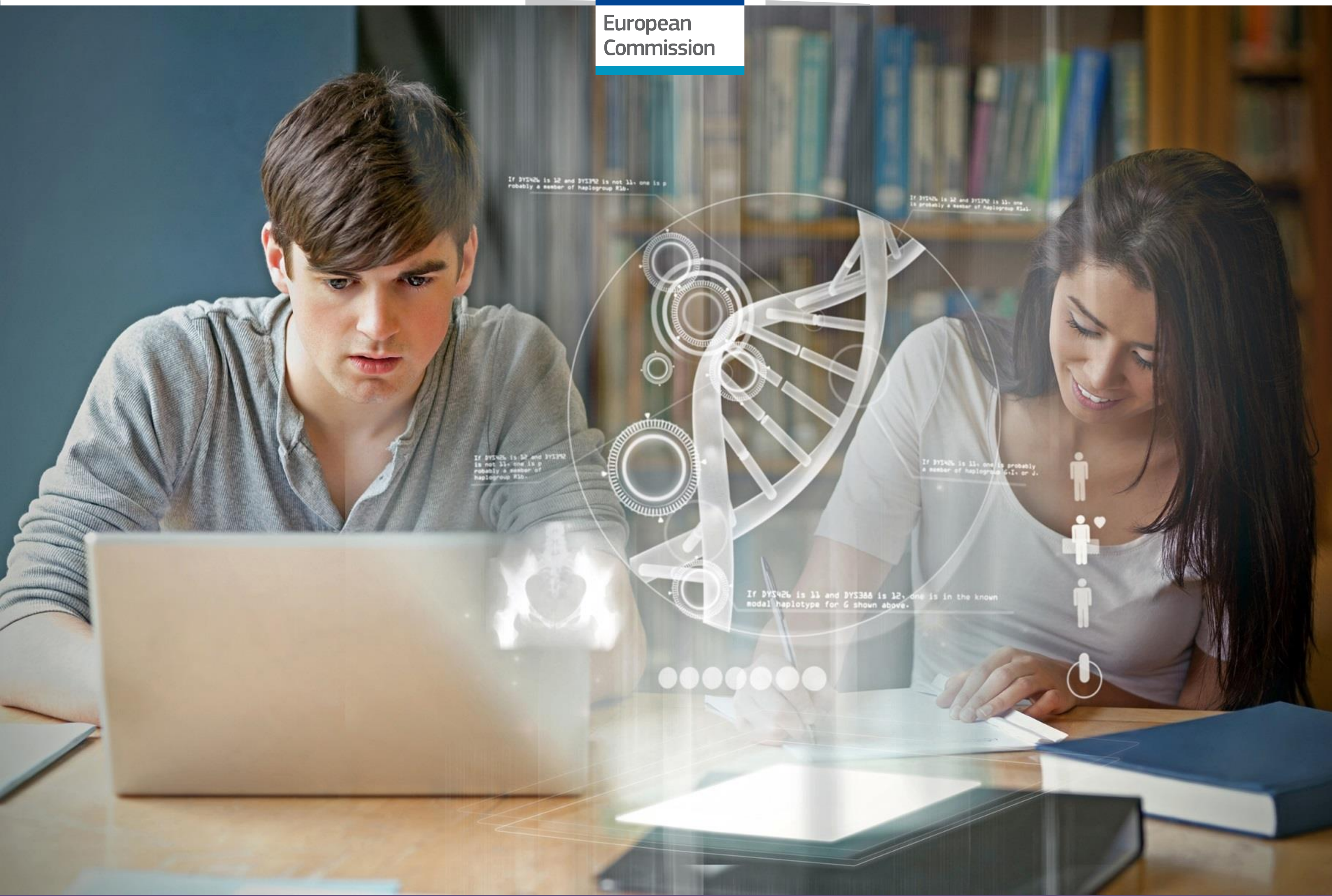
European Commission



Beyond achievement

A comparative look into 15-year-olds' school engagement, effort and perseverance in the European Union

Education and Training

Disclaimer

Acknowledgements

The authors wish to thank all those who in different ways contributed to this study. In particular, we wish to acknowledge the helpful advice, support and guidance of study's Steering Group of Commission officials, in particular Dr. Susanne von Below, Marco Montanari, Geir Ottestad and Oliver Heiden. We are also grateful to the members of the study's advisory board – Rolf Vegar Olsen, Hans-Peter Blossfeld and Sandra McNally – for their comments and suggestions, which were critically important in several phases of the study. We would also like to thank Francesco Avvisati, Francesca Borgonovi and Myako Ikeda of the OECD and all those who participated in the seminar that the authors held at the OECD in April 2018. The comments received and the opportunity to interact with the OECD researchers were very useful in refining the research. We are also thankful to Christian Monseur (University of Liège) and Elodie Pools (University of Liège) for their inestimable work in computing the indices based on log-files. Finally, we are also grateful to Daniela Biancardi, Sonia Marzadro, Letizia Monti, Chiara Munari and Enrico Rettore for their valuable input during the initial phases of the study.

EUROPEAN COMMISSION

Directorate-General for Education, Youth, Sport and Culture
Directorate A — Policy Strategy and Evaluation
Unit A4 – Evidence-Based Policy and Evaluation

E-mail: eac-unite-a4@*ec.europa.eu*

European Commission
B-1049 Brussels

EUROPEAN COMMISSION

Beyond achievement

A comparative look into 15-year-olds' school engagement, effort and perseverance in the European Union

Authors

Davide Azzolini, Nicola Bazoli, Ilaria Lievore, Antonio Schizzerotto, Loris Vergolini

Directorate-General for Education, Youth, Sport and Culture

2019 EN

Europe Direct is a service that answers your questions about the European Union.

- by freephone: **00 800 6 7 8 9 10 11**
(certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

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

Luxembourg: Publications Office of the European Union, 2019

EN PDF: ISBN: 978-92-79-93947-1 doi: 10.2766/98129 NC-06-18-138-EN-N

Table of Contents

Executive summary

Besides assuring the transmission of traditional competences – often proxied with standardised test scores and grades – schools are expected to contribute to young people's development of non-traditional competences, such as **engagement**, **effort** and **perseverance**.

Little is known about these competences. Their theoretical definition is still disputed; their operationalisation is challenging and their measurement is heavily affected by the lack of adequate data.

This underdevelopment is mostly due to the fact that interest in this topic is very recent. However, in the past few years, awareness of the **importance of non-traditional competences for life outcomes and for full participation in society** has experienced an unprecedented surge in both the scientific literature and the policy debate.

Consequently, the need to investigate this topic has become increasingly pressing. This study has the goal of contributing to this endeavour by providing some – necessarily incomplete – answers to the following questions.

How can computer-generated data help improve the measurement of these competences? How are European students performing in terms of non-traditional competences? Are there significant country differences in how students perform? What are the most important individual and school level determinants of these competences? How can education policy effectively intervene?

The study addresses these very salient questions following the publication of computer-generated data of the **Programme for International Student Assessment (PISA) 2015**. **Log-files** are **the traces that students leave on the computer when taking the test**. Log-files store a wealth of data, concerning students' behaviour during the test (such as response time, actions taken to solve a given task, etc.), making it possible to extract meaningful information on students' non-traditional competences.

This study examines three broad non-traditional competences, namely engagement, effort and perseverance. To measure them, focus is placed on five indices.

Student engagement is measured via two indices based on students' **self-reports**:

- **school engagement**: measured through a set of variables concerning skipping school days (truancy), school classes, arriving late at school;
- **science engagement**: measured via questions about the frequency with which students carry out science-related activities at home.

Effort and **perseverance** are measured by means of three **log-file** based indices:

- **effort**: defined as the difference in response time between difficult and easy items in the first part of the test;
- **effort persistence**: defined as the difference between effort in the second and in the first part of the test;
- **perseverance**: defined as the difference in performance in the second and in the first part of the test.

Given the early stage of educational research relying on the use of computer data to measure student competences more research is certainly needed in the near future to refine the proposed **operationalisation and measurement** procedures. Moreover, future studies should pay attention to other testing settings. PISA is a low-stakes exam and this feature may have important consequences on how students behave and hence on the interpretation of the results.

This study, therefore, does not come without limitations, both in the extent to which it manages to properly measure different – but interrelated – non-traditional competences and in the extent to which the empirical findings can be generalised to contexts other than a low-stakes exam setting. With these caveats in mind, the study provides some interesting findings that can contribute to the policy debate and future research development.

The main lessons learnt from the study can be summarised as follows.

1. **Non-traditional competences correlate positively with traditional ones. Yet this correlation is weak**, suggesting that the "traditional" and "non-traditional" indicators examined in this study capture, at least in part, different dimensions of students' ability.
2. There are **noticeable country differences** in non-traditional competences across the European Union. Top-performing countries on the standardised tests are not necessarily top-performing countries in terms of non-traditional competences.
3. **What matters more** for young people's development of non-traditional competences are **individual characteristics**. Parental education and immigrant background rank among the most important ascriptive factors shaping young people's non-traditional competences. The provision of extracurricular activities and a positive school climate produce slight differences in the development of students' competence.
4. The empirical evidence shows that **some students are more in need than others** of specific programmes and attention from practitioners and policy makers.
    - If the goal is enhancing young people's engagement in science, then priority should be assigned to girls, as an attempt to also enhance gender equality in STEMs education;
    - To redress deviant behaviours, like skipping school, the first group to be targeted is made up by boys and girls who have parents with low levels of education;
    - Concerning effort, special attention should be dedicated to boys of parents with low levels of education- including boys of non-immigrant families - who seem to be the most vulnerable group on this dimension;
    - In regard to students' effort persistence and perseverance, the findings suggest that children of immigrants are the most vulnerable group, perhaps because of their lower mastery of the test language, which imposes an additional cognitive load on them.
5. **Things can be changed.** Programme evaluation literature on the effectiveness of school programmes suggest that **schools can make a difference in students' non-traditional competences**, especially if interventions are carried out at younger ages. More research is needed to ascertain whether these effects are persistent and connected to other life outcomes and to yield further insights on "what works", and "how", to enhance young people's specific competences and to reduce social disparities.

1 Introduction

1.1 Background and motivation

In addition to educational performance, **non-traditional competences**[1] such as perseverance, effort and engagement are increasingly recognised as strong predictors of young people's educational attainment and future life outcomes. Beyond the so-called traditional competences – measured by standardised reading, mathematics or science achievement tests – a growing consensus has emerged on the need to assess and analyse non-traditional competences (Farkas 2003; Kautz et al. 2014). Non-traditional competences encompass a wide array of factors that range from psychological traits to individuals' capacity to deal with problems and to engage in their wider society. This study will focus on competences that are found to be more pertinent to the population of young people (i.e., 15-year olds) and on those that are more likely to foster educational achievement, i.e., engagement, effort and perseverance. The focus on these competences is worthwhile, as they have been found to be **associated with meaningful medium- and long-term outcomes** (from educational attainment, to labour market outcomes and other life outcomes) and have been found to be malleable. In other words, **they are important and they can be nurtured by external intervention**, such as school practices and policies.

The importance of providing **young people** with a complete set of competences that go beyond traditional ones and that enable them to face the challenges of globalisation and technological change has been recognised for a long time in Europe. For example, Recommendation 2006/962/EC on **key competences for lifelong learning** urged Member States to ensure that "initial education and training offers all young people the means to develop the key competences to a level that equips them for adult life." The Recommendation was updated in 2018 and lists eight "key" competences. Some of them are more "traditional" (literacy competence, multilingual competence, mathematical competence and competence in science, technology and engineering) while others are "less-traditional" (digital competence, personal, social and learning to learn competence, citizenship competence, entrepreneurship competence and cultural awareness and expression competence) but equally important to actively participate in a knowledge society.

1.2 Aims of the study and research questions

The overall objective of this study is to **provide empirical evidence to support policy making in the field of education** and young people's competence development. Even if not based on a causal analytical framework but on **descriptive analyses**, the report seeks to identify and partially redress gaps in research about the study of non-traditional competences. In particular, the results of the report are meant to contribute to a number of purposes: *i*) highlighting country differences within the European Union; *ii*) identifying the existence of robust associations between individual student characteristics, school factors and non-traditional competences; and *iii*) shedding light on the association between traditional and non-traditional competences. To reach these goals, the study investigates three main research questions.

[1] The classification of individual competences (or, skills) is far from having gained a consolidated agreement in the scientific community. By using the term "non-traditional" competences, this report has no ambition to introduce a new competence taxonomy. Instead, the term "non-traditional competences" is meant as a pragmatic and neutral classification, which includes all kinds of competences that go beyond those that have 'traditionally' been the object of assessment in education (e.g., grades and test scores). This classification is preferred over the more widely employed classification of cognitive and non-cognitive competences. The distinction of traditional from non-traditional competences only partially overlaps with the distinction between cognitive and non-cognitive competences: competences measured by grades or test scores also entail a non-cognitive component. Likewise, non-traditional competences like school engagement could also reflect students' cognitive competences.

The **first question** concerns a **comparison of the levels and variations** in non-traditional competences among 15-year-old students **across the European Union Member States**.

The **second question** seeks to shed light on the **main determinants of non-traditional competences**. The analyses investigate the main individual, school and school-system characteristics that come into play in shaping individuals' non-traditional competences. The study seeks not only to provide policy makers with the best possible evidence on what is required for young people's non-traditional competences, but also to identify the drivers that can realistically be changed. Hence, there will be a special focus on school-level resources and practices, which could be the subject of ad hoc programmes. That said, it must be acknowledged that the present study does not tackle the issue of causality, thus providing only correlational evidence regarding what works for students' non-traditional competences. A special zoom into specialised programme evaluation literature is also carried out and presented in the concluding section, with the aim of summarising the state of the art regarding what is effective in this field.

Third, the study seeks to provide insights into the **correlation between non-traditional and traditional competences**. Again, this analysis is carried out without any causal claims, but with the firm belief that it is surely worth assessing the extent to which these different competences are related, before and after controlling for a set of individual and school level characteristics.

1.3 Data and measures

A further element of interest in the study lies in the data utilized. The study exploits *Programme for International Student Assessment* (**PISA**) data from the 2015 wave. Since 2000, PISA has become a benchmark in international large-scale educational assessment. Beyond providing standardised scores of young people's reading, mathematics and science competences, the survey also collects a wealth of information concerning students' self-reported non-traditional skills and learning behaviours. The recent release of 2015 PISA data - delivered for the first time (in almost all participating countries) as a computer-based assessment - provides researchers with an unprecedented opportunity to investigate individuals' non-cognitive skills and learning strategies. The **computer-generated information** (log-files) are the digital traces that students leave behind when taking the test. For example, these traces can contain information about the response time and the action taken to solve a complex item. They provide new measures of individuals' non-cognitive skills as compared to those based on **self-reports**. The non-traditional competences considered in the study are:

- Engagement:
  - Current engagement in school;
  - Current science engagement;
- Effort (and effort persistence);
- Perseverance.

Current school engagement is measured via classic questionnaire-based indicators of school engagement (OECD, 2013) like **truancy** and **arriving late at school**. For current science engagement, students' **science activities outside school** are considered. The focus is on science primarily because science was the major domain assessed in PISA 2015. Its relevance in relation to the goals of the study lies in the fact that this indicator could capture signs of motivation for science and hence be a predictor of both future positive attitudes towards the scientific method and future educational and occupational careers in STEM-related fields. **Effort**, **effort persistence,** and **perseverance** are, instead, measured using the public log-files.

It is important to stress that PISA is a **low-stakes exam**. This implies, first, that the above-listed competences could be considered as sub-dimensions of the more general concept of intrinsic motivation, which can be defined as “the individual’s desire to perform the task for its own sake” (Benabou and Tirole 2003, 12). Second, it implies that the results of this study cannot be generalised to high-stakes contexts.

1.4 Report structure

The report is organised as follows. In section 2, the most relevant literature is reviewed, paying especial attention to non-traditional competences and the possible influence of factors at the individual, school and education system level. Section 3 describes the data used in the analyses as well as the outcomes and the main independent variables. Sections 4-7 illustrate the main empirical results for current engagement in school (section 4); current engagement in science (section 5); effort (section 6) and perseverance (section 7). These sections follow a similar structure. They start with a descriptive analysis at the macro level (cross-country comparison) and proceed by presenting the results of multiple regression models at the micro level. Moreover, for each section a special focus is devoted to various aspects deserving in-depth investigation. Section 8 summarizes the main results and presents a profiling analysis aimed at identifying the students who are more at risk of showing low levels of non-traditional competences.

2 Literature review

Key findings

Non-traditional competences encompass a wide range of factors that are often difficult to conceptualise and measure.

This study focuses on three non-traditional competences (i.e., engagement, effort and perseverance) that are considered to be both predictive of future life outcomes and that are found to be malleable by external intervention.

The empirical literature on the main determinants of non-traditional competences is reviewed, assigning especial attention to individual background characteristics, school factors and practices, and school-system characteristics.

In this section, the scientific literature regarding the definition (sub-section 2.1), the measurement (sub-section 2.2) and the main determinants (sub-section 2.3) of non-traditional competences is reviewed. Sub-section 2.4 is devoted to the issue of correlation between non-traditional and traditional competences.

2.1 Defining non-traditional competences

Non-traditional competences encompass a wide range of factors (from psychological traits to engagement in the society) that are often difficult to conceptualise, often overlap, and on whose operational definition the scientific community has not yet reached a high degree of consensus. Considering the purpose of this study, a broad and pragmatic definition of non-traditional competences is adopted. In this report, the non-traditional competences are defined as those **individual competences that exceed those that are traditionally the object of grading and testing at school** (e.g., grades and test scores).

To avoid confusion, it has to be stressed that **non-traditional competences do not completely overlap with non-cognitive competences**. In fact, non-traditional competences also include cognitive ones (e.g., strategic thinking, problem solving or digital competences). On the other hand, achievement test scores (like those made available by PISA) reflect not only individual cognitive ability but also non-cognitive competences such as motivation and effort. A student's performance on a test is always affected by the degree of commitment that she puts in it, independently from her ability (Borghans and Schils 2012).

Non-traditional competences can thus be framed as a multidimensional concept, including different aspects such as motivation, engagement, effort, perseverance, self-efficacy and self-concept. It should be noted that these non-traditional competences are related to the so-called "Big Five" personality traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism). At the same time, it is important to stress a key distinction between competences (or character skills) and personality traits: the latter are hard to change, while **competences are modifiable by external intervention**. This motivates the decision to study engagement, effort and perseverance - competences that could be changed by education programmes (Heckman and Kautz 2013) - in this report.

According to a recent classification, students' non-traditional competences can be divided into three categories: **achieving goals**, working with others and managing emotions (OECD 2015b). The three non-traditional competences considered in this study – engagement, effort and perseverance – fall into the first category. **Engagement** refers to students' commitment at school and includes a range of behaviours and attitudes such as taking additional courses on a given topic outside school obligations, participation in extracurricular activities, and willingness to help friends in studying a given subject (Christenson et al. 2012). **Effort** refers to students' constant exertion of mental energy in learning (Wise and Kong 2005). **Perseverance** (or endurance) refers to students' self-control and ability to withstand fatigue. In other words, it is the ability to maintain concentration during the entire test (Borgonovi and Biecek 2016).

Despite the inconclusive debate on their conceptual definition and practical operationalisation, researchers consistently find non-traditional competences to be powerful predictors of important life outcomes.

Focusing on **educational attainment**, motivation and engagement are found to be crucial components in students' learning outcome (Ritterfeld et al. 2009; OECD 2013a; Schunk et al. 2014; Gutman and Schoon 2013). Heckman and Kautz (2012; 2014) found that non-traditional competences are a better predictor of high-school graduation than cognitive competences measured via standardised tests.

Moreover, several studies indicate that non-cognitive competences have direct effects on **other life outcomes** such as wages, teen pregnancy, smoking, crime, performance in achievement tests, and many other aspects of social and economic life (Bowles et al. 2001; Borghans et al. 2008; Heckman et al. 2006). Of the Big Five factors, conscientiousness is the one most strongly associated with job performance. Non-cognitive competences are also important predictors of health and longevity. Conscientiousness, again, is the strongest predictor of mortality among the Big Five personality traits, even more important than IQ and socioeconomic status (Roberts et al. 2007).

2.2 Measurement of non-traditional competences

Although it is increasingly recognised that non-traditional competences cannot be adequately captured by traditional paper-based indicators (Kautz et al. 2014), **self-reported measures** are still the most used in the literature (Schunk et al. 2014). Self-reported information is usually collected through sets of items included in questionnaires that are then used to compute indices proxying latent factors via data reduction methods (e.g., factor analysis). It must be noted that measures based on self-reports are often a mix of behaviours and attitudes. For example, in Kirsch et al. (2003), students' reading engagement is measured with PISA 2000 data on the basis of students' answers to questions covering time spent on reading, interest in and attitudes towards reading, and diversity and content of reading. Likewise, students' perseverance s – a key trait that facilitates students' capability of maintaining performance throughout a test – is often measured through self-reporting (Sundre and Moore 2002; Thelk et al. 2009). However, self-reports are potentially affected by measurement error. They require evaluative capacity and may be affected by desirability, acquiescence and interpersonal comparisons (Borgonovi and Biecek 2016). An additional critical point – which may be particularly relevant in the context of a cross-national study – is the possible existence of a "cultural bias": items can be interpreted differently across countries and cultures and require more complex factorial models to realise international comparisons (Vandenberg and Lance 2000).

Obviously, this does not mean that researchers should abandon the use of self-reported indicators, but it is important to be aware of their limitations and to try to employ further approaches, such as the approach proposed in this report that exploits computer generated data.

Alternative ways to assess students' non-traditional competences in large-scale test settings are based on two main strategies: the exploitation of test rotation designs and log-files analysis. **Test rotation designs** are often applied in student assessment studies and imply that students are (randomly) allocated to different items and/or to a different item order (like in PISA). This design has been used, for example, by Borghans and Schils (2012) in a low-stakes test setting and by Albano (2013) in a high-stakes setting. Borghans and Schils (2012) exploit the fact that individual performance substantially decreases during the test and that this performance drop differs across students and countries. They find that the magnitude of the performance drop is related to personality traits, mainly agreeableness and motivational attitudes towards learning. The motivation effect can explain nearly one-fifth of the variation in the average test scores between countries. These studies find item-position effects on test performance, which can be explained by cognitive exhaustion, fatigue effects and students' decline in motivation and concentration (Albano 2013).

Borgonovi and Biecek (2016) exploit PISA 2012's random booklet and item allocation to study students' perseverance (or, as the authors term it, academic endurance), which is defined as students' ability to maintain their baseline rates of successful test completion for the duration of the test. The results obtained by Borgonovi and Biecek (2016) reveal differences in academic endurance between countries and within countries across different subgroups. The performance drop during the test is larger for boys and children from poorer socioeconomic backgrounds. Hence, different levels of endurance could be responsible for social disparities in achievement. Evidence also points to a larger performance drop in the reading test than in mathematics and science, and a larger performance drop in constructed responses than in multiple-choice responses, as the former require more engagement, self-control and organisation to construct a response.

Approaches that overcome students' awareness are a possible alternative to these measures. Particularly promising are those methods based on **log-files**, which are computer-generated traces of students' behaviour while taking a test. As will be described in greater detail below, log-files provide a wealth of objective measures of students' behaviour such as, for example, the time taken to complete a given task and the number and types of actions taken while sitting a test.

As mentioned above, computer-generated log-files have the potential to offer rich information on users' computer behaviours. In the settings of a computer-based assessment test (such as PISA 2015), computer generated data are thus not confined to students' performance on each specific test item, but extend to track students' behaviour during the test. Hence, log-files can contain traces of every single action taken while completing the test (Bunderson et al. 1988; Williamson et al. 2006; Greiff et al. 2016). It follows that the information contained in log-files can be informative in terms of students' cognitive competences and the metacognitive strategies they employ in taking the test (Brown 1987). More specifically, log-files allow us to investigate students' navigation patterns (Berendt and Brenstein 2001), providing a wealth of information such as the timing of actions, the sequences of the actions taken, pages visited, number of links clicked, and time to first action (Greene et al. 2010; Hadwin et al. 2007). Furthermore, log-files offer the opportunity to test traditional self-reported measures of non-traditional competences, and possibly to overcome some of the above-mentioned shortcomings.

Jamieson-Noel and Winne (2003) analyse the behaviour of a sample of undergraduate students who were administered a computer-based test. The authors find that software-produced traces of learning behaviours are not always in line with self-reported measures of self-regulated learning strategies. This can be interpreted either as evidence that measures based on self-reports are more affected by error or that the indices obtained through log-files measure different sets of individual competences as compared to those that can be measured via self-reports.

Despite their potential and the diffusion of computers in education, these log-files have been exploited for research purposes only recently and many of these studies are based on small samples (Hadwin et al. 2007; Sullivan et al. 2011; Scherer et al. 2017). Small samples make it very difficult to obtain statistically robust results, to infer conclusions for larger populations, or to investigate differences across groups. As pointed out by Greiff et al. (2015), this relatively poor exploitation of log-files might be due partially to the methodological difficulties surrounding their use, analysis and interpretation. The use of log-files requires technical skills of data cleaning and matching of different datasets, and their interpretation is possible only if a deep knowledge of the test is available. Moreover, it should also be considered that most of the available log-files, even the PISA ones, are a by-product of the tests, are not meant as data to be used for research purpose, and hence have limitations in terms of their informative content that have to be considered carefully.

Some studies have focused on students' self-regulation during their learning activities - i.e., the set of cognitive, affective and behavioural actions taken while carrying out the learning activity, such as planning their goals, self-monitoring their activities, and implementing other cognitive strategies (Chen and Ford 1998; Azevedo et al. 2004; Jeske et al. 2014). The authors explore the usability of log-files to make inferences regarding students' self-regulation of their learning during online tests. Among the variables used is the number of "backward and forward" jumps that occur out of sequence and time in the e-module.

Among the studies that have been published so far, some are based on PISA 2012 data, which provide log-files containing the data of digital reading and problem-solving behaviour of a representative sample of 15-year-old students in over 40 countries and economies (OECD 2015a). One of the most valuable advantages of this data - in relation to extant research on log-files - is its sample size and country coverage.

For example, Greiff et al. (2015) use PISA log-files from the computer-based assessment of problem-solving competences to study students' strategic thinking applied to solve the task, and find a strong and positive relationship between the so-called VOTAT (vary-one-thing-at-a-time) strategy and test performance. They also highlight the existence of different profiles of non-proficient studies: including those that applied VOTAT without solving the task and those that exhibited non-strategic behaviour. The authors found that these different mastery profiles differ across countries. The authors do not investigate individual and school-level determinants of the different strategies. Hence, on this point there is room for further research.

Time on task is another relevant piece of information that can be retrieved from log-files. However, its interpretation is not straightforward. On the one hand, spending more time may be positively related to the outcome as the task is completed more carefully and the test-taker is more motivated (Wise and Kong 2005). On the other hand, the relation may be negative if working more fluently, and thus faster, reflects higher skills (Lee and Chen 2011; Goldhammer et al. 2014). Furthermore, Goldhammer et al. (2014), using data from *Programme for the International Assessment of Adult Competencies* (PIAAC), find that the associated time on task and performance changes according to the type of competence tested. More precisely, they find that in problem solving tasks – which require controlled processing – time on task is found to have a positive effect and to increase with task difficulty. On the other hand, in reading tasks – which require routine processing – the time on task effect is found to be negative, and more negative, for easier tasks. Furthermore, in problem solving, the positive time on task effect is found to decrease with increasing skill level, while in reading, the negative time on task effect increases with increasing skills.

These heterogeneous effects suggest that time on task has no uniform interpretation but is a function of task difficulty and individual skills. Greiff et al. (2016) further investigate the time on task issue by using Finnish data on a sample of ninth-grade students who were administered a complex problem solving (CPS) computer-based test. The authors study the recorded time to task completion in relation to the adopted response strategy. They find that students who observed the problem environment in a non-interfering way – in addition to actively exploring it – showed better CPS performance, whereas students who showed a high frequency of (potentially unplanned) interventions exhibited worse CPS performance. In sum, both too much and too little time spent on a given test task can be associated with poor performance.

2.3 Main determinants of non-traditional competences

The second broad area of research, illustrated in section 1.2, concerns a detailed inquiry into the **determinants** of the three non-traditional competences mentioned. The main determinants of non-traditional competences are divided into three groups, according to the relevant level: individual background characteristics; school factors and practices; school-system characteristics.

Before entering into the details, it is worthwhile recalling that the aim of the study is to provide empirical evidence useful to **support policy development** and policy monitoring in the field of education. Hence, the empirical analysis focuses on factors that can realistically be modified by policy makers. It follows, first, that the highest priority is assigned to school-level factors that can be changed by education policy. Second, that the national education institutions are included in the analysis to provide the contextual framework, as well as to test the extent to which given associations vary across different contexts. Third, that a limited – but strongly theoretically and empirically motivated – set of individual ascriptive factors is considered with the purpose of shedding light on patterns of educational inequality and the existence of heterogeneity in the studied associations.

2.3.1 Individual-family level

At the individual and family level, the focus is placed on four 'classical' factors, widely employed in sociology and economics of education to assess the existence of disparities and **inequality in educational achievement**: gender, parental education, parental occupation, and immigrant background.

After the pioneering work of Coleman and colleagues (1966), a wealth of empirical studies has confirmed that **family background factors** like parental education and family income correlate with students' scholastic engagement and academic achievement (Carneiro and Heckman 2002; Cunha and Heckman 2007; OECD 2013a). Children from less privileged backgrounds and who belong to ethnic minorities exhibit higher risks of disruptive behaviours and less well-focused work habits as early as preschool years (Farkas 2003). **Gender** differences are also prominent. For example, the higher effort of girls at school is found to account for their higher educational attainment compared to boys, even after controlling for school achievement (Jacob 2002). Demars et al. (2013) find that boys show less test-taking motivation as well as lower response time than girls. Personality traits such as conscientiousness and agreeableness seem to partly account for these gender differences. Gender is also found to be among the most important predictors of science attitudes and interest, possibly as a consequence of cultural socialization (Osborne et al. 2003).

On the basis of the literature, it is possible to form some research expectations concerning the association between the aforementioned individual and family characteristics and the non-traditional competences studied here. Concerning gender, it seems reasonable to expect girls to be more engaged in school than boys in general, and therefore to be less likely to exhibit disruptive behaviours predictive of early school leaving (such as skipping classes and arriving late at school) and to put more effort into their studies. It has to be noted that indices based both on questionnaires and on log-files are considered and therefore discrepancies could arise between the different measures. At the same time, it must be remembered that, on average, female students underperform their male counterparts in science in most EU-28 countries (OECD 2016a)[2] and, given that PISA-2015 is focused mainly on science, a male advantage could emerge on some engagement indicators.

Concerning the role played by family socioeconomic and cultural capital (**parental occupation and education**), it can be posited that both contribute to the development of the three non-traditional competences (Coleman et al. 1966).

Finally, considering the children of **immigrants**, it is well known that they tend to underperform, on average, compared with their non-immigrant counterparts on standardised tests (OECD 2016a) and other educational outcomes (Heath et al. 2008). At the same time, research shows that immigrant parents and children of certain ethnic groups possess higher educational aspirations and expectations than those of their native counterparts (Kao and Tienda 1995). This disparity explains a consistent result emerging from recent empirical studies which find higher educational attainment among children of immigrants than among natives, once social background and prior school performance are accounted for (Kristen et al. 2008; Cebolla-Boado 2011; Jackson et al. 2012).[3]

2.3.2 School level

The levels of students' non-traditional competences, as well as their variation across the above-mentioned ascriptive attributes (gender, socioeconomic and immigration backgrounds), can be changed. Schools can provide settings that are more or less favourable to young people's development of non-traditional competences. Evaluation evidence (Kautz et al. 2014) points to the importance of some elements of success: *i)* early interventions at preschool or early-primary level to increase students' belief in the importance of commitment and effort; *ii)* high-quality initial education and continuing professional development for teachers; *iii)* increasing parental involvement in children's education; *iv)* assigning students active roles within the school (e.g., as tutors for 'disadvantaged' peers or younger students).

Provision of **extracurricular activities** (e.g., sports, art and drama clubs), especially if carried out at early ages, is often found to be related to students' development of non-cognitive traits and competences like self-esteem, self-efficacy, perseverance, discipline, ability to work in teams, and curiosity (OECD 2015, Covay and Carbonaro 2010; Winner et al. 2013; Farkas 2003; Carneiro and Heckman 2002; Howie et al. 2010).

A further well-established fact is that **teacher quality** is crucial not only in raising student achievement, but also for children's self-esteem, motivation and emotional stability (Chetty et al. 2011; OECD 2015). For example, Jackson (2012) finds that teacher quality has a causal impact on behaviours that are not measured by testing, such as absences, suspensions, marks, and grade progression. So far, no study has been able to disentangle the specific teacher characteristics that are

[2] More precisely, girls obtain significantly better results in comparison to boys only in Finland, Latvia, Slovenia, Greece and Bulgaria.

[3] Of course, children of immigrants are a very heterogeneous population across the different EU Member States. A more in-depth analysis is out of the scope of this report, but future studies should pay closer attention to this heterogeneity. A further interesting comparison would be between immigrant students and their counterparts in the country of origin, due to the self-selection occurring in the migration process.

conducive to students' development of traditional and non-traditional competences, but it is reasonable to assume that teachers' participation in continuing professional development and their involvement in schools' decision-making could indirectly lead to better student outcomes.

The education evaluation literature also points to the importance of programmes that **involve parents** for students' engagement and other educational outcomes (OECD 2015; Avvisati et al. 2010). In particular, experimental evidence has shown that simple programmes of parent-school meetings aimed at increasing parental involvement in their children's education have positive effects on students' school-related behaviour and attitudes, particularly reducing truancy and misbehaviour (Avvisati et al. 2013).

A further important aspect is connected with the extent to which schools promote **student involvement** and participation in school governance and classroom management. Student involvement in school processes can allow them to acquire non-traditional competences related to teamwork and negotiation, thus raising their self-efficacy and sense of responsibility (Taylor and Johnson 2002).

Last but not least, **school climate** is an additional factor that may affect students' engagement and motivation (OECD, 2013). School climate can be measured both via indicators related to students' behaviour as well as by the quality of student–teacher relationships. School climate and fair student–teacher relationships have also been found to foster citizenship competences (Geboers et al. 2013) and trust (Morris and Klesner 2010).

2.3.3 School-system level

Specific research questions regarding school systems have not been formulated and these systems are analysed in an explorative way. This choice is motivated by the fact that institutional features like tracking and the level of selectivity need complex political discussion at nation-state level.

As is evident from the literature, **educational institutions and settings** (e.g., the existence of tracking or the availability of high-quality childcare) are often found to be strongly associated with country differences in education performance (Wößmann 2000) and the magnitude of education inequality (Van de Werfhorst and Mijs 2010). Therefore, even though causality claims need to be carefully verified empirically, it seems reasonable to expect that given configurations of the school systems can not only enhance students' development of cognitive skills but also foster their non-traditional competences. Evidence supporting this statement comes from experimental studies assessing the impact of educational programs, mostly in the United States (Chetty et al. 2011; Heckman and Kautz, 2013). Education systems can be characterised by different degrees of **horizontal differentiation** such as: the number of available tracks (e.g., general vs vocational schools) and the age of selection into the different tracks. According to the available research, the more horizontally differentiated an education system is, the higher its social inequality in educational attainment will be (Wößmann 2009; Hanushek and Wößmann 2006; Brunello and Checchi 2007; Pekkarinen et al. 2012), while there is mixed evidence on whether this system increases or reduces labour market attainment, especially among students from disadvantaged social backgrounds (Brunello and Checchi 2007). The horizontal differentiation of a system has also been found to be negatively correlated with students' instrumental motivation to learn, even after accounting for performance levels (OECD 2013b, 86). Considering that instrumental motivation is an important factor in pursuing longer-term educational and occupational goals, finding which institutional features facilitate or obstruct its development becomes a very policy-relevant question.

In addition to horizontal differentiation, **vertical differentiation** is an important feature of school systems. Vertical differentiation can be defined as the ways in which students progress through the education system (OECD 2016b). The main indicator of this type of differentiation is the extent to which students of the same age are enrolled in different grade levels. The distribution of same-age students across different grades could be a consequence of both different school starting ages and different rates of grade repetition. While there is evidence suggesting that the older children are when they first start school, the lower their achievement levels and the wider social disparities (Deming and Dynarski 2008; OECD 2016b), the empirical evidence on the effect of grade repetition is more mixed (Ikeda and García 2014). However, it is often found that children from less advantaged backgrounds face a higher risk of grade retention, which can then turn into early school leaving (Agasisti and Cordero 2015).

A third important set of school-system characteristics concerns the **level of standardisation**, that is, the extent to which single schools are autonomous in decision making versus having to follow centrally determined rules. In this regard, typical indicators refer to schools' autonomy in terms of curriculum, in managing resources, and the existence of standardised exit exams (OECD 2016b, Below et al. 2013). Current research shows that school autonomy seems to exert beneficial effects on student achievement where accountability systems (e.g., external exit exams) are in place and properly implemented (Wößmann 2016).

Finally, as emerges from recent comparative reviews, most EU countries have legislation and **national/regional curricula** that promote social and transversal competences in the student population, thereby recognising social and emotional competences as being among the educational goals of schools (Eurydice 2012; OECD 2015). Among the more widely promoted competences are children's autonomy, responsibility and their ability to cooperate with others. Moreover, school curricula include subjects that are aimed at developing students' social and emotional competences, such as physical and health education or civic and citizenship education. However, few education systems provide detailed guidance on how to enhance social and emotional development (OECD 2015). Hence, it is likely that centrally determined frameworks are not always evenly implemented throughout the country, but rather leave ample autonomy to single schools and teachers. This is why, as illustrated in section 3.4, the study does not place great emphasis on these national-level policies and instead focuses more on school-level initiatives and practices.

2.4 Correlation between non-traditional and traditional competences

The third area of investigation, introduced in section 1.2, concerns a thorough analysis of the **correlation** between the three **non-traditional competences** of interest and the science, mathematics and reading competences.[4] This is achieved, first, by the means of bivariate analysis and, second, through the implementation of more sophisticated econometric analyses. Among the latter, multiple regression models are implemented to investigate the association between traditional and non-traditional competences while controlling for other factors such as family background and school factors.

Moreover, the relationship between traditional and non-traditional competences may differ across countries, schools and contexts, hence the analysis takes these sources of heterogeneity into account. It is not possible to assess any causal links for the issue of reverse causality. It nevertheless remains worthwhile to analyse the correlation between the various competences in order to understand ways in which they are related and how this relationship changes across countries.

[4] The report chiefly examines science because science is the major assessment domain in PISA 2015.

3 Data and measurement

Key findings

The 2015 computer-based assessment of PISA makes it possible to measure non-traditional competences by exploiting both students' self-reports and computer-generated log-files.

Self-reports are used to measure engagement in school and in science, while the log-files are used to measure effort, effort persistence and perseverance.

Effort is the difference in the response time in easy vs. difficult items at a given point of the test. Effort persistence is a measure of the extent to which students can keep up their effort and is measured as the difference between effort measured in the first part and in the second part of the test. Perseverance captures the effect of fatigue by comparing performance at different points of the test.

The analytical strategy followed in this report to identify the main individual and school determinants of non-traditional competences involves the estimation of a series of multiple regression models.

This section provides an overview of the data and the measurement procedures adopted to construct and validate the main variables employed in the analysis. First, sub-section 3.1 offers a brief overview of the data available in PISA 2015 and the country samples. Second, it provides a description of the operationalisation of the outcome variables. More precisely, sub-section 3.2 illustrates the questionnaire-based indices (i.e., engagement in school and engagement in science), while sub-section 3.3 gives an account of the procedures adopted to derive the effort and perseverance indices from the log-files. Finally, sub-section 3.4 describes the analytical strategy and the independent variables considered in the analyses.

3.1 The PISA data

The PISA 2015 survey focused on science, with reading and mathematics as minor areas of assessment.[5] Moreover, for the first time in PISA 2015, the main mode of assessment was computer-based (CBA) tests, instead of the traditional paper-based test (PBA). The CBA tests produce a set of digital traces (**log-files**) that can be exploited for educational research.

The study is based on the sample of 15-year-old students in the EU 28 Member States (N = 182,114). However, among the European Member States, two (Malta and Romania, as shown Table 3.1) administered the assessment using the paper-based format. These two countries did not administer some key questions concerning engagement either, therefore they cannot be included in any of the analyses presented in this report. Table 3.1 shows an overview of the assessment modes in the participating countries in the EU as well as the size of the samples of students involved. It includes only the EU-28 countries that are the focus of this report.

[5] PISA 2015 also included the assessment of collaborative problem solving (CPS) and financial literacy (FL), which are not considered for the study. More details on the PISA 2015 survey can be found at http://www.oecd.org/pisa/.

Table 3.1 Overview of the PISA 2015 assessment mode, data and sample size.

Countries	Code	Mode	Sample size
Austria	AT	CBA	7,007
Belgium	BE	CBA	9,651
Bulgaria	BG	CBA	5,928
Cyprus	CY	CBA	5,771
Czech Republic	CZ	CBA	6,894
Germany	DE	CBA	6,504
Denmark	DK	CBA	7,161
Spain	ES	CBA	6,736
Estonia	EE	CBA	5,587
Finland	FI	CBA	5,882
France	FR	CBA	6,108
United Kingdom	UK	CBA	14,157
Greece	EL	CBA	5,532
Croatia	HR	CBA	5,809
Hungary	HU	CBA	5,658
Ireland	IE	CBA	5,741
Italy	IT	CBA	11,583
Lithuania	LT	CBA	6,525
Luxembourg	LU	CBA	5,299
Latvia	LV	CBA	4,869
Malta	MT	**PBA**	3,634
Netherlands	NL	CBA	5,385
Poland	PL	CBA	4,478
Portugal	PT	CBA	7,325
Romania	RO	**PBA**	4,876
Slovak Republic	SK	CBA	6,350
Slovenia	SI	CBA	6,406
Sweden	SE	CBA	5,458

3.2 Questionnaire-based measures of student engagement

In this study, **school engagement** is understood as the extent to which students identify with and value schooling outcomes, and participate in academic and non-academic school activities (OECD, 2000). In this vein, two main aspects are covered in PISA-2015 data:

- *Current engagement in school*: questions about truancy and arriving late for school
- *Current science engagement*: questions on students' science activities outside school.

Current engagement is measured via classical indicators like **truancy** and **arriving late at school** (OECD 2013b). More specifically the following three indicators are considered:

- In the last two full weeks of school, how often did I skip a whole school day?
- In the last two full weeks of school, how often did I skip some classes?
- In the last two full weeks of school, how often did I arrive late for school?

The students' possible responses are "none", "one or two times", "three or four times" and "five or more times".[6]

To measure students' **current science engagement**, **science activities outside school** are examined. This dimension has been measured using nine items included in the student questionnaire, which have then been used to compute a unique index (**SCIEACT**):

- Watch TV programmes about broad science.
- Borrow or buy books on broad science topics.
- Visit websites about broad science topics.
- Read broad science magazines or science articles in newspapers.
- Attend a science club.
- Simulate natural phenomena in computer programmes/virtual labs.
- Simulate technical processes in computer programmes/virtual labs.
- Visit websites of ecology organisations.
- Follow news of science, environmental, or ecology organisations via blogs and microblogging.

To compute the SCIEACT index, the OECD-EDU team implemented a latent trait approach based on a Generalised Partial Credit Model that makes it possible to compute a unidimensional index. For more details, see the PISA 2015 technical report (OECD 2017).[7] Higher SCIEACT index values correspond to higher engagement levels, while lower values correspond to lower engagement levels.

Because the above-described indices are based on students' self-reports through the questionnaires, there is an unavoidable issue of missing values due to non-response. As shown in Appendix 3.1, the missing-value issue is more accentuated on the SCIEACT index, because it cumulates student non-responses across the nine items used to compute this index, while the amount of missingness for the other three indicators of school engagement is less problematic. Appendix 3.1 presents a set of statistical checks aimed at ascertaining the potential bias caused by the presence of missing values, which is more accentuated in some countries than in others.

3.3 Indices derived from the test log-files

Two further indices of non-traditional competences (i.e., **effort** and **perseverance**) are constructed exploiting two sources of information: the **log-files** and the PISA **rotation scheme** employed in the test administration.[8]

For each student taking the CBA test, the system recorded and stored a rich set of data describing all the actions taken in answering each test item. This information is vast and varies depending on the specific item type. For example, more complex and interactive tasks entail more information (e.g., clicks, browsing, corrections, etc.), while simple ones (e.g., multiple-choice questions) contain less information (e.g., response time and correctness of the single item). Of the total amount of digital traces, a subset was made publicly available alongside the main

6 The possibility of constructing an index applying factor analysis techniques has been explored, but it emerges that the latent construct cannot be compared across countries. Therefore, it has been decided to analyse the three items separately.

7 As described in the PISA 2015 technical report (OECD 2017, pg. 319), SCIEACT is a very reliable index (Cronbach's Alpha is close to 1.0 in almost all countries).

8 With the available data, many alternative indices of effort and perseverance can be computed. In the main text, only those that were considered more appropriate are analysed in this report. Alternative ways in which these indices can be operationalised are reported in Appendix A3.2 (Table A3.6).

PISA files. In this study, the publicly available log-files are used.[9] As will be explained in greater detail when describing the operationalisation of the indices of non-traditional competences based on the log-files, the log-file information utilised concerns time spent on answering a single item (i.e., milliseconds elapsed before sending the answer) and correctness of the item.

The PISA tests are administered following a so-called "item rotation scheme" (Figure 3.1). All developed items are grouped into assessment **clusters**, lasting around 30 minutes each. Then, one-hour **sessions** are formed containing different combinations of two clusters and two-hour **booklets** are the result of a combination of two sessions. Each student is randomly assigned to a given booklet; hence each student gets different combinations of test items and receives them in a different order. Students have a break between the two sessions. This random variation in the items and their order has been exploited in the literature to carry out analyses of the link between non-cognitive and cognitive competences (Borghans and Schils 2012; Borgonovi and Biecek 2016).

Figure 3.1 The PISA 2015 test design: Booklet, sessions and clusters.

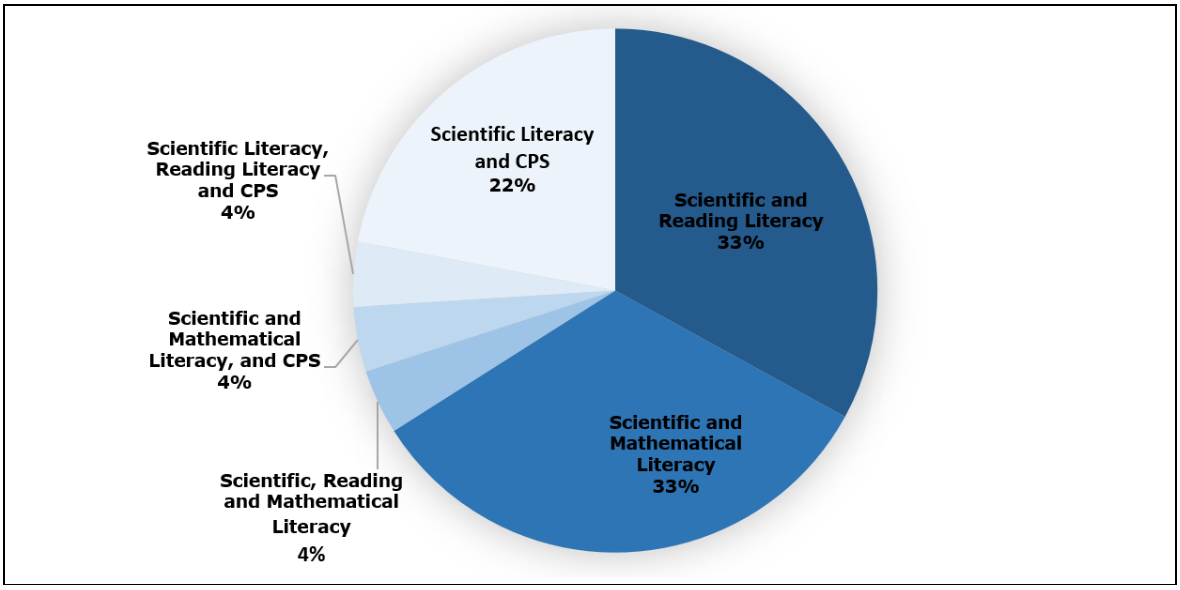
Each cluster is domain-specific (i.e., it covers only one of the assessed domains). While most students responded to items from two domains, a minority of students received a booklet assessing three different domains. The analyses carried out for this report are based on the exclusion not only of a small proportion of students who received the one-hour booklet (in some cases students with special needs), but also of students who received more than two domains in the two hours of testing. The reason for excluding the former group is that this group does not allow for the construction of comparable indices due to the limited duration of the test. The second group is dropped because it has been observed that receiving an additional domain

[9] In addition to these publicly available log-files, there also exist non-public log-files, which could not be used within the present study due to time constraints. Although these data are potentially very informative (see Greiff et al. 2015 for an example), their use requires complex and very time-consuming data processing operations. These data are stored in formats which cannot be immediately used for statistical purposes and a series of data-cleaning operations is needed before analysing them.

deteriorates students' overall performance. Indeed, the percentage of right answers is lower when the booklet contains three domains rather than two and hence alters their comparability with the rest of the sample.

Figure 3.2 shows how the total sample of students assessed via computer is split according to the different combination of domains. Between 66% and 88% of the students – depending on whether the country administered the Computer Problem Solving (CPS) module or not – are assessed on two domains (and hence will be included in our study on log-file based measures). Thanks to the random allocation of students to the different booklets, these students are a representative sample of the original sample and, hence, of the reference population of 15-year-olds in the countries considered.

Figure 3.2 Main survey computer-based assessment design for countries where the Computer Problem Solving (CPS) domain was administered.



Source: OECD 2017, page 40.

As mentioned above, Malta and Romania administered the paper-based version of the test and hence are excluded from the analyses. Moreover, for Cyprus only the effort index could be computed as the item score database was processed with a different statistical package, making it difficult to recover the type of missing item which is key to compute student performance.[10]

In total, from the remaining cases, 55,000 were lost due to "student-level deletions": students who were administered a booklet with 3 domains; students who were administered collaborative problem solving (CPS) clusters; and students who received the one-hour (UH) booklet. A further 2,000 cases were dropped for three reasons: the presence of two very easy clusters in Bulgaria, the presence of three special response format items, and the presence of some skipped/unreached items.[11]

[10] The database made available on the Cyprus national portal is a SPSS dataset and SPSS and SAS define missing values in different ways. In SAS, the software used for the operationalisation of effort and perseverance, along with general missing values, the user can distinguish different missing codes for different types of missing values (here, omission, not reached, invalid, not administered). In SPSS, all these missing values are embedded in a general category of a user-defined 'missing value'.

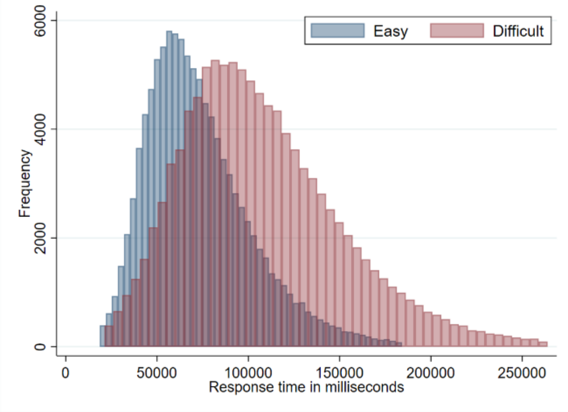
[11] The data of very easy clusters M06b and R06b administered in Bulgaria were not scaled and were therefore not used for constructing the perseverance index. Special response format items (i.e., R219Q01, R404Q10 and S641Q04) were also deleted. Moreover, the log-file indices could not be computed if a student did not answer all items of the cluster (or, in the case of too easy and/or too hard items).

3.3.1 Effort

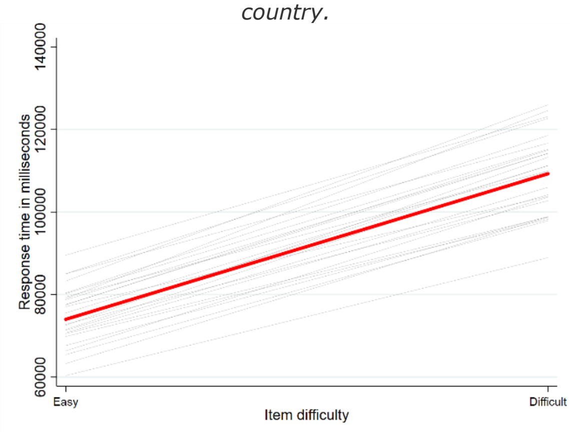
In this study, effort is understood as the activation of mental power to perform a task and, building upon previous literature (Wise and Kong 2005), this study measures this by exploiting information on response time. More precisely, the adopted measure of effort is the within-individual and within-cluster difference between the response time (RT, measured in milliseconds) when answering difficult items and when answering easy items. Because the index is constructed at the **individual** and the **cluster** level, it is possible to **overcome the two most important sources of bias** in empirical analyses of this kind. The first of these sources of bias is **student ability bias**, which occurs when comparing response time in easy vs. difficult items located in a given point of the test but across different students. The second source of bias is **item test positioning**, which becomes critical when comparing a given student's response time in easy vs. difficult items, which are located in different points of the test. The proposed method fixes student ability and item test positioning and allows only item difficulty to change.

Figure 3.3. Response time in easy and difficult items in EU Member States (cluster 1).

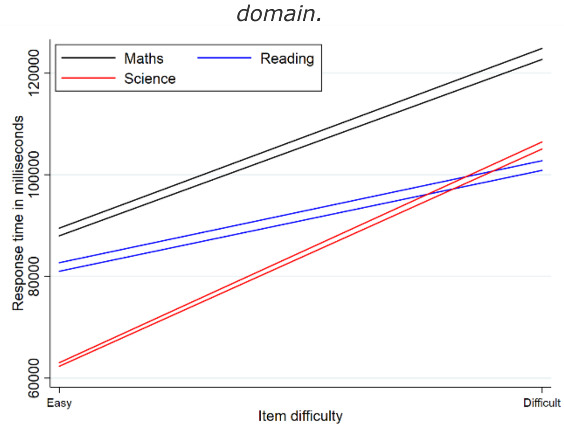
Panel a: *Response time distribution (all domains).*



Panel b*: Mean response time by item difficulty and country.*



Panel c*: Mean response time by item difficulty and domain.*



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. Extreme values (1st and 99th percentiles of response time are not included). In panel b, red line represents the average response time for EU Member States. In panel c, double lines show the 95% CI lower and upper bounds. The information is not available for Malta and Romania.

Item difficulty is established according to the item parameters from the **international IRT scaling** (OECD 2017). The main analyses are based on the entire set of items of the clusters, hence they do not distinguish multiple-choice from open-ended ones. In principle, this could introduce some noise in the analysis, as open-ended questions may be more time consuming than multiple-choice ones, independently of their difficulty. Additional analyses performed separately on multiple-choice and open-ended items suggest that the obtained results are by and large comparable to those presented in this report.[12]

Within each cluster, the individual's RT for the **five easiest** items is compared to the same individual's RT for the **five hardest** items. This approach is meant to avoid relying on only one item (the easiest and the hardest), which would result in having a less reliable indicator. Five is the maximum number that can be used for all clusters.

The main **assumption** underlying this operationalisation of effort is that more difficult items require more effort to be answered (i.e. more time is spent on them). Figure 3.3 shows a number of statistical analyses that were performed on the PISA data in order to test this assumption. Figure 3.3 (panel a) shows the distribution of mean RT in easy and difficult items overall. The figure exhibits two noteworthy patterns. First, the difficult item distribution is shifted to the right as compared to the easy item distribution, indicating that, on average students spend more time on responding to difficult items than they do on easy items. Second, response time variability is visibly higher for difficult items, as the distribution is always flatter than the distribution estimated for easy items. Additional analyses, not shown here, indicate that the two patterns just-described are apparent for the three domains, but are more pronounced for science, where the RT estimate for easy items is especially precise and low.

Figure 3.3 (panels b and c) shows the average RT differences between easy and difficult items. The figures in panel b (red line) show that, on average, students spend more time on difficult items than they do on easy items (109 vs. 74 secs). The pattern is qualitatively similar across countries (thin grey lines in the left panel). Overall, the index works similarly across the three tested domains, although it is stronger in science (43.1 seconds) than in maths (35.0 secs) and reading (19.9 secs). In all cases, the average response time on difficult items is always statistically significantly higher than on easy ones.

3.3.2 Effort persistence

For each individual student, effort can be computed four times, one for each administered cluster. These four effort measures allow computation of a second effort-related index: effort persistence. This index captures students' ability to maintain effort throughout the test. The index is operationalised as the difference between effort in cluster 2 and effort in cluster 1.[13] Given that within a session, students are tested on one domain only (cases of students receiving two domains are dropped from the analyses), the only meaningful change between cluster 1 and cluster 2 is time. Hence, under the hypothesis that as time passes, students feel more tired, this index represents a reliable indicator of effort change and can be interpreted as the effort drop occurring when fatigue increases. The index can assume either negative values,

[12] Yet it should be stated that performing distinct analyses by item type implies a substantial loss of statistical precision of the index, due to the need to rely on only one easy and one difficult item, relative to using more items as proposed below.

[13] Effort measures obtained from clusters 3 and 4 are not considered for the effort persistence index computation. This is motivated by the fact that effort levels in the second session are more affected by fatigue, as can be seen by comparing the mean values of effort in the four clusters (Table 3.2). Also, in the second session a non-negligible fraction of students receives two domains instead of only one, thus altering the comparison of the effort measures.

when effort in cluster 2 is lower than effort in cluster 1, or positive values, when the opposite pattern takes place. When the index shows values around zero, it means that the student did not change her effort through session 1.

Table 3.2 shows the effort estimates in the four clusters. On average, effort is always higher in the first part of the session (hence, positions clusters 1 and 3 relative to positions clusters 2 and 4, respectively). Also, effort is higher in cluster 1 than in cluster 3, probably because students at the very beginning of the test are not yet affected by fatigue. This evidence also suggests that, in order to maintain the same level of effort throughout the test, students need to be persistent and not affected by fatigue.

Table 3.2. Effort by cluster position.

		Effort	
Session	Cluster	Mean	S.E.
1	1	35314.6	242.1
	2	27610.4	263.2
2	3	29052.5	274.1
	4	24123.3	228.5

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. Effort is measured in milliseconds. The information is not available for Malta and Romania.

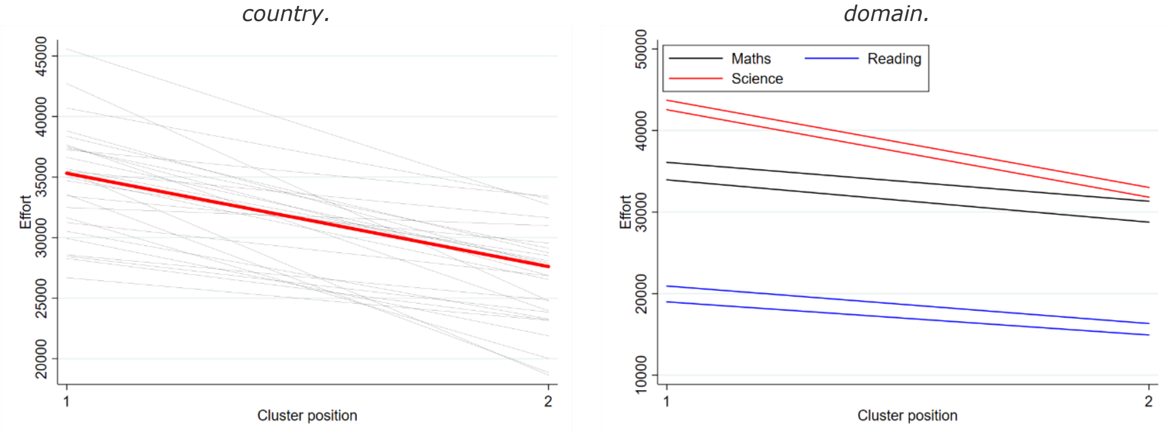
The above patterns motivate the choice *i)* to analyse effort in the first cluster only, in order to assess students' effort without it being biased by fatigue (which could vary systematically among students, e.g., the lowest-performing students could be more affected by it); and *ii)* to analyse effort persistence throughout the test as a separate and independent outcome. Because, as explained above, the test domain administered to student *j* changes between session 1 and 2, the analysis is restricted to session 1.

On average (red line in Figure 3.4), effort drops from 35.3 to 27.6 seconds: a statistically significant decrease of about 7.7 seconds. This means that, on average, students' effort in responding to the test drops as time in front of the computer passes. When looking at the three domains separately, science shows the largest effort drop (10.7 seconds) while the effort drops in maths and reading are 4.9 and 4.3 seconds, respectively. Across all domains, the effort drop is statistically significant.

Figure 3.4 Effort change between cluster 1 and cluster 2.

Panel a: *Effort change from cluster 1 to cluster 2 by country.*

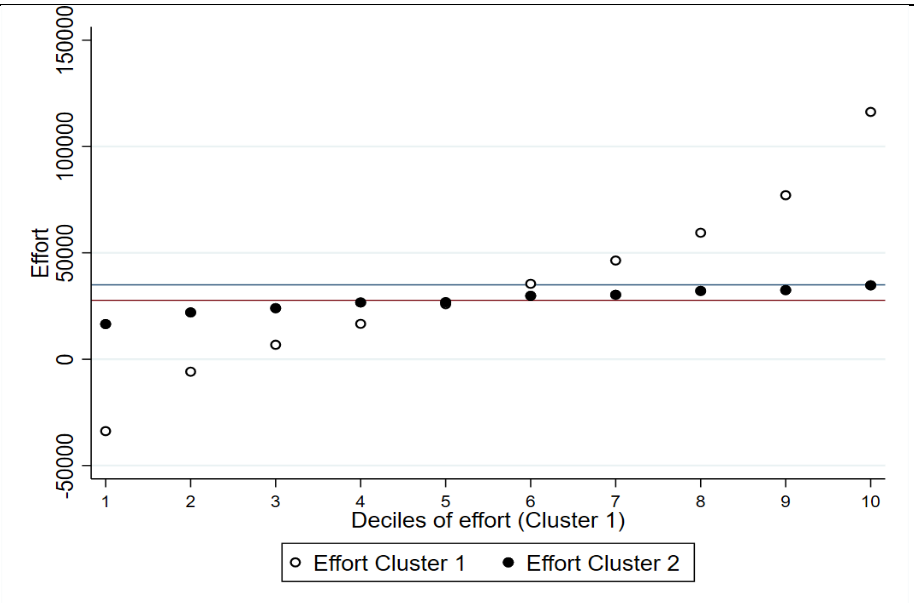
Panel b: *Effort change from cluster 1 to cluster 2 by domain.*



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. Effort is measured in milliseconds. In panel a, the red line represents the average effort for EU Member States. The information is not available for Malta and Romania.

To shed further light on the link between effort in cluster 1 and 2, and to learn more about students' test-taking behaviour, the sample of students is now divided into deciles of effort in cluster 1 (Figure 3.5). The figure shows that among those who start by making low effort, there is the largest increase from cluster 1 to cluster 2, while the largest drop occurs among those who start with more effort. These patterns can be partially explained by the existence of *ceiling* and *floor* effects. At the centre of the distribution students' behaviour is instead more 'balanced' between the two clusters.[14]

Figure 3.5. Effort in cluster 1 and effort in cluster 2, by effort deciles in cluster 1.



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. Effort is measured in milliseconds. Horizontal lines represent the average effort in cluster position 1 (blue) and the average effort in cluster position 2 (red) for EU Member States. The information is not available for Malta and Romania.

Therefore, by comparing effort in cluster 1 with effort in cluster 2, four different student profiles can be identified:

- Students who make a higher than average effort in cluster 1 and persist in cluster 2 ("**hard working**");
- Students who make a high effort in cluster 1, but lose effort in cluster 2 ("**hasty**");
- Students who make little effort in cluster 1, but increase their effort in cluster 2 ("**Slow starter**");
- Students who make little effort in cluster 1 and further reduce or do not improve their effort in cluster 2 ("**work-shy**").[15]

Table 3.3 confirms that the majority of students are those who decrease (or do not increase) their effort (the sum of "hasty" and "work-shy" is 51.1%). However, it also

[14] Students who make a lot of effort in the first cluster also register higher numbers of unreached items by the end of the first session (Table A3.2 in Appendix 3). This can be explained by the fact that they spent too much time on cluster 1 and did not have enough time to complete the second cluster.

[15] To avoid overestimating non-persistence, students are labelled as "non-persistent" if their effort persistence value is at least one tenth of a standard deviation below zero.

shows that the single profile showing the highest frequency in the sample is made up of those students who show an increase in their effort ("slow starter", 38.1%), who number slightly more than "hasty" (35.6%). The smallest category is "hard-working" students (10.8%), while "work-shy" students represent a non-negligible category comprising about 15% of the sample.

Table 3.3. Effort profile distribution.

Profile	%
Hard working	10.8
Hasty	35.6
Slow starter	38.1
Work-shy	15.5

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

3.3.3 Perseverance

Perseverance (or endurance) is defined as "students' motivation and the impact that motivation has on self-control and the ability to withstand fatigue" (Borgonovi and Biecek 2016, 128). To measure this concept, the authors exploit the random allocation of test booklets in PISA 2006, 2009 and 2012 to obtain measures of perseverance at an aggregate level. More precisely, they compute the difference in the percentage of correct answers at the beginning and at the end of the test. Borgonovi and Biecek's decision to compute the index at an **aggregate level** was motivated by the consideration that no student receives the same item twice during the PISA test.

The measure of perseverance proposed in this report is instead an **individual level** one and expresses the **difference in performance at different points of the test**. The idea is to investigate the extent to which students perform equally well throughout the test. More precisely, the measure of students' perseverance is based on the comparison of their performance on cluster 1 vs their performance on cluster 2.[16] Rather than computing a simple mean score by cluster, Weighted Likelihood Estimates (WLE) of students' ability were derived for each cluster and for each domain by using the international item parameters (OECD 2017). In contrast to the simple mean score, **the computation of WLE integrates items difficulty**, so that **WLEs are comparable from one cluster to another**, independent of the cluster average difficulty (Muraki 1992). The WLE scores have been rescaled to have a mean of 100 and a standard deviation of 10 to facilitate the interpretation of the results.

Table 3.4 shows the average WLE scores computed in each of the four clusters and it presents the overall average, because the results found separately domain by domain yield the same conclusions. It emerges that there is a drop in the WLE score when passing from the first to the second cluster. The same pattern is also found between cluster 3 and 4. The drop in the WLE score from the first (third) to second (fourth) cluster can also be interpreted as a sign of fatigue. The increase in the score from the second to the third cluster can be explained by the break at the end of the first session or by the domain change that occurs from session 1 to session 2.

[16] The same procedure is implemented for the second session as well, comparing performance on cluster 3 and 4, and leads to comparable results.

Table 3.4. Mean and standard error (S.E.) for the WLE scores according to cluster position.

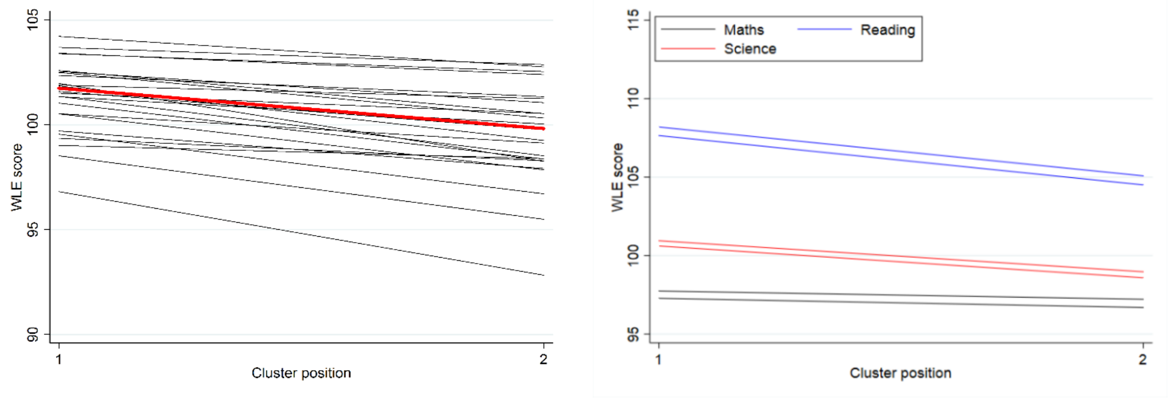
		WLE	
Session	Cluster	Mean	S.E.
1	1	101.75	0.078
	2	99.83	0.085
2	3	100.91	0.085
	4	99.48	0.085

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Romania and Malta.

As Figure 3.6 shows, the decrease in performance across clusters takes place in all countries included in the analysis. Figure 3.6, panel a (red line), shows that, on average, EU students perform more poorly in the second panel than in the first one (as stated above) and that the pattern is qualitatively similar across countries (thin grey lines in the left panel). Looking at the differences between the three domains (panel b), the index works as expected for reading and science, while for maths the WLE score does not drop significantly from the first to the second cluster. Thus, no significant performance drop is detected for maths. Further analyses, which go beyond the aim of this report, are needed to shed light on these diverging patterns.

Figure 3.6. Change in the WLE scores according cluster position, by country (left panel) and by domain (right panel).

Panel a*: Change in WLE from cluster 1 to cluster 2 by country.* ***Panel b****: Change in WLE from cluster 1 to cluster 2 by domain.*

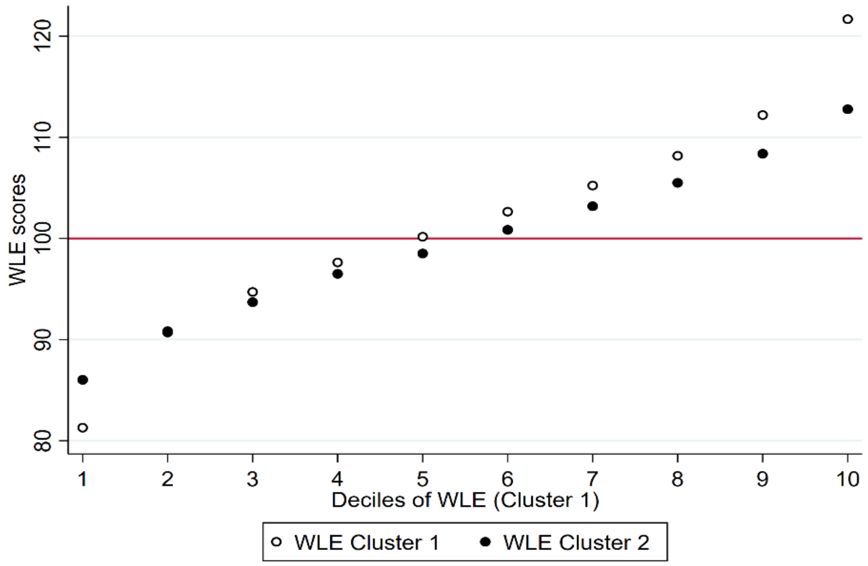


Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. In panel a, red line represents the average WLE score for EU Member States. The information is not available for Cyprus, Romania and Malta.

It should be noted that the perseverance index, constructed as the difference between performance in cluster 2 and performance in cluster 1, can assume either positive or negative values. Positive values signal that students do not worsen their performance but, on the contrary, improve it when moving from the first to the second cluster. Negative values indicate students whose performance drops, meaning that not all students perform more poorly in the second cluster than in the first. Finally, students with a perseverance index around 0 are students who perform equally well in the two clusters.

To gain a better understanding of how students actually perform throughout the test, it is useful to compare the changes in the WLE scores across clusters. Figure 3.7 shows that students who perform very well at the beginning of the test have, on average, a lower score in the second cluster, and vice versa. It should be stressed that, in any case, those belonging to the highest deciles show results that are always above the average (100). On the other hand, the worst students in the first cluster, who improve their performance, still remain below the average. Therefore, the interpretation of perseverance – like effort persistence – is not straightforward because it is influenced by ceiling and floor effects. Indeed, higher values on this index can be achieved by low performer students, i.e., those students who perform very badly in the first cluster and, for whom, therefore, it is very difficult to perform even more poorly in the second cluster (floor effect). The opposite is true for the high performer students: they start with a very high score, and it is difficult for them to maintain the same standard in the second cluster (ceiling effect). At first, this result is unexpected, because it has been hypothesised that more persevering students should perform better on the tests. The fact that the overall test score correlates strongly with the WLE score in cluster 1 suggests that what really matters for the test score is the performance at the beginning of the test.

Figure 3.7. WLE scores in clusters 1 and 2 by WLE score deciles in cluster 1.



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania.

Because the perseverance index strongly depends on the starting point, it is difficult to interpret when used in its continuous form. A possible solution is to use the information in the WLE scores in the different clusters to build a typology to take the complexity of the perseverance concept into account. Four types of students are identified:

- Students who perform above the average in the first cluster and either improve or remain constant in the second one ("Persistently good");
- Students who perform above the average in the first cluster and worsen in the second one ("Starts well but drops");
- Students who perform below the average in the first cluster and improve in the second one ("Slow starter");
- Students who perform below the average in the first cluster and worsen in the second one ("Persistently weak").

Table 3.5. Average values of the scores on science, maths and reading according to the perseverance typology.

Profile	%
Persistently good	16.1
Starts well but drops	33.8
Slow starter	26.2
Persistently weak	24.0

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania.

Table 3.5 shows that the majority of students are those who start well and then decline in performance, while the smallest category is made up of the "persistently good" students. Note that the "persistently weak" students represent a fairly large category, comprising about one quarter of the sampled students.

3.4 Analytical strategy and independent variables

As stated in Section 2, one of the aims of the study is to identify the determinants of non-traditional competences at the individual, school and school-system levels.

Before describing the independent variables used in detail, it is worthwhile outlining the analytical strategy employed in the empirical sections of this report (i.e., sections 4-7). The models presented in this report follow the same logic independent of the outcome considered. The idea is to estimate a **series of multiple regression models** by inserting the variables at the different level according to a stepwise strategy. More precisely, four models are estimated[17]:

- **Model 1** considers a set of background variables measured at the individual level.
- **Model 2** adds the school level variables to M1.
- **Model 3** adds the traditional competences measured via the first plausible value (standardised) of science.
- **Model 4** includes school-system characteristics.

The logic of this stepwise strategy is graphically depicted in Figure 3.8. M1 and M2 are meant to estimate the main determinants of the non-traditional competences, while M3 yields an estimate of the relationship between traditional and non-traditional competences, controlling for M1 and M2 covariates. According to the scheme in Figure 3.8, M1 provides an estimate of the *total* influence of each individual background variable, while M2 yields information on the *direct* influence of these background variables (i.e., the influence of individual background variables, net of school factors effects). The arrow linking background characteristics to school factors indicates the possibility that, as demonstrated by a large bulk of empirical studies, students from advantaged socioeconomic backgrounds enrol in higher-quality schools (e.g., those with the best teachers or classmates). For this reason, M2 also provides the *total* influence of school level variables. The same logic can be applied to M3 with reference to the competence in science: for the individual variables the *direct* influence is estimated, while for science it is the *total* that is estimated. One additional model (M4)

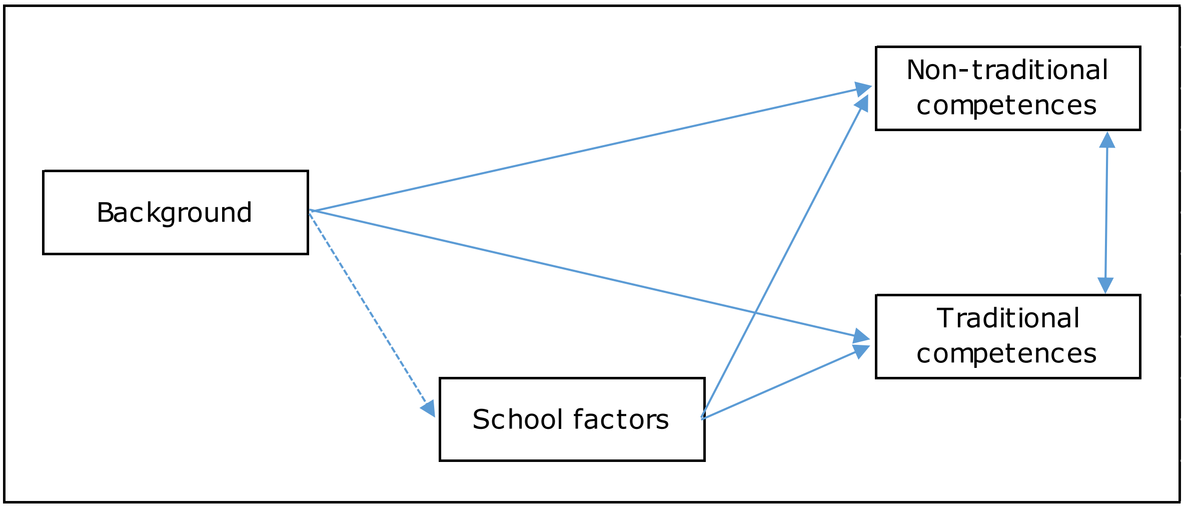
[17] All models include country-fixed effects in order to account for factors at the national level that could possibly affect the results and are estimated using the complex sampling design weights - operation made possible by the Stata routine *repest* (Avvisati and Keslair 2017). Moreover, the models consider the same analytic sample (i.e., the one generated by Model 3) and all the continuous variables are standardised in order to compare the magnitude of the coefficients. The models are estimated using OLS for the continuous and binary outcomes, while for the typology a multinomial logistic regression is applied. For the binary indicators the results are the same if logistic regression is used.

is also estimated with the aim of analysing the influence of school-system characteristics on the non-traditional competences.[18]

It is worthwhile stressing that this analytical strategy does not lead to the identification of causal effects. Hence, even if, when commenting on the empirical results, terms that recall a causal logical framework (e.g., effect) are sometimes employed, they are by no means to be interpreted in causal terms.

The full list of variables employed to carry out this analysis is shown in Table 3.6. For each level, the variables are considered as either "main variables of interest" or "control variables." The interpretation will focus only on the former, while the latter will be used only as statistical controls in the multiple regression models. The choice of the variables used depends, first, on the findings of the current literature summarised in section 2 and, second, on the actual availability and quality (e.g., the incidence of missing values) of the indicators collected through the student and school questionnaires carried out within PISA 2015.

Figure 3.8 Linkages between background, school factors and student traditional and non-traditional competences.



At the individual level, the following main variables were employed: **gender**, migration background and social origin (parental education and occupation). **Migration background** is measured through four categories[19]: natives (when both parents were born in the test country), mixed-parentage children (when one parent was born in a foreign country), second-generation children (when both parents were born abroad but the child was born in the test country), first-generation children (when both parents and the child were born abroad). **Parental education** takes different values, depending on the highest level of education completed by either of the two parents: up to lower secondary; upper secondary, tertiary education and above. **Highest parental ISEI** comes as an already computed index, which measures the standard international socioeconomic index of occupational status (Ganzeboom et al 1992). Additional variables, such as age in months and attended grade (relative to the country mode), are included as controls.

18 In this model, country-fixed effects are not included due to possible problems of multicollinearity. Of course, this implies that model 4 does not make it possible to distinguish between the influence of school-system characteristics and other factors that might vary across countries.

19 The exploration of country-of-origin differences in immigrant origin student competences is made difficult by the small sample size of the immigrant-origin students and by the different aggregations of countries of origins adopted in the different EU Member States. To further investigate this topic, an ad hoc analysis – which goes beyond the scope of this project – would be needed.

At the school level, the focus is on four indices that previous studies found to be correlated to students' non-traditional competences: extra-curricular activities, teacher involvement in school decision-making, parental involvement and school climate. The four indicators are all based on the questionnaire completed by school principals. The index of **extra-curricular activities** is an additive index that refers to activities that schools offer to students, such as musical groups, school magazines, chess clubs and so on. **Teacher involvement** refers to the extent to which teachers are assigned responsibility for curriculum and resources. For **parental involvement**, an indicator that considers parents' participation in a specific school-related activity (as reported by school principals) is used. More precisely, the variable expresses the proportion of parents who participate in "discussions about child's progress with a teacher on their own initiative". The question places the emphasis on parents' initiative, hence it can be understood as an indicator of parental involvement in their children's schooling. To proxy **school climate**, a PISA-constructed indicator on the school's learning environment is used. It indicates the diffusion of student behaviours that, according to the principal, hinder learning in the school. Among the items that make up the indicator are bullying, truancy, alcohol and drug use, and so on. Hence, contrary to the three other indicators, in this indicator higher values correspond to negative school conditions. The factors mentioned do not, of course, cover the entire spectrum of school factors and resources; therefore, all analyses will have to control for a list of other variables. Several variables are considered as school-level controls and measure dimensions such as school autonomy and management and school capacity. Two variables are added at the school level to take instruction time into account, as this could affect the outcomes and, at the same time, vary across countries. More precisely, the total number of hours per week and the percentage of hours dedicated to science are considered. School tracking (ISCED orientation) is also employed. School tracking (ISCED orientation) is a further control, included to consider the existence of different school programmes – i.e., academic or vocational oriented curricula. The ISCED level at which the student is enrolled, the student/teacher proportion and the average science score of the school are also included as controls.

Table 3.6 Main independent variables.

Level	Main variables	Control variables
Individual and Family	• Gender • Highest parental ISEI (HISEI) • Parental occupation • Immigrant background • Science, maths and reading competence	• Age • Grade
School	• Extracurricular activities • Proportion of parents involved in school activities • Teacher involvement in school decision-making • School climate	• School autonomy and management (educational leadership, responsibility for curriculum, responsibility for resources, school type) • School capacity (educational material shortage, staff shortage) • Student/teacher proportion • Weekly instruction hours) • ISCED level • ISCED orientation • School average science score

School-system	• Horizontal differentiation • Vertical differentiation • System autonomy • Average performance (average science score) • Compulsory school start and end ages	• Country GDP

Finally, at the country level, horizontal and vertical differentiation, the starting and leaving ages of compulsory schooling, the level of school autonomy, and the overall average science score are considered as main factors, while the country's GDP is included to control for other factors that might possibly confound country comparisons.[20]

As shown in Appendix 3.1, several independent variables are affected by small to moderate incidences of missing values. As was the case for the outcome variables illustrated in the previous section, the method adopted to deal with missing values was that of deleting all cases with missing observations on the main variables. However, observations with missing values in control variables are retained in the sample and assigned to a 'missing value' category, to avoid losing too many observations. Appendix 3.1 shows all the robustness checks that have been performed in order to make sure that these missing values do not hamper the solidity of the results.

Table 3.7 shows the distribution of all main dependent and independent variables in the overall sample. Because the PISA sample is representative of the population of 15-year-olds, the table then provides a comprehensive look at 15-year-olds in Europe. These estimates, like all estimates that are presented in this report, take account of the complex PISA sampling design, by incorporating the PISA-provided replicate weights and the final student weights.

All main continuous independent variables (HISEI, and the four school-level factors) have been standardised to have a mean of 0 and a standard deviation of 1 upon their inclusion in the regression models. The regression coefficients therefore express changes in the outcome variables when incrementing the value of the independent variable by one standard deviation. The same standardisation procedure was applied to the first plausible value of the three domain test scores (reading, mathematics and science) in order to speed up the computation and facilitate the interpretation of the results. Models including all plausible values yielded qualitatively identical results.

[20] This information was collected from Eurydice and OECD reports. More precisely, vertical differentiation is defined according to the country incidence of students repeating a grade, while horizontal differentiation considers three variables: the number of tracks; the age of first selection into these tracks and the proportion of selective schools in the school system.

Table 3.7 Descriptive statistics for the main variables.

	N	%	Mean	Sd	Range	Min	Max
INDEPENDENT VARIABLES							
Individual\Family							
Gender							
Male	87,793	50.41					
Female	85,811	49.59					
Parental education							
Primary or lower secondary	14,819	12.06					
Upper secondary	58,792	34.63					
Post-secondary\tertiary	94,199	53.31					
Immigrant background							
Native	129,223	78.67					
Mixed parentage	17,198	9.75					
II generation	10,639	6.65					
I generation	8,001	4.93					
Highest parental ISEI	158,175		51.16	21.78	78.00	11.00	89.00
Science competence	173,604		496.97	97.67	769.39	102.37	871.76
Reading competence	173,604		496.68	99.61	847.38	6.28	853.66
Maths competence	173,604		494.39	91.95	767.81	81.15	848.96
School							
Extracurricular activities	157,501		5.59	2.38	10.00	0.00	10.00
Proportion of parents involved in school activities	155,361		45.50	27.43	100.00	0.00	100.00
Teacher involvement in school decision-making	160,688		3.90	1.86	12.00	0.00	12.00
Negative school climate	157,210		-0.02	0.96	6.28	-2.39	3.89
School-system							
Horizontal differentiation							
Low	71,956	38.00					
Medium	36,354	14.69					
High	65,294	47.31					
Vertical differentiation							
Low	126,596	50.16					
High	47,008	49.84					
DEPENDENT VARIABLES							
Science engagement	156,388		-0.02	1.12	5.13	-1.77	3.36
Truancy							
At least one time	36,163	21.12					
Truancy (2)							
At least one time	49,814	28.72					
Late							
At least one time	75,040	44.00					
Effort	116,113		35314.59	47221.17	7792155.00	-1771760.0	6020395.00
Persistence	114,589		-7319.40	60217.59	7176828	-6011018.0	1165810.00
Perseverance	112,156		-1.92	9.61	137.82	-75.82	61.99

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights.

4 Current engagement in school

Key findings

School engagement is measured using three indicators: skipping at least one entire day of school; skipping some classes; and arriving late to school.

School engagement indicators (especially skipping one school day or some classes) show substantial variation across countries. Arriving late to school is the most common student behaviour across EU Member States.

Overall, boys and children of immigrants show the least regular school attendance patterns. Family background, as measured via parental education and parental socioeconomic status, is strongly associated with the three engagement indicators, but its influence happens to be fully captured by academic performance. The correlation between the indicators of current engagement and students' academic performance is consistently negative.

School climate (proxied by the presence of student misbehaviour) is found to be a strong determinant of current engagement: i.e., over and above individual characteristics, a negative school climate is associated with a higher risk of truancy.

This section is devoted to the analysis of three classic indicators of current engagement in school:

- **Truancy**: skipping vs not skipping at least a whole day of school in the past two weeks;
- **Truancy (2)**: skipping vs not skipping some classes in the past two weeks;
- **Lateness**: arriving late vs never arriving late at school in the past two weeks.

Even if these three outcomes are related to a similar underlying phenomenon, i.e., students' engagement in school, confirmatory factor analysis showed that these three indicators cannot be summarised as a single indicator in all countries, suggesting that they refer partially to different dimensions of students' school engagement. It can be argued that skipping a whole day of school is connected with students' school disaffection and negligence, and that it can be interpreted as a choice to ignore school duties. Skipping some classes, in contrast, may be due to a conscious choice that induces students to avoid particular classes – perhaps because they have difficulties in a specific subject or with a certain teacher. Other factors that facilitate delays in reaching school or class absenteeism may be certain familial or residential conditions, such as the home-school distance or the use of public transportation. For these reasons – and according to some previous PISA reports (OECD, 2014; OECD, 2013) – these three indicators of current engagement are considered separately.

This section consists of three main sub-sections. The first presents some descriptive results of cross-country comparisons in the proportion of individuals experiencing truancy, truancy (2) and late arrivals. The second sub-section shows the results of a set of multiple regression models aimed at exploring the main determinants at the individual level and at the school level – controlling also for students' achievement – of the three indicators of current engagement in school. Finally, the third sub-section presents the results of multiple regression models that include the features of the national education systems.

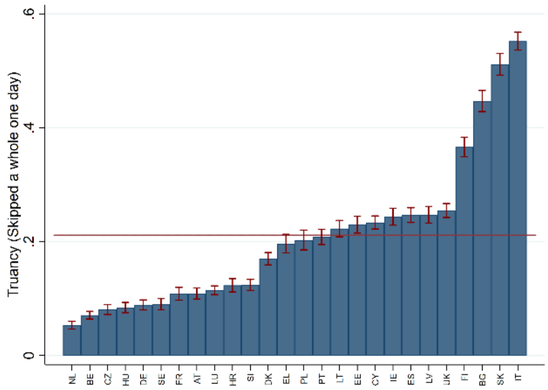
4.1 Cross-county comparisons

Figure 4.1 shows cross-country variation in the three indicators of current engagement in school. Panels a, b and c show average country values of the outcomes of interest and the relative standard errors.

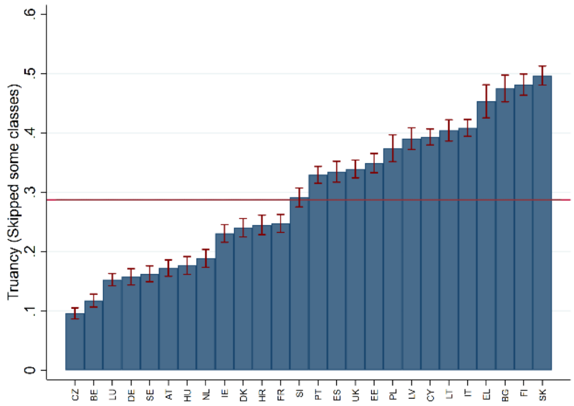
Skipping a whole day of school is probably the strongest indicator of a rather low degree of school engagement. As is evident, the first indicator of truancy has lower values than the other two indicators. In panel a, average country values for truancy (skipping a whole day) show that, overall, there are no marked differences among EU Member States. However, there are four exceptions: Finland, Bulgaria, Slovakia and Italy register somewhat higher average values. This pattern is also apparent in panels b and c, which display the other current engagement indicators.

Figure 4.1. School engagement indicators in EU Member States.

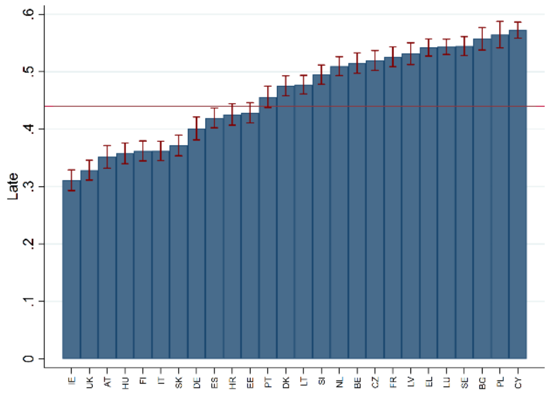
Panel a*: Truancy (skipped a whole day).*



Panel b*: Truancy (skipped some classes).*



Panel c*: Arriving late.*



Note: FBK-IRVAPP analysis of PISA 2015 data; the horizontal red line represents the average values for EU Member States. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

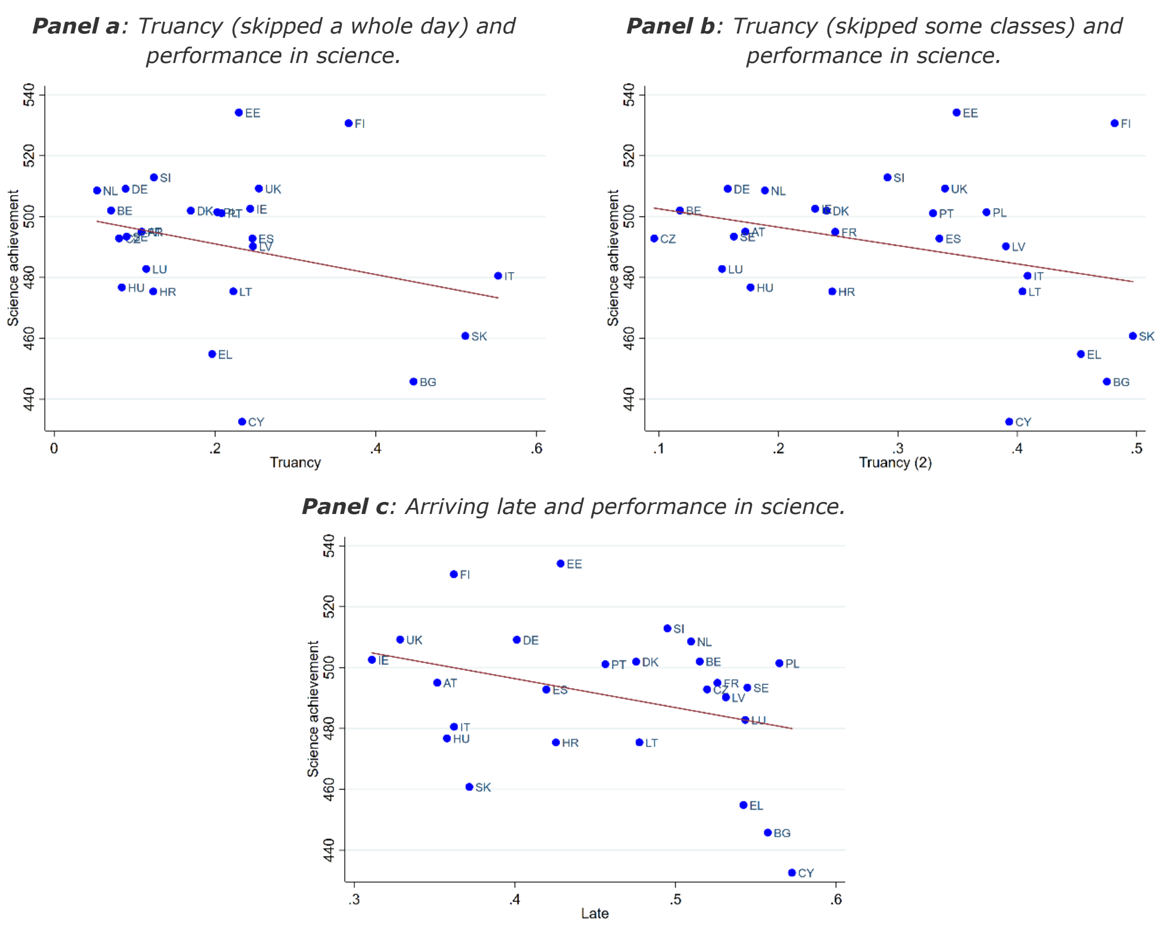
It should be borne in mind that these measures are somewhat imperfect, and their accuracy may depend on: *i)* whether students are honest in answering the survey, especially if they believe that their answers can be viewed by the teachers; and *ii)* whether students who are more likely to be absent actually answered the survey. Most importantly, there may be important details that make a difference when translating questions in different languages. In Finland, Bulgaria and Slovakia, for instance, the questions about truancy include absenteeism for reasons of illness. Therefore, cross-country comparisons may be difficult. As noted above, disengagement in school has many facets and questionnaire-based, self-reported information can be affected by a number of measurement errors.

Finally, the "arriving late" outcome presents the highest values for all countries – i.e., it is the most common among the three 'misbehaviours'. The average is also more similarly distributed among EU member countries than the truancy indicators, suggesting that arriving late at school is an indicator for both disengaged students and/or students who regularly encounter obstacles of various kinds in arriving at school.

4.2 Country correlations with achievement score

In order to properly understand the relationship among the three indicators of current engagement in school and students' achievement, scatterplots at the macro level are drawn (Figure 4.2). Unsurprisingly, the correlation at the macro level between the three indicators of current engagement in school and the average performance in science is negative. Countries with higher overall science achievement are also the countries with the lowest levels of students' school disengagement. The same pattern is observed when considering performance in maths and reading.[21]

Figure 4.2. Scatter plot between school engagement indicators and achievement in science in EU Member States.



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

[21] These additional results will be made available upon request.

4.3 The individual and school level determinants of school engagement

This section presents the results of a series of three linear multiple regression models. The aims of these analyses are: *i)* establishing the gross and net associations between individual characteristics and the engagement indicators; *ii)* assessing the association between school factors and the same outcome variables over and beyond student characteristics; *iii)* investigating the link between the three engagement indicators and student cognitive competence. Even if the three outcome indicators are coded as dummy variables, taking either value 1 or 0, linear regression models were estimated.[22]

The results (presented in Table 4.1) show that some key **factors at the individual level** are strongly associated with the outcomes of interest. On average, **boys show a higher risk** of skipping a whole day of school, skipping classes, or arriving late at school. Moreover, **children of immigrants appear to be more at risk** than native students of both skipping whole days of school or some classes and of arriving late to school. This gap slightly decreases for all three outcomes when controlling for science performance, even if it remains highly significant. It must be stressed that these native/immigrant gaps are always controlled for family background characteristics, which could confound the comparison. They therefore indicate a migration-specific disadvantage in regard to school attendance regularity. The results clearly show that the largest risk is found among children of immigrants who were born abroad (i.e., the so called first generation), indicating that the group of immigrant newcomers requires special attention and support from schools and authorities.

With regard to parental education and parental socioeconomic status as determinants of the current engagement in school, the results are mixed. Overall**, the higher the parental status and the parental education level, the lower the probability that students will skip days or classes and arrive late at school**. This is particularly apparent when looking at models 1 and 2. However, controlling for performance in science reverses this situation, suggesting that the initial advantage of students from higher socioeconomic backgrounds is almost entirely explained by their performance levels. Once this is taken into account, all differences disappear. In some cases, the effect of socioeconomic background changes from positive to negative, and vice versa, depending on the outcome considered. This further reinforces the idea that, **any influence of parental background on attendance is captured by performance levels**.[23]

Among **school level determinants**, only the indicator of a **negative school climate** has a significant and positive effect on the outcomes of current engagement in school in both model 2 and model 3, controlling for performance in science. Parental involvement – measured as the proportion of students' parents who participated in teacher-parent meetings on their own initiative – appears irrelevant in determining students' engagement.

The **extra-curricular activity** index has a small negative effect on students' propensity to skip a whole day of school, while it has a positive effect when considering students' propensity to skip some classes or arrive late. This finding could

---

[22] Results obtained via logistic or probit regression models yielded very comparable results, hence linear regression model coefficients are shown as they are more easily interpretable. Moreover, the models have been re-estimated removing the three countries in which the questionnaire differed (Finland, Bulgaria and Slovakia) and the results change only marginally.

[23] The association between individual level characteristics and performance in science is substantially strong (see Table A4.1 in Appendix A4), therefore it is obvious that the parameters indicating the influence of individual level variables change dramatically after including science performance in the model. Questioning which is the most important dimension in shaping non-traditional competence would require further analyses that go beyond the aim of this report.

be explained by the fact that students in schools that offer many extracurricular activities feel somehow ‘overwhelmed’ by the many activities they are required to attend. In other words, offering extra-curricular activities is not harmful for students’ engagement in general, but it can have some minor negative consequences on the attendance of some classes and on punctuality. Schools with higher rates of **teacher involvement in school decision-making** are also schools in which students have a lower likelihood of skipping an entire day of school.

To summarise, among the indicators of current engagement in school, “skipping at least a whole day of school in the previous two weeks” could be considered as the best indicator of school disengagement and low school attendance. Indeed, the patterns of the determinants of school engagement, measured as skipping a whole day, are slightly different, compared to those of skipping some classes and arriving late. These results are also consistent with the cross-country comparison graphs shown in section 1.

Additional models were run to investigate a) whether the above-mentioned results were homogeneous across the EU countries and b) the role played by some institutional features of the national education systems. The results of these models are not shown here – but are reported in Appendix A4 – because no clearly interpretable patterns, but rather erratic results, were found, whose investigation is beyond the scope of this report.

Table 4.1. Linear regression models for the school engagement indicators, according to individual characteristics (model 1), school-level factors (model 2) and performance in science (model 3). Selected parameters.

	TRUANCY (a whole day)			TRUANCY (some classes)			Arriving late		
	M1	**M2**	**M3**	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**
Individual level									
Male (ref. Female)	0.012***	0.008**	0.016***	0.023***	0.016***	0.025***	0.066***	0.058***	0.068***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Migrant Background (ref. Natives)									
Mixed parentage	0.032***	0.031***	0.026***	0.035***	0.034***	0.029***	0.068***	0.067***	0.061***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Second-generation	0.031***	0.031***	0.018*	0.064***	0.061***	0.046***	0.107***	0.106***	0.088***
	(0.01)	(0.01)	(0.01)	(0.012)	(0.012)	(0.012)	(0.014)	(0.014)	(0.014)
First-generation	0.042***	0.034***	0.019*	0.075***	0.067***	0.051***	0.116***	0.104***	0.083***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)
Parental education (ref. Low secondary)									
Upper secondary	-0.037***	-0.032***	-0.023***	-0.014	-0.009	<.001	-0.016*	-0.009	0.003
	(0.008)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Tertiary	-0.031***	-0.024***	-0.014*	-0.001	0.007	0.018**	-0.002	0.009	0.023**
	(0.008)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Highest parental ISEI (std)	-0.021***	-0.015***	-0.003	-0.012***	-0.004	0.008***	-0.006**	0.004	0.019***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
School level									
Extra-curricular activity index (std)		-0.005*	<.001		0.006*	0.011***		<.001	0.006*
		(0.003)	(0.003)		(0.003)	(0.003)		(0.004)	(0.004)
Proportion of parents involved in school activities (std)		0.002	0.002		0.001	0.002		<.001	0.001
		(0.002)	(0.002)		(0.003)	(0.003)		(0.003)	(0.003)
Teacher involvement in school decision-making (std)		-0.004*	-0.005**		-0.003	-0.004		-0.006	-0.006*
		(0.002)	(0.002)		(0.003)	(0.003)		(0.004)	(0.003)
Negative school climate (std)		0.011***	0.007***		0.030***	0.026***		0.025***	0.020***
		(0.003)	(0.003)		(0.003)	(0.003)		(0.004)	(0.004)
School achievement									
Performance in science (std)			-0.055***			-0.059***			-0.072***
			(0.002)			(0.003)			(0.003)
Constant	-0.393***	-0.471***	-0.522***	-0.343***	-0.384***	-0.443***	0.124	0.024	-0.045
	(0.106)	(0.108)	(0.108)	(0.127)	(0.127)	(0.126)	(0.115)	(0.109)	(0.108)
N	131436	131436	131436	131032	131032	131032	131387	131387	131387
R^2	0.132	0.138	0.150	0.057	0.065	0.076	0.039	0.047	0.062
Country clusters	26	26	26	26	26	26	26	26	26

Note: FBK-IRVAPP elaboration on PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects; model 2 also controls for a rich set of school-level factors. All continuous variables are standardized; only the first PISA plausible value (standardized) is included. All models incorporate the PISA complex sampling design weights. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

5 Current engagement in science

Key findings

Students' engagement in science is defined as the extent to which students engage in science-related out-of-school activities. It is intended to be an indicator of motivation and interest for science-related subjects and hence a possible predictor of future positive attitudes towards science and science-related careers.

On average, Southern and Eastern EU countries score higher on this indicator than Northern states. This pattern is partly due to cross-country variation in the practice of assigning homework after school, as the indicator does not distinguish spontaneous engagement in science-related activities from activities that are part of homework.

Across countries, boys, children of immigrants and children of highly-educated parents show higher engagement in science, even when controlling for science performance.

Among the school practices considered, the empirical findings point to the importance of a positive school climate as a predictor of students' science engagement.

Beyond the practice of homework, some features of national education systems are correlated with science engagement. Students in horizontally stratified countries and in countries where compulsory school ends later exhibit higher levels of engagement, while country system autonomy is negatively correlated with engagement in science.

This section examines students' engagement in science, which is defined as the extent to which students engage in science-related out-of-school activities. Importantly, this index of engagement in PISA 2015 does not relate to school engagement in general but is specifically related to science only, as science was the major domain in the 2015 assessment. This implies that the results should be interpreted keeping in mind this subject-specificity of the indicator. Sub-section 5.1 presents some aggregate statistics, with the aim of exploring how the indicator varies across the EU Member States and how it correlates with average science performance. Sub-section 5.2 presents the main findings obtained through a series of multiple regression models and information about the main determinants of engagement in science at the individual and school level. In sub-section 5.3, the role of some features of the national systems on current engagement is examined. Finally, in section 5.4, a focussed analysis of the achievement-engagement link is pursued.

5.1 Cross-country comparisons

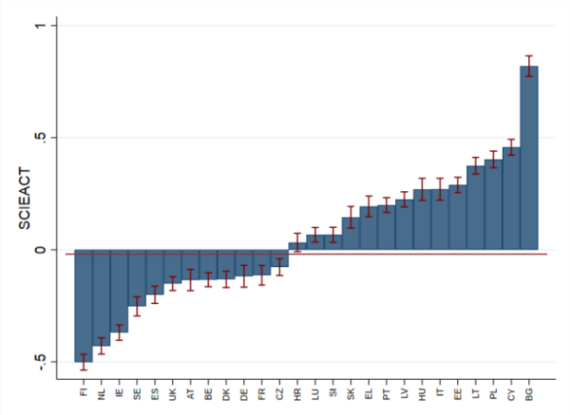
Figure 5.1 shows cross-country variation in the **students' science activities index** (SCIEACT). Panel a presents the SCIEACT index country average values and the relative standard errors. Bulgaria, Cyprus, Poland and Lithuania are the countries with the greatest engagement in science. Conversely, Finland, the Netherlands, Ireland and Sweden present lower values on the students' science activities index. These results suggest that science engagement is higher in the Southern and Eastern EU countries, and lower in the North.

A counterintuitive finding emerges when looking at the association between aggregate levels of the SCIEACT index and the average country science performance (Figure 5.1, panel b). Indeed, countries with higher engagement in science (such as Cyprus, Bulgaria and Lithuania) exhibit lower average scores in science and, conversely, countries with lower engagement in science (such as Finland, the Netherlands and Ireland) show higher performance in science. Panel b shows a negative correlation (r = -0.61) between engagement in science and science performance (this negative correlation is also present for reading (-0.69) and maths (-0.62)). Therefore, at the macro level, as students' science activities index score increases, science performance decreases.

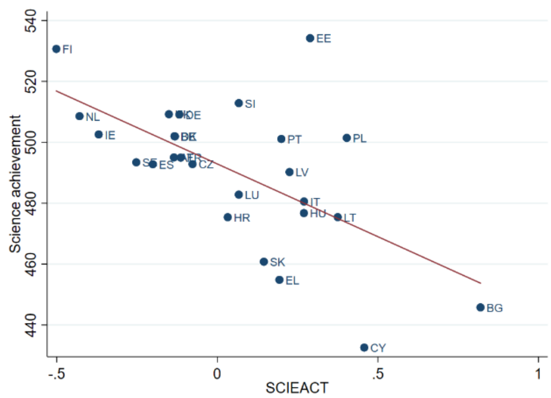
The link between science competence and science engagement is further addressed in sub-sections 5.4, in order to try to supply a plausible explanation for this result. In any case, it is possible to start looking at Figure 5.2, which reports the SCIEACT index **quartile distribution**, in order to consider its variability. It shows that the five countries with the lowest engagement in science (Finland, the Netherlands, Ireland, Sweden and Spain) are also those with the largest interquartile range, hence they are the countries in which there is the highest variability in the indicator. Moreover, the indicator distribution in these countries is clearly stretched towards the bottom, suggesting that the low average values in these countries are due mainly to the presence of students with very low values.[24]

Figure 5.1. Students' science activities and their relationship with science achievement in EU Member States.

Panel a: *Country averages for students' science activities.*



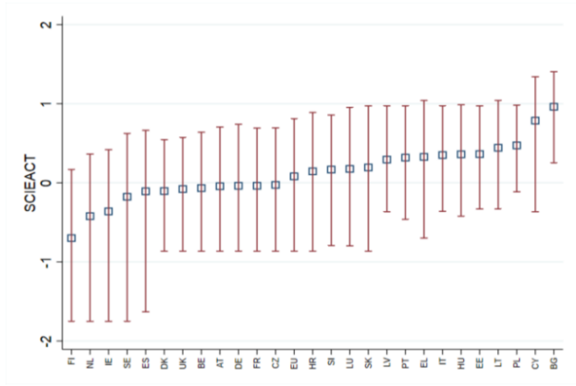
Panel b: *Scatter plot between students' science activities and achievement in science in EU Member States.*



Note: FBK-IRVAPP analysis of PISA 2015 data. Horizontal red line represents the average for EU Member States. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

24 The country order does not change significantly if the country rankings are calculated with the median index rather than the mean.

Figure 5.2. Quartile students' distribution for science activities.



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

5.2 Individual and school determinants of science engagement across EU MSs

Table 5.1 presents the results of three nested multiple regression models aimed at exploring the main determinants of engagement in science at the **individual** and **school level**.

On looking at Model 1 estimates, individual characteristics are found to be strongly associated with engagement in science. In line with extant research on gender differences in attitudes towards science (Osborne et al. 2003), boys exhibit significantly greater engagement than girls. This **gender gap** seems quite substantial as it accounts for more than one-quarter of a standard deviation of the SCIEACT index. Considering how widespread gender differences in relation to STEM disciplines are, this result does not come as a surprise.

Children of immigrants show higher engagement in science than children of natives. This is true for all immigrant-origin children, but especially true for the first generation. The gap between the first generation and natives has nearly the same size as the gap estimated between boys and girls.

Finally, pupils whose parents have a tertiary education degree or a high HISEI score show greater engagement in science. However, the family background coefficients are substantially smaller than those detected for gender and immigrant background.

On adding **school characteristics** (Model 2), individual characteristics effects do not disappear and remain qualitatively unchanged with respect to M1. Hence, schools do not account for any of the observed individual-level associations. Concerning the role played by the school-level factors of interest, neither extracurricular activities, nor parental involvement nor teacher involvement in school decision-making affect engagement in science at all. Instead, pupils in schools with worse school climates (higher student behaviour issues) tend to have less engagement. This result echoes past research findings, which emphasize the importance of classroom environment for students' attitudes towards science (Osborne et al. 2003)

Finally, controlling for **pupils' achievement** (Model 3), all individual and school characteristics coefficients on engagement remain highly significant, except for parental HISEI, which loses size and statistical significance. Pupils' science achievement is positively correlated with engagement in science. This result contrasts with the negative macro-level correlation and in some countries the micro-level correlation has a negative sign.

Table 5.1. Linear regression models for students' science activities, according to individual characteristics (model 1), school level factors (model 2) and performance in science (model 3). Selected parameters.

	M1	M2	M3
Individual level			
Male (ref. Female)	0.393***	0.400***	0.386***
	(0.010)	(0.011)	(0.011)
Migrant background (ref. Natives)			
Mixed parentage	0.049***	0.047***	0.055***
	(0.018)	(0.018)	(0.018)
Second-generation	0.147***	0.146***	0.168***
	(0.031)	(0.031)	(0.031)
First-generation	0.310***	0.311***	0.335***
	(0.033)	(0.033)	(0.033)
Parental education (ref. Low secondary)			
Upper secondary	0.018	0.015	-0.001
	(0.020)	(0.020)	(0.019)
Tertiary	0.158***	0.148***	0.130***
	(0.018)	(0.019)	(0.019)
Highest parental ISEI (std)	0.038***	0.029***	0.010
	(0.006)	(0.006)	(0.006)
School level			
Extracurricular activities (std)		0.012	0.005
		(0.008)	(0.008)
Proportion of parents involved in school activities (std)		-0.002	-0.003
		(0.007)	(0.007)
Teacher involvement in school decision-making (std)		0.012	0.012
		(0.009)	(0.009)
Negative school climate (std)		-0.023***	-0.017**
		(0.007)	(0.007)
Student achievement			
Performance in science (std)			0.091***
			(0.007)
Constant	-0.365	-0.393	-0.324
	(0.288)	(0.289)	(0.285)
N	124469	124469	124469
R^2	0.094	0.097	0.102
Country clusters	26	26	26

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects; Model 2 also controls for a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; * p<0.10; ** p<0.05; *** p<0.01.

The next sub-section seeks to gain a better understanding of the complex relationship between science achievement and science engagement.

Figure A5.1 in Appendix 5 shows the main individual regression coefficients of M2 **models run separately by country**. The gender gap is consistent across countries: in all EU countries, boys exhibit more engagement in science than girls. In addition, migration background and parental education coefficients present substantial variation across countries. Finally, parental HISEI coefficients do not show clearly consistent patterns across countries.

5.3 The role of education systems

Table 5.2 reports the coefficients of a set of country-level characteristics on engagement in science.

Table 5.2. Linear regression model of country characteristics on students' science activities.

School-system level	**M4**
Horizontal differentiation (ref. Low)	
Medium	0.219***
	(0.031)
High	0.169***
	(0.024)
Vertical differentiation (ref. Low)	
High	-0.211***
	(0.021)
Compulsory starting age (ref. 5 years old)	
6 years old	-0.081
	(0.049)
7 years old	0.055
	(0.043)
Compulsory leaving age (ref. 15 years old)	
16 years old	0.167***
	(0.019)
18 years old	0.314***
	(0.030)
Country system autonomy	-0.409***
	(0.078)
Country science achievement (MEAN)	-0.005***
	(0.000)
Country Gross domestic product (GDP)	-0.000***
	(0.000)
Constant	2.720***
	(0.345)
N	124469
R^2	0.091

Note: FBK-IRVAPP analysis of PISA 2015 data. The model controls for the same list of variables included in M3 with the exception of country fixed effects. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; * p<0.10, ** p<0.05, *** p<0.01. Data on compulsory starting and leaving ages come from *Eurydice 2018;* GDP comes from *OECD 2016a.*

The table shows that, net of the set of individual and school-level characteristics, some **features of national systems** are significantly correlated with current engagement in science. More precisely, higher levels of science engagement are found in countries with medium or high levels of horizontal differentiation. In contrast, science engagement is lower in countries with higher levels of vertical differentiation. Moreover, systems characterised by longer compulsory school duration tend to favour science engagement, while this is lower in systems with higher autonomy. To summarise, this correlational analysis indicates that the level of selectivity is negatively correlated with science engagement, while the latter is positively correlated with strong central governance and by the presence of school tracking.

5.4 Unfolding the macro/micro gap in the correlation between achievement and engagement in science

A focus analysis is presented to reconcile the contradictory results on the achievement-engagement link at the macro – where a negative correlation is found – and at the micro – where, instead, there is a strongly positive correlation – levels. A possible explanation is that the available measure of engagement in science reflects national school characteristics that can confound the results.

To delve deeper into this question, table 5.3 shows achievement-engagement coefficients of science in the two groups of countries with highest (BG, CY, PL, LT, EE) and lowest (FI, NL, IE, SE, ES) science engagement. The coefficients are obtained through the same Model 3 shown above, run separately country by country. This strategy is employed to try to understand whether the link between science performance and science engagement varies depending on the context (i.e., the national average science engagement levels). The results show that in countries with low average levels of science, pupils with higher engagement in science outperform pupils with lower engagement. Conversely, in countries with high levels of science engagement, pupils with higher performance present lower engagement in science (with the sole exception of Poland).

Table 5.3. Student achievement coefficients of model 3 for the five countries with low and high engagement in science.

Countries with high science engagement				
BG	CY	PL	LT	EE
-0.205***	-0.067***	-0.020	-0.143***	-0.043**
Countries with low science engagement				
FI	NL	IE	SE	ES
0.128***	0.225***	0.394***	0.175***	0.181***

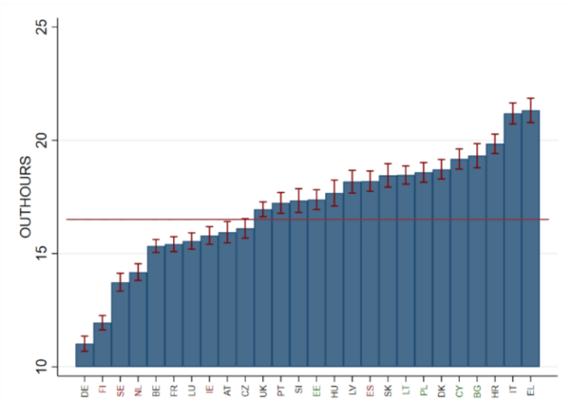
Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. * p<0.10, ** p<0.05, *** p<0.01.

Among the factors that can be invoked to explain such diverging patterns, the practice of **assigning homework** deserves particular attention. As shown in Figure 5.3 (panel a), the time students declare that they spend studying outside school hours differs widely across countries, revealing the existence of quite substantially different approaches in this respect across EU Member States. For our purposes, it is even more important to look at the macro correlation between homework and engagement in science (panel b). This correlation is strongly positive (r = 0.67): as time spent studying outside school increases, engagement in science increases. Thus, the current engagement index seems to reflect the time spent outside school learning. In other words, the engagement in science index is formed by students' science-related activities, but some of these activities can comprise students' homework, as the question administered to students does not explicitly distinguish 'voluntary' engagement in science-related activities from science-related activities that are part of homework. Indeed, students may watch TV shows, borrow books or visit websites about science topics as part of their homework. Moreover, as shown in panel c, the practice of homework is much more widespread in countries with low average test performance, which explains the negative association at the country level between test performance and science engagement (shown in Figure 5.2).

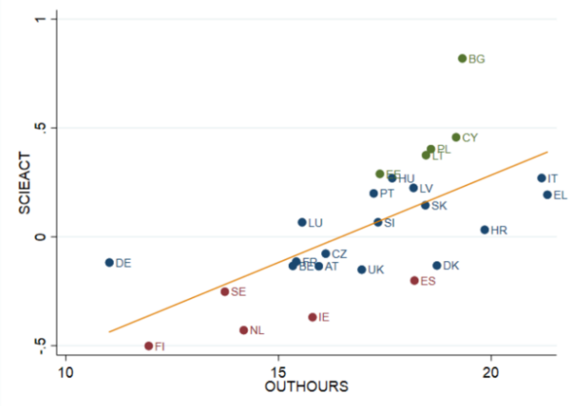
In contexts characterised by low levels of homework, the indicator is likely to better capture the existence of spontaneous and intrinsic interest in science, while in countries where the practice of homework is more widespread, the indicator can partly capture homework assignments. Hence, it shall be borne in mind that indicators of student out-of-school engagement could capture different aspects of students' learning activities, and that proper consideration of the context is required for accurate interpretation.

Figure 5.3. Time spent studying outside school and macro correlation with engagement in science.

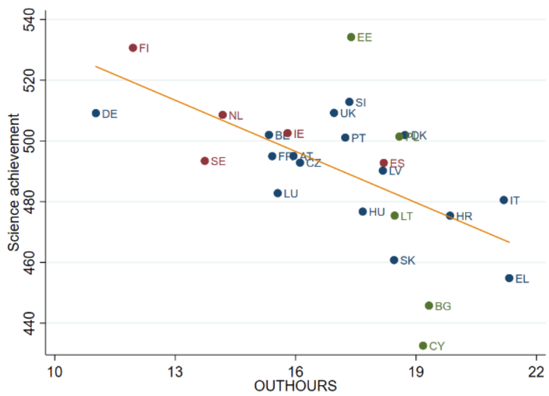
Panel a: *Country averages for time spent studying outside school.*



Panel b: *Scatter plot between times spent studying outside school and students' science activities.*



Panel c: *Scatter plot between time spent studying outside school and students' science competences.*



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Green dots: Countries with high science engagement; red dots: Countries with low science engagement; blue dots: countries with medium science engagement.

6 Effort and effort persistence

Key findings

Student effort is conceived as the extent to which students are committed to answering the questions in the test. It is measured as the difference in the response time in easy vs. difficult items at a given point of the test.

Effort persistence is a measure of the extent to which students are able to maintain their effort as they proceed through the test. It is measured as the difference between effort measured in the first and the second parts of the test.

Across EU Member States, there is high heterogeneity in both effort and effort persistence and there is a weak country-level correlation between the measures.

Boys show less effort than girls, and only partially recover thanks to their higher ability to withstand fatigue (i.e., higher effort persistence). Children of immigrants show higher difficulties in maintaining effort, especially in the reading test. Students from higher social backgrounds exhibit higher levels of effort and persistence, but when controlling for performance these associations disappear. The science score is a very strong and positive predictor of both effort and effort persistence.

Some features of the national systems are linked to student effort. Students living in countries with high vertical school stratification and higher school starting age show higher levels of effort, while highly horizontally stratified systems correlate negatively with student effort.

This section addresses the topic of students' **effort** and **effort persistence** when taking the PISA test. Sub-section 6.1 shows how student effort varies across EU Member States. Sub-section 6.2 presents the results of a series of multiple regression models aimed at establishing the relative weight of the main individual- and school-level determinants of effort. Section 6.3 sheds light on country variations in the role played by the individual and school determinants and investigates the role played by country-level education system features in affecting individuals' effort. Finally, section 6.4 provides a joint analysis of effort and effort persistence with the aim of shedding light on students' test-taking behaviour.

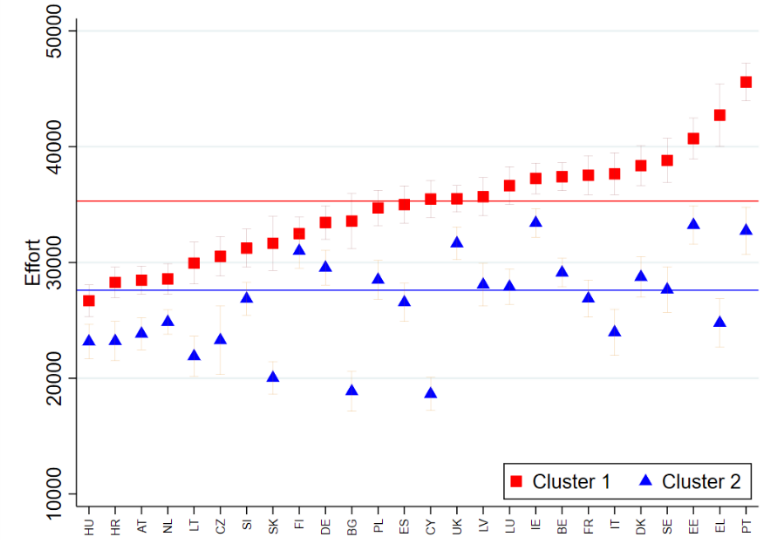
6.1 Cross-country comparisons

This section shows cross-country variations of the effort indices computed – as illustrated in section 3.3 - in the first and second clusters, as well as correlations with country average performances on the tests.

Figure 6.1 exhibits a number of interesting patterns. First, the levels of **effort** (cluster 1, red square symbols) are **heterogeneous across countries**. The country with the lowest level of effort (Hungary) records an average value of 27 seconds, while the country with the highest value (Portugal) records 46 seconds. A significant number of countries (10) show below-average effort levels and 9 other countries score well above the EU level (35 seconds).

When considering effort in cluster 2 (blue triangles), and its comparison with effort in cluster 1 (i.e., effort persistence), it is apparent that all countries but one (Finland) show a significant average drop in effort. The situation is again heterogeneous across countries, and, quite interestingly, there is a loose correlation between the two clusters' average efforts, as the ranking of the countries changes visibly when considering effort in cluster 2 rather than 1.

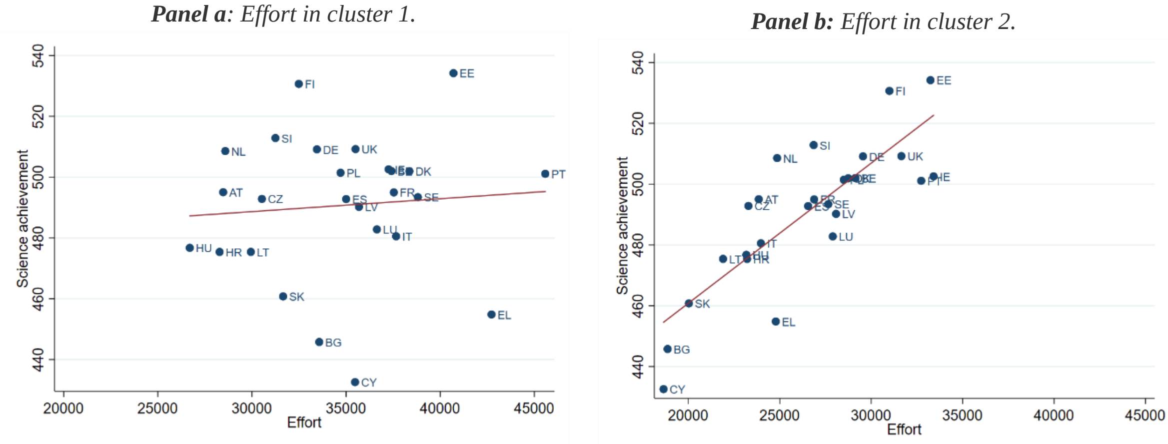
Figure 6.1. Effort in cluster 1 and in cluster 2 in EU Member States.



Note: FBK-IRVAPP analysis of PISA 2015 data. Bars are 95% CI. Horizontal lines represent the average for EU Member States in cluster 1 (red line) and in cluster 2 (blue line). Estimates obtained using PISA provided replicate weights and final student weights. Effort is measured in milliseconds. The information is not available for Malta and Romania.

Figure 6.2. Scatter plots between effort in cluster 1 and achievement in science and correlation between effort in cluster 2 and achievement in science in EU Member States.

***Panel a**: Effort in cluster 1.* ***Panel b:** Effort in cluster 2.*



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. Effort is measured in milliseconds. The information is not available for Malta and Romania.

Figure 6.2 shows the correlation between country average performance scores and effort in clusters 1 (panel a) and 2 (panel b). Both correlations are positive. The correlation between effort in cluster 1 and performance is weaker than the one between effort in cluster 2 and performance. However, being country-level correlations, they cannot be interpreted as a sign of the link between cognitive and non-cognitive competences at the individual level. This individual-level analysis is the object of the next sub-section.

6.2 The individual and school determinants of effort

Table 6.1 shows the results of three nested models on effort and effort persistence, while Table A6.1 in Appendix A6 shows the final model (model 3) separately by domain. At the individual level, boys make less effort at the start of the test than girls, but then partially 'recover' when moving from the first to the second cluster position. More precisely, depending on the model, boys exhibit about a 2.8 to 3.6 second negative effort relative to girls, but they are more persistent than girls when passing from the first to the second cluster (about 1.2 seconds). Hence, **overall boys display lower levels of effort than girls**. These patterns seem to be particularly evident in science (Table A6.1).

Even if statistical precision is reduced to the smaller sample size, children of immigrants record the same average starting levels of effort as natives, but **second-generation students are less persistent in effort than natives**. Hence, migration background, net of all other factors included in the model, seems not to affect the level of effort but does seem to affect students' ability to withstand fatigue. Analysis by domain (Table A6.1) suggests that this higher drop in effort is more pronounced in the reading test, pointing to the possible role played by poor mastery of the language of the country of residence in inducing more 'fatigue' for children of immigrants.

Children of well-educated and higher-occupation households display higher effort (tertiary educated + 4.1 secs, 1std HISEI +3.8 secs). Lower social background children do not recover in the second cluster (no statistically significant coefficients are found for effort persistence), so the gap remains. This 'advantage' is partially accounted for by <u>school characteristics</u> (e.g., the coefficient of high parental education drops from 4.1 to 2.8 seconds), signalling the importance of contextual factors and school quality in mediating the link between family background and children outcomes. Interestingly**, once performance in science is controlled for, no social-background coefficients are significant**. Science score is among the strongest predictors of effort and effort persistence, even after accounting for the entire set of individual and school level covariates. Without trying to disentangle the causal direction in the relationship between effort and performance, the latter finding clearly indicates that socioeconomic background is strongly correlated to both science performance and effort.

The test score is indeed a very strong predictor of effort. One standard deviation increase in the score is associated with an 11-second increase in effort and a 1.2-second increase in effort persistence. Hence, **high-performing students both make more effort and are more persistent**. This finding holds also when breaking down the analysis by domain. However, it appears that the net association between test scores and effort is stronger for maths (19 seconds) than for science (10.2) and reading (8.8 seconds). Moreover, when looking at effort persistence, only mathematics is statistically associated with this second index (+4.4 seconds).

Concerning the role played by schools, some statistically significant associations are found and are worth commenting on. Schools that offer more **extracurricular activities** are schools where students, on average, put in more effort (+1.6 seconds). However, this association is only marginally significant when controlling for students' performance and negative associations are detected when looking at effort persistence. Thus, students in schools that offer a lot of extracurricular activities may be more motivated but also more prone to 'fatigue'. As shown in Table A6.1 in the Appendix, this seems to be true for science and reading.

Schools with a high incidence of **student behavioural problems** are also schools where students make less effort (-1.2 seconds) and this gap persists when progressing through the test (as the coefficients related to effort persistence are small and non-significant). Once students' performance is included in the models, however, the coefficients of this school index lose significance. Teacher and parent participation never show any statistically significant association with the two effort indices.

Table 6.1. Linear regression models for effort and effort persistence, according to individual characteristics (model 1), school level factors (model 2) and performance in science (model 3). Selected parameters.

	Effort			Effort persistence		
	M1	M2	M3	M1	M2	M3
Individual level						
Male (ref. Female)	-2787.148*** (518.431)	-1973.938*** (511.416)	-3639.770*** (510.394)	1211.066* (689.238)	1343.194** (683.595)	1165.625* (693.48)
Migrant Background (ref. Natives)						
Mixed parentage	37.331 (901.184)	29.956 (893.95)	1087.826 (898.015)	-916.964 (1318.14)	-910.415 (1315.926)	-797.486 (1310.979)
Second-generation	-1396.76 (1297.314)	-1268.683 (1305.456)	1467.62 (1284.883)	-3890.603** (1892.165)	-3858.449** (1890.96)	-3560.386* (1900.135)
First-generation	-2678.919 (1689.946)	-1564.485 (1652.88)	2009.336 (1540.897)	-1934.763 (2141.477)	-1766.998 (2137.683)	-1390.357 (2165.798)
Parental education (ref. Low secondary)						
Upper secondary	2087.872* (1082.69)	1255.56 (1042.6)	-766.805 (994.064)	822.353 (1529.141)	825.908 (1516.074)	611.267 (1517.443)
Tertiary	4131.187*** (1103.053)	2820.873*** (1057.468)	319.85 (994.44)	443.266 (1437.097)	426.952 (1426.106)	161.537 (1424.852)
Highest parental ISEI (std)	3851.736*** (374.975)	2758.211*** (373.704)	345.748 (371.002)	257.392 (438.735)	227.345 (453.638)	-30.736 (464.109)
School level						
Extra-curricular activity index (std)		1566.411*** (379.208)	610.276* (339.882)		-1244.452*** (438.299)	-1346.223*** (445.96)
Proportion of parents involved in school activities (std)		432.877 (318.678)	328.797 (261.055)		-329.944 (341.931)	-340.249 (342.984)
Teacher involvement in school decision-making (std)		467.066 (384.089)	545.418 (341.004)		374.993 (416.098)	383.729 (416.226)
Negative school climate (std)		-1272.370*** (349.686)	-397.354 (331.668)		-272.666 (357.794)	-179.297 (352.65)
School achievement						
Performance in science (std)			11544.004*** (331.348)			1231.281*** (427.877)
Constant	8457.035 (13841.807)	21535.87 (13734.735)	31046.932** (13214.933)	-23099.45 (19421.61)	-23074.37 (19212.66)	-22059.92 (19295.29)
N	90006	90006	90006	88841	88841	88841
R^2	0.017	0.027	0.068	0.005	0.006	0.006
Country clusters	26	26	26	26	26	26

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects; model 2 also controls for a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

Figure A6.1 (panels a-e) in the Appendix shows the main individual regression coefficients of model 2 on effort (Table 6.1 above), run separately by country. The figure shows that boys consistently register lower levels of effort than girls, although in some countries the difference is not statistically significant. The migration and socioeconomic background gaps exhibit more variability, but they point to overall non-significant associations, in line with what was found in the pooled model shown above. Panel f of the figure shows the science competence regression coefficients of model 3 on effort run separately by country: the effects of test scores are always strongly significant.

Table 6.2. Linear regression model for the likelihood of spending more time on easy items than on difficult items according to individual characteristics, school level factors and performance in science. Selected parameters.

	Spending more time on easy items than on hard items
Individual level	
Male (ref. Female)	0.022*** (0.003)
Migrant Background (ref. Natives)	
Mixed parentage	-0.002 (0.008)
Second-generation	0.008 (0.01)
First-generation	-0.004 (0.011)
Parental education (ref. Low secondary)	
Upper secondary	0.01 (0.01)
Tertiary	0.008 (0.009)
Highest parental ISEI (std)	<0.001 (0.003)
School level	
Extra-curricular activity index (std)	-0.008*** (0.003)
Proportion of parents involved in school activities (std)	-0.004* (0.002)
Teacher involvement in school decision-making (std)	-0.002 (0.002)
Negative school climate (std)	0.004* (0.002)
School achievement	
Performance in science (std)	-0.065*** (0.002)
Constant	0.224** (0.106)
N	83894
R^2	0.037
Country clusters	26

Note: FBK-IRVAPP analysis of PISA 2015 data. Model controls for age of the student, grade retention and ISCED-level attended, country fixed effects; a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

Let us now consider a specific and partly unexpected test-taking behaviour. While, on average – as shown in section 3.3.1 – students spend more time on difficult items than on easy items, there is a non-negligible fraction of students who do the opposite: i.e., they spend more time on easy items than on difficult items. These students are labelled **"contrarian" students** and **account for nearly 16% of the sample**.[25] It is

[25] "Contrarians" are students who register an effort value lower than one tenth of a standard deviation of effort (- *SD_eff * 0.10,* where *SD_eff* stands for the effort standard deviation in the three domains and

important to consider how this behaviour correlates with other student and school characteristics, since it may be induced by very poor test-taking motivation (or non-seriousness); or it could also be an indicator of poor ability (e.g., students spend less time on difficult items because they do not know how to answer them and they simply use 'guessing' as a dominant strategy).[26] Table 6.2 shows the results of a model of the probability of being "contrarian". The model shows that this behaviour is more frequent among boys and lower performing students. The model also shows that some school practices may reduce non-seriousness in test taking (i.e., providing extracurricular activity and, marginally, also involving parents in teacher-parent meetings, positive school climate).

6.3 The role of national education systems

Table 6.3 sets out the coefficients of country-level characteristics on effort. It is important to recall that this report studies statistical correlations without making any attempt at causal inference. That said, the table shows that, net of the set of individual and school-level characteristics listed included in model 3, **some features of the national systems** seem to be statistically correlated with effort. In particular, students in highly horizontally stratified countries exhibit lower levels of effort. On the other hand, in countries with high levels of student selection across grades (i.e., high levels of grade retention) and higher school starting age, students tend, on average, to display higher effort levels.

ranges between 39 and 47 seconds). The share of contrarians varies across domains: it is lower in science and higher in mathematics and reading.

[26] Akyol et al. (2018) provide various estimates of the fraction of non-serious students by looking at response time and missing item patterns. However, their estimates vary widely (between 4 and 25%) according to the different criteria employed.

Table 6.3 Linear regression model for effort according to country system characteristics. Selected parameters.

	M4
School-system level	
Horizontal differentiation (ref. Low)	
Medium horizontal diff	328.89 (1286.41)
High horizontal diff	-6588.56*** (1046.78)
High vertical differentiation (ref. Low)	4451.85*** (973.40)
Compulsory starting age (ref. 5 years old)	
6 years old	11132.32*** (1776.25)
7 years old	13527.86*** (1609.44)
Compulsory leaving age (ref. 15 years old)	
16 years old	-87.10 (792.60)
18 years old	735.42 (1377.69)
Country system autonomy	-2288.09 (2700.22)
Country science achievement (mean)	-3.932 (15.43)
Country GDP	0.090* (0.050)
Constant	7650.10 (15305.85)
N	90006
R^2	0.025

Note: FBK-IRVAPP analysis of PISA 2015 data. The model controls for the same list of variables included in M3 with the exception of country fixed effects. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01. Data on compulsory starting and leaving ages come from *Eurydice 2018;* GDP comes from *OECD 2016a*

6.4 A focus on effort persistence and student test-taking behaviour

This section investigates which student and school characteristics are predictive of the effort persistence typology, as defined in section 3.3.2. A set of multinomial logistic regressions is run and the M1 and M3 results are shown below in Table 6.4. **Boys are less likely to be "hard working" and "hasty"** than girls. Conversely, boys are more likely to be work-shy. These gaps remain significant even after controlling for science performance. Hence, it is once again confirmed that boys make less effort than girls in answering the test.

Children of immigrants are less likely to fall within the first two profiles. While for first- and the second-generation children the only strongly significant difference is found in regard to "hard-working", mixed-parentage children are also less likely than natives to be hasty. Taken together, these results confirm that **children of immigrants struggle more than natives in maintaining their effort throughout the test**. However, once test performance is controlled for, all differences lose statistical significance.

Children of parents with higher education levels and higher socioeconomic positions exhibit a clear advantage, as they are more likely than their counterparts to be hardworking and hasty. However, these differences also disappear once test performance is modelled. This pattern is explained by the correlation of science performance with student test-taking behaviour. It is largest for profile 1 (hard working) and lowest for profile 4 (work-shy).

Among school-level variables, it is confirmed that **school climate** is associated with student test-taking behaviour: the presence in the school of students exhibiting deviant behaviours reduces students' likelihood of being hard working relative to being work-shy. The analysis shows that **teacher involvement** could also play a role. Although the statistical significance of the estimated coefficients is generally weaker, it seems that students attending schools where teachers are actively involved in school management are less likely to be work-shy in comparison to their counterparts attending schools with lower teacher involvement.

Table 6.5 shows how the **different profiles achieve, on average, different competence levels in science**. There is a clear order, even for profiles in the middle - "slow starter" and "hasty" - whose position was not theoretically obvious. Students who start with high levels of effort also perform better on the test. Results are similar (but not shown here) for the other domains.

Table 6.4 Multinomial logistic regressions for students' test-taking behaviour according to individual characteristics (model 1), school level factors and performance in science (model 3). Selected log-odds parameters.

	M1			M3		
	Slow starter	**Hasty**	**Hard working**	**Slow starter**	**Hasty**	**Hard working**
Individual level						
Male (ref. Female)	0.003 (0.039)	-0.108*** (0.04)	-0.114** (0.047)	-0.009 (0.04)	-0.146*** (0.042)	-0.186*** (0.048)
Migrant Background (ref. Natives)						
Mixed parentage	-0.133* (0.068)	-0.135** (0.058)	-0.219** (0.09)	-0.112 (0.069)	-0.089 (0.06)	-0.147 (0.092)
Second-generation	-0.078 (0.086)	-0.04 (0.085)	-0.311** (0.127)	-0.013 (0.087)	0.102 (0.091)	-0.09 (0.138)
First-generation	-0.084 (0.077)	-0.174* (0.098)	-0.377*** (0.123)	0.024 (0.074)	0.051 (0.093)	-0.018 (0.119)
Parental education (ref. Low secondary)						
Upper secondary	0.049 (0.062)	0.105 (0.068)	0.262*** (0.075)	-0.018 (0.064)	-0.034 (0.07)	0.044 (0.077)
Tertiary	0.065 (0.066)	0.209*** (0.073)	0.292*** (0.08)	-0.009 (0.067)	0.036 (0.075)	-0.014 (0.081)
Highest parental ISEI (std)	0.115*** (0.021)	0.195*** (0.02)	0.310*** (0.024)	0.034 (0.021)	0.025 (0.02)	0.033 (0.024)
School level						
Extra-curricular activity index (std)				-0.042* (0.022)	-0.006 (0.023)	-0.016 (0.026)
Proportion of parents involved in school activities (std)				0.006 (0.02)	0.017 (0.019)	0.011 (0.028)
Teacher involvement in school decision-making (std)				0.039* (0.023)	0.053** (0.025)	0.059* (0.03)
Negative school climate (std)				-0.018 (0.019)	-0.02 (0.022)	-0.059** (0.024)
School achievement						
Performance in science (std)				0.278*** (0.021)	0.586*** (0.024)	0.950*** (0.03)
Constant	-0.515 (0.935)	-0.984 (0.983)	-2.125* (1.276)	0.044 (0.926)	0.179 (0.972)	-0.239 (1.255)
N		88841			88841	
Country clusters		26			26	

Note: FBK-IRVAPP analysis of PISA 2015 data. "Work-shy" is the reference category. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects; model 3 also controls for a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01. Reference category is profile 4 (work-shy).

Table 6.5 Average science score by effort profile.

Profiles	Average science score
Work-shy	459.0
Slow starter	486.7
Hasty	515.7
Hard working	549.2

Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

7 Perseverance

Key findings

Student perseverance is computed as the difference between performance at different points in the test. Based on this index, four student profiles have been identified: "persistently good", "starts well but drops", "slow starter", and "persistently weak".

The countries with a higher proportion of "persistently good" students are the Nordic and Central European countries, while the "persistently weak" pupils are more likely to live in Eastern or Southern Europe. Countries with more "persistently good" students or with more students that "start well but drop" are also the best performing countries, suggesting that the starting point is what matters for overall performance.

Students from higher socioeconomic backgrounds, boys and native students have an advantage and show higher perseverance; however, once the correlation is controlled for science performance, the individual characteristics lose influence.

Extracurricular activities offered by schools are shown to be positively correlated with students' perseverance.

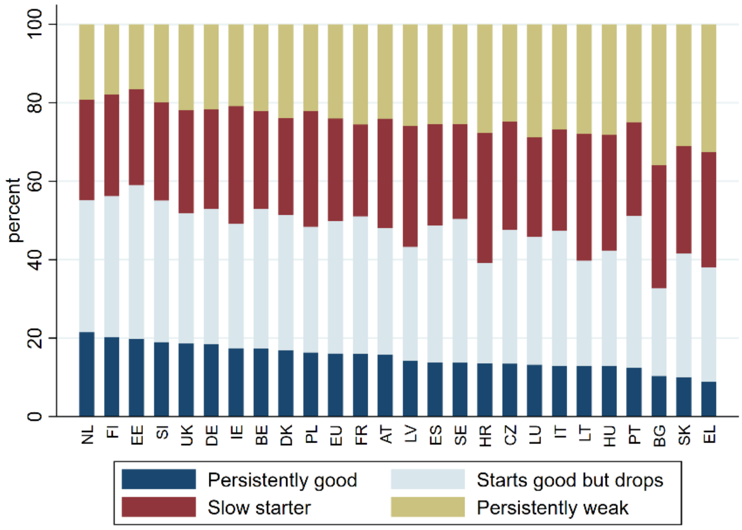
This section is devoted to the analysis of perseverance. In sub-section 7.1, some cross-country comparative figures are shown and commented. In sub-section 7.2, the focus is on the main determinants at the individual and school levels, while the role played by the school-system factors is explored in sub-section 7.3.

7.1 Cross-country comparisons

This section reports how the **perseverance typology**, introduced in section 3.3, is distributed across countries and how this typology correlates with country average performance on the tests.

Figure 7.1 shows the **proportion of students** belonging to each group by country of residence. The countries with a higher proportion of "persistently good" students are above all countries in North and Northwest of Europe (The Netherlands, Finland and Estonia), together with two of the Central European countries (Germany and Slovenia) and the United Kingdom, while the "persistently weak" pupils are more likely to live in Eastern or Southern Europe (and Luxembourg). Two extreme examples are the Netherlands and Greece: in the first, the share of students who perform persistently well is over 20%, dropping to about 10% in Greece. Looking at 'persistently weak' students, this fraction is highest in Bulgaria (nearly 38%) and lowest in Estonia (less than 20%).

Figure 7.1. Proportion of students belonging to the four perseverance categories in EU Member States.



Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania.

In general, countries with more "persistently good" students or with more students that "start well but drop" are also the best-performing countries, as can be seen from Figure 7.2. Even with the "slow starter" students the correlation is negative, suggesting that what matters most for overall performance is the starting point. For the sake of brevity, only the science test score is considered here, but the results are qualitatively the same when the other two domains are considered. This result is not surprising, because the typology is built according to the performance in science in the different clusters. What is interesting is the negative association between the proportion of slow starters and the science score. This means that the recovery of these students in the second cluster is not enough to prevent them from being low-performers.

The relationship between the typology and the test scores clearly emerges from a simple analysis that looks at the average performance of the four groups. Table 7.1 shows that those who start above the average (i.e., "persistently good" and "starts well but drops") achieve the highest test scores. "Slow starter" students are unable to close the gap with students who perform very well in the first cluster, confirming the results that emerged from the macro analysis.

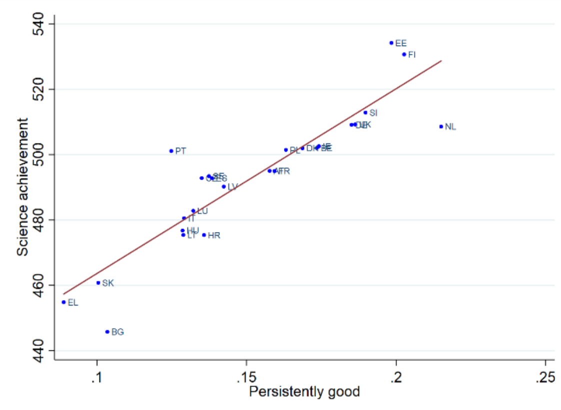
Table 7.1 Average science score by the perseverance typology.

Profiles	**Average science score**
Persistently good	583.2
Starts well but drops	549.8
Slow starter	453.3
Persistently weak	422.8

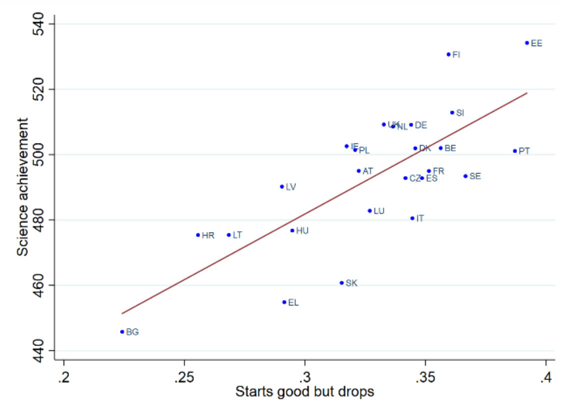
Note: FBK-IRVAPP analysis of PISA 2015 data. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania.

Figure 7.2 Scatter plots between the average science score and the proportion of students belonging to each of the four categories of the perseverance typology.

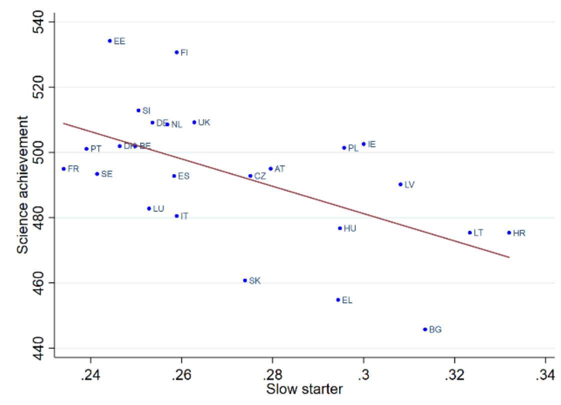
***Panel a**: Science achievement and "persistently good" students.*



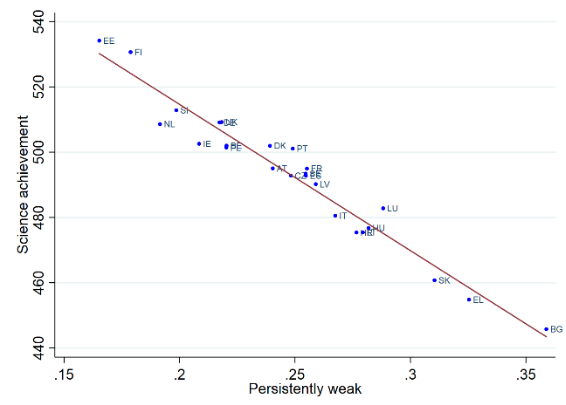
***Panel b**: Science achievement and "starts well but drops" students.*



***Panel c**: Science achievement and "slow starter" students.*



***Panel d:** Science achievement and "persistently weak" students.*



Note: FBK-IRVAPP analysis of PISA 2015 data. The information is not available for Cyprus, Malta and Romania.

7.2 Individual and school determinants of perseverance in PISA test taking

Table 7.2 shows the results of a multinomial logistic regression analysis using "persistently weak" as reference category. The first model considers only individual characteristics; the second also controls for school level variables; in the third, the score on the science test is added as a further covariate. The first noteworthy result is that, considering individual level characteristics, **students from higher socioeconomic backgrounds display higher chances of being "persistently good"** than "persistently weak" in comparison with students from less well-off and less educated families. Moreover, the size of the coefficients decreases from "persistently good" (showing the largest coefficients) to "slow starter" (showing the smallest coefficients). Looking, for example, at the influence of parental HISEI in M1, the parameter expressing the chance of being "persistently good" instead of "persistently weak" is 0.600; this becomes 0.471 for the "starts well but drops" and 0.129 for the "slow starter". The same pattern is also observable for parental education, gender and migration background.

Boys systematically outperform girls. This result can in large part be explained by the overall higher proficiency of boys in science, which is also confirmed by M3, where the inclusion of the science test scores eliminates any gender difference.

The results clearly show that **children of immigrants underperform natives** in all parts of the test. Contrary to what was found when looking at effort persistence (6.4), however, the most fragile group is no longer mixed-parentage children, but the first generation. Overall, these results are largely in line with all evidence on the underperformance of immigrants' children on PISA test scores.

It is important to stress that the influence of these individual characteristics remains statistically significant once school-level characteristics are added to the model. What makes the difference is, rather, the inclusion of test performance among the covariates (Model 3). **Science performance wipes away all statistically significant associations** except two coefficients related to parental HISEI. This result is a consequence of the very strong link that exists between the perseverance typology and the test performance (as shown earlier in Table 7.1) and the role played by test performance as an 'intervening' variable between the socio-demographic factors and perseverance.

The role played by school-level factors is examined in model 2. The results are substantially in line with the evidence that emerged from the analyses of effort and truancy. **Extracurricular activities positively influence student perseverance**, i.e. they increase the chance of being in the first two categories of perseverance, which are also the categories featuring better performance in science ("persistently good" and "starts well but drops"). The same pattern is found for the **school climate** index. In this case, the sign of the parameters is negative, indicating that **schools with behavioural problems negatively influence their students' perseverance**. On the other hand, neither parental involvement nor teacher involvement exerts any significant influence on perseverance.

Additional models were estimated in order to investigate whether the results described above were **homogeneous** across the **EU countries**. Due to computational problems the cross-country analyses are run by recoding the four-category outcome in a dummy variable that takes value 1 for the "Persistently good" and 0 otherwise. The results of these models are not shown here, but reported in Figure A7.1 Appendix 7, because the emerging pattern substantially confirms what is found in the pooled model: on average, boys, native students and those from advantaged socioeconomic backgrounds are more likely to be "persistently good" than less privileged students.

Table 7.2 Multinomial logistic regression for the perseverance typology, according to individual characteristics (model 1), school level factors (model 2) and performance in science (model 3). Selected log-odds parameters.

	M1			M2			M3		
	Persistently good	**Starts well but drops**	**Slow starter**	**Persistently good**	**Starts well but drops**	**Slow starter**	**Persistently good**	**Starts well but drops**	**Slow starter**
Individual level									
Male (ref. Female)	0.178*** (0.038)	0.106** (0.04)	-0.026 (0.038)	0.314*** (0.036)	0.229*** (0.039)	0.013 (0.038)	0.062 (0.038)	0.058 (0.042)	-0.01 (0.037)
Migrant Background (ref. Natives)									
Mixed parentage	-0.123 (0.067)	-0.094 (0.057)	-0.071 (0.068)	-0.126 (0.066)	-0.102 (0.058)	-0.076 (0.068)	0.088 (0.075)	0.068 (0.064)	-0.038 (0.067)
Second-generation	-0.471*** (0.114)	-0.258*** (0.077)	-0.088 (0.081)	-0.479*** (0.112)	-0.279*** (0.076)	-0.1 (0.08)	0.002 (0.131)	0.103 (0.095)	-0.02 (0.083)
First-generation	-0.722*** (0.112)	-0.612*** (0.085)	-0.290*** (0.073)	-0.582*** (0.11)	-0.502*** (0.085)	-0.249*** (0.075)	0 (0.129)	-0.033 (0.106)	-0.141 (0.076)
Parental education (ref. Low secondary)									
Upper secondary	0.391*** (0.08)	0.313*** (0.055)	0.072 (0.065)	0.311*** (0.084)	0.237*** (0.058)	0.048 (0.065)	-0.038 (0.099)	-0.046 (0.071)	-0.024 (0.066)
Tertiary	0.494*** (0.083)	0.370*** (0.065)	0.052 (0.066)	0.349*** (0.088)	0.248*** (0.067)	0.021 (0.067)	-0.13 (0.102)	-0.098 (0.082)	-0.029 (0.067)
Highest parental ISEI (std)	0.600*** (0.025)	0.471*** (0.02)	0.129*** (0.019)	0.473*** (0.026)	0.363*** (0.02)	0.094*** (0.019)	0.123*** (0.027)	0.083*** (0.022)	0.03 (0.018)
School level									
Extra-curricular activity index (std)				0.170*** (0.031)	0.129*** (0.024)	0.028 (0.024)	0.009 (0.029)	0.011 (0.022)	0.006 (0.023)
Proportion of parents involved in school activities (std)				0.049 (0.027)	0.012 (0.024)	0.011 (0.018)	0.04 (0.021)	-0.001 (0.02)	0.005 (0.017)
Teacher involvement in school decision-making (std)				-0.023 (0.032)	0.007 (0.023)	0.002 (0.018)	-0.002 (0.028)	0.026 (0.023)	0.004 (0.018)
Negative school climate (std)				-0.146*** (0.03)	-0.111*** (0.026)	-0.018 (0.021)	-0.006 (0.027)	<.001 (0.024)	0.006 (0.02)
School achievement									
Performance in science (std)							2.826*** (0.043)	2.219*** (0.033)	0.483*** (0.028)
Constant	-4.164*** (1.063)	-3.162*** (0.884)	0.197 (0.927)	-2.457** (1.114)	-1.637 (0.878)	0.63 (0.937)	-1.454 (1.195)	-0.498 (1.011)	1.041 (0.929)
N		86895			86895			86895	
Country clusters		25			25			25	

Note: FBK-IRVAPP analysis of PISA 2015 data. "Persistently weak" is the reference category. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects; model 3 also controls for a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Romania or Malta. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01. Reference category is profile 4 (persistently weak).

7.3 The role played by the national education systems

Finally, Table 7.3 shows a model including **country institutional characteristics**. The emerging patterns indicate that horizontal differentiation is detrimental for student perseverance. The parameter expressing the relationship between this variable and the probability of being in the better categories of perseverance with respect of "persistently weak" is always negative. On the other hand, the results for the vertical differentiation are less clearly interpretable. Other variables that are positively correlated with perseverance are compulsory start age, system autonomy and overall country science performance.

Table 7.3 Multinomial logistic regression for the perseverance typology, according to the characteristics of the education system. Selected parameters.

		M4	
School-system level	**Persistently good**	**Starts well but drops**	**Slow starter**
Horizontal differentiation (ref. Low)			
Medium horizontal diff	-0.611*** (0.094)	-0.475*** (0.089)	-0.251** (0.089)
High horizontal diff	-0.0562*** (0.099)	-0.674*** (0.081)	-0.231** (0.085)
High vertical differentiation (ref. Low)	0.028 (0.078)	0.198*** (0.055)	-0.069 (0.051)
Compulsory starting age (ref. 5 years old)			
6 years old	1.986*** (0.169)	1.901*** (0.136)	0.572*** (0.148)
7 years old	1.981*** (0.167)	1.762*** (0.135)	0.599*** (0.137)
Compulsory leaving age (ref. 15 years old)			
16 years old	-0.027 (0.073)	-0.107* (0.054)	-0.031 (0.043)
18 years old	0.283* (0.112)	0.101 (0.082)	0.144* (0.067)
Country system autonomy	1.206*** (0.279)	1.405*** (0.239)	0.469* (0.220)
Country science achievement (mean)	0.025*** (0.001)	0.017*** (0.001)	0.001*** (0.001)
Country GDP	<.001 (<.001)	<.001 (<.001)	<.001 (<.001)
Constant	-18.663*** (1.233)	-13.284*** (0.907)	-4.757*** (1.061)
N		86895	

Note: FBK-IRVAPP analysis of PISA 2015 data. "Persistently weak" is the reference category. The model controls for the same list of variables included in M3 with the exception of country fixed effects. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01. Data on compulsory starting and leaving ages come from *Eurydice 2018;* GDP comes from *OECD 2016a.*

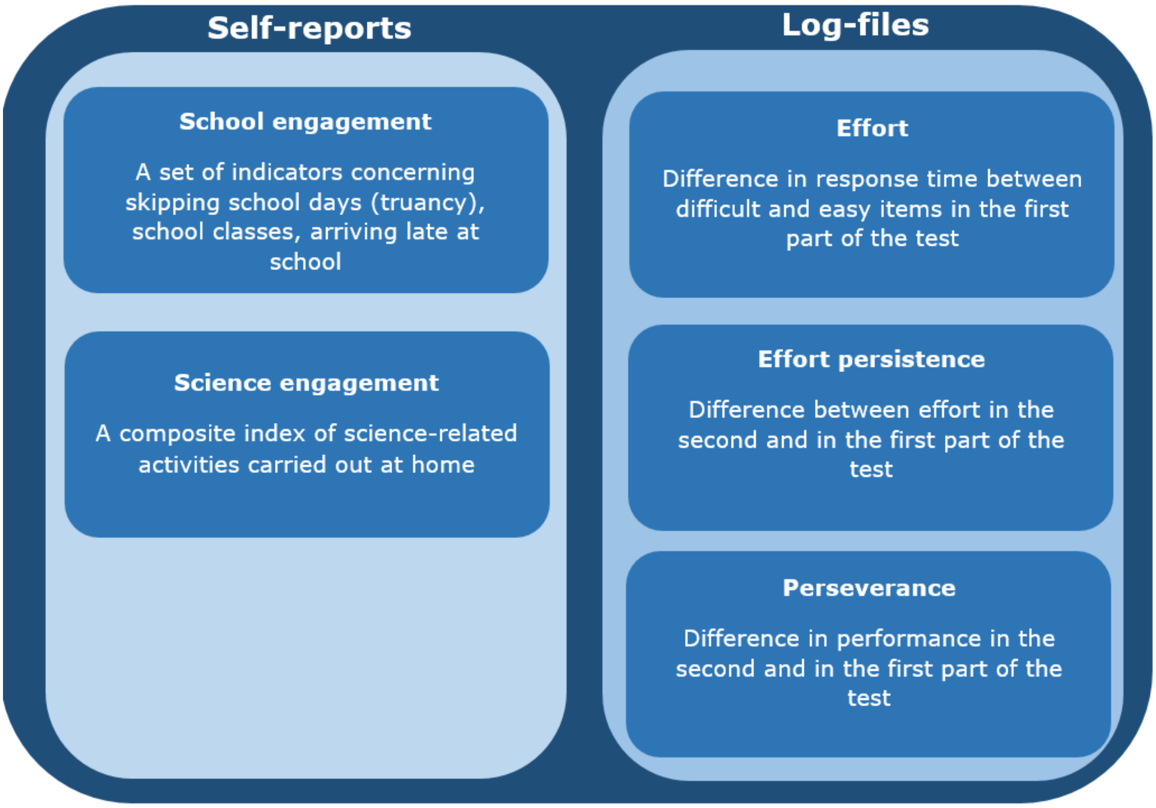
8 Summary and conclusions

This final section summarises the main empirical findings of the study regarding the determinants of engagement, effort and perseverance across the EU Member States (sub-section 8.1). The empirical results clearly point to the prominence of individual factors over those of the schools and education systems in shaping young people's non-traditional competences. For this reason, a "profiling" exercise is carried out to identify the groups of students with the lowest competences (sub-section 8.2) and thereby to identify the students who are more in need of external support. In the last sub-section, 8.3, the correlational evidence produced in this report and the most solid and recent programme evaluation literature are jointly considered, with the goal of deriving some policy suggestions.

8.1 A comprehensive look into students' non-traditional competences

This study is among the very few that exploit computer-generated data to attempt new methodological approaches for measuring students' non-traditional competences. This has been done by using the large dataset from the PISA 2015 standardised tests. The study has focused on three main non-traditional competences, namely engagement, effort and perseverance. The set of indices built and examined in this study is summarised in Figure 8.1. The figure differentiates the indices based on students' self-reports (i.e., school engagement and science engagement) from those elaborated exploiting the computer log-files (i.e., effort, effort persistence, and perseverance).

Figure 8.1 The non-traditional competence indicators measured in the study.



Given the early stage of educational research on the use of computer data to measure competences, more studies will certainly be needed in the near future to refine the operationalisation and measurement proposed in this report as well as to delve deeper into the links between non-traditional and traditional competences and their respective determinants at the individual and school levels. Moreover, future studies should pay attention to other testing settings. PISA is a low-stakes exam, and this feature may have important consequences with regard to how students behave and hence on the interpretation of the results (Akyol et al. 2018).

With these caveats in mind, this study has provided some interesting findings that are worth summarising here insofar as they can inform policy making and contribute to future research development.

First take-away message

Non-traditional competences (i.e., engagement, effort, perseverance) correlate positively with traditional competences (i.e., test performance). However, this correlation is not very strong. This suggests that the “traditional” and “non-traditional” indicators examined in this study - even if all partly related to the same latent concept of student motivation - capture different sub-dimensions of this concept.

First, the empirical analysis has pointed out the existence of weak, though significant, correlations between traditional and non-traditional competences and between the different non-traditional competences under study (Table 8.1). Although these correlations are generally weak – possibly because of imprecise measurement – they match theoretical expectations. Students who perform well on the standardised test also display higher levels of school participation and higher engagement in out-of-school study-related activities. High-performing students also show higher effort when answering the test and seem to be less affected by fatigue, as their effort and their performance is maintained at higher levels throughout the test. The correlation between truancy and other non-traditional competences is very weak and, when significant, is negative: students who skip school also make less effort and have less perseverance on the test. Science engagement correlates positively with effort, but negatively with effort persistence and perseverance. These correlations are statistically significant, but very small and therefore of little substantive significance. In line with expectations, students who make more effort on the test are also those whose performance levels decrease less during the test. Finally, the negative correlation found between effort and effort persistence is explained by the fact that a decrease in effort is more likely among those who put more effort into the initial phases of the test.

Table 8.1 Correlation between traditional and non-traditional competences and among non-traditional competences.

Pair		Correlation
Science achievement	Truancy	-
Science achievement	Science engagement	+
Science achievement	Effort	+
Science achievement	Effort persistence	+
Science achievement	Perseverance	++
Truancy	Science engagement	0
Truancy	Effort	-
Truancy	Effort persistence	0
Truancy	Perseverance	-
Science engagement	Effort	+
Science engagement	Effort persistence	-
Science engagement	Perseverance	-
Effort	Effort persistence	--
Effort	Perseverance	+
Effort persistence	Perseverance	+

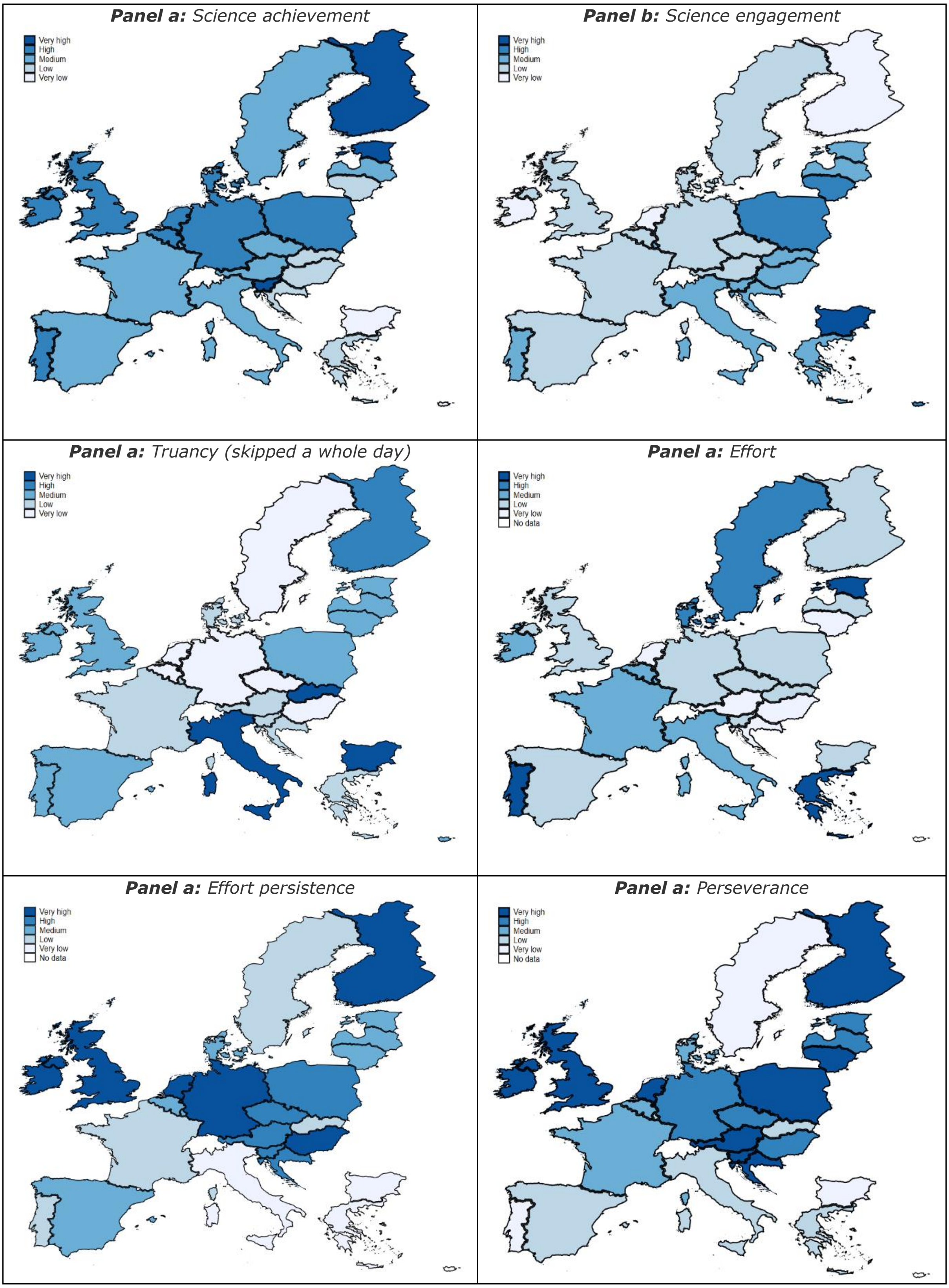
Note: +/- weak correlation (<.30); ++/-- medium correlation (>=.30 and <.70); +++/--- strong correlation; 0 no correlation (at 5% level). Perseverance is measured as a dummy variable taking value 1 for students who do not worsen their performance and 0 for those who worsen their performance throughout the test.

Second take-away message

There is **pronounced country heterogeneity** in non-traditional competences across the European Union. Top-performing countries on the standardised tests are not necessarily top-performing countries in terms of non-traditional competences.

The second important result emerging from the analyses presented in the report is that non-traditional competences show some noticeable **variation across Europe** (Figure 8.2). While average science competence is higher in Northern Member States compared with Southern ones, science engagement is highest in South-Eastern countries. The other indicators (truancy, effort, effort persistence and perseverance) are spread across the continent without clearly noticeable patterns. Hence, no clear-cut association between traditional and non-traditional competences is found when performing country-level analyses.

Figure 8.2 Maps of the studied competences across EU-28 countries



Note: Darker colours indicate higher levels of the indicators.

Third take-away message

What matters most for young people's development of non-traditional competences are **individual characteristics**. Parental education and immigrant background rank among the most important ascriptive factors shaping youths' non-traditional competences. Even if school characteristics play a smaller role, the provision of extracurricular activities and a positive school climate have the potential to make a difference for students' competences.

The third result deserving special attention is the role played by student characteristics, school factors and education system features. A variance-decomposition analysis has been performed in order to estimate what portion of variance of each of the competence indicators is explained by country, school factors and student characteristics. The results of this analysis are reported in Table 8.2. They show that the widest proportion of the variation in non-traditional competences derives from **individual-level** variation within schools. Schools account, at best, for 5 percent of the total variation. The country heterogeneity shown in the maps above does not contradict the low proportion of explained variance at the country level. The variance decomposition indicates that the variance between countries (or schools) is simply much lower than the variance within countries (or schools). In any case, on some indicators (those collected via the questionnaire), a non-negligible country-level variation is detected (ranging between 7.2% and 10.6%).

Table 8.2 Variance decomposition by country, school and individual levels (%).

Variance	Science engagement	Truancy	Effort	Effort persistence	Perseverance	Science achievement
Country	7.2	10.6	1.1	0.5	1.1	7.5
School	4.2	3.7	3.5	<0.1	1.1	35.5
Student	88.6	85.7	95.5	99.5	97.8	57.1
Total	100.0	100.0	100.0	100.0	100.0	100.0

The results above reveal a substantially higher influence of schools on science scores (35.5% of the total variance) in comparison to the other outcomes considered. This result calls into question the role played by schools with regard to students' overall development, beyond the cognitive dimension. Before addressing this point, let us briefly summarise which specific individual and school factors matter for students' non-traditional competences.

Table 8.3 shows a summary of the results presented in the previous sections regarding the role of individual and school-level factors. In general, **boys** are more likely to skip school. On average, they devote less effort to the test, but they also seem to be more persevering and more interested in science activities outside school. **Non-native students** show a greater engagement in science, but also have low engagement in school and a very low ability to maintain effort and persist in performance. The covariates related to **socioeconomic background** (parental education and HISEI) go in the same direction: students with more educated parents and parents in higher-level occupations display higher school and science engagement, stronger effort and greater perseverance.

Table 8.3. Summary of the main results.

	Individual level				School level			
	Male	Children of immigrants	Education	Highest parental ISEI	Extra-curricular activities	Parental involvement	Teacher involvement in school decision-making	Negative school climate
Truancy (whole day)	+	+	-	-	-	-	-	+
Science engagement	+++	++	+	+				-
Effort	-		+	+	+			-
Effort persistence	+	-			-			
Perseverance	+++	---	+++	+++	+			-
Science	+	--	++	++	+			-

Note: the table reports the sign of the parameters estimated through model 2 in the various sub-section. One sign (+/-) indicates a significant but modest effect size (up to 0.15), two signs (++/--) indicate a relevant effect size (0.15-0.30) and three signs (+++/---) are reserved for a huge effect size (greater than 0.30). The effect size is calculated by dividing the coefficient by the standard deviation of the outcome.

For school-level variables, the most important factors are **extra-curricular activities** and **school climate** (i.e., the share of students with forms of misbehaviour that can hinder learning). The influence of the latter goes in the direction predicted: it is positively correlated with the likelihood of truancy and is negatively correlated with effort, science engagement and persistence. On the other hand, extra-curricular activities are positively correlated with effort and perseverance, but are negatively associated with the risk of skipping a whole day of school and with effort persistence. These results lead to an important conclusion. The factors linked to **classic dimensions of social inequalities** (gender, immigrant and social backgrounds) play a crucial role not only in traditional competences, but also in non-traditional ones.

8.2 Who are the most-in-need students?

Fourth take-away message

The empirical evidence gathered in this study points to the fact that **some specific student profiles are in greater need than others** of external intervention.

- If the goal is enhancing young people's engagement in science, then priority should be assigned to girls, as an attempt to also enhance gender equality in STEM education.
- To improve school engagement and reduce truancy, the first group to be targeted should be boys and girls whose parents have low levels of education.
- Regarding effort, special attention should again be devoted to children of parents with low levels of education, including natives, who seem to be the most vulnerable group in this dimension.
- With regard to students' effort persistence and perseverance, the findings suggest that children of immigrants are the most vulnerable group: in part this may be a consequence of their low mastery of the test language, which imposes an additional cognitive load on them.

The individual-level variables affecting non-traditional competences can be exploited to identify **profiles of students who are more at risk** of showing low levels of non-traditional and traditional competences. Table 8.4 presents the average values of non-traditional and traditional competences for a set of profiles that jointly consider gender, parental education and migration background.[27] White cells indicate student profiles displaying scores in traditional and non-traditional competences that are close to or substantially higher than the average. Blue cells indicate profiles of students who score substantially below the average. Dark-blue cells identify students who score significantly lower than the average on the different indicators, while the light-blue ones identify students scoring below the average – though not significantly. The latter are highlighted because they mainly identify children of immigrants. The sample size for the non-native population is limited, and this is the main reason that the results turned out to be statistically non-significant.

Table 8.4 A profile of European 15-year-olds' non-traditional and traditional competences.

Profiles	Science engagement	Skipping one school day	Effort	Effort persistence	Perseverance	Science
Female non-native/ low parental education						
Female non-native/ high parental education						
Female native/ low parental education						
Female native/ high parental education						
Male non-native/ low parental education						
Male non-native/ high parental education						
Male native/ low parental education						
Male native/ high parental education						
Pooled average	-0.02	0.21	35314.5	-7319.40	-1.92	496.9

Note: Blue-highlighted cells indicate student profiles who perform less well than the pooled average by at least one-fifth of a standard deviation. Dark-blue cells indicate statistically significant differences. Light blue cells indicate non-statistically significant differences. White cells indicate student profiles who perform as well as the pooled average or above.

All in all, the table highlights that some student profiles perform worse than others: this is the case for **children of parents with low levels of education**, regardless of migration status, while children of high-educated parents tend to perform relatively well on all indicators. Some peculiarities emerge when looking at single indicators separately. With regard to science engagement, the gender dimension stands out: **girls systematically show lower-than-average science engagement**, regardless of the other background characteristics. **Truancy**, on the other hand, reveals the

[27] For this exercise, parental education and migration background are coded as dummy variables, where highly-educated parents are those with a tertiary degree and the non-native category considers children of two foreign-born parents. For the sake of simplicity and the limited sample size, parental HISEI is not included in this exercise. It can be assumed, though, that it would by and large yield the same results as those obtained using parental education, given the strong correlation between the two background indicators.

prominence of parental education, and underlines the **critical situation of native boys whose parents have low levels of education**.

When it comes to effort, what matters most is once again parental education. Specifically, **native boys whose parents have low levels of education show**, on average, **low effort levels**, drawing attention once again to the specific fragility of native working-class boys. **For effort persistence and perseverance** – the two indicators that measure students' capacity to withstand fatigue during the test – **what matters most is undoubtedly immigrant background**. A possible explanation is that language may constitute a further "cognitive load" that affects persistence and perseverance throughout the test. Finally, concerning the **science test**, it is a combination of immigrant background and parental education that makes the difference: **the students who perform substantially better**

8.3 Policy implications and future research

Fifth take-away message

Programme evaluation literature on the effectiveness of school programmes supports the idea that **schools could be important for students' non-traditional competences**, especially if interventions are carried out at early ages and if they are well targeted to the students most in need. More research is needed to ascertain whether these effects are persistent and linked to other life outcomes. More *ad hoc* policy experimentations could yield further insights into effective approaches to enhancing specific young people's competences.

Recalling the ultimate goal of this study - providing empirical evidence to support policy making - this last sub-section discusses some **policy indications for future EU and national programmes**. *What can schools do to enhance students' non-traditional competences?*

Engagement, effort and perseverance are not immutable personality traits. Rather, they are competences that are acquired and that, as such, can be modified by external intervention. Hence, **schools have the potential to make a difference** to young people's non-traditional competence development.

The findings presented and discussed in this report (sub-section 8.1) depict a situation in which **schools currently play a limited role in developing young people's non-traditional competence**, compared with the role they play in developing traditional competences. However, this low influence of the school-level variables does not mean that possible programmes at this level are doomed to fail. Some limitations in the empirical analyses presented could, at least in part, account for the observed low influence of schools. First, only correlations are shown and no causal effects are identified. Second, the choice of variables to be modelled was heavily affected by the availability and quality of the information gathered via school questionnaires in the different countries. Third, PISA data entail a body of information about a set of activities implemented at the school level, but little is known about the precise contents of these activities.

With these limitations, this study found that some school practices are positively correlated with students' non-traditional competences. Chiefly, the provision of **extracurricular activities** and a **positive school climate** were found to systematically and positively correlate with students' engagement, effort and perseverance.

Even if the existing evaluation research on effective ways to promote students' non-traditional competence is still in its infancy, some **robust causal studies support the idea that schools can make a difference** (see Table 8.5).

Combining the descriptive evidence from this study and the empirical evidence from the programme evaluation literature, may give policy makers a clearer picture of the lines of action that could be prioritised and help them elaborate future policy experiments that could be pursued to test the precise effects of programmes on specific outcomes of interest. An important recommendation - in light of the substantial heterogeneity that exists among students and the prominent role of family background (see Table 8.4) - is that **school-based interventions, to be effective, should target the students who are most in need**.

Table 8.5 Lessons from counterfactual impact evaluation studies.

A brief overview of some specific programmes and practices that have proven successful in developing children's non-cognitive competences is presented in this box. A recent review of robust empirical studies, mainly conducted in the United States, concludes that short-term, school-based interventions can enhance a range of students' non-cognitive skills (e.g., social skills, emotional well-being, motivation, self-efficacy and self-regulation) (Siddiqui and Ventista 2018). The reviewed studies generally report low- to medium-level effect sizes. The most effective interventions involve schools and **parent collaboration**, **freedom for students to communicate and express their feelings** and regular implementation of the interventions. Overall, the available empirical evidence suggests that **investments in non-cognitive skills should start early** – between childhood and adolescence: "Prevention is more effective than remediation" (Heckman and Kautz 2013, p. 89). While it is generally agreed that the existing evidence concerning the **persistence of these effects** is insufficient, **high-quality early childhood programmes** have demonstrated lasting and beneficial effects on several non-cognitive competences (Chetty et al. 2011; Heckman and Kautz, 2013). Kautz et al. (2014) provide three examples of US-based programmes, albeit with only short-term evaluations. The first is *Tools of the Mind*, which attempts to teach **preschool and early-primary school children to regulate their social and cognitive behaviour**. The second is a programme designed to enhance the *Mindset* of children in such a way that **children believe that competences are malleable**. This programme aims to instil the idea that achievement is the result of hard work rather than innate intelligence. The third is the *OneGoal* programme, which selects and trains high-school teachers to help students apply to colleges, improve grades and test scores, and persist through college by cultivating social and emotional skills. Evidence from a large-scale meta-analysis of school-based interventions on social and emotional learning in the United States documents that: 1) the reviewed programmes have significant positive effects on social and emotional skills such as goal setting, conflict resolution and decision making; 2) **classroom teachers and other school staff** are critical for programme success; 3) interventions can be incorporated into standard educational practices; 4) interventions can be successful at all educational levels (from primary to upper secondary); and 5) effective programmes need to incorporate learning practices with **training, active forms of learning, time and attention focussed on** skill development tasks, and explicit learning objectives (Durlak et al. 2011).

Recent examples of policies implemented in **Europe** are the *Sure Start Programme* in the United Kingdom and *Entrepreneurs for Social Inclusion* in Portugal. The first programme targeted children aged 3-4 years old, with the aim of improving **social behaviour** and **child independence**. The second programme targeted teenagers (age 13-15) and aimed to reduce dropout rates and violent behaviour by working on motivation, self-control and social skills. A recent evaluation of the Portuguese programme was carried out by Martins (2017). The study showed that the programme led to a significant increase in the probability of grade progression.

Impact evaluation studies also demonstrate that measures aimed at increasing non-cognitive skills should be targeted at all stakeholders, **including families and schools**, and should also include a **training component for parents** (Avvisati et al. 2013).

Finally, there is evidence that early childhood interventions have the potential to **reduce social disparities** in cognitive and non-cognitive skill development.

In general, more research is needed to assess the extent to which specific programmes may also effectively redress gender and migration-background gaps.

References

Agasisti, T. and Cordero, J.M. (2015). "The determinants of repetition rates in European countries: insights from an empirical analysis using PIRLS 2011 data", Instituto Nacional de Evaluación Educativa-INEE.

Akyol, Ş. P., Krishna, K., & Wang, J. (2018). Taking PISA Seriously: How Accurate are Low Stakes Exams? (No. w24930). National Bureau of Economic Research.

Albano, A. D. (2013). "Multilevel modelling of item position effects". *Journal of Educational Measurement* 50 (4): 408–426.

Avvisati, F., Besbas, B. and Guyon, N. (2010). "Parental involvement in school: A literature review", *Revue d'économie politique*, 120(5), 759-778.

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E. (2013). "Getting parents involved: A field experiment in deprived schools", *Review of Economic Studies*, 81(1), 57-83.

Avvisati, F., & Keslair, F. (2017). REPEST: Stata module to run estimations with weighted replicate samples and plausible values.

Azevedo, R., Guthrie, J. T., and Seibert, D. (2004). "The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia". *Journal of Educational Computing Research* 30, 87–111.

Below, S. von, Powell, J.J.W. and Roberts, L.W. (2013). "Educational Systems and Rising Inequality: Eastern Germany after Unification", *Sociology of Education* 86(4), 362–375.

Benabou R. & Tirole J. (2003). "Intrinsic and extrinsic motivation", *The Review of Economic Studies*, 70(3), 489-520.

Berendt, B., and Brenstein, E. (2001). "Visualizing individual differences in Web navigation: STRATDYN, a tool for analyzing navigation patterns". *Behavior Research Methods, Instruments, & Computers* 33(2), 243-257.

Borghans, L., and Schils, T. (2012). "The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components". Unpublished manuscript.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). "The economics and psychology of personality traits". *Journal of human Resources* 43 (4), 972-1059.

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

Bowles, S., Gintis, H. and Osborne, M. (2001). "Incentive-enhancing preferences: Personality, behavior, and earnings". American Economic Association Papers and Proceedings 91, 155–158.

Brown, A. L. (1987). "Metacognition, executive control, self-regulation, and other more mysterious mechanisms". In F.E. Weinert and R.H. Kluwe eds., Metacognition, motivation, and understanding, 65-116. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Brunello, G. and Checchi, D. (2007). "Does school tracking affect equality of opportunity? New international evidence", *Economic Policy*, 22, 781-861.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). "The four generations of computerized educational measurement". In R. L. Linn Ed., Educational Measurement (367–407). Washington, DC: American Council on Education.

Carneiro, P., and Heckman, J.J. (2002). "The evidence on credit constraints in post-secondary schooling", *The Economic Journal* 112(482), 705-734.

Cebolla-Boado, H. (2011). "Primary and secondary effects in the explanation of disadvantage in education: The children of immigrant families in France", *British Journal of Sociology of Education*, 32(3), 407-430.

Chen, S. Y. and Ford, N. J. (1998). “Modelling user navigation behaviours in a hypermedia-based learning system: An individual differences approach”. *Knowledge and Organization* 25, 67–78.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. The Quarterly Journal of Economics, 126(4), 1593-1660.

Christenson, S. L., Reschly, A. L. and Wylie, C. eds. (2012). Handbook of research on student engagement. Springer Science & Business Media.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., and Weinfeld, F. D. (1966), Equality of Educational Opportunity, Washington, DC: U.S. Dept. of Health Education and Welfare Office of Education.

Covay, E. and Carbonaro, W. (2010). “After the bell: Participation in extracurricular activities, classroom behavior, and academic achievement”, *Sociology of Education*, 83(1), 20-45.

Cunha, F. and Heckman, J.J. (2007). “The Technology of Skill Formation”, *The American Economic Review*, 97(2), 31-47.

Demars, C., Bozhidar E., Bashkov M. and Alan B. Socha. (2013). “The Role of Gender in Test-Taking Motivation under Low-Stakes Conditions”, *Research & Practice in Assessment* 8, 69-82.

Deming, D. and Dynarski, S. (2008). “The lengthening of childhood”, *Journal of Economic Perspectives*, 22, 71–92.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., and Schellinger, K. B. (2011). “The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions”. *Child development*, 82(1), 405-432.

Eurydice (2018). Compulsory Education in Europe 2016/2017. Eurydice Report. Luxembourg: Publications Office of the European Union. Eurydice (2012). Developing Key Competences at School in Europe: Challenges and Opportunities for Policy. Eurydice Report. Luxembourg: Publications Office of the European Union.

Farkas, G. (2003). “Cognitive skills and noncognitive traits and behaviors in stratification processes”, *Annual Review of Sociology*, 29(1), 541-562.

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). “A standard international socio-economic index of occupational status”. *Social science research*, 21(1), 1-56.

Geboers, E., Geijsel, F., Admiraal, W. and Ten Dam, G. (2013). “Review of the effects of citizenship education”, *Educational Research Review*, 9, 158-173.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., and Klieme, E. (2014). “The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment”. *Journal of Educational Psychology* 106 (3), 608–626.

Greene, J. A., Costa, L. J., Robertson, J., Pan, Y., and Deekens, V. M. (2010). “Exploring relations among college students’ prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment”. *Computers & Education* 55(3), 1027-1043.

Greiff, S., Niepel, C., Scherer, R., and Martin, R. (2016). “Understanding students' performance in a computer based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files”. *Computers in Human Behavior* 61, 36-46.

Greiff, S., Wüstenberg, S., and Avvisati, F. (2015). “Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving”. *Computers & Education* 91, 92-105.

Gutman, L. M., and Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. The Education Endowment Foundation.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., and Winne, P. H. (2007). "Examining trace data to explore self-regulated learning". *Metacognition and Learning* 2, 107–124.

Hanushek E.A. and Wößmann, L. (2006). "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries", *The Economic Journal*, 116, C63-C76.

Heath, A.F., Rothon, C. and Kilpi, E. (2008). "The second generation in Western Europe: Education, unemployment, and occupational attainment", *Annual Review of Sociology*, 34, 211-235.

Heckman, J. J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition (No. w19656). National Bureau of Economic Research.

Heckman, J. J., & Kautz, T. (2014). Achievement tests and the role of character in American life. The myth of achievement tests: The GED and the role of character in American life, 1-71.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *Journal of Labor Economics* 24(3), 411-482.

Heckman, J.J., and Kautz, T. (2012). "Hard evidence on soft skills". *Labour economics* 19(4), 451-464.

Howie, L.D., Lukacs, S.L., Pastor, P.N., Reuben, C.A. and Mendola, P. (2010). "Participation in activities outside of school hours in relation to problem behavior and social skills in middle childhood", *Journal of School Health*, 80(3), 119-125.

Ikeda, M. and García, E. (2014). "Grade repetition", *OECD Journal: Economic Studies*, 1, 269-315.

Jackson, C.K. (2012). "Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina", National Bureau of Economic Research working paper no. 18624.

Jackson, M., Jonsson, J. O. and Rudolphi, F. (2012). "Ethnic inequality in choice-driven education systems: A longitudinal study of performance and choice in England and Sweden", *Sociology of Education*, 85(2), 158-178.

Jacob, B.A. (2002). "Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education", *Economics of Education Review*, 21(6), 589-598.

Jamieson-Noel, D., and Winne, P. H. (2003). "Comparing self-reports to traces of studying behavior as representations of students' studying and achievement", *Zeitschrift für Pädagogische Psychologie* 17, 159–171.

Jeske, D., Backhaus, J. and Stamov Roßnagel C.Rl. (2014). "Self- regulation during e-learning: using behavioural evidence from navigation log files", *Journal of Computer Assisted Learning* 30(3), 272-284.

Kao, G., and Tienda, M. (1995). "Optimism and achievement: The educational performance of immigrant youth", *Social science quarterly*, 76(1), 1-19.

Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B. and Borghans, L. (2014). "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success", National Bureau of Economic Research working paper No. 20749.

Kirsch, I., De Jong, J., LaFontaine, D., McQueen, J., Mendelovits, J. and Monseur, C. (2002). Reading for change: Performance and engagement across countries: Results of PISA 2000. OECD Publishing.

Kristen, C., Reimer, D. and Kogan, I. (2008). "Higher education entry of Turkish immigrant youth in Germany", *International Journal of Comparative Sociology*, 49(2-3), 127-151.

Lee, Y. H., & Chen, H. (2011). "A review of recent response-time analyses in educational testing", *Psychological Test and Assessment Modeling* 53(3), 359–379.

Martins, J., Vale, A., & Mouraz, A. (2017). “All-day schooling: improving social and educational Portuguese policies”, *International Electronic Journal of Elementary Education*, 7(2), 199-216.

Morris, S.D., and Klesner, J.L. (2010). “Corruption and trust: Theoretical considerations and evidence from Mexico”, *Comparative Political Studies*, 43(10), 1258-1285.

Muraki, E. (1992). “A generalized partial credit model: Application of an EM algorithm”, ETS Research Report Series 1, i-30.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM algorithm. Princeton, New Jersey: Educational Testing Service.

OECD (2003), Student engagement at school. A sense of belonging and participation. Results from Pisa 2000.

OECD (2006), Where immigrant students succeeds. A comparative review of performance and engagement in Pisa 2003, Paris, Oecd.

OECD (2013a). What makes schools successful? Resources, policies and practices – Volume IV.

OECD (2013b). Ready to learn: students’ engagement, drive and self-beliefs – Volume III.

OECD (2014). Who are the school truants?, PISA in Focus, 35.

OECD (2015a). Students, Computers and Learning: Making the Connection, PISA, OECD Publishing.

OECD (2015b). Skills for Social Progress: The Power of Social and Emotional Skills, PISA, OECD Publishing.

OECD (2016a). PISA 2015 results (Volume I): Excellence and Equity in Education, PISA, OECD Publishing, Paris.

OECD (2016b). PISA 2015 results (Volume II): Policies and Practices for Successful Schools, PISA, OECD Publishing, Paris.

OECD (2018), The resilience of students with an immigrant background. Factors that shape well-being, OECD Publishing, Paris.

OECD. (2017). PISA 2015 technical report. OECD Publishing, Paris.

Osborne, J., Simon, S., & Collins, S. (2003). “Attitudes towards science: A review of the literature and its implications”. International journal of science education, 25(9), 1049-1079

Pekkarinen, T., Uusitalo, R. and Kerr, S. (2012). “School Tracking and Development of Cognitive Skills”, IZA Discussion Paper, 4058.

Pepinsky, Thomas B. (2016) “A note on listwise deletion versus multiple imputation”, *Political Analysis*, 1-9.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A. and Goldberg, L. R. (2007). “The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes”, *Perspectives on Psychological Science* 2(4), 313-345.

Scherer, R., Greiff, S., and Kirschner, P. A. (2017). “Editorial to the special issue: Current innovations in computer-based assessments”, *Computers in Human Behaviour* 76, 604-606.

Schunk, D. H., Meece, J. L., and Pintrich, P. R. (2014). “Motivation in education: Theory, research, and applications” (4th ed.). Upper Saddle River: Pearson.

Siddiqui, N. and Ventista, O.M. (2018) “A review of school-based interventions for the improvement of social emotional skills and wider outcomes of education”, *International Journal of Educational Research*, 90, 117-132.

Sullivan, S., Gnesdilow, D., and Puntambekar, S. (2011). “Navigation behaviours and strategies used by middle school students to learn from a science hypertext”, *Journal of Educational Multimedia and Hypermedia* 20, 387–423.

Sundre, D. L. and Moore, D. L. (2002). “The Student Opinion Scale: A measure of examinee motivation”, *Assessment Update* 14(1), 8-9.

Taylor, M. and Johnson, R. (2002). "School Councils: Their Role in Citizenship and Personal and Social Education", National Foundation for Educational Research, Slough.

Thelk, A., Sundre, D. L., Horst, J. S., and Finney, S. J. (2009). "Motivation matters: Using the student opinion scale (SOS) to make valid inferences about student performance" *The Journal of General Education* 58(3), 128–151.

Van de Werfhorst, H.G. and Mijs, J.J.B. (2010). "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective", *Annual Review of Sociology*, 36, 407-428.

Vandenberg, R. J., and Lance, C. E. (2000). "A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research", *Organizational research methods* 3(1), 4-70.

Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (2006). Automated scoring of complex tasks in computer-based testing. Hillsdale, NJ: Lawrence Erlbaum.

Winner, E., Goldstein, T., and Vincent-Lancrin, S. (2013). Educational Research and Innovation: Art for art's sake? The impact of arts education. OECD Publishing, Paris.

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

Wößmann, L. (2000). "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence", Kiel Working Paper No. 983.

Wößmann, L. (2009). "International Evidence on School Tracking: A Review", CESIfo DICE Report 1/2009, 26-34.

Wößmann, L. (2016). "The Importance of School Systems: Evidence from International Differences in Student Achievement", *Journal of Economic Perspectives* 30(3), 3-32.

Appendices

This section includes the methodological appendices for some of the sections of the study. Because there was no need for an appendix to the introduction (section 1) or the literature review (section 2), the appendices start with the appendix to the third section, on data and methods.

Appendix A3.1 Missing values in PISA 2015 questionnaire data

Table A3.1 shows the incidence of missing values on the engagement indicators. The same statistics are then calculated for all individual (Table A3.2) and all school-level (Table A3.3) variables. When looking at the situation in the single EU Member States, Germany stands out negatively in all the engagement indicators, with about 50% of values missing in the SCIEACT index and about 18% missing in the school engagement items.

Table A3.1 Missing values on the engagement indicators, by EU Member State.

	Current science engagement		Current engagement in school					
Country	**SCIEACT index**		**Skipped a whole school day**		**Skipped some classes**		**Arrived late for school**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
AT	737	10.52	107	1.53	131	1.87	104	1.48
BE	1243	12.88	541	5.61	585	6.06	545	5.65
BG	1206	20.34	408	6.88	459	7.74	447	7.54
CZ	397	5.76	215	3.12	275	3.99	218	3.16
DE	3217	49.46	1163	17.88	1180	18.14	1155	17.76
DK	1046	14.61	568	7.93	575	8.03	554	7.74
ES	515	7.65	111	1.65	119	1.77	98	1.45
EE	173	3.10	134	2.40	126	2.26	112	2.00
FI	408	6.94	125	2.13	137	2.33	134	2.28
FR	667	10.92	264	4.32	264	4.32	239	3.91
UK	1112	7.85	575	4.06	616	4.35	583	4.12
EL	293	5.30	125	2.26	126	2.28	118	2.13
HR	290	4.99	126	2.17	129	2.22	131	2.26
HU	748	13.22	156	2.76	170	3.00	172	3.04
IE	194	3.38	94	1.64	129	2.25	121	2.11
IT	697	6.02	341	2.94	424	3.66	377	3.25
LT	456	6.99	249	3.82	277	4.25	261	4.00
LU	669	12.63	148	2.79	177	3.34	138	2.60
LV	196	4.03	94	1.93	103	2.12	99	2.03
NL	306	5.68	226	4.20	219	4.07	215	3.99
PL	73	1.63	23	0.51	38	0.85	25	0.56
PT	269	3.67	163	2.23	174	2.38	183	2.50
CY	599	10.75	256	4.60	279	5.01	276	4.95
SK	625	9.84	303	4.77	356	5.61	352	5.54
SI	418	6.53	231	3.61	246	3.84	238	3.72
SE	662	12.13	231	4.23	238	4.36	197	3.61

Table A3.2 Missing values on the main individual-level variables, by EU Member State.

Country	Immigrant background		Parental education		Parental HISEI	
	N	%	N	%	N	%
AT	194	2.77	161	2.3	421	6.01
BE	513	5.32	374	3.88	781	8.09
BG	286	4.82	130	2.19	878	14.81
CZ	166	2.41	166	2.41	491	7.12
DE	893	13.73	1079	16.59	1175	18.07
DK	386	5.39	214	2.99	1050	14.66
ES	271	4.02	104	1.54	416	6.18
EE	182	3.26	101	1.81	277	4.96
FI	179	3.04	92	1.56	252	4.28
FR	252	4.13	241	3.95	578	9.46
UK	1014	7.16	1179	8.33	1892	13.36
EL	227	4.1	40	0.72	447	8.08
HR	276	4.75	82	1.41	416	7.16
HU	126	2.23	106	1.87	480	8.48
IE	564	9.82	106	1.85	343	5.97
IT	558	4.82	283	2.44	740	6.39
LT	306	4.69	223	3.42	835	12.8
LU	253	4.77	252	4.76	527	9.95
LV	104	2.14	63	1.29	420	8.63
NL	266	4.94	89	1.65	301	5.59
PL	78	1.74	68	1.52	247	5.52
PT	323	4.41	142	1.94	400	5.46
CY	382	6.86	101	1.81	459	8.24
SK	236	3.72	91	1.43	766	12.06
SI	171	2.67	70	1.09	371	5.79
SE	337	6.17	237	4.34	466	8.54

Table A3.3 Missing values on the main school-level variables, by EU Member State.

Country	Extra-curricular activities (ACTIV)		Proportion of parents involved in school activities (SCH_SC064Q01TA)		Teacher involvement in school decision-making (SCH_TEACHPART)		Negative school climate (SCH_STUBEHA)	
	N	%	N	%	N	%	N	%
AT	104	1.48	470	6.71	0	0	174	2.48
BE	521	5.4	916	9.49	350	3.63	799	8.28
BG	133	2.24	290	4.89	0	0	199	3.36
CZ	202	2.93	121	1.76	116	1.68	121	1.76
DE	1439	22.12	1682	25.86	1494	22.97	1626	25
DK	1481	20.68	1660	23.18	1486	20.75	1560	21.78
ES	10	0.15	150	2.23	10	0.15	35	0.52
EE	43	0.77	43	0.77	0	0	0	0
FI	75	1.28	189	3.21	40	0.68	153	2.6
FR	807	13.21	728	11.92	390	6.39	598	9.79
UK	3154	22.28	3963	27.99	3050	21.54	3706	26.18
EL	125	2.26	2	0.04	0	0	12	0.22
HR	0	0	43	0.74	0	0	43	0.74
HU	393	6.95	444	7.85	327	5.78	379	6.7
IE	322	5.61	477	8.31	338	5.89	413	7.19
IT	3323	28.69	3614	31.2	3151	27.2	3503	30.24
LT	0	0	8	0.12	0	0	11	0.17
LU	165	3.11	0	0	0	0	0	0
LV	49	1.01	110	2.26	11	0.23	55	1.13
NL	1810	33.61	2344	43.53	1656	30.75	2375	44.1
PL	75	1.67	35	0.78	0	0	35	0.78
PT	1304	17.8	204	2.78	29	0.4	62	0.85
CY	80	1.44	9	0.16	0	0	6	0.11
SK	0	0	55	0.87	0	0	9	0.14
SI	488	7.62	654	10.21	468	7.31	520	8.12
SE	0	0	32	0.59	0	0	0	0

In the scientific literature on the handling of missing values, two main techniques have been developed to limit their impact on the analysis. The first refers to the use of multiple imputation (see Gelman and Hill 2006, for an introduction), a statistical technique that tries to overcome this problem by imputing the missing values according to a statistical model that must be able to predict the missingness well. In contrast, the second technique, listwise deletion, refers to a process in which all observations with at least one missing value are dropped. Both multiple imputation and listwise deletion are unbiased when missing values on the dependent variable are randomly distributed in the sample; if missing values are not at random (i.e., not purely missing at random), there is no evidence that multiple imputation performs better (Pepinsky, 2016). In this report, therefore, the listwise deletion technique has been implemented as a robustness check.

Table A3.4 Models with missing "99 category".

	SCIEACT	TRUANCY	TRUANCY (2)	LATE
Individual level				
Male (ref. Female)	0.383***	0.015***	0.025***	0.065***
	(0.010)	(0.003)	(0.004)	(0.004)
Migrant Background (ref. Natives)				
Mixed parentage	0.049***	0.030***	0.027***	0.063***
	(0.017)	(0.006)	(0.007)	(0.008)
Second-generation	0.159***	0.017*	0.041***	0.095***
	(0.031)	(0.010)	(0.012)	(0.013)
First-generation	0.343***	0.016	0.048***	0.084***
	(0.029)	(0.011)	(0.010)	(0.012)
Parental education (ref. Low secondary)				
Upper secondary	0.003	-0.024***	0.005	0.006
	(0.018)	(0.007)	(0.008)	(0.009)
Tertiary	0.147***	-0.013**	0.026***	0.030***
	(0.019)	(0.007)	(0.008)	(0.009)
Highest parental ISEI (std)	0.023**	-0.005	0.010**	0.029***
	(0.011)	(0.004)	(0.005)	(0.004)
School level				
Extra-curricular activity index (std)	-0.001	<.001	0.011***	0.007**
	(0.008)	(0.003)	(0.003)	(0.003)
Proportion of parents involved in school activities (std)	-0.005	0.002	0.004	0.003
	(0.007)	(0.002)	(0.003)	(0.003)
Teacher involvement in school decision-making (std)	0.013	-0.005**	-0.003	-0.005
	(0.008)	(0.002)	(0.003)	(0.003)
Negative school climate (std)	-0.017**	0.009***	0.026***	0.021***
	(0.007)	(0.003)	(0.003)	(0.004)
School achievement				
Performance in science (std)	0.079***	-0.057***	-0.059***	-0.072***
	(0.006)	(0.002)	(0.003)	(0.003)
Constant	-0.375	-0.510***	-0.476***	-0.006
	(0.285)	(0.106)	(0.122)	(0.104)

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects and a rich set of school-level factors. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

More precisely, a series of models is run to investigate the amount of possible bias introduced into the analytical models by the missing values. First, the same models shown in the report were run, including as "missing category" all missing values on the independent variables. This analysis showed similar results to those obtained from models run excluding missing values (Table A3.4). The same analysis has been performed for effort, effort persistence and perseverance missing values. Here, once again, results suggested randomness of missing values both at individual and at school level.

Second, limited to the countries with the highest incidence of missing values (i.e., Germany, Denmark and Bulgaria), linear probability models were run to study the probability of having a missing value on any of the four dependent variables referred to pupils' engagement (Table A3.5). The risk of having a missing value with respect to the relevant individual, family and school characteristics available in the data was estimated. The results exhibited that the missing problem is not particularly dramatic for the main individual and school variables: even when it is significant, the coefficients tend to be rather small.

Table A3.5. Linear models for the probability of being missing according to individual and school level for Germany, Denmark and Bulgaria.

	DE				DK				BG			
	Science engagement	Truancy	Truancy (2)	Late	Science engagement	Truancy	Truancy (2)	Late	Science engagement	Truancy	Truancy (2)	Late
Individual level												
Male (ref. Female)	0.023	0.013***	0.015***	0.010*	0.015*	0.004	0.006	0.006	0.036***	0.009**	0.005	0.002
Migrant Background (ref. Natives)												
Mixed parentage	-0.078***	-0.009	-0.003	0.002	0.005	-0.011	-0.014**	-0.011	0.019	0.013	0.026	0.022
Second-generation	0.023	0.011	0.020*	0.012	-0.001	0	0.001	-0.006	0.054	0.065	0.109	0.115*
First-generation	0.045	0.028	0.034	0.032*	0.037	-0.008	-0.016	-0.007	0.058	-0.031**	0.011	0.016
Parental education (ref. Low secondary)												
Upper secondary	-0.017	-0.018*	-0.011	-0.014	-0.005	-0.018	-0.024**	-0.024**	-0.025	0.006	0.019	0.011
Tertiary	-0.036*	-0.026***	-0.025***	-0.024***	-0.014	-0.012	-0.017*	-0.020*	-0.056	0.002	0.027	0.016
Highest parental ISEI (std)	-0.019*	0.004	0.004	0.003	-0.010*	-0.007*	-0.005	-0.002	-0.004	-0.001	-0.007***	-0.005*
School level												
Extra-curricular activity index (std)	-0.002	-0.008**	-0.007*	-0.008**	-0.017	-0.001	-0.002	-0.003	-0.030**	<.001	-0.002	-0.006
Proportion of parents involved in school activities (std)	0.038*	0.005	0.006	0.005	<.001	-0.002	-0.001	-0.003	-0.011	<.001	-0.002	-0.001
Teacher involvement in school decision-making (std)	-0.004	0.002	0.001	0.001	-0.012	0.004	0.003	0.003	-0.013	0.002	-0.001	<.001
Negative school climate (std)	-0.007	-0.006	-0.005	-0.007	0.003	-0.001	-0.001	-0.001	0.015	0.003	0.003	0.003
Constant	0.511	-0.274*	-0.350**	-0.369**	-0.476**	-0.15	-0.208*	-0.137	0.725**	0.091	0.068	-0.008

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, model 2 for a rich set of school-level factors. Estimates obtained using PISA provided replicate weights and final student weights. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

Appendix A3.2 Indices based on log-files

Starting from the same theoretical definition, effort, effort persistence and perseverance can be operationalised and measured in different ways. Table A3.6 presents an overview of all possible measures that have been computed for the study, beyond those presented in the main text.

Effort has been measured in the main text as the difference in the mean response time between difficult and easy items in cluster 1. The same index can be computed using differences in median values or resorting to the ratio of means (or medians) between difficult and easy items. It is also possible to use medians or ratios for persistence in effort, which, in the main text, has been measured as the difference between effort in cluster 2 and effort in cluster 1. Moreover, alternative operationalisation can arise using different clusters, instead of the first for effort and the first and the second for persistence of effort. In general, the differences in mean have been preferred for two reasons: 1) the mean is a better-known concept among a non-technical audience; 2) the results employing means or medians were qualitatively the same.

Concerning perseverance - that is, the difference in WLE scores in the second and first cluster - the main alternatives are: either using the percentage of correct answer instead of WLE scores or, exploiting the random allocation of the booklets, the difference in the correctness of the same item in different clusters. The first option is not very reliable, because it does not take into account the difficulty of the items or the ability of the students, thus making less comparable measures of performances across clusters. The second one is feasible only at the aggregate level (e.g., at country level), because it is not possible to observe the same item for the same students in two different clusters. The results using the latter approach are shown in appendix A7.

The choice of considering only the first two clusters for measuring the indices based on log-files is motivated by the fact that the first part of the test is less affected by fatigue. The latter, as shown in the empirical sections of the main text, exerts an effect on both effort and performance. A cross-session comparison could be problematic because each session is based on different competence domains.

Table A3.7 shows that students who make a lot of effort in the first cluster also register higher numbers of non-reached items in the first session, but at the same time register a higher science score compared with those put in less effort at the beginning of the test.

Table A3.6 Alternative measures of the indices based on log-files.

Indices	Information	Operationalisation	Alternatives
Effort	Response time (RT)	Difference in average RT between 5 most difficult and 5 easiest items located in cluster 1	• Differences in medians • Ratio of means • Ratio of medians
Persistence in effort	Response time (RT)	Difference between Effort in cluster 2 and Effort in cluster 1	• Differences in medians • Ratio of means • Ratio of medians • The same measures calculated using cluster 4 and cluster 3 • The same difference calculated between cluster 3 (or 4) and cluster 1, (NB: different competence domains)
Perseverance (endurance)	Weighted Likelihood Estimate score (WLE)	Difference in WLE score in the second and first cluster	• The same measures calculated using cluster 4 and cluster 3 • The same difference calculated between cluster 3 (or 4) and cluster 1, (NB: different competence domains) • The same difference calculated using the percentage of correct answers instead of WLE scores • Difference in the correctness answer on the same items in different clusters (feasible only at aggregate-level)

Table A3.7 Average number of unreached items and science score by deciles of effort in cluster position 1.

Deciles in effort (cluster 1)	Non-reached items	Science score
1	0.307	454.9
2	0.267	470.3
3	0.239	483.0
4	0.192	492.6
5	0.201	501.5
6	0.211	507.6
7	0.215	516.7
8	0.201	522.8
9	0.246	527.8
10	0.332	538.2

Note: FBK-IRVAPP analysis of PISA 2015 data.

Appendix A4 School engagement

Table A4.1 Linear regression models for science performance according to individual characteristics.

	Science
Individual level	
Male (ref. Female)	7.587***
	(0.997)
Migrant Background (ref. Natives)	
Mixed parentage	-9.714***
	(1.468)
Second-generation	-27.404***
	(2.239)
First-generation	-38.401***
	(2.446)
Parental education (ref. Low secondary)	
Upper secondary	21.794***
	(1.698)
Tertiary	26.848***
	(1.753)
Highest parental ISEI (std)	28.181***
	(0.529)
Constant	265.791**
	(0.106)
N	151673
R^2	0.162
Country clusters	26

Note: FBK-IRVAPP analysis of PISA 2015 data. The model control for age of the student and for country fixed effects. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

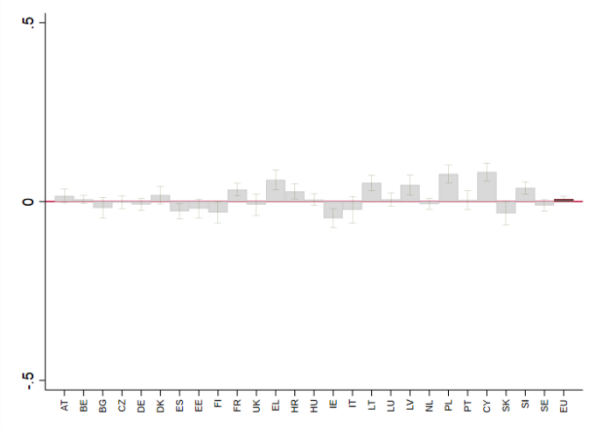
Table A4.2 Linear regression models for truancy (a whole day), truancy (some classes) and arriving late at school according to country's school-system characteristics. Selected parameters.

School-system level	TRUANCY (a whole day)	TRUANCY (some classes)	Arriving Late
Horizontal differentiation (ref. Low)			
Medium horizontal diff	0.189***	-0.009	-0.034***
	-0.011	-0.011	-0.011
High horizontal diff	-0.049***	-0.121***	-0.009
	-0.008	-0.01	-0.011
High vertical differentiation (ref. Low)	-0.079***	-0.012	0.160***
	-0.008	-0.011	-0.011
Constant	0.633***	0.322**	-0.175
	-0.135	-0.162	-0.152
N	131436	131032	131387
R^2	0.127	0.065	0.054

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, a rich set of school-level factors and other country's school-system characteristics. Only the first PISA plausible value (standardized) is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

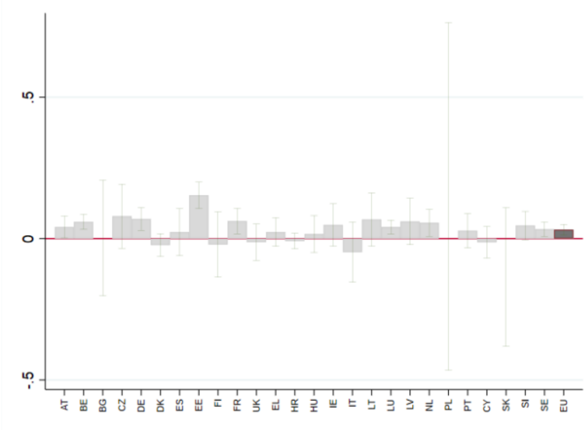
Figure A4.1 Individual-level coefficients in truancy across EU Member States.

***Panel a**: Gender gap in truancy.*



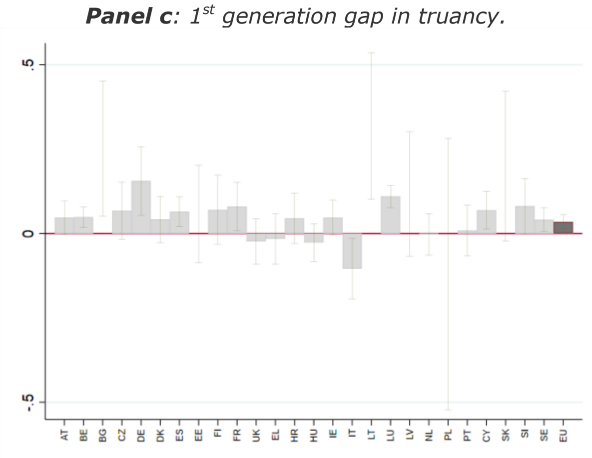
Note: reference category: females.

***Panel b**: 2nd generation gap in truancy.*



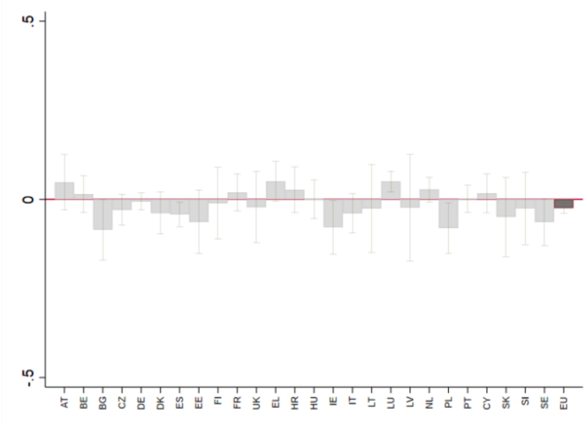
Note: reference category: natives.

***Panel c**: 1st generation gap in truancy.*



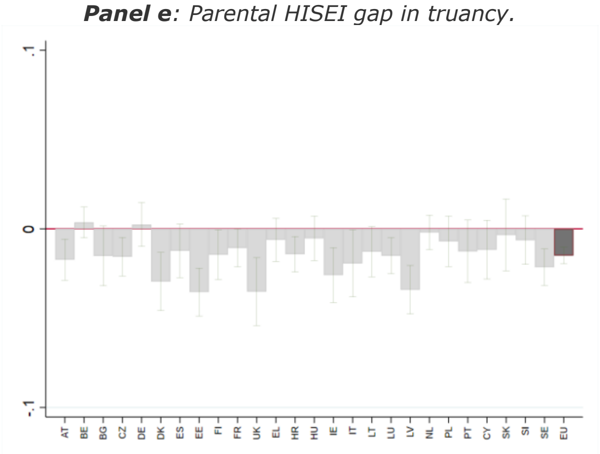
Note: reference category: natives.

***Panel d**: Parental education gap in truancy.*



Note: reference category: low educated parents.

***Panel e**: Parental HISEI gap in truancy.*

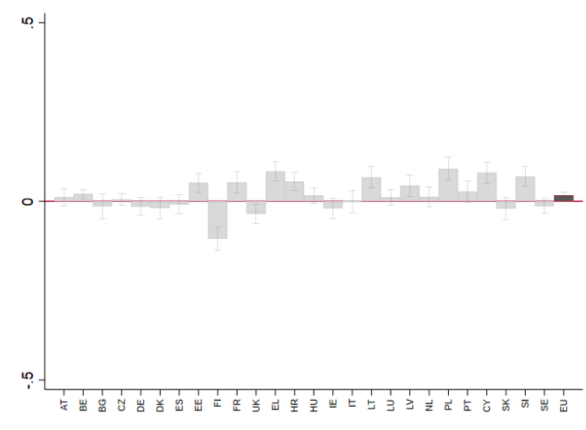


Note: reference category: change in 1SD of HISEI.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries – BG, PL, SK (1st and 2nd generation), EE, LT, LV (1st generation) – immigrant status is not shown because of the small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

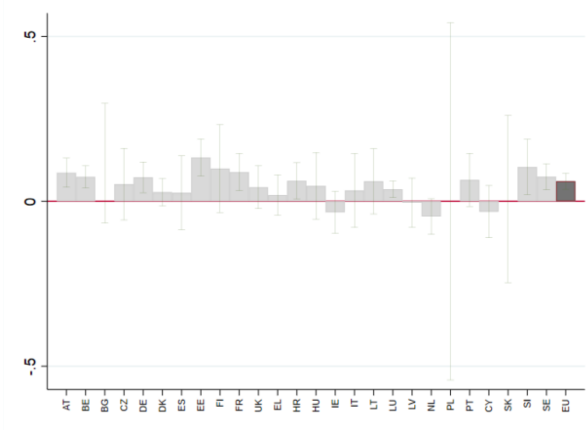
Figure A4.2 Individual-level coefficients in truancy (2) across EU Member States.

***Panel a**: Gender gap in truancy (2).*



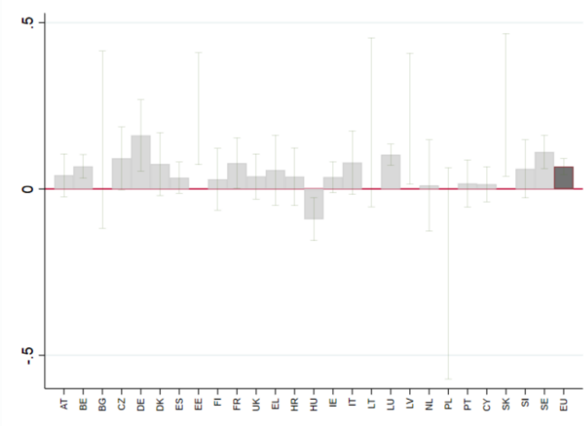
Note: reference category: females.

***Panel b**: 2nd generation gap in truancy (2).*



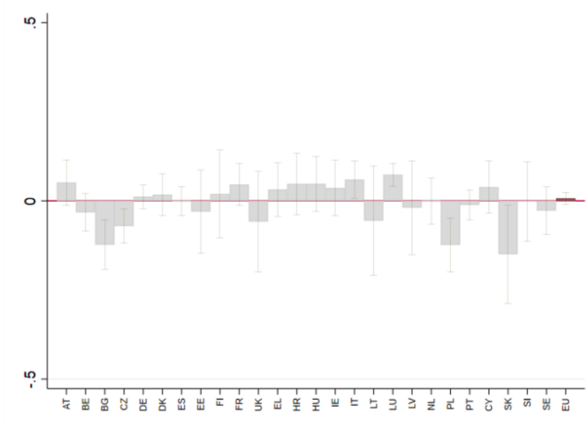
Note: reference category: natives.

***Panel c**: 1st generation gap in truancy (2).*



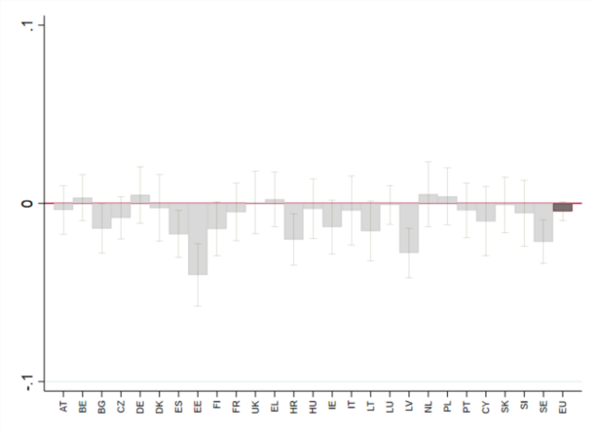
Note: reference category: natives.

***Panel d**: Parental education gap in truancy (2).*



Note: reference category: low educated parents.

***Panel e**: Parental HISEI gap in truancy (2).*

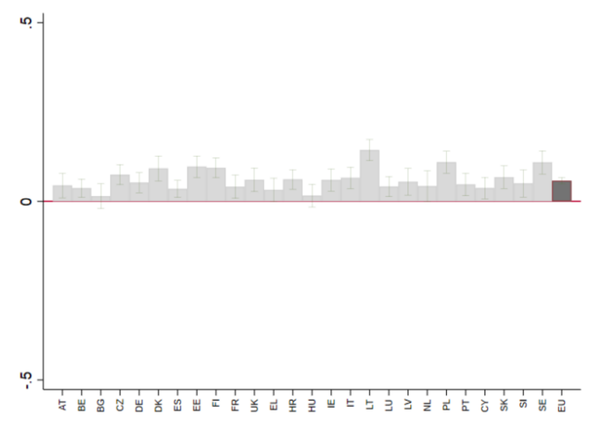


Note: reference category: change in 1SD of HISEI.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries – BG, PL, SK (1st and 2nd generation), EE, LT, LV (1st generation) – immigrant status is not shown because of small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

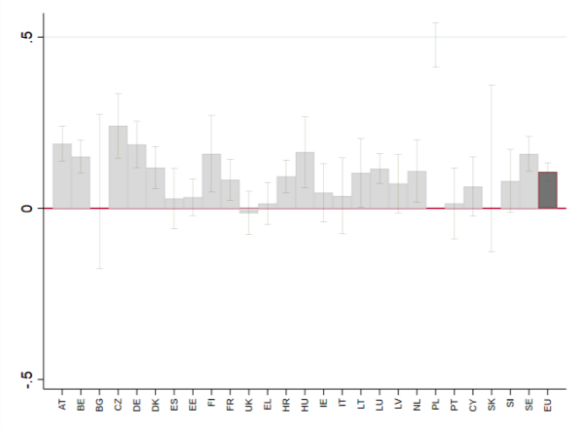
Figure A4.3 Individual-level coefficients in arriving late across EU Member States.

***Panel a**: Gender gap in arriving late.*



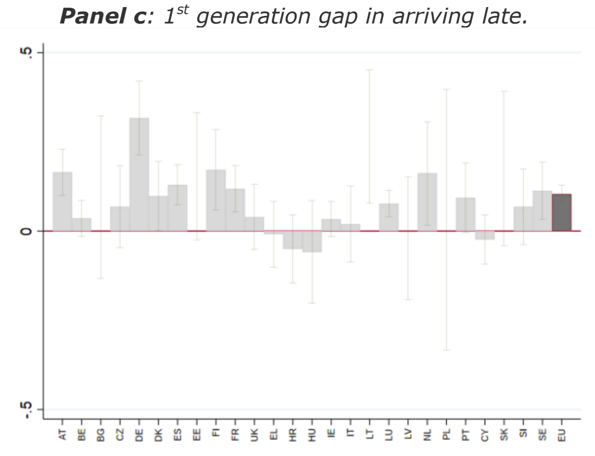
Note: reference category: females.

***Panel b**: 2nd generation gap in arriving late.*



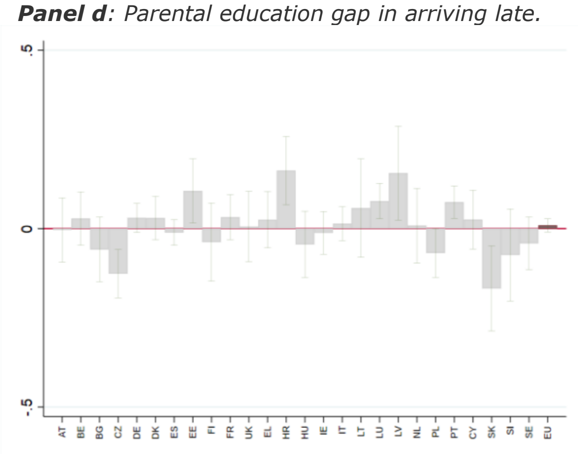
Note: reference category: natives.

***Panel c**: 1st generation gap in arriving late.*



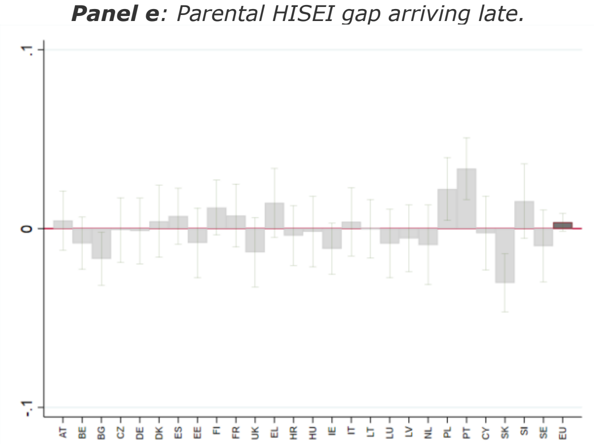
Note: reference category: natives.

***Panel d**: Parental education gap in arriving late.*



Note: reference category: low educated parents.

***Panel e**: Parental HISEI gap arriving late.*



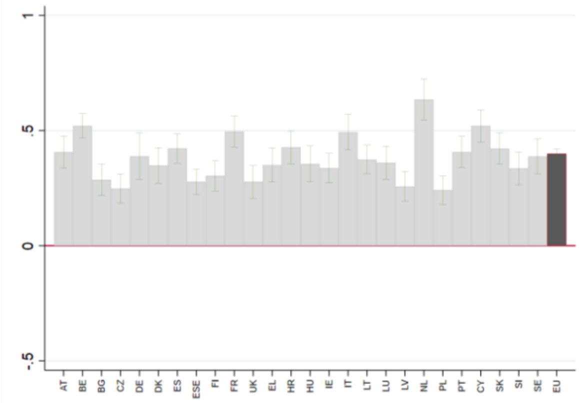
Note: reference category: change in 1SD of HISEI.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries – BG, PL, SK (1st and 2nd generation), EE, LT, LV (1st generation) – immigrant status is not shown because of small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

Appendix A5 Science engagement

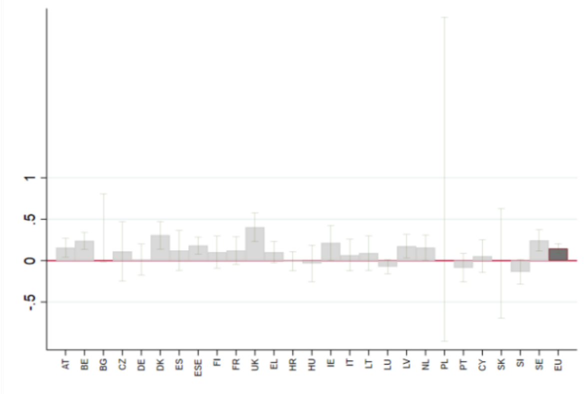
Figure A5.1 Individual-level coefficients in students' science activities across EU Member States.

Panel a*: Gender gap in students' science activities.*



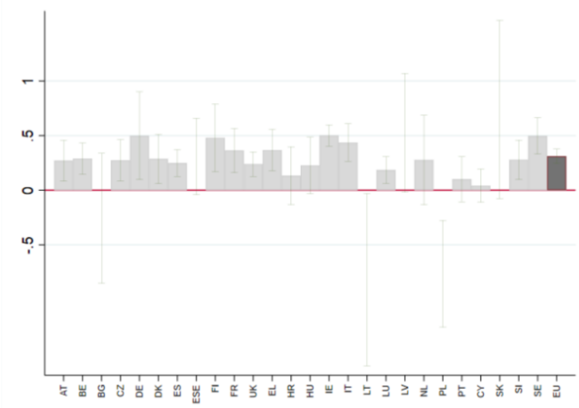
Note: reference category: females.

Panel b*: 2nd generation gap in students' science activities.*



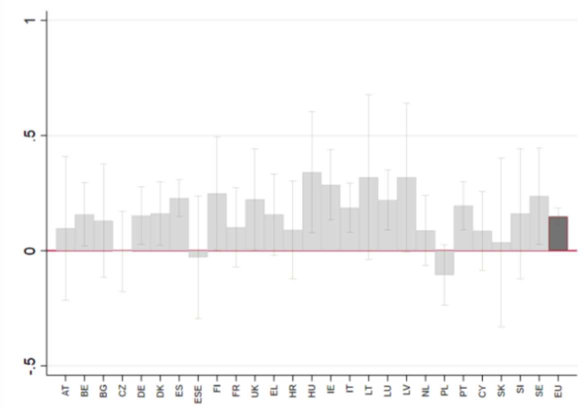
Note: reference category: natives.

Panel c*: 1st generation gap in students' science activities.*



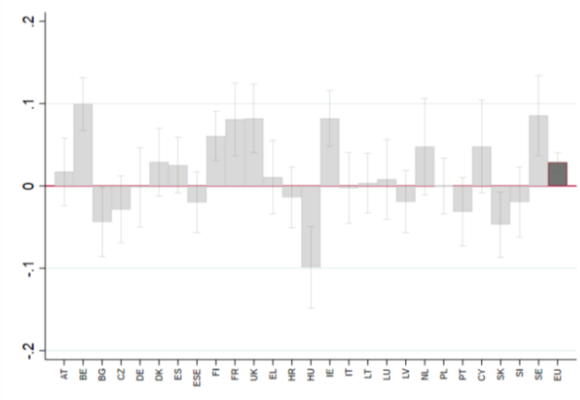
Note: reference category: natives.

Panel d*: Parental education gap in students' science activities.*



Note: reference category: low educated parents.

Panel e*: Parental HISEI gap in students' science activities.*



Note: reference category: change in 1SD of HISEI.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries – BG, PL, SK (1st and 2nd generation), EE, LT, LV (1st generation) – immigrant status is not shown because of small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

European Commission

Appendix A6 Effort and effort persistence

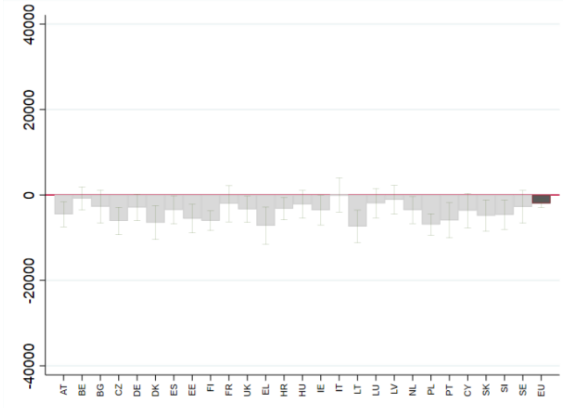
Table A6.1 Linear regression models for effort and effort persistence by domains (maths, reading, science), according to individual characteristics, school level factors and performance in the relative domain. Selected parameters. Coefficients display milliseconds.

	Effort			**Effort persistence**		
	Maths	**Reading**	**Science**	**Maths**	**Reading**	**Science**
Individual level						
Male (ref. Female)	-1762.175	-119.646	-5387.227***	720.793	-119.973	1522.004*
	(1334.663)	(1103.218)	(680.791)	(1800.456)	(1484.966)	(872.414)
Migrant Background (ref. Natives)						
Mixed parentage	-2841.439	-774.43	4656.379***	5542.423*	-97.675	-4651.835***
	(2065.507)	(1689.198)	(1128.687)	(2887.138)	(2495.442)	(1493.243)
Second-generation	2376.138	3014.627	570.203	-4841.925	-7118.065*	-1330.878
	(2983.81)	(2692.55)	(1404.081)	(4403.739)	(4039.602)	(1728.175)
First-generation	3268.842	-536.35	1979.122	1968.155	-2542.406	-2453.373
	(3040.469)	(3244.411)	(2163.484)	(4301.99)	(4245.483)	(2598.721)
Parental education (ref. Low secondary)						
Upper secondary	-358.7	-1863.955	-157.027	913.655	502.083	430.218
	(2718.761)	(2095.747)	(1498.579)	(3278.921)	(2816.767)	(1863.323)
Tertiary	1575.607	234.786	-106.489	2105.317	-2295.723	199.612
	(2477.095)	(2315.058)	(1485.074)	(3095.44)	(2946.51)	(1857)
Highest parental ISEI (std)	1262.390*	-539.944	232.899	-628.37	915.319	-79.594
	(722.185)	(615.761)	(400.067)	(937.379)	(842.275)	(545.095)
School level						
Extra-curricular activity index (std)	1095.503	602.387	395.689	-1163.684	-2112.362**	-973.766*
	(722.319)	(638.193)	(421.92)	(998.13)	(913.659)	(550.116)
Proportion of parents involved in school activities (std)	703.742	623.033	-144.239	-1560.880**	-112.141	88.029
	(520.868)	(469.792)	(359.28)	(785.723)	(668.945)	(452.24)
Teacher involvement in school decision-making (std)	416.728	462.159	635.656*	1136.495	874.838	-125.514
	(712.064)	(678.034)	(350.688)	(983.586)	(901.856)	(489.835)
Negative school climate (std)	-236.075	-2.964	-450.531	-597.208	-75.565	-24.798
	(717.972)	(545.001)	(454.854)	(883.664)	(768.4)	(492.513)
School achievement						
Performance in math (std)	19079.715***			4401.231***		
	(676.434)			(1061.208)		
Performance in reading (std)		8794.751***			-821.094	
		(775.619)			(1204.463)	
Performance in science (std)			10240.076***			666.784
			(407.326)			(540.43)
Constant	69493.551**	-2075.564	39534.791**	-60587.91	-30412.13	2009.215
	(31822.408)	(28491.354)	(16528.767)	(51225.24)	(37206.79)	(23258.094)
N	22370	22424	45212	22169	22139	44533
R^2	0.120	0.038	0.084	0.012	0.005	0.011
Country clusters	26	26	26	26	26	26

Note: FBK-IRVAPP analysis of PISA 2015 data. All models control for age of the student, grade retention and ISCED-level attended, country fixed effects, and a rich set of school level factors. Only the first PISA plausible value is included. Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania. The lowest and the highest percentiles were included in this analysis. Standard error in brackets; *p<0.10; **p<0.05; ***p<0.01.

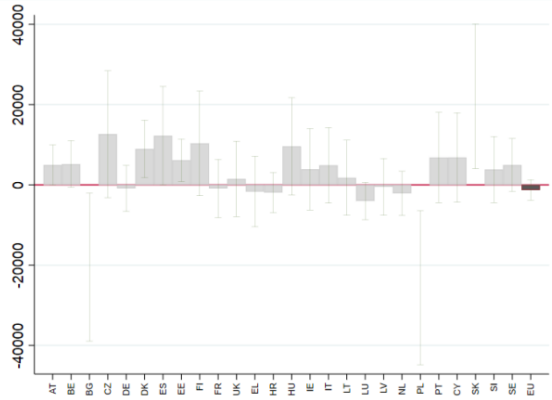
Figure A6.1 Individual-level coefficients in effort across EU Member States.

***Panel a**: Gender gap in effort.*



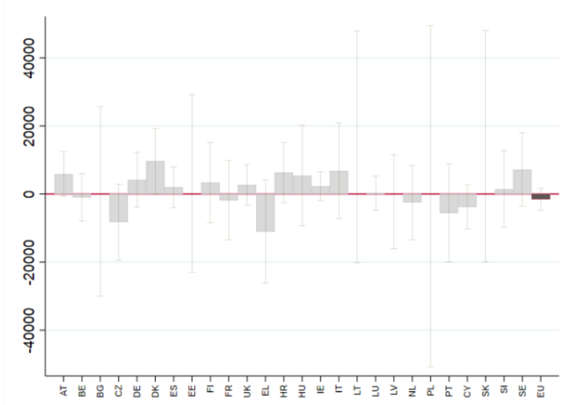
Note: reference category: females.

***Panel b**: 2nd generation gap in effort.*



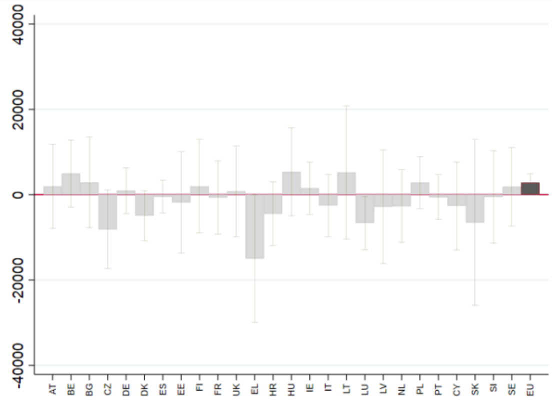
Note: reference category: natives.

***Panel c**: 1st generation gap in effort.*



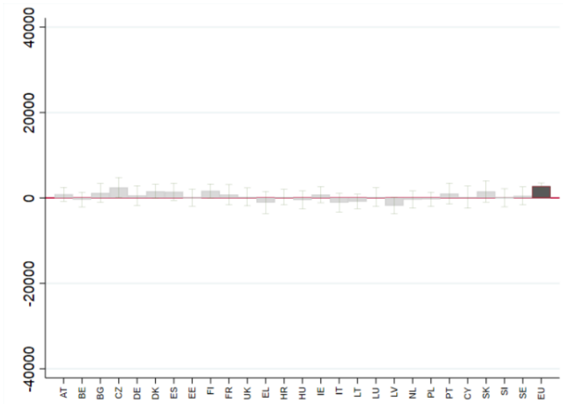
Note: reference category: natives.

***Panel d:** Parental education gap in effort.*



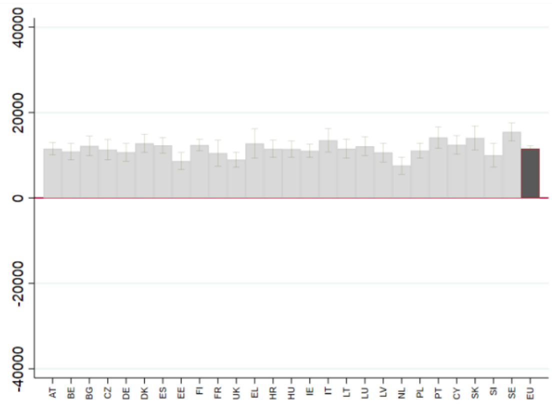
Note: reference category: low educated parents.

***Panel e**: Parental HISEI gap in effort.*



Note: reference category: change in 1SD of HISEI.

***Panel f**: Science achievement gap in effort.*



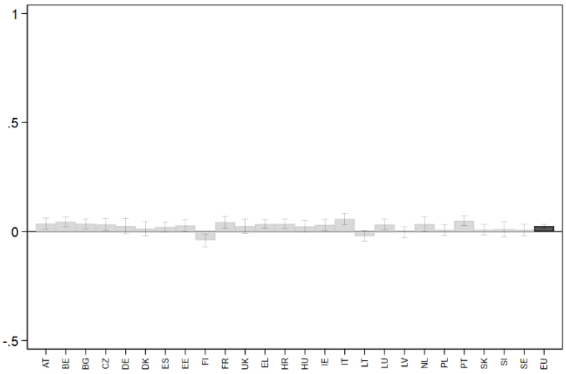
Note: references category: Change in 1SD of science.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries - BG, PL, SK (1st and 2nd generation), EE, LT, LVA (1st generation) - immigrant status is not shown because of small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Malta and Romania.

Appendix A7 Perseverance

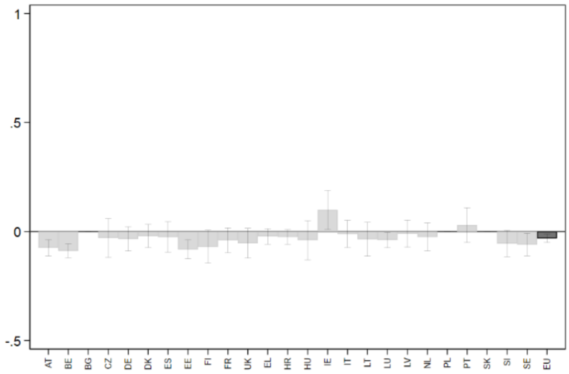
Figure A7.1 Individual-level coefficients in students' perseverance across EU Member States.

***Panel a**: Gender gap in students' perseverance.*



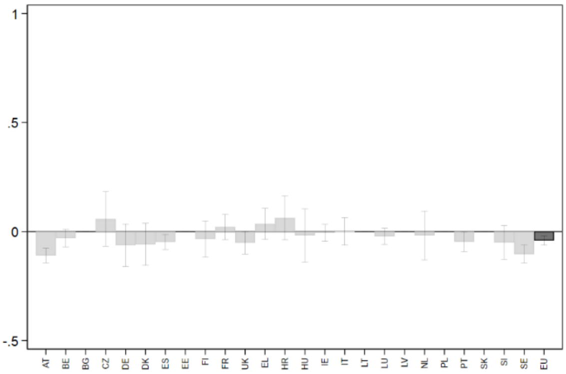
Note: reference category: females.

***Panel b**: 2nd generation gap in students' perseverance.*



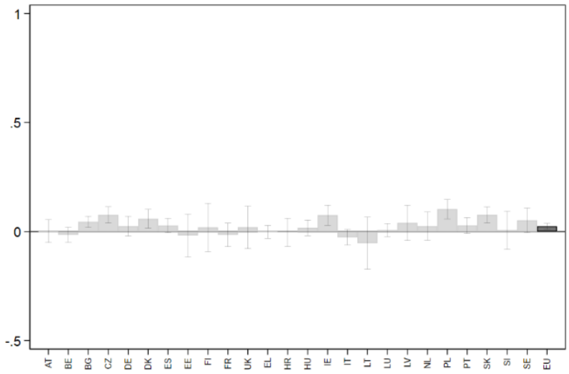
Note: reference category: natives.

***Panel c**: 1st generation gap in students' perseverance.*



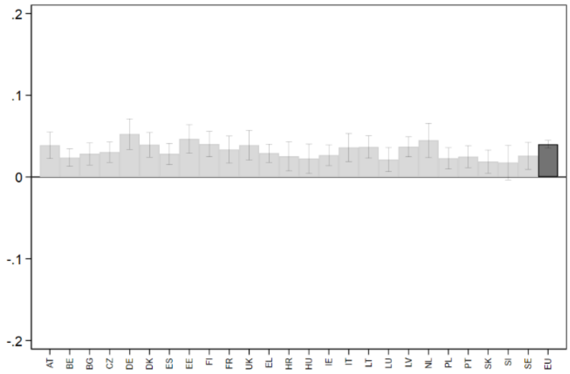
Note: reference category: natives.

***Panel d**: Parental education gap in students' perseverance.*



Note: reference category: low educated parents.

***Panel e**: Parental HISEI gap in students' perseverance.*



Note: reference- category: change in 1SD of HISEI.

Note: FBK-IRVAPP analysis of PISA 2015 data. The parameters and the standard errors plotted in the figure come from model 2. In some countries - BG, PL, SK (1st and 2nd generation), EE, LT, LV (1st generation) - immigrant status is not shown because of small sample size (<50). Estimates obtained using PISA provided replicate weights and final student weights. The information is not available for Cyprus, Malta and Romania.

Following the approach of Borgonovi and Biecek (2016), perseverance can also be measured at aggregated level. These authors exploit the random allocation of students to the different booklets to assess the effect of item position on performance. In this specific case, the comparison is performed on the same set of items, meaning that it is not possible to calculate a measure at individual level, but only at aggregated level.

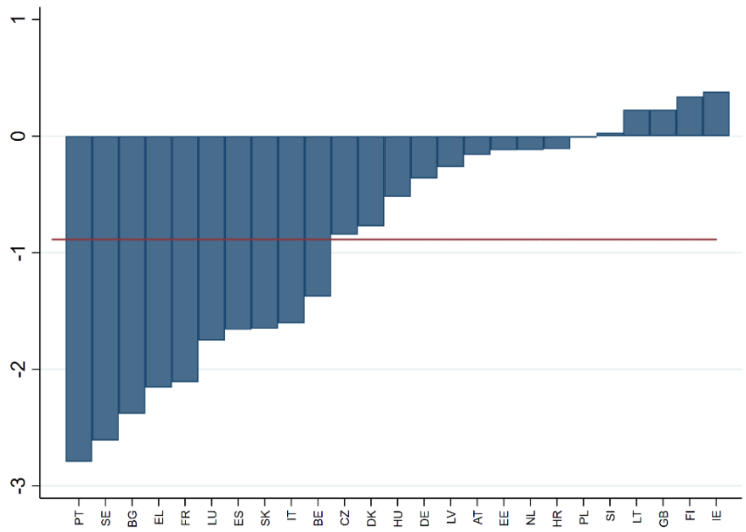
Following the work of Borgonovi and Biecek (2016), "endurance" is computed as 1-E_i for each country only for science (the major domain in PISA 2015):

$$E_i = \frac{1}{12}\sum(WLE_1 - WLE_2)$$

This index is simply the average, within each country, of the differences in the WLE score in clusters 1 and 2 calculated across the twelve set of items about science administered to students[28]. Figure A7.2 reports the ranking of the countries according to this index, while Figure A7.3 shows a negative relationship between endurance and the overall score in science.

If the correlations between the different indices related to the concept of perseverance are considered, it is possible to note that the correlations at country level between the perseverance indices in Sessions 1-2 and the endurance index are very high. Moreover, the correlation with the proportion of "persistently good" is also definitely remarkable. Our analysis shows that our measure, developed at the individual level, is comparable with the one used by Borgonovi and Biecek (2016).

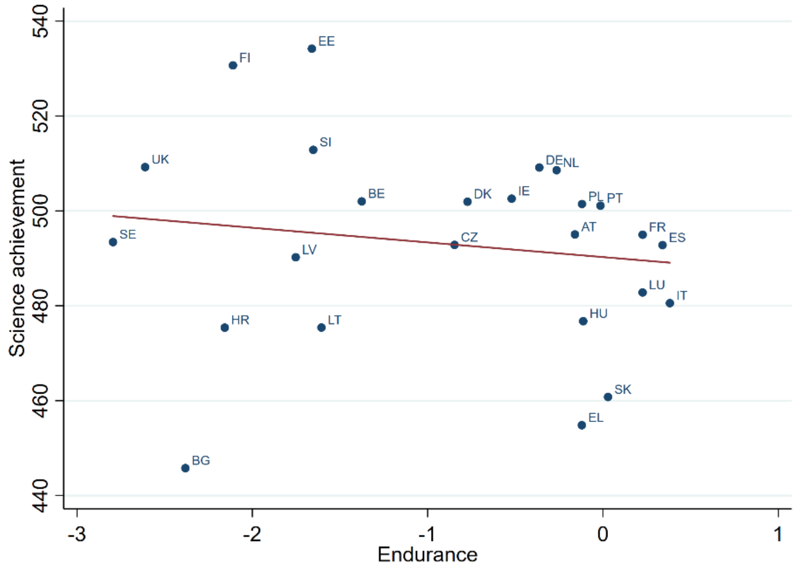
Figure A7.2 Endurance index in EU Member States.



Note: FBK-IRVAPP analysis of PISA 2015 data. The information is not available for Cyprus, Romania and Malta.

[28] Borgonovi and Biecek (2016) computed endurance comparing cluster 1 with cluster 3. It was decided to rely on the comparison between clusters 1 and 2 to facilitate the comparison with our perseverance indices, which are calculated within session.

Figure A7.3 Scatter plot between achievement in science and endurance in EU Member States.



Note: FBK-IRVAPP analysis of PISA 2015 data. The information is not available for Cyprus, Romania and Malta.

Table A7.4 Correlation between the various indices of perseverance at country level.

	Endurance	Persistently good	Perseverance (Session 1)	Perseverance (Session 2)
Endurance	1.000			
Persistently good	0.639	1.000		
Perseverance (Session 1)	0.973	0.656	1.000	
Perseverance (Session 2)	0.881	0.646	0.889	1.000

Note: FBK-IRVAPP analysis of PISA 2015 data. The information is not available for Cyprus, Romania and Malta.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

NC-06-18-138-EN-N

Publications Office

ISBN: 978-92-79-93947-1