which we do not observe in the studies) may also be associated with individual's labour market prospects, like job opportunities and wages. At least the evidence we find here seems in line with the results of other skill development programmes that indicate that skills in general are easier and faster developing in younger than in older persons but still non-cognitive skills are more malleable than cognitive skills at later ages (Kautz et al., 2014).

Gender, measured by the share of participants being women (standardised), is never significant in the four models. Combining these insights with those of the previous section, it appears that it is more the characteristics of the training programme, wherein the female population participated, that matters for the effectiveness of the programmes for women only, rather than study population itself.[5] For example, women more likely participated in training programmes having a longer duration; which clearly positively influenced their non-cognitive skills and subsequent career. While this conclusion can obviously be influenced by several features and characteristics of the programmes considered in the meta-analysis, and, therefore, cannot be generalised as such, it does call for further attention in newly developed randomized control trials.

Besides the four models estimated in Table 7, we have done several other analyses. For example, we have estimated a model where the year of publication of the study was also incorporated as a linear trend or as dummies for each year. For a few years (2012, 2013), this variable was significant, however, this significance was most likely driven by the fact that, in those years we could collect more estimates. Other variables that we have added to the models are the impact factor and H-index associated with the journals wherein the studies were published (described in the next section). The variable impact factor is equal to -0.03 SD in all four models estimated in Table 7. This finding implies that studies published in higher ranked journals, according to the impact factor, generally estimate lower effect sizes, than studies in lower ranked journals. Even though the estimated coefficient of -0.03 SD is small, this seems relevant information, whereas the overall programme effectiveness expressed in effect sizes is small (at most 0.1 to 0.2 SD). The variable H-index, on the other hand, yields a very small and positive coefficient, however, only in model 1, because in models 2-4 this variable is no longer significant.

The fact that higher impact journals yield smaller effect sizes is an important caveat for statistical inference. It is generally acknowledged that estimated program effectiveness, at least partly, can be ascribed to improperly organised and evaluated programs. It is very difficult to assess, however, which part of the estimated effect sizes should be ascribed to those issues of evaluation. In the next section we further assess the quality of evidence retrieved by looking at the impact of publication bias on the estimated effect sizes.

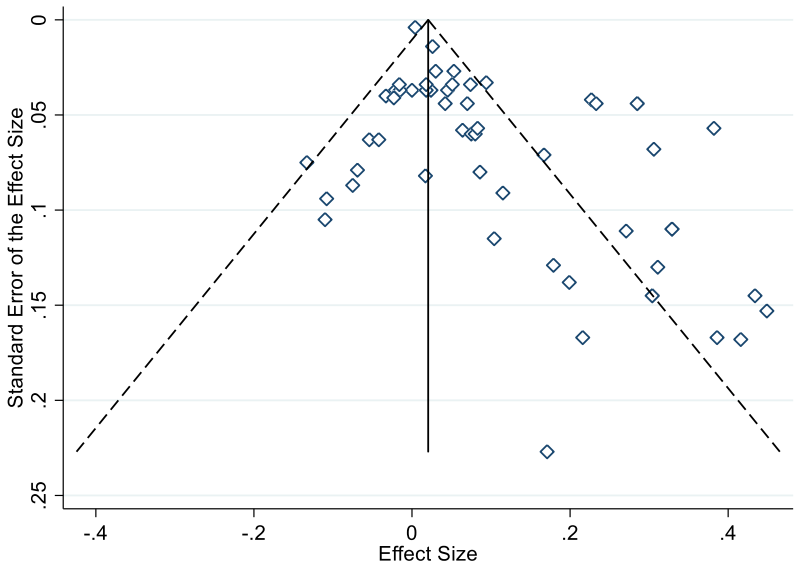
[5] Adhvaryu et al. (2018) and Adoho et al. (2014) specifically report on a training programme developed for women only. The female participants received life skills training (Adhvaryu et al., 2018) and workplaced-based soft skills training (Adoho et al., 2014) followed in both studies by a job placement with support.

9. Publication bias

In general, studies are more likely to be published when a significant effect size can be shown, leading to an upward bias in the estimated mean effect size $\hat{\beta}_0$. Therefore, it is a still a possibility that β_0 from equation (3) presented in section 5 is not the 'true' effect size, for example, when the studies included in database II were more likely to be published.

In order to give an idea of whether publication bias matters for our sample of studies, we plotted the effect sizes against the standard error of the effect size in Figure 4 (i.e. a funnel plot, see Sterne & Egger, 2001). In absence of publication bias, we should observe a symmetric plot around the blue bar. Furthermore, the tendency to only publish studies with significant (positive) effect sizes, leads to missing dots at the left of the blue bar, because this is where the non-significant studies would be included. We conclude from Figure 4 that publication bias indeed plays a role.

Figure 4: Effect size versus standard error of the effect size

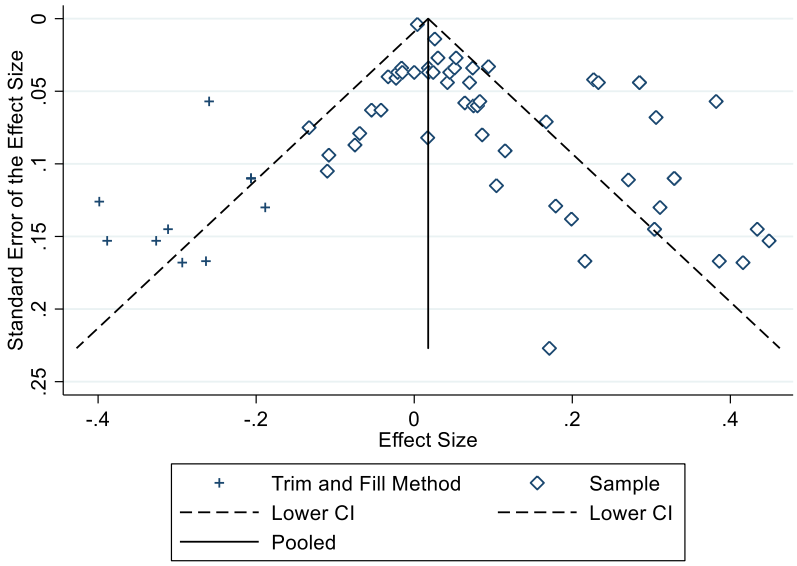


Notes: Blue bar denotes the weighted mean effect size using 59 effect sizes from database°II. The dotted bars are 95% confidence intervals around the mean effect size. The Egger test of the null hypothesis of 'no small-study effects' is rejected at 1-percent-significance level.

There are methods for correcting for publication bias, like the "trim and fill method" that adjusts for censored (missing) studies (Steichen, 2001). In our case, that would mean that we fill-in the missing information of studies with negative effect sizes at the left of the blue bar in Figure 4, with information on studies at the right-hand side of the blue bar.

Figure 5 presents this new funnel plot with an application of the "trim and fill method". In total, we estimate 12 filled effect sizes (+), so that the total number of estimates in database II is equal to 71.

Figure 5: Effect size versus standard error of the effect size with an application of the "Trim and Fill Method"



Notes: Figure estimated by using the nonparametric "trim and fill method" (Steichen, 2001). Blue bar denotes the mean weighted effect size using 59 effect sizes from database°II and 12 'filled effect sizes' estimated with the "trim and fill method". The dotted bars are 95% confidence intervals around the mean effect size.

Table 8 presents the weighted mean effect sizes without and with the "trim and fill method". We do not distinguish here between unconditional/conditional effect sizes, but present the overall picture of what a correction for publication bias would do with the weighted mean effect sizes. The effect sizes corrected for publication bias somewhat decline as compared to the first row in Table 8. However, the overall conclusion remains the same: training programmes are effective in increasing the private return of study participants on the labour market. The effect sizes are positive, significant and small.

Table 8: Impact of training programmes in interpersonal and soft skills on earnings with an application of the "Trim and Fill method" (Database II)

	Fixed effects (Model A)		**Random effects (Model B)**	
All effect sizes (59)	0.021	***	0.096	***
All effect sizes, incl. trim&fill (71)	0.018	***	0.063	***

Notes: (a) Based on 59 effect sizes from database II. Number of effect sizes between brackets.
(b) Null hypothesis that the weighted mean effect size is equal to zero rejected at significance levels 10% (*); 5% (**); and 1% (***).
(c) Models estimated by using the nonparametric "trim and fill method" (Steichen, 2001).

To sum up, the influence of using the "trim and fill method" on the estimated weighted mean effect size (using a random effects model) seems rather limited. There are intuitive reasons to believe that publication bias indeed plays only a minor role in our results and conclusions. We particularly miss studies with negative effect sizes, which also appear to be non-significant, and, therefore, usually did not get published. In practical terms, this means that the training programme in non-cognitive skills did not work and, additionally, may suffer from small sample size and lack of statistical power to prove any small significant negative effects. Indeed, the standard error relative to the effect size is quite large for the filled effect sizes. These two features together, (1) a training programme in non-cognitive skills that would like to yield a positive return on the labour market, but which turns out negatively for the study participants; and (2) small sample sizes, should warn an anonymous reviewer for quality issues with the research design. This kind of studies, indeed, should not alter the overall conclusion on the effectiveness of training programmes in non-cognitive skills. One could even argue that it is not necessary to correct for the likely publication of studies that meet the aforementioned two criteria.

10. Discussion

We have started our analysis with first sharing our concerns that heterogeneity in the measurement of non-cognitive skills across research projects may hinder us from the application of meta-analysis. However, while other competitive personality traits models exist (for example, HEXACO), in the identified papers discussing relationship between personality traits and labour market outcomes published in the last decade, still the Big Five model was the mostly used to measure non-cognitive skills. Yet, despite growing interest in the research on importance of personality traits and labour marker outcome, the body of evidence identified in systematic literature review was still too scarce to extend our discussion in this paper on other labour market outcomes than earnings. Therefore, in this paper, we mainly focused on the presentation of meta-analysis results of relationship between non-cognitive skills and individual earnings. We have identified 29 relevant studies from which 333 estimates were extracted and assigned to two different databases. Evidence from the first database produced estimates based on observational data, whilst evidence from the second one focused on the effectiveness of training programmes aiming at development of non-cognitive skills.

The analysis of first database reveals that, in general, personality traits are rewarded on the labour market with an increase in the level of earnings. However, this positive impact is only observed to a limited extent compared to the level of returns observed for cognitive skills. At the same time, there are also many differences in the sign of the estimates underlying the non-cognitive skills measured by Big Five model. For example, conscientiousness and openness are two personality traits that bring higher earnings, while agreeableness and neuroticism (low emotional stability) are associated with receiving lower earnings. We also observed some indication towards more importance of non-cognitive skills for women, but due to a great heterogeneity of analysed studies the confirmation of this result still requires more evidence. In particular, more studies are needed in order to analyse the relationship between non-cognitive skills and earnings for separate gender groups, as well as to include other important moderating variables (e.g occupational status, economic sector, and education level). Other researchers (e.g. Collischon, 2020) encourage also to look at the comparison of high-wage and low-wage employees separately, in order to avoid, for example, the problem of cross-country differences in minimum wages settings that could overall compress the impact of non-cognitive skills on wages in low income groups.

Based on the analysis of the second database, including results of studies discussing trainings effectiveness, we observe that training programmes with a particular strong (and only) focus on non-cognitive skills are able to yield a positive and significant effect on earnings. However, when training components are focusing on both cognitive and non-cognitive skills, this observed increase is much smaller. Among other factors, this might be related to the design of the training programme. As Durlak et al. (2011) shows the presence of implementation problems may significantly influence the outcomes of the training programmes. Nevertheless, among many aspects in design of analysed training programmes that could have impacted this observed relationship, we were only in the position to control for the impact of duration of the programme, which as expected was significantly positive. We may want to mention here that the heterogeneity in reporting of the results by researchers made the extracting information from single papers quite challenging, and at the same time, made more difficult this analysis, hindering us to include more control variables and to make further recommendations about this relationship. Based on this

analysis, we could only confirm the positive relationship with age, which we interpret in the line with other existing results, which shows that: although in general it is easier to learn new skills when being young, it is also possible to increase the non-cognitive skills at the later stages of life with the overall positive impact on individuals' earnings.

References

Acemoglu, D., & Autor, D. (2010). Skills, Tasks and Technologies (Working Paper 16082). NBER Working Paper Series. National Bureau of Economic Research. http://www.nber.org/papers/w16082

Acosta, P., Muller, N., & Sarzosa, M. A. (2015). Beyond qualifications: returns to cognitive and socio-emotional skills in Colombia. The World Bank. https://doi.org/10.1596/1813-9450-7430

Adhitya, D., Mulyaningsih, T., & Samudro, B. R. (2019). The Role of Cognitive and Non-Cognitive Skills on Labour Market Outcomes in Indonesia. Jurnal Ekonomi Malaysia, 53(1), 3-16. http://dx.doi.org/10.17576/JEM-2019-5301-1

Adhvaryu, A., Kala, N., & Nyshadham, A. (2018). The skills to pay the bills: Returns to on-the-job soft skills training (Working paper 24313). NBER Working Paper Series. National Bureau of Economic Research. https://www.nber.org/papers/w24313

Adoho, F., Chakravarty, S., Korkoyah, D. T., Lundberg, M., & Tasneem, A. (2014). The impact of an adolescent girls employment program: The EPAG project in Liberia (Policy Research Working Paper 6832). The World Bank. https://openknowledge.worldbank.org/handle/10986/17718

Albandea, I., & Giret, J. F. (2018). The effect of soft skills on French post-secondary graduates' earnings. International Journal of Manpower, 39(6), 782-799. https://doi.org/10.1108/IJM-01-2017-0014

Almlund, M., Duckworth, A., Heckman, J., & Kautz, T. (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds), Handbook of the Economics of Education Volume 4 (pp. 1-181). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Alvares de Azevedo, T., Davis, J., & Charles, M. (2013). Testing What Works in Youth Employment: Evaluating Kenya's Ninaweza Program. Volume 1: A Summative Report. The International Youth Foundation and The World Bank. https://www.iyfnet.org/sites/default/files/library/GPYE_KenyaImpactEval_V1.pdf

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. Personality and Social Psychology Review, 11(2), 150-166. https://doi.org/10.1177%2F1088868306294907

Arntz, M., Gregory, T. & Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis (OECD Social, Employment and Migration Working Papers. No. 189). OECD Publishing. https://doi.org/10.1787/5jlz9h56dvq7-en

Autor, D. H., & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. Journal of labor Economics, 31(S1), S59-S96. https://doi.org/10.1086/669332

Balcar, J. (2016). Is it better to invest in hard or soft skills?. The Economic and Labour Relations Review, 27(4), 453-470. https://doi.org/10.1177/1035304616674613

Barrick, M.R., & M.K. Mount (1991). The Big Five Personality Dimensions And Job Performance: A Meta-Analysis. Personnel Psychology, 44(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Becker, G.S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. University of Chicago Press.

Berlin, J. A., & Golub, R. M. (2014). Meta-analysis as evidence: building a better pyramid. Jama, 312(6), 603-606. https://doi.org/10.1001/jama.2014.8167

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating skilled self-employment in developing countries: Experimental evidence from Uganda. The Quarterly Journal of Economics, 129(2), 697-752. https://doi.org/10.1093/qje/qjt057

Block, J. (2010). The Five-Factor framing of personality and beyond: Some ruminations. Psychological Inquiry, 21, 2–25. https://doi.org/10.1080/10478401003596626

Borenstein, M., Hedges, L.V., Higgins, J.PT., & Rothstein, H.R.. (2009). Introduction to Meta-Analysis. John Wiley & Sons, https://doi.org/10.1002/9780470743386

Borghans, L., Ter Weel, B., & Weinberg, B. A. (2014). People skills and the labor-market outcomes of underrepresented groups. ILR Review, 67(2), 287-334. https://doi.org/10.1177/001979391406700202

Borghans, L., Lee Duckworth, A., Heckman, J. J., & B. ter Weel (2008). The Economics and Psychology of Personality Traits. Journal of Human Resources, vol. 43 (4), 972-1059. https://doi.org/10.3368/jhr.43.4.972

Bowles, S., Gintis, H., & M. Osborne. (2001). The Determinants of Earnings: A Behavioral Approach. Journal of Economic Literature, 39 (4), 1137-176. https://doi.org/10.1257/jel.39.4.1137

Broecke, S. (2015). Experience and the returns to education and skill in OECD countries: Evidence of employer learning? OECD Journal: Economic Studies, vol. 2015/1. https://doi.org/10.1787/eco_studies-2015-5jrs3sqrvzg5.

Cairns, R., & B., Cairns (1994). Lifelines and risks: Pathways of youth in our time. Cambridge University Press.

Calero, C., & Rozo, S. V. (2016). The effects of youth training on risk behavior: the role of non-cognitive skills. IZA Journal of Labor & Development, 5(1), 12. https://doi.org/10.1186/s40175-016-0058-6

Calero, C., Diez, V. G., Soares, Y. S., Kluve, J., & Corseuil, C. H. (2017). Can arts-based interventions enhance labor market outcomes among youth? Evidence from a randomized trial in Rio de Janeiro. Labour Economics, 45, 131-142. https://doi.org/10.1016/j.labeco.2016.11.008

Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D., & Soares, Y. (2011). The labor market impacts of youth training in the Dominican Republic. Journal of Labor Economics, 29(2), 267-300. https://doi.org/10.1086/658090

CEDEFOP. (2019). The Skills Employers Want! (Briefing Note) https://www.cedefop.europa.eu/files/9137_en.pdf

Cho, Y., Kalomba, D., Mobarak, A. M., & Orozco, V. (2013). Gender differences in the effects of vocational training: Constraints on women and drop-out behavior (Policy Research Working Paper 6545). The World Bank. http://documents.worldbank.org/curated/en/882971468272376091/Gender-differences-in-the-effects-of-vocational-training-constraints-on-women-and-drop-out-behavior

Chowdhury, S., Ham, J. C., Dhabi, N. A., & Schildberg-Hörisch, H. (2017). Self-Employment, Family Production and the Returns To Cognitive and Non-Cognitive Skills in Rural Bangladesh, conference paper, http://conference.iza.org/conference_files/VWPreferences_2017/6591.pdf.

Cinque, M., Carretero, S., & Napierala, J. (forthcoming). Non-cognitive skills and other related concepts: towards a better understanding of similarities and differences. JRC Working Papers on Labour, Education and Technology.

Collischon, M. (2020). The Returns to Personality Traits Across the Wage Distribution. Labour, 34 (1), 48–79. https://doi.org/10.1111/labr.12165

Cook, K.W., Vance, C.A., & Spector, P.E. (2000).The Relation of Candidate Personality With Selection-Interview Outcomes. Journal of Applied Social Psychology, 30, 867-885. https://doi.org/10.1111/j.1559-1816.2000.tb02828.x

Cunha, F., Heckman, J.J., & Schennach, S.M. (2010), Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica, 78: 883-931. https://doi.org/10.3982/ECTA6551

Cunningham, W., Torrado, M. P., & Sarzosa, M. (2016). Cognitive and non-cognitive skills for the Peruvian labor market: Addressing measurement error through latent skills estimations (Policy Research Working Paper 7550). The World Bank. https://openknowledge.worldbank.org/handle/10986/23725

De Coulon, A., Vahé, N., & Speckesser, S. (2020). The long-term impact of improving non-cognitive skills of adolescents: Evidence from an English remediation programme (Research Discussion Paper 028). Centre for Vocational Educational Research, London School of Economics & Political Science. https://cver.lse.ac.uk/textonly/cver/pubs/cverdp028.pdf

Deming, D. J. (2017). The growing importance of social skills in the labor market. The Quarterly Journal of Economics, 132(4), 1593-1640. https://doi.org/10.1093/qje/qjx022

Díaz, J. J., Arias, O., & Tudela, D. V. (2013). Does perseverance pay as much as being smart? The returns to cognitive and non-cognitive skills in urban Peru. Unpublished paper, http://conference.iza.org/conference_files/worldb2014/arias_o4854.pdf.

Duckworth, A.L., & Yeager, D.S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. Educational Researcher, 44(4):237-251. https://doi.org/10.1177/1088868306294907

Durlak, J. A., Weissberg, R.P., Dymnicki, A. B., Schellinger, K. B., & Taylor, R,.D. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. Child Development, 82 (1), Pages 405–432.

https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eurofound (2020). Game-changing technologies: Transforming production and employment in Europe. Publications Office of the European Union. https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef19047en.pdf

Farrington, C.A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T.S., Johnson, D.W., & Beechum, N.O. (2012). Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review. University of Chicago Consortium on Chicago School Research. https://consortium.uchicago.edu/sites/default/files/2018-10/Noncognitive%20Report_0.pdf

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. The Journal of Abnormal and Social Psychology, 44(3), 329–344. https://doi.org/10.1037/h0057198

Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. Journal of Economic Behavior & Organization, 89, 122-135. https://doi.org/10.1016/j.jebo.2013.02.004

Frey, C.B., & Osborne, M.A. (2013). The Future of Employment: How Susceptible are Jobs to Computerization? Oxford Martin Programme on Technology and Employment. https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf

Funder, D. C. (2001). Personality. Annual Review of Psychology, 52, 197-221. https://doi.org/10.1146/annurev.psych.52.1.197

Furukawa T.A., McGuire H., & Barbui C. Meta-analysis of effects and side effects of low dosage tricyclic antidepressants in depression: systematic review. BMJ, 2, 325(7371):991. https//doi.org/10.1136/bmj.325.7371.991

Gardner, D. G., & Cummings, L. L. (1988). Activation theory and task design: Review and reconceptualization. In B. M. Staw, & L. L. Cummings (Eds.), Research in Organizational Behavior (Vol. 10). Elsevier

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them? Applied Measurement in Education, 29(4), 245-249. https://doi.org/10.1080/08957347.2016.1209207

Gelissen, J., & Graaf, P. M. D. (2006). Personality, social background, and occupational career success. Social Science Research, 35(3), 702. https://doi.org/10.1016/j.ssresearch.2005.06.005

Gensowski, M. (2014). Personality, IQ, and lifetime earnings (IZA Discussion Papers No. 8235). Institute for the Study of Labor (IZA). http://ftp.iza.org/dp8235.pdf

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeerch, C., Walker, S., ... & Grantham-McGregor, S. (2013). Labor market returns to early childhood stimulation: A 20-year followup to an experimental intervention in Jamaica (Research Policy Paper No. 6529). The World Bank. http://documents1.worldbank.org/curated/en/388661468040444014/pdf/WPS6529-REVISED.pdf

Girtz, R. (2012). The effects of personality traits on wages: a matching approach. Labour, 26(4), 455-471. https://doi.org/10.1111/j.1467-9914.2012.00556.x

Goldberg, L.R., Sweeney, D., Merenda, P. F., & Hughes, J. E. (1998). Demographic variables and personality: the effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. Personality and Individual Differences, 24 (3), p. 393-403. https://doi.org/10.1016/S0191-8869(97)00110-4.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. Journal of Personality and Social Psychology, 59(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Gonzalez Vazquez, I., Milasi, S., Carretero Gomez, S., Napierala, J., Robledo Bottcher, N., Jonkers, K., Goenaga, X. (eds.), Arregui Pabollet, E., Bacigalupo, M., Biagi, F., Cabrera Giraldez, M., Caena, F., Castano Munoz, J., Centeno Mediavilla, C., Edwards, J., Fernandez Macias, E., Gomez Gutierrez, E., Gomez Herrera, E., Inamorato Dos Santos, A., Kampylis, P., Klenert, D., López Cobo, M., Marschinski, R., Pesole, A., Punie, Y., Tolan, S., Torrejon Perez, S., Urzi Brancati, C., Vuorikari, R. (2019). The changing nature of work and skills in the digital age. Publications Office of the European Union. https://doi.org/10.2760/679150

Groh, M., Krishnan, N., McKenzie, D., & Vishwanath, T. (2016). The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan. IZA Journal of Labor & Development, 5(1), 9. https://doi.org/10.1186/s40175-016-0055-9Hampf, F., Wiederhold, S., & Woessmann, L. (2017). Skills, earnings, and employment: exploring causality in the estimation of returns to skills. Large-scale Assessments in Education, 5, 1-30. https://doi.org/10.1186/ s40536-017-0045-7

Hanushek, E., Schwerdt, G., Wiederhold, S., & Woessman, L. (2015). Returns to Skills around the World: Evidence from PIAAC. European Economic Review, 73/C, pp. 103-130. https://doi.org/10.1016/j.euroecorev.2014.10.006

Hanushek, E.A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. Journal of Economic Literature, 46 (3): 607-68. https://doi.org/10.1257/jel.46.3.607

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. Labour economics, 19(4), 451-464. https://doi.org/10.1016/j.labeco.2012.05.014

Heckman, J. J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition (Working Paper 19656). NBER Working Papers Series. National Bureau of Economic Research. http://www.nber.org/papers/w19656

Heineck, G. (2011). Does it pay to be nice? Personality and earnings in the United Kingdom. ILR Review, 64(5), 1020-1038. https://doi.org/10.1177/001979391106400509

Heineck, G., & Anger, S. (2010). The returns to cognitive abilities and personality traits in Germany. Labour economics, 17(3), 535-546. https://doi.org/10.1016/j.labeco.2009.06.001

Heckman, J.J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. Journal of Labor Economics, 24 (3). https://doi.org10.3386/w12006

Hilger, A., Nordman, C. J., & Sarr, L. R. (2015). Non-cognitive skills, social networks and labor market outcomes in Bangladesh. Unpublished manuscript.

Hogan, R., & Hogan, J. (2007). Hogan Personality Inventory. Manual. Hogan Assessment Systems, Inc. http://www.hoganassessments.com/sites/default/files/HPI%20Tech%20Manual%20-%20S.pdf

Ibarraran, P., Ripani, L., Taboada, B., Villa, J. M., & Garcia, B. (2014). Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. IZA Journal of Labor & Development, 3(1), 10. https://doi.org/10.1186/2193-9020-3-10

Jaimovich, Ni.and Siu, H., (2018). The "End of Men" and Rise of Women in the High-Skilled Labor Market (CEPR Discussion Paper No. DP13323). https://ssrn.com/abstract=3287070

Kankaras, M. (2017).Personality matters: Relevance and assessment of personality characteristics (OECD Education Working Papers 157). OECD Publishing. https://doi.org/10.1787/8a294376-en

Kautz, T., Heckman, J.J., Diris, R., ter Weel, B., & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success (OECD Education Working Papers No. 110). OECD Publishing. https://doi.org/10.1787/5jxsr7vr78f7-en

Kluve, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., & Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. World Development, 114, 237-253. https://doi.org/10.1016/j.worlddev.2018.10.004

Kluve, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., Weidenkaff, F. & Witte, M. (2016). Do Youth Employment Programs Improve Labor Market Outcomes? A Systematic Review (IZA Discussion Paper No. 10263). IZA Discussion paper series. http://ftp.iza.org/dp10263.pdf

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. American Economic Journal: Applied Economics, 3(1), 101-28. https://doi.org/ 10.1257/app.3.1.101

Lordan, G. (2018). Robots at work: A report on automatable and non-automatable employment shares in Europe. Publications Office of the European Union.

Messick, S. (1978). Potential Uses of Noncognitive Measurement in Education. ETS Research Bulletin, 1, i–25. https://doi.org/10.1002/j.2333-8504.1978.tb01156.x

Montenegro, C. E., & Patrinos, H. A. (2014). Comparable estimates of returns to schooling around the world (Policy Research Working Paper 7020). The World Bank. https://openknowledge.worldbank.org/handle/10986/20340

Mueller, G., & Plug, E. (2006). Estimating the Effect of Personality on Male and Female Earnings. ILR Review, 60(1), 3–22. https://doi.org/10.1177/001979390606000101

Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training (OECD Social, Employment and Migration Working Papers No. 202). OECD Publishing. https://doi.org/10.1787/2e2f4eea-en

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. The Journal of Abnormal and Social Psychology, 66(6), 574–583. https://doi.org/10.1037/h0040291

Nyhus, E. K., & Pons, E. (2005). The effects of personality on earnings. Journal of Economic Psychology, 26 (3), 2005, 363-384. https://doi.org/10.1016/j.joep.2004.07.001.

OECD (2019). OECD Employment Outlook 2019: The Future of Work. OECD Publishing. https://doi.org/10.1787/9ee00155-en.

O'Connell, M., & Sheikh, H. (2011). 'Big Five' personality dimensions and social attainment: Evidence from beyond the campus. Personality and Individual Differences, 50 (6), 828-833. https://doi.org/10.1016/j.paid.2011.01.004

Paul, M., & Leibovici, L. (2014). Systematic review or meta-analysis? Their place in the evidence hierarchy. Clinical Microbiology and Infection, 20(2), 97-100. https://doi.org/10.1111/1469-0691.12489

Patrinos, H. A. (2016). Estimating the return to schooling using the Mincer equation. IZA World of Labor, 278. https://doi.org/ 10.15185/izawol.278

Poole, C., & Greenland, S. (1999). Random-effects meta-analyses are not always conservative. American Journal of Epidemiology,1;150(5), 469-75. https//doi.org/ 10.1093/oxfordjournals.aje.a010035

Premand, P., Brodmann, S., Almeida, R., Grun, R., & Barouni, M. (2016). Entrepreneurship education and entry into self-employment among university graduates. World Development, 77, 311-327. https://doi.org/10.1016/j.worlddev.2015.08.028

Sánchez Puerta, M., Valerio, A., & Bernal, M.G. (2016). Taking Stock of Programs to Develop Socioemotional Skills: A Systematic Review of Program Evidence (English). The World Bank Group. http://documents.worldbank.org/curated/en/249661470373828160/Taking-stock-of-programs-to-develop-socioemotional-skills-a-systematic-review-of-program-evidence

Rammstedt, B., Danner, D,. & Lechner, C. Personality, competencies, and life outcomes: results from the German PIAAC longitudinal study. Large-scale Assessments in Education, 5, 2. https://doi.org/10.1186/s40536-017-0035-9

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and

Cognitive Ability for Predicting Important Life Outcomes. Perspectives on psychological science : a journal of the Association for Psychological Science, 2(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. American Economic Journal: Applied Economics, 4(4), 121-39. https://doi.org.10.2307/23269744

Schwarzer, G., Carpenter, J.R., & Rücker, G. (2015). Meta-Analysis with R, Springer. https://10.1007/978-3-319-21416-0

Sidik, K., & Jonkman, J.N. (2005). A note on variance estimation in random effects meta-regression. Journal of Biopharmaceutical Statistics, 15, 823-838. https://doi.org/10.1081/BIP-200067915

Smith, G.M. (1967). Usefulness of Peer Ratings of Personality in Educational Research. Educational and Psychological Measurement, 27(4), 967-984. https://doi.org.10.1177/001316446702700445

Spielman, R.M., Jenkins, W.J., & Lovett, M. D (2014). Psychology 2e. OpenStax. https://assets.openstax.org/oscms-prodcms/media/documents/Psychology2e-WEB_0eRvAre.pdf

Stanley, T.D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5, 60–78. https://doi.org/10.1002/jrsm.1095

Steichen, T.J. (2001). Nonparametric trim and fill analysis of publication bias in meta-analysis: erratum. Stata Technical Bulletin, 10, Article STB58. https://www.stata-press.com/journals/stbcontents/stb58.pdf

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. Journal of clinical epidemiology, 54(10), 1046-1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. Personnel Psychology, 44(4), 703-742. https://doi.org/10.1111/j.1744-6570.1991.tb00696.x

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. Human Performance, 12(1), 1-29. https://doi.org/10.1207/s15327043hup1201_1

Eijck, C.J.M. van; & Graaf, P.M. de (2004). The Big Five at school: The Impact of Personality on Educational Attainment. The Netherlands' Journal of Social Science, 40 (1), 24-42.

https://research.tilburguniversity.edu/en/publications/the-big-five-at-school-the-impact-of-personality-on-educational-a

Viechtbauer W. (2005)- Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics, 30(3). 261-293. https://doi.org/10.3102/10769986030003261

Vooren, M., Haelermans, C., Groot, W., & Maassen van den Brink, H. (2019). The Effectiveness of Active Labor Market Policies: A Meta-Analysis. Journal of Economic Surveys, 33 (1), 125–149. https://doi.org/10.1111/joes.12269

Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. Review of Economics and Statistics, 96(4), 849-861. https://doi.org/10.1162/REST_a_00449

World Economic Forum (2015). New Vision for Education: Unlocking the Potential of Technology. World Economic Forum. http://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf

Annex A: Descriptive Statistics

We have pooled the two databases together in order to give an overall overview of the number of estimates dealing with the relationship between non-cognitive skills and earnings. The table with descriptive statistics is provided in Annex B. We briefly discuss this table as follows.

First, we present the estimates by geographical coverage. Studies dealing with the U.S. and Germany deliver many estimates (29.8% and 12.6%, respectively). The reason for this observation is that these studies (for Germany: Heineck & Anger, 2010, 2011; for the U.S.: Fletcher, 2013; Gensowski, 2014; and Girtz, 2012) discuss the relationship between the Big Five personality traits and earnings, and, doing so, the authors present per component of the Big Five an estimate. Furthermore, except for Fletcher (2013), these authors also present an estimate separately for the male only and female only population.

In fact, data on the population under consideration have been gathered for all studies included in both databases. The majority (42.3%) of estimates come from a male and female (mixed) population; and 26.4% (31.1%) from a female (male) only population. This allows us to present results separately by male, female and mixed populations (see Section 5 and Section 6 for further details).

Whereas there is always a time lag between writing the publication and actually publishing it in a journal, we observe more articles with a year of publication closer to 2010 than to 2020 (further discussion on this is provided in Section **Error! Reference source not found.**). As from 2013, however, the number of articles published on the relationship between non-cognitive skills and earnings fluctuates around 9 percent. This could reflect a decreased attention for the topic over the past few years (which appears to hold for OECD countries (Kluve et al., 2019)). It could also reflect difficulties with programme implementation and/or evaluation in schools or colleges; or simply be the result of financial difficulties to support research in education in the aftermath of the financial crisis.

Further, we present the estimates by methodology used by the authors. At the level of the study, 13 studies are randomised controlled trials; 9 studies use quasi-experimental regression techniques (and 6 of these 9 combine correlational estimates with quasi-experimental techniques); and 7 studies deliver correlational estimates. A total of 13 studies produce 48.6 percent of all correlational estimates. These studies are solely collected in database I and often deal with the estimation of Mincer equations (or equivalents) that correlate the personality traits to earnings and controlling for factors like age, experience (four studies: Adhitya, 2019; Heineck & Anger, 2010; Hilger et al., 2018; and Lindqvist & Vestman, 2011), and educational attainment. Authors in database I often present estimates controlled for educational attainment. In many cases, they also present the returns to non-cognitive skills *without* controlling for educational attainment. This feature can be used in Section 5 as a way to present estimates on the returns to personality traits *not* controlled and controlled for educational attainment.

With regard to the studies included in database II, all estimates are controlled for educational attainment as a consequence of the randomized controlled trials. Nonetheless, authors still present relevant differences in estimates conditional on employment (or only looking at earnings of those who are employed on the labour market), and estimates unconditional on employment (or looking at the full population whether or not they are employed on the labour market). Section **Error! Reference source not found.** accounts for this.

Finally, we collected information on the way the authors dealt with linearity of the relationship between earnings and non-cognitive skills. In fact, all authors but one (Gensowski, 2014) included in database I transformed the earnings variable with its logarithm as to approximate for a linear specification. On the contrary, as much as 7 authors from database II chose for the earnings variable as a monetary outcome (instead of its logarithm).

The transformation of the outcome variable earnings with its logarithm, in combination with the standardised values of the non-cognitive skills in the regression, yields the advantage that we can easily compare estimates from database I with each other. There is indeed a considerably low between-study heterogeneity in the way authors have estimated the returns to personality traits in database I. We exploit this advantage in Section 5.

With regard to the database II, we follow the traditional literature on how to compute effect sizes from estimates, standard errors, sample size and/or standard deviations collected from the articles. These effect sizes can then be compared across studies using random effects meta-regression techniques.

Table A1: Estimates collected in Database I and Database II together by study characteristics

	Frequency	**Percentage**	**Cumulative**
(a) by geographical coverage			
Bangladesh	42	12.61	12.61
Brazil	8	2.4	15.02
Colombia	28	8.41	23.42
Czech Republic	2	0.6	24.02
Dominican Republic	9	2.7	26.73
England	2	0.6	27.33
France	3	0.9	28.23
Germany	42	12.61	40.84
India	1	0.3	41.14
Indonesia	30	9.01	50.15
Jamaica	6	1.8	51.95
Jordan	6	1.8	53.75
Kenya	2	0.6	54.35
Liberia	9	2.7	57.06
Malawi	2	0.6	57.66
Peru	33	9.91	67.57
Sweden	6	1.8	69.37
Tunesia	2	0.6	69.97
U.S.	66	19.82	89.79
Uganda	4	1.2	90.99
United Kingdom	30	9.01	100

Table A1 (Continued)

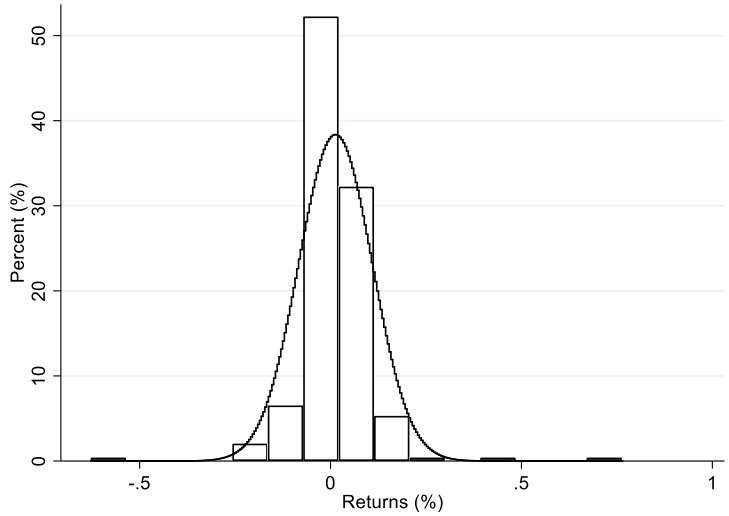
	Frequency	Percentage	Cumulative
(b) by year of publication			
2010	44	13.21	13.21
2011	39	11.71	24.92
2012	40	12.01	36.94
2013	30	9.01	45.95
2014	39	11.71	57.66
2018	28	8.41	66.07
2016	21	6.31	72.37
2017	30	9.01	81.38
2018	32	9.61	90.99
2019	30	9.01	100

	Frequency	Percentage	Cumulative
(c) by method applied in the study			
RCT	57	17.12	17.12
(Quasi)-Experimental	114	34.23	51.35
Correlational	162	48.65	100
(d) by gender			
female only	88	26.43	26.43
male only	104	31.23	100
male and female	141	42.34	68.77
(e) by specification of the outcome variable			
Monetary value	63	18.92	18.92
Logarithm	270	81.08	100

Notes. Based on 333 estimates from 29 studies and 2 databases.

The analysis of the distribution of the returns to non-cognitive skills plotted on Figure A1 show that overall, we observe that many estimates of returns to non-cognitive skills are either equal to zero (over 50%) or slightly positive (almost 40%), while only few appear to be negative (less than 10%).

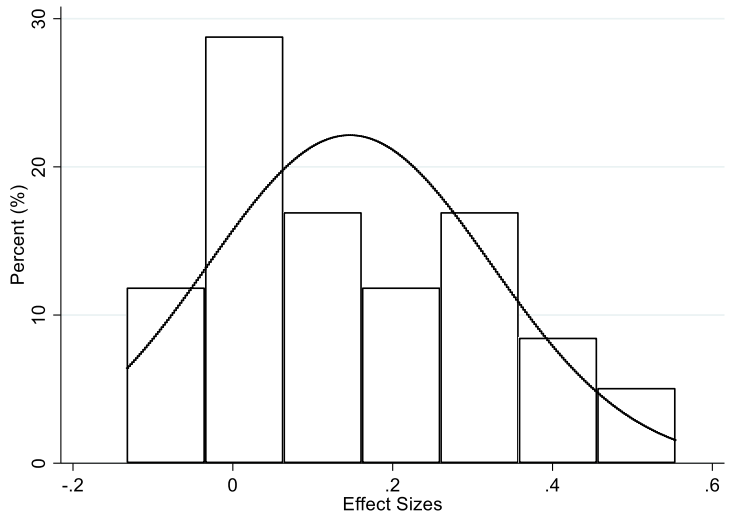
Figure A1: Distribution of the returns to non-cognitive skills (%) in Database I



Note: Based on 245 estimates for non-cognitive skills. We use a histogram with a plot for the normal distribution. On the X-axis we plot the estimated returns. On the Y-axis, we present the share of estimates from the studies (%) that fall into a range of estimates, as denoted on the X-axis.

In Figure A2, we present the distribution of the effect sizes observed in our studies. Based on this histogram, we expect an average effect size across all studies approximately equal to 0.15 SD.

Figure A2: Distribution of the effect sizes in Database II



Note: Based on 59 effect sizes from database II. The Y-axis denotes the percent share of studies with an effect size estimated in the corresponding bins on the X-axis. We added a plot of the normal distribution to the histogram

Annex B: Journal and Impact

Table B1: Journal, impact factor and H-index by study included in the meta-analysis

First author	Year	Journal	Impact Factor	H-index
Acosta	2015	IZA Discussion Papers	0.97	81.00
Adhitya	2019	Jurnal Ekonomi Malaysia	0.71	5.00
Adhvaryu	2018	NBER WORKING PAPER SERIES	1.58	243.00
Adoho	2014	Policy Research Working Paper World Bank	0.74	77.00
Albandea	2018	International Journal of Manpower	1.00	49.00
Azevedo	2013	International Youth Foundation Report	N/A	N/A
Balcar	2016	The Economics and Labour Relations Review	2.26	15.00
Blattman	2014	The Quarterly Journal of Economics	11.78	228.00
Calero	2016	Labour Economics	1.47	6.00
Calero	2017	Labour Economics	1.25	78.00
Card	2011	Journal of Labor Economics	3.61	94.00
Cho	2013	IZA Discussion Papers	0.97	81.00
Chowdhury	2017	IZA conference paper	N/A	N/A
Cunningham	2016	Policy Research Working Paper World Bank	0.74	77.00
De Coulon	2010	Working Paper	N/A	N/A
Deming	2017	The Quarterly Journal of Economics	11.78	228.00
Diaz	2013	IZA Discussion Papers	0.97	81.00
Fletcher	2013	Journal of Economic Behavior & Organization	1.59	101.00
Gensowski	2014	IZA Discussion Papers	0.97	81.00
Gertler	2013	NBER WORKING PAPER SERIES	1.58	243.00
Girtz	2012	LABOUR	N/A	N/A
Groh	2016	IZA Journal of Labour & Development	1.47	6.00
Heineck	2010	Labour Economics	1.25	78.00
Heineck	2011	Industrial and Labor Relations Review	2.64	69.00
Hilger	2018	IZA Discussion Papers	0.97	81.00
Ibarraran	2014	IZA Journal of Labour & Development	1.47	6.00
Lindqvist	2011	American Economic Journal: Applied Economics	4.04	74.00
Premand	2016	World Development	3.87	164.00
Rodriguez-Planas	2012	American Economic Journal: Applied Economics	4.04	74.00

Note: Latest information available in November 2020 used.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:
- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:
- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub