Fynch MEYNENT

ENSAE 2nd year
Internship Application
Year 2022-2023

Analyze of the Intergenerational Transmission of Skills

Trinity College Dublin, Department of Sociology
Dublin, IRELAND

Course Supervisor: Jan SKOPEK
June 5th 2023 – August 11th 2023

Contents

1 Introduction **4**

2 Literature Review **5**

2.1 Introduction and Thinking 5

2.2 What is a skill ? 5

2.2.1 Definition 5

2.2.2 Various Typologies 5

2.3 How to measure skills ? 6

2.3.1 Direct measurement 6

2.3.2 Indirect measurement 7

2.4 Which skills are transmitted and how? 8

2.4.1 The types of skills which are transmitted 8

2.4.2 The effects of the educational skills that select occupation 8

2.4.3 The effect of the second sociability by education 8

2.4.4 The role of gender and other demographic variables 8

2.4.5 A little reflection about income 9

3 Construction of a theoretical framework **9**

3.1 The role of parenting style 9

3.2 The role of financial resources 9

3.3 The importance of the cultural capital 10

3.3.1 The embodied cultural capital 10

3.3.2 The objectified cultural capital 10

3.3.3 The institutionalized cultural capital 11

3.4 The secondary socialization of parents at work and parenting style 11

4 Construction and Analysis of the data **11**

4.1 Presentation of the data 11

4.1.1 The O*NET Database 11

4.1.2 The National Educational Panel Study 12

4.2 Construction of the database 12

4.2.1 Adaptation of the O*NET Database 12

4.2.2 Merging with NEPS 13

4.3 Some descriptive statistics 14

4.3.1 Outcome and explanatory variables 14

4.3.2 Missing values and control variables 17

5 Modelization **19**

5.1 Vocabulary outcome 19

5.2 Mathematical outcome 24

6 Analysis of the results **28**

7 Conclusion **28**

A PFA on the occupational skills **32**

B Code of the construction of the database **33**

Acknowledgements

I would like to thank Jan SKOPEK for his welcome and his pedagogy. It was a pleasure to work with him. I also want to thank Chris LYSAGHT, the administrator of the Trinity Research in Social Sciences (TRiSS) who was helpful with logistics issues. Finally, I thank Nour BOULAHCEN who accommodated me for two weeks in Dublin because I didn't have an apartment.

1 Introduction

This internship was realized in the Department of Sociology of Trinity College Dublin. I was invited by Jan SKOPEK, the head of the Sociology Department. Therefore, it was a research internship in an academic context.

One of sociology's main problems is studying inequalities and their intergenerational transmission. Indeed, sociologists established a strong correlation between parents' and children's financial and cultural capital outcomes. Some theories try to explain this: the mechanisms of social reproduction by schooling or the transmission of economic resources. However, not many theories try to approach this phenomenon by analyzing the intergenerational transmission of skills.

With a literature review, I thought about what is a skill, how I can measure it, and if a correlation exists. Then, I saw that the correlation between the skills of parents and their children exists. Therefore, I tried to see what is the "black box" that explains this correlation. It was the problem of this work: trying to understand what does happen in this black box. Thus, I produced a conceptual framework to explain where I talked about two theories. One was about the selection theory, which means that the transmission of skills is linked to the social selection of parents that brings them in their social position and therefore, helps them to transmit their skills and reproduce this selection and then the hierarchical hierarchy. Another was about secondary socialization with the idea that in a profession, a parent can attain some attitudes, values, and skills that they can teach their children. Hence, I tried to establish if a theory was dominating another, precision for which skills and which kind of children outcome it can help.

To answer this problem, I merged an O*NET Skills database with an extract of the National Educational Panel Study (NEPS). In the O*NET Database, for each skill, the level of requirements and the importance of the skill are evaluated in two different situations. The idea was to estimate the skills of parents by their occupational skills and compare them to children's outcomes in cognitive skills. Therefore, I will model it by linear regressions to see, firstly, the correlation between specific parents' skills and children's outcomes. Then, I will compare the effects of the level's skills and the importance's skills to see what's dominant. Then, I will analyze the dominant framework and suggest some way to pursue the research process.

2 Literature Review

A large part of my work was to analyze the literature review to think about the methodology I will use and the global context of the problem.

2.1 Introduction and Thinking

In the theory of social mobility and education, I establish that there's a correlation between the situation social of a parent and their children. To analyze this correlation, they used the transmission of economic and cultural capital. It means that I analyze how SES and education influence the social situation of their children. However, the literature didn't explore the transmission of skills.

2.2 What is a skill ?

2.2.1 Definition

In VandenBos 2007, a skill is defined as "an ability or proficiency acquired through training and practice.". It means that it's something that a person can do which is not innate but attained. It's not about specific fields. However, these definitions are very general. It can be adapted to explore some typologies.

2.2.2 Various Typologies

Clearly, talking just about skills, in general, can be reductive. I will try here to establish typologies of various types of skills. All the articles I read don't use the same typology, but I can see some common features.

Cognitive Skills

Firstly, the searchers often talk about cognitive skills. Usually, they refer to verbal, quantitative, or mathematics skills and reasoning skills.

Liu Y. 2013, Barg K. 2023 and Lundborg P. 2018 mention verbal skills. Coneus K. 2009 also uses this term but not in a general categorization such as "cognitive skills". It corresponds to all the skills linked to expression and knowing vocabulary. Another way to allude to this sub-category is to talk about language skills. That's the words chosen in A. Jacobs B. 2021, v. d. V. R. Jacobs B. 2021, and Hanushek E. 2018. About the characteristics of the mother, Barg K. 2023 make a difference between literacy (reading, writing) and verbal (teaching, persuading).

Another aspect is quantitative skills. They are mentioned in Liu Y. 2013 and are more or less equivalent to math skills in Lundborg P. 2018,A. Jacobs B. 2021,v. d. V. R. Jacobs B. 2021, and Hanushek E. 2018. Moreover, it can have a sort of proximity with what Barg K. 2023 call spatial awareness if math skill measures include geometrical tests.

Finally, another aspect is the ability to reason. Lundborg P. 2018 uses this term. On the other hand, Barg K. 2023 talks about the ability to link concepts and Liu Y. 2013 the analytic skills.

Lundborg P. 2018 made just an opposition between cognitive skills and non-cognitive skills. On the other hand, other articles are more precise about the other type of skills.

Social Skills

A lot of articles mention social skills. Liu Y. 2013 considers that they are the skills relative to management skills and nurturing ones. In Okumura T. 2014, it is linked to all the skills in communication, interpersonal interactions, and leadership. Eventually, Coneus K. 2009 refers also to the term of "social skill". However, if the term is similar, it doesn't always refer to the same thing because in Liu Y. 2013, they talk about the labor market so it's professional-related whereas in Coneus K. 2009 they talk about the skills of 3-year-old children.

Other type of skills

However, our articles are from various academic fields: Economics (for example, Lundborg P. 2018, Coneus K. 2009), Sociology (A. Jacobs B. 2021, and Liu Y. 2013) or psychology (Duppong Hurley K. 2014)

Therefore, they can have various types of skills they can use in a precise context.

Liu Y. 2013 tries to understand the mechanisms of the labor market. Therefore, they talked in their typology about creative skills (architecture, design, art...) and technical skills (for example, computing skills). It's because these two types of skills aren't rewarded in the same way as other cognitive skills. So their typology was a bit different.

In Coneus K. 2009, the author refers also to motor skills. It can be explained by the context that motor function is a skill perceived as an important sign of development during childhood. In Hershbein B. 2018, the economists talk about "physical capital" that can be considered as a skill because it's capital that you can apply to job tasks. In Barg K. 2023, they also talk about the physical skill of the mother.

Last, because I focus on transmission, the concept of parenting skills can be interesting because it can influence the way the skill is transmitted. In Duppong Hurley K. 2014, they define parenting skills or practice as behaviors, beliefs, coping mechanisms, and reactions to stress or discipline. They didn't make a clear distinction between skills and practice, but it can be a first idea of an approach I can make to analyze the intergenerational transmission of skills.

2.3 How to measure skills ?

2.3.1 Direct measurement

Declaration

To measure the social skills of children, Okumura T. 2014 use the data collected in the NLSY79 and ask people who are between 20 and 28 years old how they were sociable at age 6 and early adulthood. However, self-declaration of skills has a lot of limits. For example, Devaux 2014 shows that respondents often underestimate their consumption of alcohol in surveys. Indeed, it would be devaluing to accept a substantial consumption in front of pollsters. In the case of Okumura T. 2014, it's possible to assume that people over-estimate their social abilities in order to put themselves forward.

For example, considering these biases, measuring by psychometric methods can be a solution.

Psychometric methods

A first way to measure the skills can be IQ. In Anger S. 2009, they used cognitive data from the German Socio-Economic Panel Study measured by sub-modules of the WAIS test, which is the standard test to measure IQ of adults. They measured word fluency or speed at solving tasks. They controlled by the education of the parents, family characteristics, type of geographic area of growth, and size as a health indicator. On these subsets, Sternberg 2006 the WAIS satisfies some strong reliability and validity properties. However, it measures just some type of abilities (more memory-analytical abilities and less synthetic-creative or practical-contextual abilities).

Another psychometric way is used to measure children's cognitive skills in Barg K. 2023. They tested their cognitive skills with some modules of the BAS II test. They chose to measure the ability to verbal actions, inductive reasoning, and spatial awareness. However, they didn't proceed with a psychometric test to determine parents' skills.

Before, I looked for the general skills of children and parents. Nevertheless, I can consider that parenting skills can have an influence on the transmission of the skills mentioned above. Therefore, I have to find a way to measure it. In Duppong Hurley K. 2014, they provided a review of various measures of parental behaviors, beliefs, copying mechanism, reaction to stress or discipline. The idea is to determine which measure are adapted to estimate parenting skills considering their properties of reliability, validity and factor structure for those which were concerned. Only five measures satisfied all criteria, and in these measures, only two can be related to parenting skills: Alabama Parenting Measure which measured: positive involvement with children, supervision and monitoring, use of positive discipline techniques, consistency in the use of such discipline and use of corporal punishment.(Frick 1999) and Parent Child Relationship Inventory. This measure yields scores on 7 content scales: parental support, satisfaction with parenting, involvement, communication, limit setting, autonomy and role orientation. (Gerard 1994) Other concerned alliance between parents or child abuse but it was not clearly more about skills.

To measure skills of the children, Lundborg P. 2018 use the data collected by the Swedish army when 18 year's old people do their compulsory military service. For the cognitive skills, they used the Enlistment Battery 80, which contains four separated tests: Instructions, Synonyms, Metal Folding and Technical Comprehension. The idea is to measure mathematical, reasoning skills and verbal abilities. Then, the scores are aggregated in a standard composite measure calculated by the military enlistment service. To estimate non-cognitive skills, they used interviews carried out by certified psychologists employed by the Swedish Army. They estimate their psychological abilities and endurance, capability of taking initiatives, responsibility and social skills. A composite score is estimated. Therefore, they estimate outcomes for children based on these skills and their health.

Therefore, I can see that some methods are based on "school" tests. It's also the case for A. Jacobs B. 2021. Indeed, they decided to compare the results of parents and children to the CITO test, an assessment gave to dutch pupils at the end of their primary school. These tests measure ability in math and language. Therefore, to measure parents' skills and children's ones, they used the results to these tests. In Hanushek E. 2018, they produce the same strategy of similar tests about math and languages.

2.3.2 Indirect measurement

In the previous articles, I realized that some of them don't use cognitive skills about parents. I assume that it's not always easy in a research protocol to obtain these data on large scales. Therefore, some searchers try to proxy parental skills.

<u>The education</u>

First, the educational level is used as a measure of skills.

Coneus K. 2009 want to analyze the intergenerational transmission of human capital that they assimilate to skill. To make a proxy of the human capital of the mother, they consider that education is a part of it with innate ability, educational values, and other favorable environmental characteristics.

However, it's not necessarily a measure of skill, but it can be used as a control variable. Barg K. 2023 consider mother's highest level of education as a part of family resources. The idea is to use the family resources as a control variable.

<u>The occupation</u>

However, it doesn't offer a large dataset about the skills of parents. Therefore, I can try to estimate it by using the occupation. It can be a good proxy. Indeed, Autor D. 2013 suggest a Roy model to give a theoretical framework to the following empirical observation: job tasks are a good predictor of a wage and workers characteristics are significantly related to these tasks. Therefore, with this model, they propose that workers self-select themselves the job that offers the highest rewind for the bundle of tasks they can do. The tasks performed are linked to their skills, so I can conclude that occupational skills should be a good proxy.

To measure skills, Barg K. 2023 used the UK Millennium Cohort Study (MCS) and merged it with British Skills Survey (BSS) so as to generate the mother's occupation-specific skill variables. In the BSS, the respondents evaluated the importance of being able to resolve some tasks in their job. Therefore, they conducted a Principal Component Analysis to estimate eight indices. They calculated the mean of each index for respondents in 79 SOC-occupation groups. They also merged with mother characteristics available on MCS. Then, they conserved five indices (social skills were not correlated enough, and teamwork and planning were strongly correlated with other factors). They kept: literacy, numeracy, problem-solving, verbal and physical. So, BSS allows to proxy the mother's skills. The same type of strategy is used in Liu Y. 2013. In order to evaluate cognitive, creative, technical, and social skills, they used the data at level occupation from the O*NET database which specifies skill requirements in the job...

Similarly, in order to make a proxy of the parents social skills, Okumura T. 2014 used the DOT (Dictionary of Occupational Titles) to have some elements about social skills: talking, adaptability with other people, and a preference for a relational job.

Then, I will analyze how and which skills are transmitted by parents to children.

2.4 Which skills are transmitted and how?

2.4.1 The types of skills which are transmitted

I will check if the assumption that parents transmit their skills to children is acceptable. It can be coherent with the fact that Hanushek E. 2018 children with parents having higher maths skills are more likely to choose STEM during high school or a field in tertiary education.

In Barg K. 2023, they established a positive relationship between the verbal abilities of their mother and the verbal abilities of the child. Moreover, there's an inverse correlation between the physical skill of the mother and the verbal skill of the child. Furthermore, the education of the mother can structure the transmission of skills but also the familial context. In fact, Coneus K. 2009 show that reading stories is strongly linked to the verbal abilities of three years old children. This correlation is linked to a higher level of education of the mother, and this level is also linked with the presence of a father's support that can influence the development of verbal and social skills. However, I don't know if it's just a correlation or a causal mechanism.

In Hanushek E. 2018, searchers produce a strong correlation between parents' and children's results. They eliminated the influences that don't affect the skills. By instrumenting the difference between the competence of parents and their classmates, they establish a causal effect of the parents' skills on children one's. v. d. V. R. Jacobs B. 2021 showed that Inequalities in Educational Opportunities are explained firstly by key skills (cognitive), then financial resources, and finally by soft skills (familiarity with school culture for example).

Notwithstanding, I have to highlight that there's a likely difference between educational skills of occupational skills. Therefore, the process of transmission of skill can be different. I can observe that there are some differences in Liu Y. 2013. Indeed, the searcher suggests making a distinction between general cognitive skills learned at school and those which are specialized (cognitive or not) learned in the workplace. It means that I can have two approaches to the transmission of skills: the role of the general skills linked to education and the role of the skills obtained during the professional trajectory of parents.

2.4.2 The effects of the educational skills that select occupation

The first idea is that I can consider educational skills that select an occupation. It's the selection theory. Indeed, Baudelot C. 1974 explain that school depends on the economic system. Therefore, by the limit of occupation, there's necessary a form a selection in school. Then, I can consider that a lot of individuals reach an occupation about this educational luggage and skills. The occupation shouldn't be seen exactly as a reflection of skills by requirements but as the results of the selection process that can explain the inequalities in intergenerational transmissions of skills.

In A. Jacobs B. 2021, they made a cohort study to compare the results of the CITO test between parents and their children who passed the test in 2014/2015. They establish a bias because some data were from the 1977 cohort whereas others were from the 1989 ones. The results of children who have older parents had better results, because the higher a person is educated, the older they will have a child. There's a link with selection theory: by having longer studies and higher education, they have children later and it biases the result. In Hanushek E. 2018, they highlight that the educational experience has more effect than genetic inheritance to explain the transmission of skills. Therefore, it's important to consider the selection theory.

2.4.3 The effect of the second sociability by education

By the approach of the secondary socialization produced by Berger P. 1966. They consider that there's secondary socialization during adulthood where the norms and attitudes incorporated during childhood can be reassessed. Therefore, social agents continue to learn norms, values, and attitudes at work, and also skills... It means that I can't reduce their skills and way of parenting as their educational trajectory.

2.4.4 The role of gender and other demographic variables

Moreover, some social and demographic characteristics can affect the transmission of skills.

In Anger S. 2009, the authors establish that gender has an impact on the transmission of skills. Indeed, for sons and daughters, their coding speed is more determined by the mother whereas verbal fluency by the own-gender parent. The result is a bit different in Lundborg P. 2018. They establish that the father's education has more importance for children's skills (cognitive and non-cognitive ones) than those of the mother (even if the correlation stays positive) but the mother's education has more importance for health. In Okumura T. 2014, the searchers show that there's a positive correlation between the social skills of a father and his son, but the other parent-child couples don't check this property.

The structure of the family can also have a small influence. Marjoribanks 1975 show that the inverse of family size is positively correlated with verbal, reasoning, and spatial ability scores of 11-year-old boys. Furthermore, I see in Lundborg P. 2018 that for twins, the effect of the mother's education is larger but small in the children's outcome than for adopted sons.

2.4.5 A little reflection about income

To analyze the transmission of skills, it's important to don't forget the role of income. Indeed, Brooks-Gunn J. 1997 establish that fighting against extreme poverty during childhood can reduce the inequality in school income of children. Moreover, to explain Inequalities of Education Opportunities IEO, v. d. V. R. Jacobs B. 2021 consider that the second factor that explained IEO was the financial resources. Therefore, I can consider that income has an income on results. Therefore, I have to take into account the mechanisms that underlie income inequalities.

Firstly, Williams M. 2018 show that not only occupations explain inequality but also occupational tasks. Moreover, Liu Y. 2013 showcase that increase in skill implies a larger payoff of schooling, especially in analytical tasks. These tasks need a high-level education of the worker to be performed. Then, performing rewarding tasks is linked to higher education, and it explains a part of inequalities. Therefore, people with higher education can have more income to invest in the children's education, hence the skill transmission.

3 Construction of a theoretical framework

With the previous literature review, I realized that there's a clear correlation between the skills of parents and their children. Some of them even suggest there's a causal effect between parents' skills. However, I don't really know what happens in the "black box" of the intergenerational transmission of skills.

3.1 The role of parenting style

Parenting style affects children's outcomes. Indeed, Baumrind 1967 and Baumrind 1971 considers that parenting style is on two axes: warmth and control. She shows that parents with high warmth but also a consistent applying rule and regulation method have well-adjusted children. Authoritarian parents (high control, low warmth) and permissive parents (low control, high warmth) demonstrate less positive outcomes, especially in achievement motivation. Moreover, children of authoritarian parents are less skilled in social interaction. However, these results should be moderate because Bowes J. 2009 consider they are ethnocentric and an outcome can differ from various cultural contexts. Moreover, they highlight that affection is the most important aspect to make them successful in adaptation.

Therefore, I see that parenting style is important in the intergenerational transmission of skills. Nevertheless, parenting is influenced by social contexts and capital available.

3.2 The role of financial resources

Poverty and low income are risk factors for children. Indeed, Radke-Yarrow R. 1988 shows that poverty affects children through the stress it implies on parents and the parental "irritability and anger [that] may fester in chronic conditions.".

In the UK, Ball 2006 made a typology of the circuit of school that varies by the social class of the parents. He established that there's a circuit of local schools which recruit the local community, cosmopolitan and high-profile elite schools, and independent schools (private and reached by the parents whose children cannot enter the 'cosmopolitan elite'). Private fee-paying schools can be very prohibitive. More children of the middle class go to private schools and the elite system, where they are given more resources and support. This situation can help children to develop skills.

However, Lasne 2010 shows that the results of children at the national exam at the end of middle school are higher for teachers' children than for executive's children. Indeed, they more often validate and their grades are on average higher, whereas the average monthly net income of a secondary teacher is 2800 euros in 2021 (Bour R. 2022) while the average net income for an executive is 4331 euros (Sanchez Gonzalez J. 2023). Therefore, I can't reduce the intergenerational transmission of skills as a consequence of economic income.

3.3 The importance of the cultural capital

Bourdieu P. 1970 highlight the importance of cultural capital to explain the inequalities at school. Bourdieu and Passeron consider they have three forms: embodied, objectified, and institutionalized.

3.3.1 The embodied cultural capital

The embodied cultural capital is part of this capital linked to socialization and the habitus that refers to the way to be in the body, intellectual dispositions which are linked to social origin. Some dispositions are more valuated than others. Indeed, the impact of the habitus can be seen in Bourdieu P. 1964. Pedagogic methods are more compatible with senior executives' children: dissertation, oral participation, and vocabulary in keeping with school culture... The teachers often reproach the popular classes' children their "vulgarity", or the fact they are "too much school" whereas senior executive's children demonstrate "delicacy". It's because there's a proximity to the code of school and the social codes of the upper class. Indeed, Bernstein 1971 explores this with a sociolinguistic point of view. He explains that in school, only the "produced" code is evaluated, whereas the "restricted" one is rejected. However, children from the popular class are less often socialized to this sociolinguistic code than upper classes children, which gives them disadvantages at school. It's coherent with the article Coneus K. 2009. The searchers show that there's a strong positive correlation between reading stories to a child and his development of verbal skills. It's linked to socialization about books.

To confirm that this theoretical framework is stilling relevant by comparing two children's interviews in Lahire 2019. The searchers did a study of the life conditions of 18 children from various social classes. I consider here the cases of Libertad, a young Rom girl living in a very precarious situation, and Valentine, a child of French gentility. The investigators show that Libertad has strong difficulties with the French language and had to learn the "pupil job" whereas Valentine is "well-behaved" at school and has a rich vocabulary in French and is also able to construct sentences in English.

Therefore, this analysis shows that habitus and socialization have an impact in school, which is a tool a selection as Baudelot C. 1974 highlight. Indeed, the school will select children with parents who have cultural dispositions near these of schools. Therefore, socialization allows the transmission of skills that are legitimated by school institutions.

However, financial resources matter. Indeed, it can be a tool to preserve a "between oneself". Indeed, Lahire 2019 shows that Valentine goes to the Racing Club (a prestigious sports club in Paris where the entrance fees are equal to 7000 euros, excluding an annual subscription). Thus, the aim is not the sport but this "between oneself" that can maintain solidarity by creating an endogenous social network that is an important social capital. Therefore, transmitting cultural capital and social codes is linked to these resources.

3.3.2 The objectified cultural capital

The objectified cultural capital is the set of the materials of cultural objects (books, paintings...) that a person can appropriate tout themselves. In Coneus K. 2009, children have more picture books if their mothers are high-educated, which gives material support to the development of verbal skills reached above. Therefore, I can conclude that having cultural objects can help the transmission of skills. (For example, in the case of Libertad cited in Lahire 2019, she doesn't have a book or exercise book at home).

3.3.3 The institutionalized cultural capital

The institutionalized cultural capital refers to all the things that legitimate the cultural capital as an individual attribute (for example diplomas). By considering what I mentioned above, children from social classes with high cultural capital are more likely to go to the most valuated places with the selection operated by the school system. Bourdieu P. 1970 explain that the cultural capital of the parents implies social inequalities. These social inequalities are converted into school inequalities that become social inequalities by a process of legitimating inferred by the school system. Therefore, these inequalities influence the labor market outcomes.

The Bourdieusian theory cans offer an idea of what does happen in the black box of transmission. Cultural capital is transmitted in the habitus that will affect the nature and the way of skills that a child will incorporate. Firstly, they will receive through this socialization some skills related to the family. By having a school code, upper-class children are more compatible with the schooling system so they will integrate more of the skills given by school in a culturally evaluated way. Thus, they have more chances to become a high-skill worker and develop abstract and cognitive skills than children from other social classes.

However, socialization doesn't occur only during childhood. That's why I should analyze also the secondary socialization of parents. One of the main places can be work. That's why I will consider secondary socialization at work as an important aspect of our theoretical framework.

3.4 The secondary socialization of parents at work and parenting style

In Flanagan 1990, the searchers show that employed parents bring to children information about the world of work. Therefore, children learn from their work ethic and are influenced by their aspirations for employment and further education. It can be seen as a form of informational and normative capital that the child cans use.

To go further, I should consider that occupation can have an impact on socialization. Indeed, Hughes 1958 analyzes the socialization of doctors. In order to integrate the role of "doctor", a strong duality between the profane culture and the professional culture appears. So as to outperform this duality, individuals look to a reference group and construct their personality about this reference group. It functions as an anticipating socialization. One of the valuated attitudes is the "sang-froid". This result is coherent with the results about the teacher. Being socialized in school cans help to have the social codes linked to them. Thus, I can reasonably assume that socialization at work affects parenting. The incorporation of norms, skills, and values from work can affect child socialization.

4 Construction and Analysis of the data

4.1 Presentation of the data

4.1.1 The O*NET Database

I used the O*NET database version 27.3. It gathers data about workers' characteristics, requirements, and occupation-specific characteristics in the United States and talks about around 1000 jobs. I took the part about skills. In this database, for each skill in each job, they measured a score. They distinguish two levels: importance and level requirements. There were around 35 types of skills. To measure the level, they made a continuous scale between 1 and 7. To measure importance; they produced a continuous scale between 1 and 5. In both cases, the higher is the score, the bigger the level and importance of occupation. (Fleisher M. 2012)

4.1.2 The National Educational Panel Study

The National Educational Panel Study (NEPS) is a German database that contains longitudinal data about educational processes and competence development. It's provided by the Research Data Center at the Leibniz Institute for Educational Trajectories (RDC-LIfBi). In seven starting cohorts, they collected data about the development of skills of more than 70,000 participants from early childhood to old age.

I worked with an extract of the NEPS and give for 3481 children their scores in various tests: vocabulary (at 3, 5, 7, and 9 years old), sensomotoric development (taken at 6 months old), basic cognitive skills (at 6 and 9 years old), mathematics (4, 6 and 8 years old) and scientific competences (at 5,7 and 9 years old). Moreover, for every child, I have access to their ISEI(International Socio-Economic Index), occupation, and number of years of education of each parent.

I will focus on the outcome of vocabulary and mathematics.

- For vocabulary, at 37-39 months old, they measure the ability to understand the word, sentence, and discourse that the child listens to. After three years old, reading tests are also provided. It's measured by a German version of the Peabody Picture Vocabulary Test (PPVT-IV) (Lenhard A. 2015). The test is adapted by age. The score corresponds to the number of items solved by the respondent.

- For mathematics abilities, it's measured being inspired by the concept of "mathematical literacy" defined in PISA (Organisation for Economic Co-Operation and Development [OECD] 2003). The framework distinguishes four axes:

  – Sets, number, and operations. It's the ability to understand numbers and contexts of calculation.

  – Unit and measuring. It's related to all quantification aspects.

  – Space and shape. It's about all geometrical issues.

  – Change and relationships. It's the ability to see some relationships and patterns.

  The score is obtained using the item response theory. The idea of this method is to give more weight to the difficult questions.

4.2 Construction of the database

4.2.1 Adaptation of the O*NET Database

Firstly, I wanted to reduce the number of rows. Indeed, a row was about skill in one specific job (one apparition for level score, one for importance score). I reshaped the dataframe in order to obtain the following structure:

Table 1: Reduced Reshaped O*NET Database

Occupation (O*NET-SOC Code)	Scale Type	Score skill 1	Score skill 2	Score skill 3
Occupation 1	LV			
Occupation 1	IM			
Occupation 2	LV			
Occupation 2	IM			
Occupation 3	LV			

I conserved measures by level requirement and by importance. The idea is that scores for level requirements give an idea of how people are selected for their job, whereas importance can reflect the skills developed during the profession and more generally during secondary socialization.

Therefore, I used some crosswalk files to convert the O*NET-SOC code into ISCO-08 code for a future merger. However, the O*NET-SOC code was more precise so various codes corresponded to a unique ISCO-08. Therefore, I aggregated by ISCO-08 code the occupations assimilating their score to the average of the values associated with each code. It was the easiest method to do it. Another idea was to consider the number of people who are in each O*NET-SOC code in order to balance the score, but it was more difficult.

Therefore, in reality, there were around 35 kinds of skills. I thought about doing a Principal Factor Analysis (PFA) to reduce the dimensionality. However, I have to check if the correlation between variables was satisfying enough to use this method. I cut the database into one that conserves the importance's scores and the other the level's scores in order to don't mix their respective intra-correlation.

In both databases, I estimated that Cronbach's alpha number was equal to 0.96. It's greater than 0.7, which is the acceptance criteria for coherence of the data's correlation. Having this information, I decided to do a PFA. I rotated the data with a varimax method in order to simplify the interpretation. The idea is to maximize the number of variables correlated with a principal factor.

In Appendix 1, there is the table of correlation between rotated factors and the initial skills variables. For both regressions, I obtained that four factors were the best option. Indeed, only these factors had an associated eigenvalue greater or equal to 1.

I concluded that the four factors were the following categories. They were similar between the level scale and importance scale:

- Non-STEM cognitive and decision skills: it regrouped abilities in verbal competencies, learning, analysis, and people management.
- Manual skills: it regrouped abilities in installing, using, and operating on equipment.
- STEM skills: it regrouped abilities in programming, math, sciences, and technology.
- Management skills: it regrouped abilities in managing material and financial resources.

Finally, I obtain the following dataframe structure:

Table 2: Reduced structure of the occupational Database

Occupation (ISCO code)	Importance of Cognitive and Decision Skills	Importance of manual skills	...	Level of Manual Skills	Level of STEM Skills	Level of Managerial Skills
1111						
2111						
3111						
4211						

4.2.2 Merging with NEPS

I merged the dataframe above with the NEPS dataframe by the ISCO code of the mother and next by the ISCO code of the father. Therefore, I obtained the following dataframe:

I showed the transposed version so as to clarify the reading. I have 3841 observations.

Table 3: Reduced Transposed Final Database

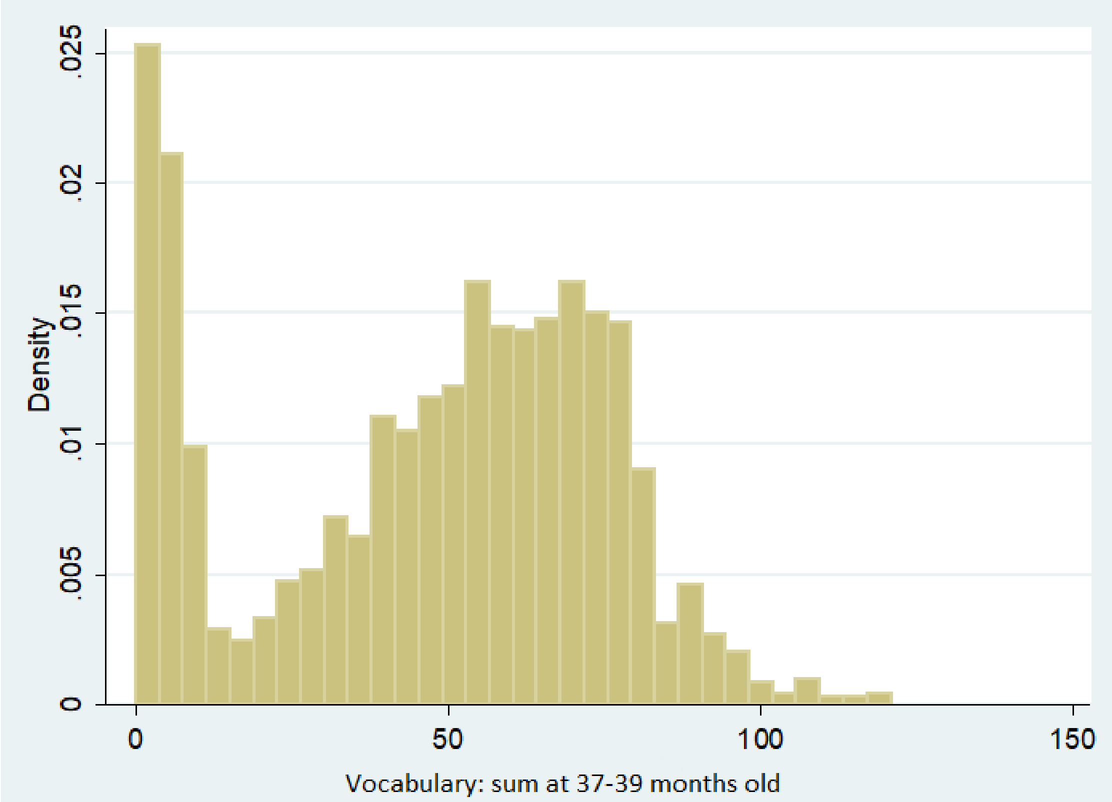
Child ID	1	2	3
Child verbal outcome at 3 years old			
Child verbal outcome at 5 years old			
Child reasoning outcome at 6 years old			
...			
ISCO Code of the mother			
ISCO Code of the father			
Number of years of education of the mother			
Number of years of education of the father			
ISEI of the mother			
ISEI of the father			
Importance of cognitive and decision skills in mother's occupation			
Importance of cognitive manual skills in mother's occupation			
...			
Level of managerial skills in mother's occupation			
Importance of cognitive and decision skills in the father's occupation			
...			

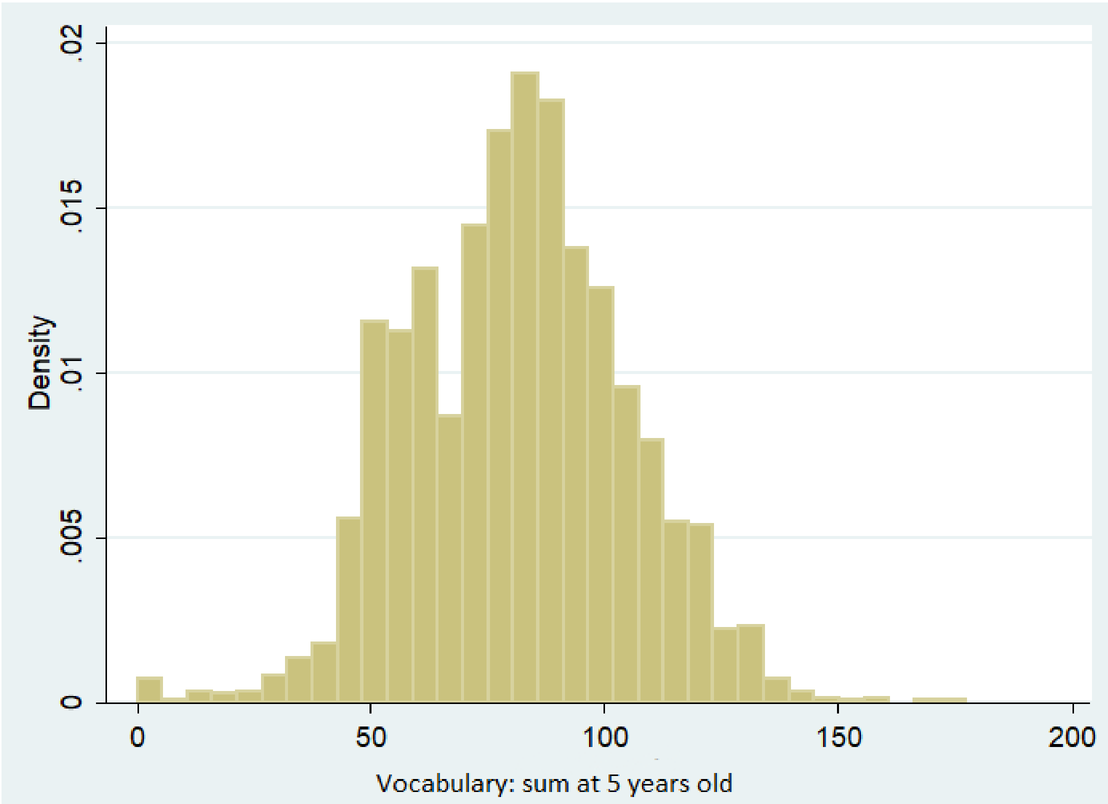
4.3 Some descriptive statistics

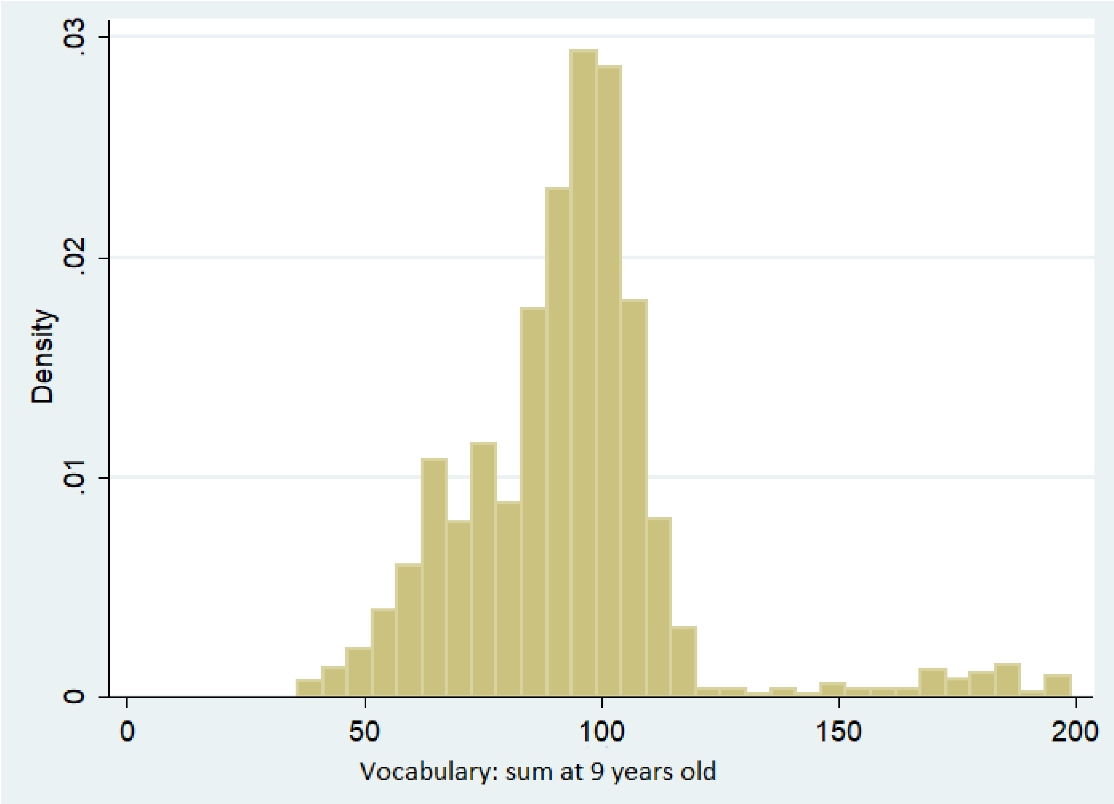
4.3.1 Outcome and explanatory variables

The occupational skills are measured with principal factors, therefore I can't make descriptive statistics easily interpretable.

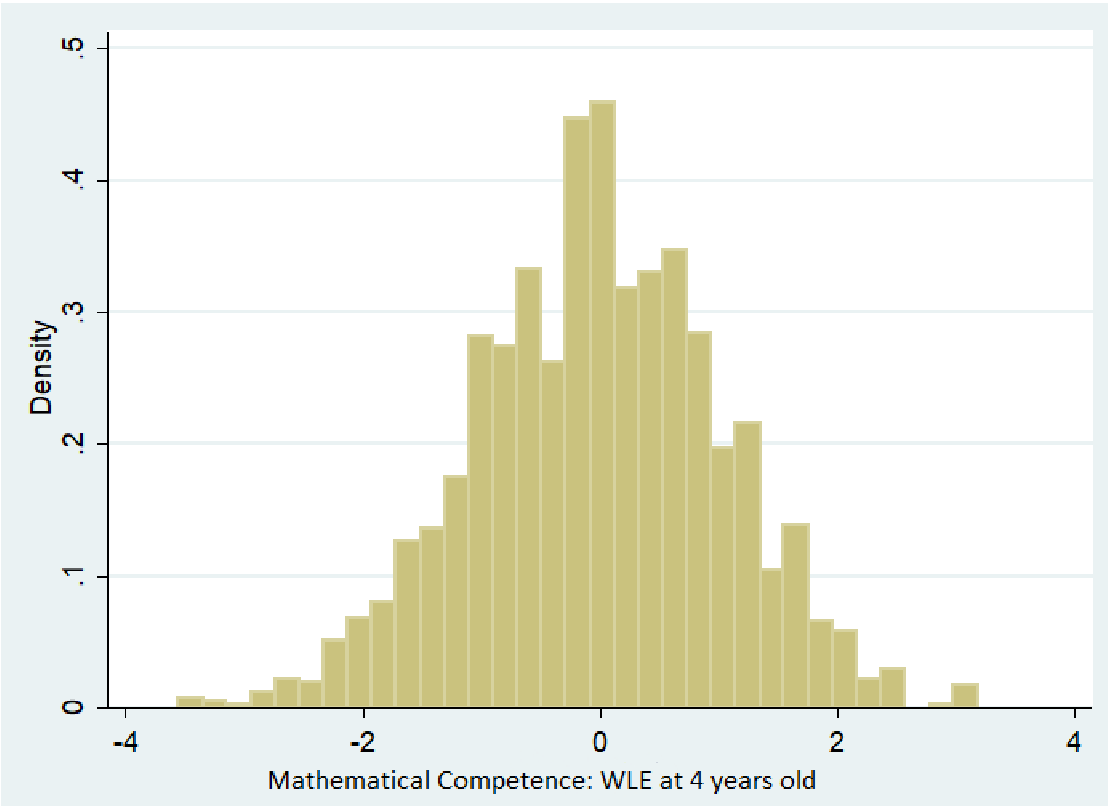
I can also analyze the children's outcome variables. I will show the score in vocabulary and mathematics abilities.

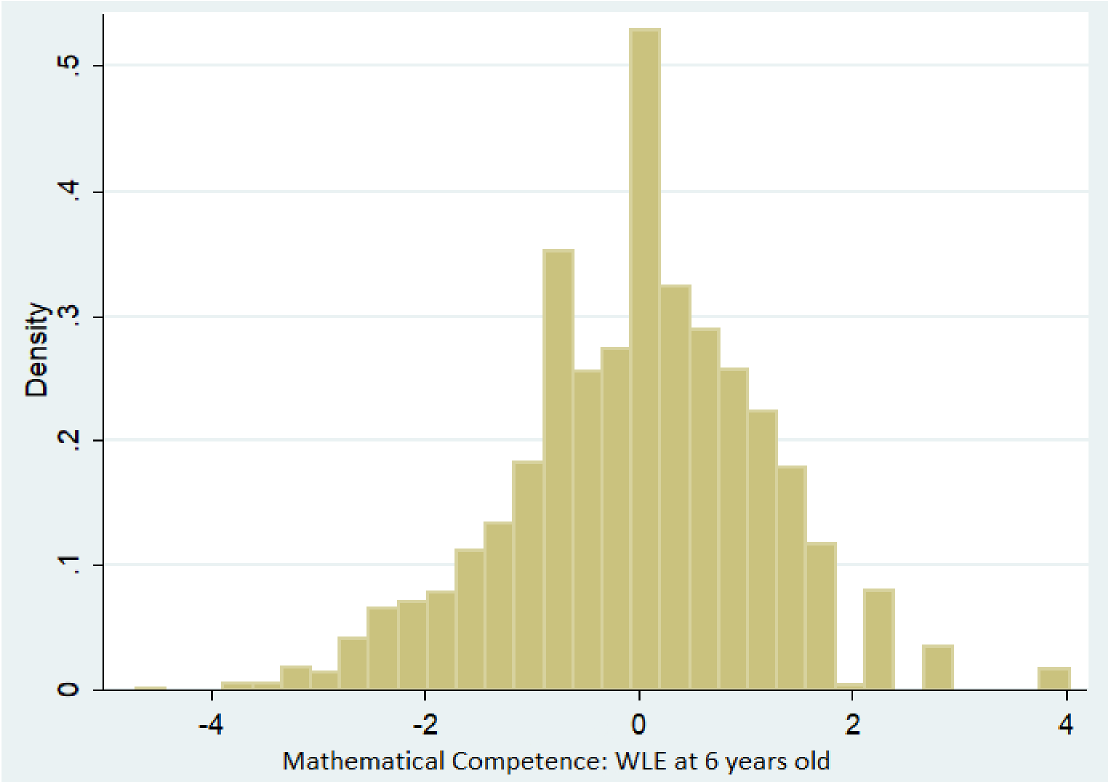

Density
0
.005
.01
.015
.02
.025
0
50
100
150
Vocabulary: sum at 37-39 months old

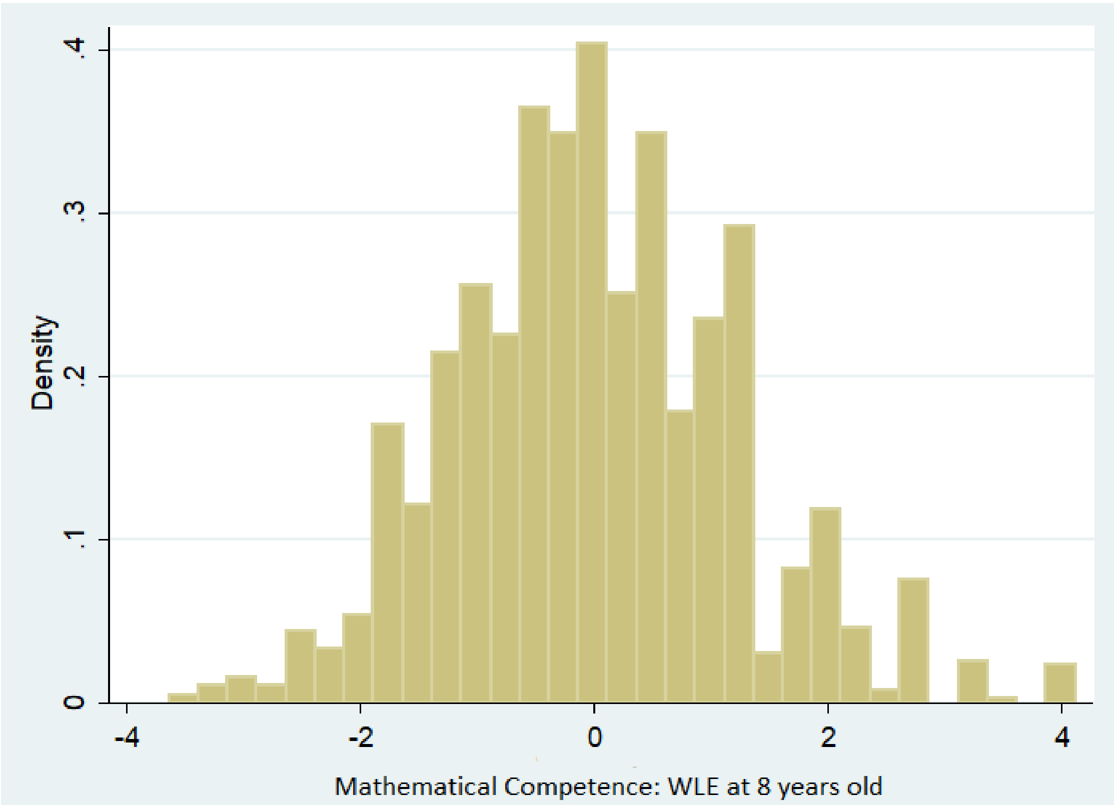

Density
0
.005
.01
.015
.02
0
50
100
150
200
Vocabulary: sum at 5 years old



In vocabulary, I see that at 3 years old, very low scores are represented. Even if the test is adapted by age, the results are higher at 5 and 9 years old. Except for the wave of 3 years old, the distribution is centered around a score of 90. At 9 years old, there are more high extreme values.







We see that the distribution seems Gaussian, but it's hard to put further because the score is not easily interpretable.

4.3.2 Missing values and control variables

Firstly, I have to see how many missing values are by category.

I see that I have a lot of missing values in the children's outcome. I thought about a way to complete the dataframe. The problem is that if I want to do it; I have to do a hypothesis of the link between children's outcomes

Table 4: Count of missing values by category

Category	Missing Values
Maternal education YEARS (earliest)	54
Paternal education YEARS (earliest)	281
Maternal occupation ISEI (earliest)	198
Paternal occupation ISEI (earliest)	345
Maternal occupation ISCO-08 (earliest)	198
Paternal occupation ISCO-08 (earliest)	345
Sensomotoric development: WLE	536
Vocabulary: sum (Wave 4: 37-39 months of Age)	1638
Vocabulary: sum (Wave 6: 5 years of age)	1418
Vocabulary: sum (Wave 8: 7 years of age)	3481
Vocabulary: sum (Wave 10: 9 years of age)	1927
Digit span: sum (Wave 4: 37-39 months of Age)	2234
Digit span: sum (Wave 7: 6 years of age)	1527
Digit span backwards: sum (Wave 8: 7 years of age)	3481
DGCF perceptual speed sum (Wave 7: 6 years of age)	1501
DGCF reasoning sum (Wave 7: 6 years of age)	1494
DGCF (reasoning): sum (Wave 10: 9 years of age)	1927
Mathematical competence: WLE (Wave 5: 4 years of age)	1467
Mathematical competence: WLE (Wave 7: 6 years of age)	1513
Mathematical competence: WLE (Wave 9: 8 years of age)	1935
WLE estimator Scientific competence (Wave 6: 5 years of age)	1421
WLE estimator Scientific competence (Wave 8: 7 years of age)	1565
WLE estimator scientific competence (Wave 10: 9 years of age)	1932
Importance mother skill cognitive decision in job	225
Importance mother skill manual in job	225
Importance mother computing skill in job	225
Importance mother management skill in job	225
Level mother cognitive decision skill in job	225
Level mother manual skill in job	225
Level mother computing skill in job	225
Level mother managing skill in job	225

and parental professional, socio-economic status, or number of years of education. In other words, I'll have to assume things that I want to test. Therefore, I only can highlight this problem but I let missing values without modifying them.

Now, I will try to see the distribution and structure of our control variables.

I see that the distribution by parents is pretty similar for years of education and ISEI. They studied on average 14.5 years and have a middle socio-economic status. The ISEI is an indicator that represents the socio-economic status by linking education and income in an occupation (Ganzeboom H. 1992). The values are between 1 and 99, so I see that I don't have the most extreme values in our sample. But there's enough variety of socioeconomic status measured by ISEI that looks satisfying to model.

Table 5: Summary statistics of control variables

Variable	**Count**	**Mean**	**Std**	**Min**
Maternal education YEARS (earliest)	3427.000000	14.514444	2.656982	9.000000
Paternal education YEARS (earliest)	3200.000000	14.594375	2.671062	9.000000
Maternal occupation ISEI (earliest)	3283.000000	54.936534	20.766057	11.560000
Paternal occupation ISEI (earliest)	3136.000000	54.532210	23.004699	11.740000
Variable	**25%**	**50%**	**75%**	**Max**
Maternal education YEARS (earliest)	13.000000	15.000000	18.000000	18.000000
Paternal education YEARS (earliest)	13.000000	15.000000	18.000000	18.000000
Maternal occupation ISEI (earliest)	39.040001	56.000000	73.910004	88.959999
Paternal occupation ISEI (earliest)	30.780001	58.059999	75.129997	88.959999

5 Modelization

I will run some linear regressions so as to establish a correlation between the skills of parents and the skills of children in vocabulary and mathematics. I will control it by education and by ISEI. In the following models, I will suppose the random selection of the sample.

5.1 Vocabulary outcome

I made the following models:

$$vocabulary_{age} = \alpha + \beta_1 * skill_{type} + \varepsilon \tag{1}$$

- Where $vocabulary_{age}$ is the standardized vocabulary outcome of the children for the $age \in \{3,5,9\}$ years old.
- Where $skill_{type}$ is a parental skill in the various types by level and importance: Non-STEM Cognitive and decision, Manual, STEM, and Managerial for each parent.
- Where ε is a error term.

I made two versions with control of this model:

$$vocabulary_{age} = \alpha + \beta_1 * skill_{type} + \beta_2 * education + \varepsilon \tag{2}$$

$$vocabulary_{age} = \alpha + \beta_1 * skill_{type} + \beta_2 * ISEI + \varepsilon \tag{3}$$

- Where $education$ is the standardized number of years of education for each parent.
- Where $ISEI$ is the ISEI for each parent.

In order to check the correlation between the explanatory variables and their non-multicollinearity, I calculated their variation inflation factors. I obtained for each pair a value between 1 and 2.5. It means that the correlation is effective and that I don't have to worry about multicollinearity. Indeed, Allison 1994 shows that there's not risk of multicollinearity for all factors smaller or equal to 2.5. The correlation of the control variables with the outcome variable is justified by the theoretical framework.

For each parent, I ran linear regressions with the models mentioned above. I chose to differentiate the parents by gender because in the literature review; I saw that there were gender-effect in the transmission of skills. In Table 6, I reported the coefficient of each skill with each vocabulary outcome with their p-value for the mother. I did the same for the father in Table 7. For each regression with control, I ran, the coefficient of education and ISEI were positive and statistically significant at the 1% level.

Table 6: Correlation between mother skills and vocabulary outcome of children

	(1)		(2)		(3)	
Skills	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Vocabulary Score at 3 years old						
Importance						
Non-STEM Cognitive and Decision	0.0508698	0.016	NS	0.188	-0.0413606	0.098
Manual	-0.1016349	0.013	-0.0745263	0.064	-0.0817147	0.044
STEM	0.136202	0.000	NS	0.163	NS	0.444
Managerial	NS	0.414	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.0926119	0.000	NS	0.510	NS	0.128
Manual	-0.1075155	0.007	-0.0731765	0.061	-0.0816759	0.038
STEM	0.0823599	0.003	NS	0.209	NS	0.463
Managerial	NS	0.745	-	-	-	-
Vocabulary Score at 5 years old						
Importance						
Non-STEM Cognitive and Decision	0.1061883	0.000	NS	0.467	NS	0.351
Manual	-0.098678	0.010	NS	0.113	-0.068873	0.062
STEM	0.1658608	0.000	NS	0.240	NS	0.358
Managerial	NS	0.231	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.1587866	0.000	NS	0.165	NS	0.157
Manual	-0.1027409	0.005	NS	0.137	-0.0645957	0.072
STEM	0.0888051	0.001	NS	0.454	NS	0.819
Managerial	NS	0.547	-	-	-	-
Vocabulary Score at 9 years old						
Importance						
Non-STEM Cognitive and Decision	0.1504129	0.000	0 .0876137	0.000	0.0540649	0.041
Manual	NS	0.692	-	-	-	-
STEM	0.1382965	0.000	NS	0.129	NS	0.261
Managerial	NS	0.145	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.2042758	0.000	0.1262184	0.000	0.0763623	0.034
Manual	NS	0.443	-	-	-	-
STEM	NS	0.196	-	-	-	-
Managerial	NS	0.398	-	-	-	-

Table 7: Correlation between father skills and vocabulary outcome of children

	(1)		(2)		(3)	
Skills	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Vocabulary Score at 3 years old						
Importance						
Non-STEM Cognitive and Decision	0.1309482	0.000	0.0680538	0.009	0.0531916	0.055
Manual	-0.0944258	0.000	NS	0.348	NS	0.520
STEM	0.0620452	0.002	NS	0.860	-0.0471103	0.057
Managerial	NS	0.405	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.1817181	0.000	0.0944448	0.002	0.0734845	0.048
Manual	-0.0882704	0.001	NS	0.328	NS	0.391
STEM	NS	0.317	-	-	-	-
Managerial	NS	0.524	-	-	-	-
Vocabulary Score at 5 years old						
Importance						
Non-STEM Cognitive and Decision	0.1912556	0.000	0.0979569	0.000	0.0744699	0.003
Manual	-0.13869	0.000	NS	0.146	NS	0.260
STEM	0.0873027	0.000	NS	0.568	-0.0802961	0.000
Managerial	NS	0.469	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.2596563	0.000	0.1229832	0.000	0.0803936	0.018
Manual	-0.1319698	0.000	-0.0403629	0.093	NS	0.101
STEM	0.0381535	0.030	NS	0.113	-0.0705197	0.000
Managerial	NS	0.449	-	-	-	-
Vocabulary Score at 9 years old						
Importance						
Non-STEM Cognitive and Decision	0.0789002	0.003	NS	0.754	NS	0.550
Manual	-0.1021761	0.000	NS	0.711	NS	0.596
STEM	0.0950726	0.000	NS	0.542	NS	0.924
Managerial	NS	0.749	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.1410322	0.000	NS	0.897	NS	0.378
Manual	-0.0944009	0.001	NS	0.656	NS	0.463
STEM	0.0652171	0.001	NS	0.524	NS	0.917
Managerial	NS	0.972	-	-	-	-

For the mother, I see that the coefficients of managerial skills are not statistically significant. The correlation between Non-STEM Cognitive and Decision, STEM skills, and child's outcome is positive whereas it's negative with manual skills. The older a child is, the stronger the positive correlation with Non-STEM cognitive and decision skills with vocabulary score. Moreover, the level's requirements are more correlated than the importance. It's the opposite for STEM skills. For the manual skills, level and importance are pretty rather similar. Without control, the manual skills are significant at 3 and 5 years old and the impact of STEM diminishes with a child's age. With the control by education, a large part of the skill effect is statistically non-significant and when it is stills significant, the coefficient is nearly divided by 2. With the control by ISEI, STEM effects become statistically non-significant. When it's statistically significant, at 3 years I observe a negative correlation between cognitive skills and children's outcome (it's statistically significant at 10% level) The penalty by manual skills is nearly divided by 2.

For the father, the coefficients of managerial skills are not statistically significant. I see that the coefficients are higher at 5 years old than at other ages. The lowest coefficients appear at 9 years old and, with control, they become non-significant. Between importance and level requirements, I observe that the importance is stronger for STEM skills and it's the opposite for Non-STEM Cognitive and Decision skills. Controlled by education, STEM, and manual (except for level coefficient at 5 years old) skills are non-statistically significant. For cognitive skills, at 3 and 5 years old, the coefficient is nearly divided by 2. When I control by ISEI, manual skills aren't significant. At 3 en 5 years old, I observe a negative correlation between vocabulary outcome and STEM. For the cognitive one, I observe that the coefficient is nearly divided by 3.

I observe pretty similar dynamics between father and mother results: the reduction of the effects by control and the non-significant coefficients of managerial skills. Moreover, the coefficients of the importance of STEM skills are greater than the coefficients of the level requirements for this type of skills, it's the opposite for Non-STEM cognitive and decision skills. For manual skills, it is stills pretty similar. For the Non-STEM cognitive and decision skills, the father's coefficients are greater than the mother's ones at 3 and 5 years old, but this trend is inverted at 9 years old. For STEM skills (except for the level coefficient at 9 years old), the coefficients of the mother are greater than the mother ones. For manual skills, the negative correlation is stronger for the mother than for the father at 3 years old, but from 5 years old, this trend is inverted. At first glance, the theory of selection can explain better the results for Non-STEM Cognitive and Decision skills. For STEM Skills, the theory of secondary socialization seems more relevant. However, I see that the loss of "significant" in coefficients when I control motivates me to check the precise coefficients for the number of years of education. I can see it in Table 8. I see that coefficients of the education are greater than those concerning skills. It means that selection theory is probably dominant in explaining the inter-generational transmission of skills. Indeed, parents who were able to have a longer education are more strongly correlated with the children's outcome than the coefficients about the skills.

Table 8: Coefficients in the regression controlled by the number of years of education of a parent

	(Mother)		(Father)	
Skills	Skill	Education	Skill	Education
Vocabulary Score at 3 years old				
Importance				
Non-STEM Cognitive and Decision	NS	0.2350076	0.0680538	0.1924096
Manual	-0.0745263	0.2188057	NS	0.2078891
STEM	NS	0.2054983	NS	0.2179333
Level Requirements				
Non-STEM Cognitive and Decision	NS	0.2307624	0.0944448	0.1692174
Manual	-0.0731765	0.2173953	NS	0.2081625
STEM	NS	0.2151286	-	-
Vocabulary Score at 5 years old				
Importance				
Non-STEM Cognitive and Decision	NS	0.3063689	0.0979569	0.2905063
Manual	NS	0.3099864	NS	0.311029
STEM	NS	0.3002808	NS	0.3280585
Level Requirements				
Non-STEM Cognitive and Decision	NS	0.2956716	0.1229832	0.2628115
Manual	NS	0.3092928	-0.0403629	0.3102967
STEM	NS	0.3091343	NS	0.3329198
Vocabulary Score at 9 years old				
Importance				
Non-STEM Cognitive and Decision	0.0876137	0.2204268	NS	0.2860109
Manual	-	-	NS	0.279819
STEM	NS	0.2423912	NS	0.276599
Level Requirements				
Non-STEM Cognitive and Decision	0.1262184	0.1944761	NS	0.2852028
Manual	-	-	NS	0.2793772
STEM	-	-	NS	0.2785673

5.2 Mathematical outcome

In order to measure the correlation between parental skills and mathematical outcomes, I made the following models:

$$math_{age} = \alpha + \beta_1 * skill_{type} + \varepsilon \quad (4)$$

- Where mat_{age} is the standardized vocabulary outcome of the children for the $age \in \{4,6,8\}$ years old.
- Where $skill_{type}$ is a parental skill in the various types by level and importance: Non-STEM Cognitive and decision, Manual, STEM, and Managerial for each parent.
- Where ε is a error term.

I made two versions with control of this model:

$$math_{age} = \alpha + \beta_1 * skill_{type} + \beta_2 * education + \varepsilon \quad (5)$$

$$math_{age} = \alpha + \beta_1 * skill_{type} + \beta_2 * ISEI + \varepsilon \quad (6)$$

- Where $education$ is the standardized number of years of education for each parent.
- Where $ISEI$ is the ISEI for each parent.

The non-multicolinearity and correlation between the variables of control and the explanatory variables are verified because their variance inflation factor are between 1 and 2.5. In the Table 9, I reported the coefficient of each skill with each mathematical outcome with their p-value for the mother. I did the same for the father in Table 10. For each regression with the control I ran, the coefficient of education and ISEI were positive and statistically significant at the 1% level.

For the mother, I see that manual skill coefficient at 3 years old and managerial's ones at 4 and 6 years old are not statistically significant. The correlation between Non-STEM Cognitive and STEM skills is positive. It's the opposite for manual and managerial skills. The highest coefficients, in absolute value, appear at 6 years old. Level requirements coefficients for Non-STEM Cognitive and Decision and Manual Skills are higher than Importance ones. For STEM skills, it's the opposite. When I control the model by the number of years of education, the coefficient of Non-STEM cognitive and managerial skills are non-statically significant. It's also the case for manual ones except at 8 years old where the effect is a bit divided. For the STEM one, when it's statistically significant, the coefficients are divided by 2 or 3. When I control by ISEI, only non-STEM cognitive skills at 4 years old and level requirements in manual skills at 8 years old are statistically significant. For non-STEM cognitive skills, the coefficients are very reduced. The diminution of the impact is lower for manual skills.

For the father, managerial skills are not significantly correlated with children's outcomes. The sense of correlation is similar to those of the mother and also the age dynamics. Importance's coefficient is higher for STEM skills than level ones. It's the opposite for manual and Non-STEM cognitive skills. When I control it by education, all coefficients become none significant except for the cognitive ones at 4 and 8 years old where they are divided by 2 or 3. When I control by ISEI, all coefficients are none significant except one of them: the Level Requirements of STEM when the child is 4 years old. I observe an inversion of the sense of correlation.

By comparing the results of the mother and the father, I observe some common dynamics: the same sense of comparison of coefficients between importance and level requirements, manual coefficients, and age dynamics. At 6 years old, the highest coefficients appear. However, STEM values are higher for the mother and non-STEM cognitive ones are higher for the father.

Table 9: Correlation between mother skills and mathematical outcome of children

	(4)		(5)		(6)	
Skills	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Mathematical Score at 3 years old						
Importance						
Non-STEM Cognitive and Decision	0.0775272	0.000	NS	0.925	-0.040764	0.084
Manual	NS	0.132	-	-	-	-
STEM	0.1780492	0.000	0.0799042	0.003	NS	0.299
Managerial	NS	0.444	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.13111	0.000	NS	0.217	-0.0555039	0.079
Manual	NS	0.133	-	-	-	-
STEM	0.1125447	0.000	0.0572123	0.028	NS	0.212
Managerial	NS	0.115	-	-	-	-
Mathematical Score at 5 years old						
Importance						
Non-STEM Cognitive and Decision	0.0992419	0.000	NS	0.833	NS	0.238
Manual	-0.0748172	0.056	NS	0.363	NS	0.213
STEM	0.2016714	0.000	0.0789205	0.004	NS	0.213
Managerial	NS	0.822	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.1576634	0.000	NS	0.236	NS	0.215
Manual	-0.0862494	0.023	NS	0.310	NS	0.170
STEM	0.1163517	0.000	0.0502338	0.054	NS	0.304
Managerial	NS	0.214	-	-	-	-
Mathematical Score at 9 years old						
Importance						
Non-STEM Cognitive and Decision	0.0648164	0.006	NS	0.487	NS	0.998
Manual	NS	0.118	-	-	-	-
STEM	0.063792	0.027	NS	0.821	NS	0.166
Managerial	-0.0584058	0.017	NS	0.161	NS	0.133
Level Requirements						
Non-STEM Cognitive and Decision	0.0892044	0.001	NS	0.481	NS	0.541
Manual	-0.0949674	0.027	-0.0722952	0.089	-0.0771209	0.071
STEM	NS	0.516	-	-	-	-
Managerial	-0.0433832	0.060	NS	0.216	NS	0.227

Table 10: Correlation between father skills and mathematical outcome of children

	(4)		(5)		(6)	
Skills	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Mathematical Score at 3 years old						
Importance						
Non-STEM Cognitive and Decision	0.1278694	0.000	0.0532126	0.029	NS	0.114
Manual	-0.094907	0.000	NS	0.474	NS	0.562
STEM	0.080878	0.000	NS	0.922	NS	0.136
Managerial	NS	0.577	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0 .190068	0.000	0.0753559	0.008	NS	0.110
Manual	-0.0819633	0.001	NS	0.559	NS	0.515
STEM	0.0455603	0.009	NS	0.608	-0.0325458	0.089
Managerial	NS	0.492	-	-	-	-
Mathematical Score at 5 years old						
Importance						
Non-STEM Cognitive and Decision	0.1531039	0.000	0.0617337	0.011	NS	0.172
Manual	-0.095775	0.000	NS	0.998	NS	0.580
STEM	0.1206561	0.000	0.0401167	0.044	NS	0.534
Managerial	NS	0.794	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.2313214	0.000	0.1060635	0.000	NS	0.104
Manual	-0.0777876	0.001	NS	0.740	NS	0.555
STEM	0.0719649	0.000	NS	0.313	NS	0.327
Managerial	NS	0.858	-	-	-	-
Mathematical Score at 9 years old						
Importance						
Non-STEM Cognitive and Decision	0.0648164	0.006	NS	0.386	NS	0.980
Manual	-0.0783635	0.005	NS	0.400	NS	0.580
STEM	0.0500514	0.018	NS	0.877	NS	0.534
Managerial	NS	0.467	-	-	-	-
Level Requirements						
Non-STEM Cognitive and Decision	0.1102779	0.000	NS	0.251	NS	0.104
Manual	-0.0677197	0.012	NS	0.504	NS	0.555
STEM	NS	0.164	-	-	-	-
Managerial	NS	0.388	-	-	-	-

Table 11: Coefficients in the regression controlled by the number of years of education of a parent

	(Mother)		(Father)	
Skills	Skill	Education	Skill	Education
Mathematical Score at 3 years old				
Importance				
Non-STEM Cognitive and Decision	NS	0.2350076	0.0532126	0.2357205
Manual	-	-	NS	0.2471447
STEM	0.0799042	0.2252475	NS	0.2515678
Level Requirements				
Non-STEM Cognitive and Decision	NS	0.2411248	0..0753559	0.2163156
Manual	-	-	NS	0.2485682
STEM	0.0572123	0.2151286	NS	0.256332
Mathematical Score at 5 years old				
Importance				
Non-STEM Cognitive and Decision	NS	0.3063689	0.0617337	0.2625327
Manual	NS	0.3052158	NS	0.2840251
STEM	0.0789205	0.2755943	0.0401167	0.2646178
Level Requirements				
Non-STEM Cognitive and Decision	NS	0.2916184	0.1060635	0.2314165
Manual	NS	0.3043667	NS	0.2865537
STEM	0.0502338	0.2968553	NS	0.2776506
Mathematical Score at 9 years old				
Importance				
Non-STEM Cognitive and Decision	NS	0.1839365	NS	0.1737003
Manual	-	-	NS	0.1720935
STEM	NS	0.1842203	NS	0.1789886
Managerial	NS	0.1842203	-	-
Level Requirements				
Non-STEM Cognitive and Decision	NS	0.181199	NS	0.1643572
Manual	-0.0722952	0.1867569	NS	0.1743162
STEM	-	-	-	-
Managerial	NS	0.1871722	-	-

As I observe in the vocabulary part, at first glance, the theory of selection can explain better the results for Non-STEM Cognitive and Decision skills. For STEM Skills, the theory of secondary socialization seems more relevant. However, I see that the loss of "significant" coefficients when I control motivates me to check the precise coefficients for the number of years of education (Table 11). I see that Education coefficients are higher than those of skills.

6 Analysis of the results

Therefore, I see that there's a correlation between vocabulary, mathematical outcome, and parental skills, except for managerial ones in the majority of regressions. In both outcomes, the level requirements are more correlated for manual and Non-STEM cognitive skills and it's the opposite for STEM skills. It means that the theory of selection and the theory of secondary socialization have a conceptual interest. For Non-STEM Cognitive and Manual Skills, I can suppose that the school selection explains the result. For STEM skills, it can especially help with mathematical outcomes can a person can apply their learning to help their child's development. However, when I control by education, I see that a lot of coefficients become non-significant. Therefore, I see that the theory of selection is clearly relevant. However, when I control by ISEI, some coefficients are inverted. Therefore, selection doesn't totally explain. I can suppose that for high qualifications, parents are less available to educate their children. Therefore, socialization at the job can have an influence.

Also, I observe age and gender dynamics. I see the highest coefficients appear at 5 and 6 years old. For mathematical skills, the influence of the father is greater with non-STEM Cognitive skills and STEM ones are greater for the mother. For the vocabulary, I see also the same dynamics except at 9 years old, where the mother has more impact on non-STEM cognitive skills. For manual skills, in vocabulary, the impact of the father is greater than those of the mother except at 3 years old. In mathematics, the penalty by the father is globally stronger. For Non-STEM and Decision skills, in vocabulary, I observe, even by control, that the father is more influential at 3 and 5 years old and the mother at 9 years old. In mathematics, when I control, only the father has a positive impact at 4 and 6 years old. I don't have a clear idea to interpret that, but I can suppose that parents can have different roles in parenting and therefore be more or less influential to suit their parental tasks. For manual skills, I can suppose that the cause is the education and the ISEI because only vocabulary stays linked with the mother when I control at 3 and 5 years old for vocabulary. I can suppose, therefore, that the stronger presence of the mother in the early socialization of the child can explain that because she has less direct vocabulary skills to transmit if I link with education and ISEI. It is also coherent because I observe a similar phenomenon for the link between STEM skills and the mother and mathematical outcome at 4 and 6 years old. However, this result is not consistent with what I saw about non-STEM Cognitive skills. I can regret that I didn't know the gender of the children to make more precise the gender dynamics.

However, I have to highlight the limits of the study. Firstly, a lot of jobs had only one occurrence in my database. Therefore, there's a sample bias and I have to replicate the kind of analysis I conducted with another database. Moreover, I studied only two types of outcomes: vocabulary and mathematical skills. I will detail some tracks to pursue research in my conclusion. Last, I identified correlations and not causal effects.

7 Conclusion

Therefore, I saw that our theoretical framework allows us to compare the theory of selection and the theory of secondary socialization. The theory of selection seems dominant in explaining the inter-generational transmission of skills, even if I shouldn't reject the theory of secondary socialization, especially for STEM skills. I also have to highlight that the effect varies by the gender of the parent, type of competence, and age of the child.

There are a few tracks to pursue the research. Firstly, I should analyze the trajectories of children from early adulthood. Indeed, I observed that effects were diminished at 8 and 9 years old, so I have to check if parental skills are relevant to explain the school path. Besides, it should be a good idea to try to reproduce these results by using other databases or methods. Moreover, I didn't analyze motor, scientific, creative, or other types of skills that can be also useful to understand the inter-generational transmission of inequalities. Indeed, for example, **grusky** shows that creative skills have a specific economical return in the job market. I can also analyze by considering the gender, race, or disabilities of children and parents. Furthermore, I saw the effect of education. It should be interesting to proxy the parental competencies by the skills learned during studies to see if the correlation is better. Also, I should measure parenting skills or investment to verify their impact I highlight in the literature review.

References

Allison, P. D. (1994). *Multiple Regression: A Primer (Undergraduate Research Methods Statistics in the Social Sciences)*.

Anger S., Heineck G. (2009). "Do Smart Parents Raise Smart Children? The Intergenerational Transmission of Cognitive Abilities". In: *Journal of Population Economics* 23, pp. 1105–1132.

Angrist J., Pischke J. (2009). *Mostly Harmless Econometrics*. Ed. by Princeton.

Autor D., Handel M. (2013). "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages". In: *Journal of Labor Economics* 31.2, S59–S96.

Ball, S. (2006). *Education Policy and Social Class*. Ed. by Routledge.

Barg K., Klein M. (2023). "Maternal Occupation-Specific Skills and Children's Cognitive Development". In: *Sage*.

Baudelot C., Establet R. (1974). *L'école capitaliste en France*.

Baumrind, D. (1967). "Child care practices antecending three patterns of preschool behavior". In: *Genetic Psychology Monographs* 75, pp. 12–17.

— (1971). "Current pattern of parental authority". In: *Developmental Psychology Monographs*, pp. 1–103.

Berger P., Luckmann T. (1966). *The social construction of reality. A treatise in the sociology of knowledge*.

Bernstein, B. (1971). *Class, codes and control*. Ed. by Routledge.

Bour R., Volat G. (2022). *Les salaires dans la fonction publique de l'État*. Tech. rep. INSEE.

Bourdieu P., Passeron J. (1964). *Les Héritiers, Les étudiants et la culture*. Ed. by Les éditions de Minuit.

— (1970). *La Reproduction. Éléments pour une théorie du système d'enseignement*. Ed. by Les éditions de Minuit.

Bowes J., Grace R. (2009). *Children, Families Communities. Contexts and consequences. Third Edition*. Ed. by Oxford University Press.

Brooks-Gunn J., Duncan G. (1997). "How much does childhood poverty affect the life chances of children". In: *The Future of Children* 7.2, pp. 55–71.

Coneus K., Sprietsma M. (2009). "Intergenerational Transmission of Human Capital in Early Childhood". In: *ZEW – Centre for European Economic Research Discussion* 09.038.

Devaux, M. (2014). "Social Inequalities in Health-Related Behaviours: Is the grass greener on the other side?" PhD thesis. Université Paris-Dauphine.

Duppong Hurley K., Al. (2014). "Assessing Parenting Skills and Attitudes: A Review of the Psychometrics of Parenting Measures". In: *Journal of Child and Family Studies* 23, pp. 812–823.

Else-Quest, Al. (2010). "Cross-national patterns of gender differences in mathematics: a meta-analysis". In: *Psychological Bulletin* 136.1, pp. 103–127.

Flanagan, C. (1990). "Families and School in Hard Times". In: *New direction for child development* 46, pp. 7–26.

Fleisher M., Tsacoumis S. (2012). *O*NET® Analyst Occupational Skills Ratings: Procedures Update*. Tech. rep. Human Ressources Searching Organization (HumRRO).

Frick, P. (1999). "The Alabama Parenting Questionnaire". In: *American Psychological Association*.

Ganzeboom H., Treiman D. (1992). "A Standard Socio-Economic Index of Occupational Status". In: *Social Science Research*.

Gerard, A.B. (1994). "Parent-child relationship inventory manual". In: *Los Angeles: WPS*.

Hanushek E., Al. (2018). "The intergenerational transmission of cognitive skills". In: *ROA*.

Hershbein B., Kahn L. (2018). "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings". In: *American Economic Review* 108.7, pp. 1737–72.

Hughes, E. (1958). *Men and their work*.

Jacobs B., Al. (2021). "The Intergenerational Transmission of Skills Data Set". In: *ROA Technical Reports* 007.

Jacobs B., van der Velden R. (2021). "Exploring the Uncharted Waters of Educational Mobility: The Role of Key Skills". In: *ROA*.

Lahire, B. (2019). *Enfances de classe. De l'inégalité parmi les enfants*. Ed. by Editions du Seuil.

Lasne, A. Da Costa (2010). "La meilleure réussite scolaire des enfants d'enseignants : vers une explication causale". In: *Université de Bourgogne*.

Lenhard A. Lenhard W., Segerer R. (2015). "Peabody Picture Vocabulary Test-Revision IV German Adaption". In.

Lignier, W. (2007). "Les enfants surdoués, des "élus" de l'intelligence". In: *Mouvements* 1.49, pp. 63–72.

Liu Y., Grusky B. (2013). "The Payoff to Skill in the Third Industrial Revolution". In: *American Journal of Sociology* 133.5, pp. 1330–74.

Lundborg P., Al. (2018). "The Intergenerational Transmission of Human Capital: Exploring the Role of Skills and Health Using Data on Adoptees and Twins". In: *Journal of Population Economics* 31.4, pp. 1035–1065.

Marjoribanks, Al. (1975). "Mental Abilities: Sibling Constellation and Social Class Correlates". In: *British Journal of Social and Clinical Psychology* 14.2, pp. 109–116.

Okumura T., Usui E. (2014). "Do Parents' Social Skills Influence Their Children's Sociability?" In: *The B.E. Journal of Economic Analysis and Policy* 14.3.

Organisation for Economic Co-Operation and Development [OECD] (2003). *The PISA 2003 assessment framework – mathematics, reading, science and problem solving knowledge and skills*. Paris, France: OECD.

Pohl S., Cartensen C. (2012). *NEPS Technical Report – Scaling the Data of the Competence Tests*. Tech. rep. Federal Ministry of Education and Research.

Radke-Yarrow R., Al. (1988). "Chil Development in a Network of Relationships". In: *Relationships within families: Mutual Influence*.

Sanchez Gonzalez J., Sokhna N. (2023). *Les salaires dans le secteur privé en 2021*. Tech. rep. INSEE.

Sternberg, R., ed. (2006). *Handbook of intelligence*. Cambridge University Press.

VandenBos, G. (2007). *APA Dictionary of Psychology*. 1st ed. Washington DC: American Psychological Association. URL: https://www.apa.org/pubs/dictionaries/dict-of-psychology.

Williams M. Bol, T. (2018). "Occupations and the wage structure: The role of occupational tasks in Britain". In: *Research in Social Stratification and Mobility* 53, pp. 16–25.

A PFA on the occupational skills

Table 12: PFA on importance's score of skills

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
Active Learning	0.7521	-0.2227	0.5076	0.1361	0.1085
Active Listening	0.7373	-0.4882	0.3268	0.0216	0.1107
Complex Programming	0.6932	-0.0852	0.6072	0.2087	0.0999
Coordination	0.7874	-0.1878	0.0048	0.4343	0.1562
Critical Thinking	0.7386	-0.1484	0.5634	0.1451	0.0940
Equipment Selection	-0.1877	0.9262	-0.1134	-0.0821	0.0873
Equipment Maintenance	-0.1972	0.9180	0.0270	-0.0425	0.1159
Installation	-0.0526	0.6763	0.1796	-0.1470	0.4860
Instructing	0.8543	-0.1232	0.2186	0.0948	0.1983
Judgment and Decision Making	0.7476	-0.1371	0.4870	0.2517	0.1218
Learning Strategies	0.8098	-0.1469	0.3441	0.0743	0.1987
Management of Financial Resources	0.3489	-0.0510	0.2638	0.8438	0.0941
Management of Personnel Resources	0.7311	0.0281	0.1475	0.5599	0.1294
Mathematics	0.2458	-0.0316	0.7178	0.2922	0.3380
Monitoring	0.7848	0.0576	0.2014	0.2857	0.2586
Negotiation	0.7185	-0.3735	0.1099	0.3966	0.1750
Operation and Control	-0.1917	0.8936	-0.1841	0.0203	0.1304
Operations Monitoring	0.2748	-0.0255	0.7011	0.3646	0.2994
Operations Analysis	-0.1057	0.9339	0.0033	0.0474	0.1144
Persuasion	0.6670	-0.4381	0.1500	0.3330	0.2297
Programming	0.0696	-0.0569	0.8403	0.0288	0.2849
Quality Control Analysis	-0.1152	0.8695	0.1672	0.1584	0.1777
Reading Comprehension	0.6342	-0.3900	0.5576	0.0183	0.1345
Repairing	-0.1667	0.9256	-0.0937	-0.0771	0.1008
Science	0.3517	0.2019	0.6672	-0.0934	0.3817
Service Orientation	0.6463	-0.5112	-0.0445	0.1210	0.3044
Social Perceptiveness	0.7869	-0.4580	-0.0248	0.1532	0.1470
Speaking	0.7497	-0.5076	0.2397	0.0814	0.1162
Systems Analysis	0.6000	-0.0355	0.6794	0.2829	0.0972
Systems Evaluation	0.6440	-0.0337	0.6301	0.3261	0.0808
Technology Design	0.1768	0.2091	0.7686	0.1518	0.3113
Time Management	0.7265	-0.1352	0.1302	0.4550	0.2299
Troubleshooting	-0.1448	0.9547	0.0253	0.0195	0.0666
Writing	0.6538	-0.4173	0.5133	0.0150	0.1348

Table 13: PFA on level's score of skills

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
Active Learning	0.8709	-0.1722	0.3657	0.1418	0.0581
Active Listening	0.8547	-0.3507	0.2575	-0.0008	0.0803
Complex Programming	0.8270	-0.0434	0.4375	0.2104	0.0784
Coordination	0.8231	-0.0825	-0.0633	0.3652	0.1783
Critical Thinking	0.8661	-0.1861	0.3501	0.1074	0.0810
Equipment Selection	-0.2291	0.9204	-0.0963	-0.0836	0.0841
Equipment Maintenance	-0.1681	0.9195	0.1099	-0.0532	0.1113
Installation	0.0179	0.6826	0.1863	-0.2156	0.4526
Instructing	0.9003	-0.0045	0.1265	0.0852	0.1662
Judgment and Decision Making	0.8605	-0.1198	0.3397	0.2075	0.0867
Learning Strategies	0.8829	-0.0891	0.1702	0.1016	0.1733
Management of Financial Resources	0.5030	-0.0445	0.1537	0.8008	0.0801
Management of Personnel Resources	0.7894	0.0506	0.0056	0.5165	0.1075
Mathematics	0.5656	0.0854	0.5894	0.2649	0.2552
Monitoring	0.8824	-0.0108	0.1701	0.2551	0.1272
Negotiation	0.7904	-0.3221	-0.0217	0.3551	0.1450
Operation and Control	-0.2271	0.8987	-0.1240	0.0394	0.1238
Operations Monitoring	0.5153	0.0317	0.5942	0.3766	0.2385
Operations Analysis	-0.0742	0.9180	0.0752	0.1042	0.1352
Persuasion	0.7903	-0.3370	0.0476	0.3144	0.1608
Programming	0.3210	0.0239	0.8287	0.0072	0.2096
Quality Control Analysis	-0.0455	0.8740	0.2198	0.1797	0.1534
Reading Comprehension	0.8296	-0.2630	0.4011	-0.0123	0.0815
Repairing	-0.1925	0.9255	-0.0874	-0.0858	0.0914
Science	0.5206	0.2553	0.5637	-0.0720	0.3408
Service Orientation	0.7383	-0.4254	-0.0477	0.0977	0.2622
Social Perceptiveness	0.8527	-0.3670	-0.0688	0.1280	0.1171
Speaking	0.8713	-0.3546	0.2142	0.0600	0.0655
Systems Analysis	0.8218	0.0150	0.4186	0.2426	0.0904
Systems Evaluation	0.8359	-0.0005	0.4048	0.2681	0.0654
Technology Design	0.3624	0.3000	0.7403	0.1099	0.2185
Time Management	0.8335	-0.1075	0.0630	0.3976	0.1316
Troubleshooting	-0.1581	0.9565	0.0809	0.0293	0.0528
Writing	0.8236	-0.3293	0.3389	0.0357	0.0972

B Code of the construction of the database

Analysis of the intergenerational transmission of skills : Construction of the database (Fynch MEYNENT)

The aim of this notebook is to allow to create the database I used in my report about my analysis of the intergenerational transmission of skills. We will merge a part of a reworked O*NET Database with an extract of the NEPS database to obtain, for each child's outcome in skills, a level of skills from their parents estimated by their occupational skills.

Some of the files I used are on: https://github.com/jmeynent/TrinityResearch

Firstly, let's install and import the packages we need.

```
!pip install tslearn
!pip install h5py
!pip install kneed
!pip install openpyxl
!pip install xlrd
```

```
Requirement already satisfied: tslearn in
/opt/mamba/lib/python3.11/site-packages (0.6.1)
Requirement already satisfied: numpy in
/opt/mamba/lib/python3.11/site-packages (from tslearn) (1.24.4)
Requirement already satisfied: scipy in
/opt/mamba/lib/python3.11/site-packages (from tslearn) (1.11.1)
Requirement already satisfied: scikit-learn in
/opt/mamba/lib/python3.11/site-packages (from tslearn) (1.3.0)
Requirement already satisfied: numba in
/opt/mamba/lib/python3.11/site-packages (from tslearn) (0.57.1)
Requirement already satisfied: joblib in
/opt/mamba/lib/python3.11/site-packages (from tslearn) (1.3.0)
Requirement already satisfied: llvmlite<0.41,>=0.40.0dev0 in
/opt/mamba/lib/python3.11/site-packages (from numba->tslearn) (0.40.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/mamba/lib/python3.11/site-packages (from scikit-learn->tslearn)
(3.2.0)
Requirement already satisfied: h5py in /opt/mamba/lib/python3.11/site-
packages (3.9.0)
Requirement already satisfied: numpy>=1.17.3 in
/opt/mamba/lib/python3.11/site-packages (from h5py) (1.24.4)
Requirement already satisfied: kneed in
/opt/mamba/lib/python3.11/site-packages (0.8.5)
Requirement already satisfied: numpy>=1.14.2 in
/opt/mamba/lib/python3.11/site-packages (from kneed) (1.24.4)
Requirement already satisfied: scipy>=1.0.0 in
```

```
/opt/mamba/lib/python3.11/site-packages (from kneed) (1.11.1)
Collecting openpyxl
  Downloading openpyxl-3.1.2-py2.py3-none-any.whl (249 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 250.0/250.0 kB 5.0 MB/s eta
0:00:00a 0:00:01
lfile (from openpyxl)
  Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
Installing collected packages: et-xmlfile, openpyxl
Successfully installed et-xmlfile-1.1.0 openpyxl-3.1.2
Requirement already satisfied: xlrd in /opt/mamba/lib/python3.11/site-
packages (2.0.1)

import numpy as np
import pandas as pd
```

It's possible to find the O*NET Skills database in the following link. You have to download it in the excel format. It's also in the GitHub link I mentioned above.
https://www.onetcenter.org/dictionary/27.3/excel/skills.htm

However, the NEPS database is not in free access. If you find one, remind to convert it in an excel format to be sure that it will work.

```
#Importation of the databases
onet = pd.read_excel('Skills.xlsx')  #Replace Skills if the name is
different
neps = pd.read_excel('insert_name.xls') #Replace insert_name if the
name is different, and adapt the format
```

Construction of the occupational skills score

```
onet #Just to have a view on the database. It's not necessary to
construct.

      O*NET-SOC Code                              Title Element ID  \
0         11-1011.00                   Chief Executives    2.A.1.a
1         11-1011.00                   Chief Executives    2.A.1.a
2         11-1011.00                   Chief Executives    2.A.1.b
3         11-1011.00                   Chief Executives    2.A.1.b
4         11-1011.00                   Chief Executives    2.A.1.c
...              ...                                ...        ...
61105     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.b
61106     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.c
61107     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.c
61108     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.d
61109     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.d

                            Element Name Scale ID  Scale Name  Data
Value  \
0                   Reading Comprehension       IM  Importance
```

```
4.12
1                    Reading Comprehension       LV       Level
4.75
2                         Active Listening       IM  Importance
4.12
3                         Active Listening       LV       Level
4.88
4                                  Writing       IM  Importance
4.00
...                                    ...      ...         ...
...
61105  Management of Financial Resources       LV       Level
1.12
61106   Management of Material Resources       IM  Importance
2.00
61107   Management of Material Resources       LV       Level
1.88
61108  Management of Personnel Resources       IM  Importance
2.88
61109  Management of Personnel Resources       LV       Level
2.75

         N  Standard Error  Lower CI Bound  Upper CI Bound Recommend
Suppress  \
0      8.0            0.13            3.88            4.37
N
1      8.0            0.16            4.43            5.07
N
2      8.0            0.13            3.88            4.37
N
3      8.0            0.23            4.43            5.32
N
4      8.0            0.00            4.00            4.00
N
...    ...             ...             ...             ...
...
61105  8.0            0.13            0.88            1.37
N
61106  8.0            0.00            2.00            2.00
N
61107  8.0            0.13            1.63            2.12
N
61108  8.0            0.13            2.63            3.12
N
61109  8.0            0.16            2.43            3.07
N

      Not Relevant     Date Domain Source
0              NaN  07/2014       Analyst
```

```
1                 N  07/2014       Analyst
2               NaN  07/2014       Analyst
3                 N  07/2014       Analyst
4               NaN  07/2014       Analyst
...             ...      ...           ...
61105             N  08/2019       Analyst
61106           NaN  08/2019       Analyst
61107             N  08/2019       Analyst
61108           NaN  08/2019       Analyst
61109             N  08/2019       Analyst

[61110 rows x 15 columns]
```

Creation of a crosswalk file

To work with the NEPS, we need to convert the O*NET-SOC Code into ISCO-08 Code. Firstly, we import the crosswalk files. We will convert O*NET SOC Code 2019 into SOC code 2018, then convert SOC code 2018 into SOC code 2010, and finally convert SOC code 2010 to ISCO08. Firstly, we import the crosswalk files (they are available on my github)

```
onetsoc18 = pd.read_excel('ONET_to_Soc18.xlsx')
soc18soc10 = pd.read_excel('Soc18_to_Soc10.xlsx')
soc10isco08 = pd.read_excel('Soc10_to_ISCO08.xls')
```

Then, we construct a crosswalk to obtain a crosswalk files from O*NET-Soc Code 2019 to ISco-08 Code.

```
onetsoc18.rename(columns={'O*NET-SOC 2019 Occupation Listings' :
'ONET19', 'Unnamed: 2' : 'SOC18'}, inplace=True) #Rename columns to
make it more easy to understand
onetsoc18.drop(['Unnamed: 1', 'Unnamed: 3'], axis=1, inplace=True)
#Delete useless columns
onetsoc18.drop([0,1,2], axis=0, inplace=True) #Delete useless rows
onetsoc18 #Our final crosswalk file from ONET-2019 to SOC-2018.

          ONET19    SOC18
3     11-1011.00  11-1011
4     11-1011.03  11-1011
5     11-1021.00  11-1021
6     11-1031.00  11-1031
7     11-2011.00  11-2011
...          ...      ...
1014  55-3014.00  55-3014
1015  55-3015.00  55-3015
1016  55-3016.00  55-3016
1017  55-3018.00  55-3018
1018  55-3019.00  55-3019

[1016 rows x 2 columns]
```

```
soc18soc10.rename(columns={'U.S. Bureau of Labor Statistics' :
'SOC10', 'Unnamed: 2' : 'SOC18'}, inplace=True) #Rename
soc18soc10.drop(['Unnamed: 1', 'Unnamed: 3'], axis=1, inplace=True)
#Delete useless columns
soc18soc10.drop([0,1,2,3,4,5,6,7], axis=0, inplace=True) #Delete
useless rows
soc18soc10

       SOC10    SOC18
8    11-1011  11-1011
9    11-1021  11-1021
10   11-1031  11-1031
11   11-2011  11-2011
12   11-2021  11-2021
..       ...      ...
903  55-3015  55-3015
904  55-3016  55-3016
905  55-3017  17-3029
906  55-3018  55-3018
907  55-3019  55-3019

[900 rows x 2 columns]

soc10isco08.rename(columns={'soc10' : 'SOC10'}, inplace=True)
soc10isco08

       SOC10  isco08
0     111011    1112
1     111011    1113
2     111011    1120
3     111021    1112
4     111021    1114
...      ...     ...
1126  553015     310
1127  553016     310
1128  553017     310
1129  553018     310
1130  553019     310

[1131 rows x 2 columns]

#We convert SOC10 column of one of the crosswalk to allow a merge with
the other
soc18soc10['SOC10'] = soc18soc10['SOC10'].replace({'-': ''},
regex=True)

crosswalk = onetsoc18.merge(soc18soc10, on='SOC18') #We obtain the
crosswalk ONET19 to SOC10
crosswalk['SOC10'] =  crosswalk['SOC10'].astype(int) #We convert the
type of the column to make the joint possible
```

```
crosswalk = crosswalk.merge(soc10isco08, on='SOC10') #We obtain the
crosswalk ONET19 to ISC1008

#We delete the useless columns
crosswalk.drop(['SOC18', 'SOC10'], axis=1, inplace=True)

#we give a size of 4 characters on the ISCO code to allow future
joints.
crosswalk['isco08'] = crosswalk['isco08'].astype(str).str.zfill(4)
crosswalk

          ONET19 isco08
0     11-1011.00   1112
1     11-1011.00   1113
2     11-1011.00   1120
3     11-1011.03   1112
4     11-1011.03   1113
...          ...    ...
1595  55-3014.00   0310
1596  55-3015.00   0310
1597  55-3016.00   0310
1598  55-3018.00   0310
1599  55-3019.00   0310

[1600 rows x 2 columns]
```

We have our crosswalk file. Now, we had it to database

```
onet = onet.merge(crosswalk, left_on='O*NET-SOC Code',
right_on='ONET19', how='left')
onet

      O*NET-SOC Code                              Title Element ID  \
0         11-1011.00                   Chief Executives    2.A.1.a
1         11-1011.00                   Chief Executives    2.A.1.a
2         11-1011.00                   Chief Executives    2.A.1.a
3         11-1011.00                   Chief Executives    2.A.1.a
4         11-1011.00                   Chief Executives    2.A.1.a
...              ...                                ...        ...
89665     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.b
89666     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.c
89667     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.c
89668     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.d
89669     53-7121.00  Tank Car, Truck, and Ship Loaders    2.B.5.d

                            Element Name Scale ID  Scale Name  Data
Value  \
0                  Reading Comprehension       IM  Importance
4.12
1                  Reading Comprehension       IM  Importance
```

```
4.12
2                   Reading Comprehension       IM  Importance
4.12
3                   Reading Comprehension       LV       Level
4.75
4                   Reading Comprehension       LV       Level
4.75
...                                   ...      ...         ...
...
89665  Management of Financial Resources       LV       Level
1.12
89666   Management of Material Resources       IM  Importance
2.00
89667   Management of Material Resources       LV       Level
1.88
89668  Management of Personnel Resources       IM  Importance
2.88
89669  Management of Personnel Resources       LV       Level
2.75

         N  Standard Error  Lower CI Bound  Upper CI Bound Recommend
Suppress  \
0      8.0            0.13            3.88            4.37
N
1      8.0            0.13            3.88            4.37
N
2      8.0            0.13            3.88            4.37
N
3      8.0            0.16            4.43            5.07
N
4      8.0            0.16            4.43            5.07
N
...    ...             ...             ...             ...
...
89665  8.0            0.13            0.88            1.37
N
89666  8.0            0.00            2.00            2.00
N
89667  8.0            0.13            1.63            2.12
N
89668  8.0            0.13            2.63            3.12
N
89669  8.0            0.16            2.43            3.07
N

      Not Relevant     Date Domain Source      ONET19 isco08
0              NaN  07/2014       Analyst  11-1011.00   1112
1              NaN  07/2014       Analyst  11-1011.00   1113
2              NaN  07/2014       Analyst  11-1011.00   1120
```

```
3                    N  07/2014        Analyst  11-1011.00   1112
4                    N  07/2014        Analyst  11-1011.00   1113
...                ...      ...            ...         ...    ...
89665                N  08/2019        Analyst  53-7121.00   9333
89666              NaN  08/2019        Analyst  53-7121.00   9333
89667                N  08/2019        Analyst  53-7121.00   9333
89668              NaN  08/2019        Analyst  53-7121.00   9333
89669                N  08/2019        Analyst  53-7121.00   9333

[89670 rows x 17 columns]
```

Reshaping data

We have to separate datas indexed by level and those index by importance.

```
skill_im = onet[onet['Scale ID']=='IM']
skill_lvl = onet[onet['Scale ID']=='LV']
```

We will reshape the database in order to obtain, for each ISCO occupational level, the value of each kind of competence. Because there's less ISCO code than O*NET one, we aggregate the value by tooking the average.

```
pivot_lvl = skill_lvl.pivot_table(index='isco08', columns='Element
Name', values='Data Value', aggfunc='mean')
pivot_lvl #Just to see how does it look

Element Name  Active Learning  Active Listening  Complex Problem
Solving   \
isco08

0310                 3.375000          3.440000
3.310000
1112                 4.157500          4.220000
4.530000
1113                 4.315000          4.440000
4.560000
1114                 3.778889          3.887778
3.541111
1120                 4.083333          4.293333
4.290000
...                       ...               ...                      .
..
9621                 2.185000          3.000000
2.250000
9622                 2.873333          2.916667
2.666667
9623                 2.120000          2.940000
2.685000
9624                 2.125000          2.630000
```

```
2.315000
9629                 1.998000          2.776000
2.122000

Element Name  Coordination  Critical Thinking  Equipment
Maintenance  \
isco08

0310              3.060000           3.560000                  2.935000

1112              4.440000           4.312500                  0.000000

1113              4.500000           4.435000                  0.000000

1114              3.847778           4.068889                  0.347778

1120              4.293333           4.290000                  0.000000

...                    ...                ...                       ...

9621              2.565000           2.685000                  0.875000

9622              3.000000           2.996667                  3.460000

9623              2.370000           2.815000                  2.060000

9624              2.500000           2.565000                  2.370000

9629              2.450000           2.724000                  0.876000


Element Name  Equipment Selection  Installation  Instructing  \
isco08
0310                     2.935000         1.810     3.000000
1112                     0.217500         0.000     3.565000
1113                     0.435000         0.000     3.630000
1114                     0.527778         0.000     3.444444
1120                     0.290000         0.000     3.503333
...                           ...           ...          ...
9621                     0.310000         0.000     2.060000
9622                     2.753333         2.830     2.333333
9623                     1.620000         0.935     2.000000
9624                     1.750000         0.250     1.940000
9629                     0.524000         0.100     1.724000

Element Name  Judgment and Decision Making  ...   Science  \
isco08                                      ...
0310                              3.190000  ...  2.500000
1112                              4.437500  ...  1.435000
1113                              4.875000  ...  1.500000
1114                              3.680000  ...  1.164444
```

```
1120                              4.500000  ...  1.373333
...                                    ...  ...       ...
9621                              2.125000  ...  0.060000
9622                              2.833333  ...  1.080000
9623                              2.560000  ...  0.370000
9624                              2.120000  ...  0.125000
9629                              2.224000  ...  0.000000

Element Name  Service Orientation  Social Perceptiveness  Speaking  \
isco08
0310                     2.565000               2.940000  3.120000
1112                     3.687500               4.002500  4.312500
1113                     3.315000               4.065000  4.500000
1114                     3.234444               3.748889  3.998889
1120                     3.250000               4.043333  4.333333
...                           ...                    ...       ...
9621                     2.810000               2.380000  2.620000
9622                     2.210000               2.290000  2.793333
9623                     2.685000               2.310000  2.810000
9624                     2.125000               2.370000  2.250000
9629                     2.674000               2.526000  2.652000

Element Name  Systems Analysis  Systems Evaluation  Technology Design
\
isco08

0310                  3.000000            3.060000           1.940000

1112                  4.125000            4.060000           0.902500

1113                  4.690000            4.560000           0.935000

1114                  3.362222            3.402222           1.178889

1120                  4.126667            4.080000           0.996667

...                        ...                 ...                ...

9621                  1.940000            1.870000           0.315000

9622                  2.456667            2.376667           1.333333

9623                  2.315000            2.000000           0.310000

9624                  1.625000            1.500000           0.500000

9629                  1.074000            1.150000           0.174000


Element Name  Time Management  Troubleshooting   Writing
isco08
```

```
0310                     3.060000         3.190000  3.185000
1112                     4.065000         0.595000  4.190000
1113                     4.315000         0.000000  4.315000
1114                     3.568889         1.181111  3.902222
1120                     4.126667         0.460000  4.170000
...                           ...              ...       ...
9621                     2.500000         1.625000  2.435000
9622                     2.793333         3.540000  2.626667
9623                     2.685000         2.440000  2.625000
9624                     2.310000         2.315000  2.250000
9629                     2.172000         1.178000  2.198000

[420 rows x 35 columns]

#We do the same for importance level
pivot_imp =  skill_im.pivot_table(index='isco08', columns='Element
Name', values='Data Value', aggfunc='mean')
pivot_imp

Element Name  Active Learning  Active Listening  Complex Problem
Solving  \
isco08

0310                     3.435000          3.440000
3.250000
1112                     3.750000          4.030000
4.032500
1113                     3.875000          4.060000
4.190000
1114                     3.432222          3.876667
3.486667
1120                     3.750000          4.040000
3.960000
...                           ...               ...                      .
..
9621                     2.435000          3.435000
2.750000
9622                     2.953333          3.040000
2.953333
9623                     2.685000          2.940000
2.880000
9624                     2.310000          2.935000
2.380000
9629                     2.298000          3.174000
2.400000

Element Name  Coordination  Critical Thinking  Equipment
Maintenance  \
isco08
```

```
0310                 2.815000              3.435000                   3.065000

1112                 4.000000              4.095000                   1.000000

1113                 4.000000              4.250000                   1.000000

1114                 3.804444              3.891111                   1.193333

1120                 4.000000              4.126667                   1.000000

...                       ...                   ...                        ...

9621                 2.875000              3.000000                   1.685000

9622                 3.000000              3.206667                   3.373333

9623                 2.620000              2.940000                   2.690000

9624                 2.750000              2.815000                   2.315000

9629                 2.852000              2.752000                   1.600000

Element Name  Equipment Selection  Installation  Instructing  \
isco08
0310                     2.685000      2.000000     2.870000
1112                     1.060000      1.000000     3.310000
1113                     1.120000      1.000000     3.185000
1114                     1.362222      1.000000     3.264444
1120                     1.080000      1.000000     3.163333
...                           ...           ...          ...
9621                     1.310000      1.000000     2.190000
9622                     2.876667      2.793333     2.373333
9623                     2.060000      1.620000     2.190000
9624                     2.125000      1.190000     2.120000
9629                     1.322000      1.050000     2.024000

Element Name  Judgment and Decision Making  ...   Science  \
isco08                                      ...
0310                              3.185000  ...  2.565000
1112                              4.000000  ...  2.000000
1113                              4.190000  ...  2.000000
1114                              3.625556  ...  1.820000
1120                              3.960000  ...  1.960000
...                                    ...  ...       ...
9621                              2.815000  ...  1.060000
9622                              3.000000  ...  1.706667
9623                              2.750000  ...  1.370000
9624                              2.560000  ...  1.060000
9629                              2.576000  ...  1.000000
```

```
Element Name  Service Orientation  Social Perceptiveness  Speaking  \
isco08
0310                     2.620000               2.815000  3.185000
1112                     3.435000               3.970000  4.125000
1113                     3.185000               4.065000  4.190000
1114                     3.224444               3.611111  3.946667
1120                     3.206667               4.043333  4.126667
...                           ...                    ...       ...
9621                     3.375000               2.815000  3.435000
9622                     2.546667               2.750000  3.000000
9623                     2.940000               2.685000  2.940000
9624                     2.375000               2.500000  2.750000
9629                     3.150000               3.076000  3.324000

Element Name  Systems Analysis  Systems Evaluation  Technology Design
\
isco08

0310                  2.940000            3.000000           2.190000

1112                  3.687500            3.595000           1.782500

1113                  4.000000            4.000000           1.815000

1114                  3.277778            3.262222           1.834444

1120                  3.666667            3.666667           1.836667

...                        ...                 ...                ...

9621                  2.060000            2.000000           1.315000

9622                  2.540000            2.416667           1.873333

9623                  2.440000            2.375000           1.250000

9624                  1.880000            1.880000           1.375000

9629                  1.800000            1.900000           1.100000


Element Name  Time Management  Troubleshooting   Writing
isco08
0310                 3.000000         3.125000  3.185000
1112                 3.720000         1.470000  3.842500
1113                 3.690000         1.000000  4.060000
1114                 3.458889         1.805556  3.667778
1120                 3.710000         1.333333  3.790000
...                       ...              ...       ...
9621                 3.065000         1.940000  2.875000
9622                 2.876667         3.496667  2.666667
```

```
9623                     2.940000           2.875000  2.620000
9624                     2.560000           2.375000  2.380000
9629                     2.552000           1.876000  2.374000

[420 rows x 35 columns]
```

Principal Factor Analysis

The future objective is to make a PFA on the competences. We will do it with Stata. Therefore, we have to export it in dta format. We also conserve a copy in excel format

```
pivot_lvl.to_stata('level_pivot.dta')
pivot_lvl.to_excel('level_pivot.xlsx')
pivot_imp.to_stata('imp_pivot.dta')
pivot_imp.to_excel('imp_pivot.xlsx')
```

```
/tmp/ipykernel_425/553118153.py:1: InvalidColumnName:
Not all pandas column names were valid Stata variable names.
The following replacements have been made:

    Active Learning   ->   Active_Learning
    Active Listening   ->   Active_Listening
    Complex Problem Solving   ->   Complex_Problem_Solving
    Critical Thinking   ->   Critical_Thinking
    Equipment Maintenance   ->   Equipment_Maintenance
    Equipment Selection   ->   Equipment_Selection
    Judgment and Decision Making   ->   Judgment_and_Decision_Making
    Learning Strategies   ->   Learning_Strategies
    Management of Financial Resources   ->
Management_of_Financial_Resource
    Management of Material Resources   ->
Management_of_Material_Resources
    Management of Personnel Resources   ->
Management_of_Personnel_Resource
    Operation and Control   ->   Operation_and_Control
    Operations Analysis   ->   Operations_Analysis
    Operations Monitoring   ->   Operations_Monitoring
    Quality Control Analysis   ->   Quality_Control_Analysis
    Reading Comprehension   ->   Reading_Comprehension
    Service Orientation   ->   Service_Orientation
    Social Perceptiveness   ->   Social_Perceptiveness
    Systems Analysis   ->   Systems_Analysis
    Systems Evaluation   ->   Systems_Evaluation
    Technology Design   ->   Technology_Design
    Time Management   ->   Time_Management

If this is not what you expect, please make sure you have Stata-
compliant
column names in your DataFrame (strings only, max 32 characters, only
```

```
alphanumerics and underscores, no Stata reserved words)

  pivot_lvl.to_stata('level_pivot.dta')
/tmp/ipykernel_425/553118153.py:3: InvalidColumnName:
Not all pandas column names were valid Stata variable names.
The following replacements have been made:

    Active Learning   ->   Active_Learning
    Active Listening   ->   Active_Listening
    Complex Problem Solving   ->   Complex_Problem_Solving
    Critical Thinking   ->   Critical_Thinking
    Equipment Maintenance   ->   Equipment_Maintenance
    Equipment Selection   ->   Equipment_Selection
    Judgment and Decision Making   ->   Judgment_and_Decision_Making
    Learning Strategies   ->   Learning_Strategies
    Management of Financial Resources   ->
Management_of_Financial_Resource
    Management of Material Resources   ->
Management_of_Material_Resources
    Management of Personnel Resources   ->
Management_of_Personnel_Resource
    Operation and Control   ->   Operation_and_Control
    Operations Analysis   ->   Operations_Analysis
    Operations Monitoring   ->   Operations_Monitoring
    Quality Control Analysis   ->   Quality_Control_Analysis
    Reading Comprehension   ->   Reading_Comprehension
    Service Orientation   ->   Service_Orientation
    Social Perceptiveness   ->   Social_Perceptiveness
    Systems Analysis   ->   Systems_Analysis
    Systems Evaluation   ->   Systems_Evaluation
    Technology Design   ->   Technology_Design
    Time Management   ->   Time_Management

If this is not what you expect, please make sure you have Stata-
compliant
column names in your DataFrame (strings only, max 32 characters, only
alphanumerics and underscores, no Stata reserved words)

  pivot_imp.to_stata('imp_pivot.dta')
```

Then, run on Stata the PFA by using the PFA.do file available on the github I gave.

Finish the occupational skills database

```
#We import what we obtained from Stata
lvl = pd.read_excel('PFA_lvl.xls')
imp = pd.read_excel('PFA_imp.xls')

imp
```

```
     isco08  Active_Learning  Active_Listening
Complex_Problem_Solving  \
0       310         3.435000          3.440000
3.250000
1      1112         3.750000          4.030000
4.032500
2      1113         3.875000          4.060000
4.190000
3      1114         3.432222          3.876667
3.486667
4      1120         3.750000          4.040000
3.960000
..      ...              ...               ...                      ..
.
415    9621         2.435000          3.435000
2.750000
416    9622         2.953333          3.040000
2.953333
417    9623         2.685000          2.940000
2.880000
418    9624         2.310000          2.935000
2.380000
419    9629         2.298000          3.174000
2.400000

     Coordination  Critical_Thinking  Equipment_Maintenance  \
0        2.815000           3.435000               3.065000
1        4.000000           4.095000               1.000000
2        4.000000           4.250000               1.000000
3        3.804444           3.891111               1.193333
4        4.000000           4.126667               1.000000
..            ...                ...                    ...
415      2.875000           3.000000               1.685000
416      3.000000           3.206667               3.373333
417      2.620000           2.940000               2.690000
418      2.750000           2.815000               2.315000
419      2.852000           2.752000               1.600000

     Equipment_Selection  Installation  Instructing  ...
Systems_Analysis  \
0               2.685000      2.000000     2.870000  ...
2.940000
1               1.060000      1.000000     3.310000  ...
3.687500
2               1.120000      1.000000     3.185000  ...
4.000000
3               1.362222      1.000000     3.264444  ...
3.277778
4               1.080000      1.000000     3.163333  ...
3.666667
```

```
..                     ...           ...          ...  ...
...
415               1.310000      1.000000     2.190000  ...
2.060000
416               2.876667      2.793333     2.373333  ...
2.540000
417               2.060000      1.620000     2.190000  ...
2.440000
418               2.125000      1.190000     2.120000  ...
1.880000
419               1.322000      1.050000     2.024000  ...
1.800000

     Systems_Evaluation  Technology_Design  Time_Management
Troubleshooting  \
0              3.000000           2.190000         3.000000
3.125000
1              3.595000           1.782500         3.720000
1.470000
2              4.000000           1.815000         3.690000
1.000000
3              3.262222           1.834444         3.458889
1.805556
4              3.666667           1.836667         3.710000
1.333333
..                  ...                ...              ...
...
415            2.000000           1.315000         3.065000
1.940000
416            2.416667           1.873333         2.876667
3.496667
417            2.375000           1.250000         2.940000
2.875000
418            1.880000           1.375000         2.560000
2.375000
419            1.900000           1.100000         2.552000
1.876000

      Writing        f1        f2        f3        f4
0    3.185000  0.254217  1.677829  1.162194 -0.887223
1    3.842500  1.341973 -0.569583  0.192803  2.371664
2    4.060000  1.160001 -0.805663  0.654340  2.738459
3    3.667778  0.991501 -0.319348 -0.006640  1.456851
4    3.790000  1.115307 -0.635790  0.261679  2.741052
..        ...       ...       ...       ...       ...
415  2.875000 -0.781488 -0.570138 -0.953864 -0.232578
416  2.666667 -0.041227  2.122485  0.109372 -0.720609
417  2.620000 -0.869004  0.786314 -0.365761 -0.122075
418  2.380000 -1.300712  0.422395 -0.976066 -0.261673
419  2.374000 -0.956378 -0.664992 -1.328261 -0.868011
```

```
[420 rows x 40 columns]

lvl

     isco08  Active_Learning  Active_Listening
Complex_Problem_Solving  \
0       310         3.375000          3.440000
3.310000
1      1112         4.157500          4.220000
4.530000
2      1113         4.315000          4.440000
4.560000
3      1114         3.778889          3.887778
3.541111
4      1120         4.083333          4.293333
4.290000
..      ...              ...               ...                         ..
.
415    9621         2.185000          3.000000
2.250000
416    9622         2.873333          2.916667
2.666667
417    9623         2.120000          2.940000
2.685000
418    9624         2.125000          2.630000
2.315000
419    9629         1.998000          2.776000
2.122000

     Coordination  Critical_Thinking  Equipment_Maintenance  \
0        3.060000           3.560000               2.935000
1        4.440000           4.312500               0.000000
2        4.500000           4.435000               0.000000
3        3.847778           4.068889               0.347778
4        4.293333           4.290000               0.000000
..            ...                ...                    ...
415      2.565000           2.685000               0.875000
416      3.000000           2.996667               3.460000
417      2.370000           2.815000               2.060000
418      2.500000           2.565000               2.370000
419      2.450000           2.724000               0.876000

     Equipment_Selection  Installation  Instructing  ...
Systems_Analysis  \
0               2.935000         1.810     3.000000  ...
3.000000
1               0.217500         0.000     3.565000  ...
4.125000
2               0.435000         0.000     3.630000  ...
```

```
4.690000
3               0.527778         0.000     3.444444  ...
3.362222
4               0.290000         0.000     3.503333  ...
4.126667
..                   ...           ...          ...  ...
...
415             0.310000         0.000     2.060000  ...
1.940000
416             2.753333         2.830     2.333333  ...
2.456667
417             1.620000         0.935     2.000000  ...
2.315000
418             1.750000         0.250     1.940000  ...
1.625000
419             0.524000         0.100     1.724000  ...
1.074000

     Systems_Evaluation  Technology_Design  Time_Management
Troubleshooting  \
0              3.060000           1.940000         3.060000
3.190000
1              4.060000           0.902500         4.065000
0.595000
2              4.560000           0.935000         4.315000
0.000000
3              3.402222           1.178889         3.568889
1.181111
4              4.080000           0.996667         4.126667
0.460000
..                  ...                ...              ...
...
415            1.870000           0.315000         2.500000
1.625000
416            2.376667           1.333333         2.793333
3.540000
417            2.000000           0.310000         2.685000
2.440000
418            1.500000           0.500000         2.310000
2.315000
419            1.150000           0.174000         2.172000
1.178000

      Writing        f1        f2        f3        f4
0    3.185000  0.634280  1.851304  0.645326 -1.073805
1    4.190000  1.603612 -0.448209 -0.434995  2.328412
2    4.315000  1.771290 -0.593715 -0.273366  2.884460
3    3.902222  0.953207 -0.302851 -0.288154  1.458356
4    4.170000  1.440997 -0.514948 -0.315793  2.697041
```

```
..        ...       ...       ...       ...       ...
415  2.435000 -1.308836 -0.560847 -0.379486 -0.170033
416  2.626667 -0.131781  2.100968  0.061713 -0.983449
417  2.625000 -0.820915  0.608166 -0.208435 -0.472548
418  2.250000 -1.292948  0.539814 -0.575147 -0.359435
419  2.198000 -1.440011 -0.709149 -0.608469 -0.626591

[420 rows x 40 columns]
```

Now, we will delete useless columns and rename useful ones. We also uniformize the format of isco08 code to allow merge.

```
imp["isco08"] = imp["isco08"].astype(str).str.zfill(4)

imp = imp[['isco08', 'f1', 'f2', 'f3', 'f4']] #We only conserve
principal factors
imp.rename(columns = {'f1' : 'IM_cog_deci_skills', 'f2':
'IM_manual_skills', 'f3' : 'IM_comp_skills', 'f4' :
'IM_manag_skills'}, inplace=True)
imp

/tmp/ipykernel_425/4176696089.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  imp.rename(columns = {'f1' : 'IM_cog_deci_skills', 'f2':
'IM_manual_skills', 'f3' : 'IM_comp_skills', 'f4' :
'IM_manag_skills'}, inplace=True)

    isco08  IM_cog_deci_skills  IM_manual_skills  IM_comp_skills  \
0     0310            0.254217          1.677829        1.162194
1     1112            1.341973         -0.569583        0.192803
2     1113            1.160001         -0.805663        0.654340
3     1114            0.991501         -0.319348       -0.006640
4     1120            1.115307         -0.635790        0.261679
..     ...                 ...               ...             ...
415   9621           -0.781488         -0.570138       -0.953864
416   9622           -0.041227          2.122485        0.109372
417   9623           -0.869004          0.786314       -0.365761
418   9624           -1.300712          0.422395       -0.976066
419   9629           -0.956378         -0.664992       -1.328261

     IM_manag_skills
0          -0.887223
1           2.371664
2           2.738459
3           1.456851
4           2.741052
```

```
..               ...
415        -0.232578
416        -0.720609
417        -0.122075
418        -0.261673
419        -0.868011

[420 rows x 5 columns]
```

```
#We do the same for level database
lvl["isco08"] = lvl["isco08"].astype(str).str.zfill(4)
lvl = lvl[['isco08', 'f1', 'f2', 'f3', 'f4']]
lvl.rename(columns = {'f1' : 'LV_cog_deci_skills', 'f2':
'LV_manual_skills', 'f3' : 'LV_comp_skills', 'f4' :
'LV_manag_skills'}, inplace=True)
lvl
```

```
/tmp/ipykernel_425/581462118.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  lvl.rename(columns = {'f1' : 'LV_cog_deci_skills', 'f2':
'LV_manual_skills', 'f3' : 'LV_comp_skills', 'f4' :
'LV_manag_skills'}, inplace=True)
```

```
    isco08  LV_cog_deci_skills  LV_manual_skills  LV_comp_skills  \
0     0310            0.634280          1.851304        0.645326
1     1112            1.603612         -0.448209       -0.434995
2     1113            1.771290         -0.593715       -0.273366
3     1114            0.953207         -0.302851       -0.288154
4     1120            1.440997         -0.514948       -0.315793
..     ...                 ...               ...             ...
415   9621           -1.308836         -0.560847       -0.379486
416   9622           -0.131781          2.100968        0.061713
417   9623           -0.820915          0.608166       -0.208435
418   9624           -1.292948          0.539814       -0.575147
419   9629           -1.440011         -0.709149       -0.608469

     LV_manag_skills
0          -1.073805
1           2.328412
2           2.884460
3           1.458356
4           2.697041
..               ...
415        -0.170033
416        -0.983449
417        -0.472548
```

```
418         -0.359435
419         -0.626591

[420 rows x 5 columns]

#we merge it to have level and importance in the same dataframe
job_skill = pd.merge(imp, lvl, on='isco08')
job_skill

    isco08  IM_cog_deci_skills  IM_manual_skills  IM_comp_skills  \
0     0310            0.254217          1.677829        1.162194
1     1112            1.341973         -0.569583        0.192803
2     1113            1.160001         -0.805663        0.654340
3     1114            0.991501         -0.319348       -0.006640
4     1120            1.115307         -0.635790        0.261679
..     ...                 ...               ...             ...
415   9621           -0.781488         -0.570138       -0.953864
416   9622           -0.041227          2.122485        0.109372
417   9623           -0.869004          0.786314       -0.365761
418   9624           -1.300712          0.422395       -0.976066
419   9629           -0.956378         -0.664992       -1.328261

     IM_manag_skills  LV_cog_deci_skills  LV_manual_skills
LV_comp_skills  \
0          -0.887223            0.634280          1.851304
0.645326
1           2.371664            1.603612         -0.448209        -
0.434995
2           2.738459            1.771290         -0.593715        -
0.273366
3           1.456851            0.953207         -0.302851        -
0.288154
4           2.741052            1.440997         -0.514948        -
0.315793
..               ...                 ...               ...
...
415        -0.232578           -1.308836         -0.560847        -
0.379486
416        -0.720609           -0.131781          2.100968
0.061713
417        -0.122075           -0.820915          0.608166        -
0.208435
418        -0.261673           -1.292948          0.539814        -
0.575147
419        -0.868011           -1.440011         -0.709149        -
0.608469

     LV_manag_skills
0          -1.073805
1           2.328412
```

```
2            2.884460
3            1.458356
4            2.697041
..                ...
415         -0.170033
416         -0.983449
417         -0.472548
418         -0.359435
419         -0.626591

[420 rows x 9 columns]
```

Work on the NEPS and joints

```
neps

         ID_t  PARENTS_EDUCATION  meduYRS  feduYRS  PARENTS_ISEI
mISEI  \
0     8054956                  1     16.0     16.0             1
79.489998
1     8054957                  1     18.0     16.0             1
73.910004
2     8054966                  1     16.0     13.0             1
70.500000
3     8054975                  1     13.0     13.0             1
24.530001
4     8054979                  1     15.0     15.0             1
31.080000
...       ...                ...      ...      ...           ...
...
3476  8069729                  1     13.0     13.0             1
71.550003
3477  8069741                  1     13.0     13.0             1
25.040001
3478  8069742                  1     18.0     18.0             1
82.410004
3479  8069744                  1     15.0     13.0             1
44.939999
3480  8069746                  1     18.0     13.0             1
85.410004

          fISEI  PARENTS_ISCO  mISCO_88  mISCO_08  ...
dgn7_sc3b  \
0     62.130001             1    2140.0    2100.0  ...
10
1     74.660004             1    2419.0    2432.0  ...  not
participated
2     54.549999             1    2446.0    2635.0  ...
2
3     24.530001             1    5122.0    5120.0  ...
```

```
4
4     51.560001            1    5141.0    5142.0  ...
4
...         ...          ...       ...       ...  ...               .
..
3476  36.349998            1    2460.0    2636.0  ...
NaN
3477  24.530001            1    5123.0    5131.0  ...  not
participated
3478  70.570000            1    2320.0    2330.0  ...
9
3479  28.480000            1    4115.0    4120.0  ...
8
3480  85.410004            1    2310.0    2310.0  ...
5

            dgn10_sc3b  COGNITIVE_MATH   man5_sc1          man7_sc1  \
0                   10             1.0  -.7984647           .1717753
1     not participated             1.0  -2.787582  not participated
2                    7             1.0  -.7828258          -.0418512
3                   10             1.0  -.0086852           .6254676
4     not participated             1.0  -2.226913          -2.087215
...                ...             ...        ...               ...
3476               NaN             1.0        NaN               NaN
3477  not participated             1.0  -1.813545  not participated
3478  not participated             1.0   1.903898           .3931048
3479                 7             1.0  -.9448132          -.6620925
3480                 9             1.0         .l          -.6670079

              man9_sc1 COGNITIVE_SCIENCE          scn6_sc1
scn8_sc1  \
0                -.6099               1.0           1.36372
1.85782
1     not participated               1.0          -1.81285           -
2.01072
2                .2241               1.0          -1.41709
-.72219
3                .5545               1.0          -.22023
-.20923
4     not participated               1.0            .95921
-.07942
...                ...               ...               ...
...
3476               NaN               1.0               NaN
NaN
3477  not participated               1.0  not participated  not
participated
3478            1.6124
1.0            .78287            .75734
```

```
3479             -.9287                 1.0           -.71679         -
2.01072
3480  not participated                  1.0
-.60608            .60735

              scn10_sc1
0                .72872
1     not participated
2               -.21941
3                .43187
4     not participated
...                 ...
3476                NaN
3477  not participated
3478  not participated
3479            -.34228
3480            -.70957

[3481 rows x 36 columns]

# Conversion of the type of column
neps['mISCO_08'] = neps['mISCO_08'].astype(str)
neps['fISCO_08'] = neps['fISCO_08'].astype(str)

# We make compatible the formats of ISCO code
neps['mISCO_08'] = neps['mISCO_08'].str[:4]
neps['fISCO_08'] = neps['fISCO_08'].str[:4]

     mISCO_08 fISCO_08
0        2100     3341
1        2432     2512
2        2635     3343
3        5120     5120
4        5142     1420
...       ...      ...
3476     2636     7411
3477     5131     5120
3478     2330     1213
3479     4120     5223
3480     2310     2310

[3481 rows x 2 columns]

job_skill['isco08'] = job_skill['isco08'].astype(str) #We convert the
type in the job_skill database to allow the merge

#We merge to have, for each children outcome, the skills of their
mothers.
merge_mother = pd.merge(neps, job_skill, left_on='mISCO_08', right_on
='isco08', how='left')
merge_mother
```

```
          ID_t  PARENTS_EDUCATION  meduYRS  feduYRS  PARENTS_ISEI
mISEI  \
0     8054956                  1     16.0     16.0             1
79.489998
1     8054957                  1     18.0     16.0             1
73.910004
2     8054966                  1     16.0     13.0             1
70.500000
3     8054975                  1     13.0     13.0             1
24.530001
4     8054979                  1     15.0     15.0             1
31.080000
...       ...                ...      ...      ...           ...
...
3476  8069729                  1     13.0     13.0             1
71.550003
3477  8069741                  1     13.0     13.0             1
25.040001
3478  8069742                  1     18.0     18.0             1
82.410004
3479  8069744                  1     15.0     13.0             1
44.939999
3480  8069746                  1     18.0     13.0             1
85.410004

          fISEI  PARENTS_ISCO  mISCO_88 mISCO_08  ...
scn10_sc1  \
0     62.130001             1    2140.0
2100  ...           .72872
1     74.660004             1    2419.0     2432  ...  not
participated
2     54.549999             1    2446.0     2635  ...
-.21941
3     24.530001             1    5122.0
5120  ...           .43187
4     51.560001             1    5141.0     5142  ...  not
participated
...         ...           ...       ...      ...  ...                ..
.
3476  36.349998             1    2460.0     2636  ...
NaN
3477  24.530001             1    5123.0     5131  ...  not
participated
3478  70.570000             1    2320.0     2330  ...  not
participated
3479  28.480000             1    4115.0     4120  ...
-.34228
3480  85.410004             1    2310.0     2310  ...
-.70957
```

```
      isco08  IM_cog_deci_skills  IM_manual_skills IM_comp_skills  \
0        NaN                 NaN               NaN            NaN
1       2432            1.086538         -1.342790      -0.162715
2       2635            1.977610         -0.879875      -0.261480
3       5120           -0.945854         -0.446603      -0.652797
4       5142           -0.550591         -0.947607      -0.679031
...      ...                 ...               ...            ...
3476    2636            1.452434         -1.079344      -0.519187
3477    5131           -0.965300         -0.972108      -1.180777
3478    2330            1.885723         -0.654637       0.068545
3479    4120           -0.388678         -1.396197      -0.423427
3480    2310            1.379999         -0.783826       0.781042

       IM_manag_skills LV_cog_deci_skills LV_manual_skills
LV_comp_skills  \
0                  NaN                NaN              NaN
NaN
1            -0.212655           1.025959        -1.340247       -
0.317215
2            -1.425886           2.018624        -0.896666       -
0.955235
3             1.873571          -0.795232        -0.545063       -
0.467504
4            -0.091265          -0.759187        -0.977561       -
0.316722
...                ...                ...              ...            .
..
3476          1.018556           1.650983        -0.903388       -
1.357376
3477          0.060780          -1.641216        -1.071598       -
0.385629
3478         -1.352704           1.493149        -0.831712       -
0.458447
3479         -0.413182          -0.497420        -1.373454       -
0.057375
3480         -1.714532           1.633452        -0.930233
0.357550

      LV_manag_skills
0                 NaN
1            0.052874
2           -1.785156
3            1.706339
4            0.035603
...               ...
3476         0.970316
3477         0.074008
3478        -1.263053
3479        -0.220687
```

```
3480       -1.940961

[3481 rows x 45 columns]
```

```
#We rename the wolumns in order to specify that we talk about mother's skills
merge_mother.rename(columns={ 'IM_cog_deci_skills' : 'm_IMCogDeciS', 'IM_manual_skills' : 'm_IMManualS',   'IM_comp_skills' : 'm_IMCoS',
      'IM_manag_skills' : 'm_IMManagS',     'LV_cog_deci_skills' : 'm_LVCogDeciS', 'LV_manual_skills' : 'm_LVManualS',   'LV_comp_skills' : 'm_LVCoS' ,   'LV_manag_skills' : 'm_LVManagS'}, inplace=True)
merge_mother
```

```
         ID_t  PARENTS_EDUCATION  meduYRS  feduYRS  PARENTS_ISEI
mISEI  \
0     8054956                  1     16.0     16.0             1
79.489998
1     8054957                  1     18.0     16.0             1
73.910004
2     8054966                  1     16.0     13.0             1
70.500000
3     8054975                  1     13.0     13.0             1
24.530001
4     8054979                  1     15.0     15.0             1
31.080000
...       ...                ...      ...      ...           ...
...
3476  8069729                  1     13.0     13.0             1
71.550003
3477  8069741                  1     13.0     13.0             1
25.040001
3478  8069742                  1     18.0     18.0             1
82.410004
3479  8069744                  1     15.0     13.0             1
44.939999
3480  8069746                  1     18.0     13.0             1
85.410004

          fISEI  PARENTS_ISCO  mISCO_88 mISCO_08  ...
scn10_sc1  \
0     62.130001             1    2140.0
2100  ...            .72872
1     74.660004             1    2419.0     2432  ...  not
participated
2     54.549999             1    2446.0     2635  ...
-.21941
3     24.530001             1    5122.0
5120  ...            .43187
4     51.560001             1    5141.0     5142  ...  not
participated
```

```
...          ...           ...        ...       ...  ...                ..
.
3476  36.349998             1    2460.0      2636  ...
NaN
3477  24.530001             1    5123.0      5131  ...  not
participated
3478  70.570000             1    2320.0      2330  ...  not
participated
3479  28.480000             1    4115.0      4120  ...
-.34228
3480  85.410004             1    2310.0      2310  ...
-.70957

     isco08  m_IMCogDeciS  m_IMManualS   m_IMCoS  m_IMManagS
m_LVCogDeciS  \
0       NaN           NaN          NaN       NaN         NaN
NaN
1      2432      1.086538    -1.342790 -0.162715   -0.212655
1.025959
2      2635      1.977610    -0.879875 -0.261480   -1.425886
2.018624
3      5120     -0.945854    -0.446603 -0.652797    1.873571     -
0.795232
4      5142     -0.550591    -0.947607 -0.679031   -0.091265     -
0.759187
...     ...           ...          ...       ...         ...
...
3476   2636      1.452434    -1.079344 -0.519187    1.018556
1.650983
3477   5131     -0.965300    -0.972108 -1.180777    0.060780     -
1.641216
3478   2330      1.885723    -0.654637  0.068545   -1.352704
1.493149
3479   4120     -0.388678    -1.396197 -0.423427   -0.413182     -
0.497420
3480   2310      1.379999    -0.783826  0.781042   -1.714532
1.633452

     m_LVManualS   m_LVCoS m_LVManagS
0            NaN       NaN        NaN
1      -1.340247 -0.317215   0.052874
2      -0.896666 -0.955235  -1.785156
3      -0.545063 -0.467504   1.706339
4      -0.977561 -0.316722   0.035603
...          ...       ...        ...
3476   -0.903388 -1.357376   0.970316
3477   -1.071598 -0.385629   0.074008
3478   -0.831712 -0.458447  -1.263053
3479   -1.373454 -0.057375  -0.220687
```

```
3480   -0.930233  0.357550  -1.940961

[3481 rows x 45 columns]

#We add father characteristics for each children
father_merge = pd.merge(merge_mother, job_skill, left_on='fISCO_08',
right_on='isco08', how='left')
father_merge

         ID_t  PARENTS_EDUCATION  meduYRS  feduYRS  PARENTS_ISEI
mISEI  \
0     8054956                  1     16.0     16.0             1
79.489998
1     8054957                  1     18.0     16.0             1
73.910004
2     8054966                  1     16.0     13.0             1
70.500000
3     8054975                  1     13.0     13.0             1
24.530001
4     8054979                  1     15.0     15.0             1
31.080000
...       ...                ...      ...      ...           ...
...
3476  8069729                  1     13.0     13.0             1
71.550003
3477  8069741                  1     13.0     13.0             1
25.040001
3478  8069742                  1     18.0     18.0             1
82.410004
3479  8069744                  1     15.0     13.0             1
44.939999
3480  8069746                  1     18.0     13.0             1
85.410004

          fISEI  PARENTS_ISCO  mISCO_88 mISCO_08  ...  m_LVManagS
isco08_y  \
0     62.130001             1    2140.0     2100  ...         NaN
3341
1     74.660004             1    2419.0     2432  ...    0.052874
2512
2     54.549999             1    2446.0     2635  ...   -1.785156
3343
3     24.530001             1    5122.0     5120  ...    1.706339
5120
4     51.560001             1    5141.0     5142  ...    0.035603
1420
...         ...           ...       ...      ...  ...         ...
...
3476  36.349998             1    2460.0     2636  ...    0.970316
7411
```

```
3477  24.530001             1    5123.0     5131  ...    0.074008
5120
3478  70.570000             1    2320.0     2330  ...   -1.263053
1213
3479  28.480000             1    4115.0     4120  ...   -0.220687
5223
3480  85.410004             1    2310.0     2310  ...   -1.940961
2310

      IM_cog_deci_skills  IM_manual_skills IM_comp_skills
IM_manag_skills  \
0               1.824293         -0.397477      -0.703958
1.517878
1              -0.583782          0.249249       3.276386         -
1.639028
2              -0.742728         -1.112078      -0.233284         -
1.156076
3              -0.945854         -0.446603      -0.652797
1.873571
4               1.025919         -0.296045      -0.523645
2.746239
...                  ...               ...            ...
...
3476            0.266288          1.937660       0.265476         -
0.121493
3477           -0.945854         -0.446603      -0.652797
1.873571
3478            0.987199         -0.322260       0.057986
1.295678
3479            0.286365         -0.630684      -0.938806         -
0.099073
3480            1.379999         -0.783826       0.781042         -
1.714532

     LV_cog_deci_skills LV_manual_skills LV_comp_skills
LV_manag_skills
0              0.898988        -0.549374      -0.755013
1.547815
1             -0.032607         0.195047       3.534941       -
1.681438
2             -0.666041        -1.104842       0.140653       -
1.438003
3             -0.795232        -0.545063      -0.467504
1.706339
4              0.780411        -0.357413      -0.400647
2.322201
...                 ...              ...            ...              ..
.
3476           0.484758         2.255369       0.121894       -
```

```
0.326734
3477          -0.795232        -0.545063      -0.467504
1.706339
3478           0.974806        -0.296031      -0.274092
1.350375
3479          -0.091834        -0.516989      -0.788416        -
0.317386
3480           1.633452        -0.930233       0.357550        -
1.940961

[3481 rows x 54 columns]
```

```
#We rename the columns in order to specify that we talk about father's
skills
father_merge.rename(columns={ 'IM_cog_deci_skills' : 'f_IMCogDeciS',
'IM_manual_skills' : 'f_IMManualS',   'IM_comp_skills' : 'f_IMCoS',
     'IM_manag_skills' : 'f_IMManagS',     'LV_cog_deci_skills' :
'f_LVCogDeciS', 'LV_manual_skills' : 'f_LVManualS',   'LV_comp_skills'
: 'f_LVCoS' ,    'LV_manag_skills' : 'f_LVManagS'}, inplace=True)
father_merge
```

```
         ID_t  PARENTS_EDUCATION  meduYRS  feduYRS  PARENTS_ISEI
mISEI  \
0     8054956                  1     16.0     16.0             1
79.489998
1     8054957                  1     18.0     16.0             1
73.910004
2     8054966                  1     16.0     13.0             1
70.500000
3     8054975                  1     13.0     13.0             1
24.530001
4     8054979                  1     15.0     15.0             1
31.080000
...       ...                ...      ...      ...           ...
...
3476  8069729                  1     13.0     13.0             1
71.550003
3477  8069741                  1     13.0     13.0             1
25.040001
3478  8069742                  1     18.0     18.0             1
82.410004
3479  8069744                  1     15.0     13.0             1
44.939999
3480  8069746                  1     18.0     13.0             1
85.410004

          fISEI  PARENTS_ISCO  mISCO_88 mISCO_08  ...  m_LVManagS
isco08_y  \
0     62.130001             1    2140.0     2100  ...         NaN
3341
```

```
1     74.660004             1    2419.0     2432  ...    0.052874
2512
2     54.549999             1    2446.0     2635  ...   -1.785156
3343
3     24.530001             1    5122.0     5120  ...    1.706339
5120
4     51.560001             1    5141.0     5142  ...    0.035603
1420
...         ...           ...       ...      ...  ...         ...
...
3476  36.349998             1    2460.0     2636  ...    0.970316
7411
3477  24.530001             1    5123.0     5131  ...    0.074008
5120
3478  70.570000             1    2320.0     2330  ...   -1.263053
1213
3479  28.480000             1    4115.0     4120  ...   -0.220687
5223
3480  85.410004             1    2310.0     2310  ...   -1.940961
2310

      f_IMCogDeciS  f_IMManualS   f_IMCoS  f_IMManagS f_LVCogDeciS  \
0         1.824293    -0.397477 -0.703958    1.517878     0.898988
1        -0.583782     0.249249  3.276386   -1.639028    -0.032607
2        -0.742728    -1.112078 -0.233284   -1.156076    -0.666041
3        -0.945854    -0.446603 -0.652797    1.873571    -0.795232
4         1.025919    -0.296045 -0.523645    2.746239     0.780411
...            ...          ...       ...         ...          ...
3476      0.266288     1.937660  0.265476   -0.121493     0.484758
3477     -0.945854    -0.446603 -0.652797    1.873571    -0.795232
3478      0.987199    -0.322260  0.057986    1.295678     0.974806
3479      0.286365    -0.630684 -0.938806   -0.099073    -0.091834
3480      1.379999    -0.783826  0.781042   -1.714532     1.633452

     f_LVManualS   f_LVCoS f_LVManagS
0      -0.549374 -0.755013   1.547815
1       0.195047  3.534941  -1.681438
2      -1.104842  0.140653  -1.438003
3      -0.545063 -0.467504   1.706339
4      -0.357413 -0.400647   2.322201
...          ...       ...        ...
3476    2.255369  0.121894  -0.326734
3477   -0.545063 -0.467504   1.706339
3478   -0.296031 -0.274092   1.350375
3479   -0.516989 -0.788416  -0.317386
3480   -0.930233  0.357550  -1.940961

[3481 rows x 54 columns]

father_merge.columns
```

```
Index(['ID_t', 'PARENTS_EDUCATION', 'meduYRS', 'feduYRS',
'PARENTS_ISEI',
       'mISEI', 'fISEI', 'PARENTS_ISCO', 'mISCO_88', 'mISCO_08',
'fISCO_88',
       'fISCO_08', 'CHILD_SKILLS', 'COGNITIVE', 'cdn1_sc1',
'COGNITIVE_VOCAB',
       'von4_sc3', 'von6_sc3', 'von8_sc3', 'von10_sc3',
'COGNITIVE_DIGIT',
       'dsn40001_sc3a', 'dsn70001_sc3a', 'bdn80001_sc3a',
'COGNITIVE_BASIC',
       'dgn7_sc3a', 'dgn7_sc3b', 'dgn10_sc3b', 'COGNITIVE_MATH',
'man5_sc1',
       'man7_sc1', 'man9_sc1', 'COGNITIVE_SCIENCE', 'scn6_sc1',
'scn8_sc1',
       'scn10_sc1', 'isco08_x', 'm_IMCogDeciS', 'm_IMManualS',
'm_IMCoS',
       'm_IMManagS', 'm_LVCogDeciS', 'm_LVManualS', 'm_LVCoS',
'm_LVManagS',
       'isco08_y', 'f_IMCogDeciS', 'f_IMManualS', 'f_IMCoS',
'f_IMManagS',
       'f_LVCogDeciS', 'f_LVManualS', 'f_LVCoS', 'f_LVManagS'],
      dtype='object')

#We delete useless columns
df = father_merge.drop(['PARENTS_EDUCATION','PARENTS_ISEI',
'PARENTS_ISCO', 'mISCO_88', 'fISCO_88', 'CHILD_SKILLS',
'COGNITIVE','COGNITIVE_VOCAB', 'COGNITIVE_DIGIT',
'COGNITIVE_BASIC','COGNITIVE_MATH',
'COGNITIVE_SCIENCE','isco08_x','isco08_y'], axis=1, inplace=False)
df

         ID_t  meduYRS  feduYRS      mISEI      fISEI mISCO_08
fISCO_08  \
0     8054956     16.0     16.0  79.489998  62.130001     2100
3341
1     8054957     18.0     16.0  73.910004  74.660004     2432
2512
2     8054966     16.0     13.0  70.500000  54.549999     2635
3343
3     8054975     13.0     13.0  24.530001  24.530001     5120
5120
4     8054979     15.0     15.0  31.080000  51.560001     5142
1420
...       ...      ...      ...        ...        ...      ...       ..
.
3476  8069729     13.0     13.0  71.550003  36.349998     2636
7411
3477  8069741     13.0     13.0  25.040001  24.530001     5131
5120
3478  8069742     18.0     18.0  82.410004  70.570000     2330
```

```
1213
3479  8069744     15.0     13.0  44.939999  28.480000     4120
5223
3480  8069746     18.0     13.0  85.410004  85.410004     2310
2310

      cdn1_sc1                                            von4_sc3  \
0      2.01033                         Practice phase not passed
1           .l                         Practice phase not passed
2      -.36671                                                25
3      1.01754  No valid statement possible: no response in al...
4      2.22342                                                 3
...        ...                                               ...
3476    .17971                                               NaN
3477   -.79319                                  not participated
3478  -3.57539                                                37
3479   -.92494  No valid statement possible: no response in al...
3480       NaN                                                 1

                                               von6_sc3  ...   m_LVCoS
\
0                                                    93  ...       NaN

1                                                    34  ... -0.317215

2                                                    47  ... -0.955235

3                                                    71  ... -0.467504

4                                                    92  ... -0.316722

...                                                 ...  ...       ...

3476                                                NaN  ... -1.357376

3477                                   not participated  ... -0.385629

3478                                                 87  ... -0.458447

3479                                                 63  ... -0.057375

3480  No valid statement possible: no response in al...  ...  0.357550


     m_LVManagS f_IMCogDeciS f_IMManualS   f_IMCoS f_IMManagS
f_LVCogDeciS  \
0           NaN     1.824293   -0.397477 -0.703958   1.517878
0.898988
1      0.052874    -0.583782    0.249249  3.276386  -1.639028    -
0.032607
2     -1.785156    -0.742728   -1.112078 -0.233284  -1.156076    -
```

```
0.666041
3      1.706339    -0.945854   -0.446603 -0.652797   1.873571     -
0.795232
4      0.035603     1.025919   -0.296045 -0.523645   2.746239
0.780411
...         ...          ...         ...       ...         ...
...
3476   0.970316     0.266288    1.937660  0.265476  -0.121493
0.484758
3477   0.074008    -0.945854   -0.446603 -0.652797   1.873571     -
0.795232
3478  -1.263053     0.987199   -0.322260  0.057986   1.295678
0.974806
3479  -0.220687     0.286365   -0.630684 -0.938806  -0.099073     -
0.091834
3480  -1.940961     1.379999   -0.783826  0.781042  -1.714532
1.633452

     f_LVManualS   f_LVCoS f_LVManagS
0      -0.549374 -0.755013   1.547815
1       0.195047  3.534941  -1.681438
2      -1.104842  0.140653  -1.438003
3      -0.545063 -0.467504   1.706339
4      -0.357413 -0.400647   2.322201
...          ...       ...        ...
3476    2.255369  0.121894  -0.326734
3477   -0.545063 -0.467504   1.706339
3478   -0.296031 -0.274092   1.350375
3479   -0.516989 -0.788416  -0.317386
3480   -0.930233  0.357550  -1.940961

[3481 rows x 40 columns]

#We export the database 'I didn't share it on github because I'm not
allowed to
df.to_stata('base_def.dta')
df.to_excel('base_def.xlsx')
```