Social Research Institute

From asking to observing. Behavioural measures of socio-emotional and motivational skills in large-scale assessments

Francesca Borgonovi
Alessandro Ferrara
Mario Piacentini

Quantitative Social Science

Working Paper No. 20-19

December 2020

Disclaimer

Quantitative Social Science

UCL Social Research Institute

University College London

55-59 Gordon Square

London WC1H 0NU

Social Research Institute

From asking to observing. Behavioural measures of socio-emotional and motivational skills in large-scale assessments

Francesca Borgonovi[1]
Alessandro Ferrara[2]
Mario Piacentini[3]

Abstract
Socio-emotional and motivational skills are routinely measured using self-reports in large-scale educational assessments. Measures exploiting test-takers' behaviour during the completion of questionnaires or cognitive tests are increasingly used as alternatives to self-reports in the economics of education literature. We use cross-sectional and longitudinal evidence to evaluate if behavioural measures can provide unbiased measures of socio-emotional and motivational skills to be used in empirical research using data from the Programme for International Student Assessment (PISA). We find that behavioural measures capture important aspects of students' academic profiles: some are importantly associated with contemporaneous performance and educational attainment. However, these measures are only limitedly correlated among themselves and have low correlations with self-report measures of the same constructs. Moreover, behavioural measures have different levels of stability over time and sensitivity to design considerations. These results suggest that more research is needed before measures of students' behaviour on a cognitive test can be used as valid indicators of socio-emotional and motivational skills.

.

Keywords: Socio-emotional and motivational skills; cross-country; PISA; large-scale assessments; behavioural measures; self-reports; education.

JEL Codes: I20; I24; I26

Acknowledgements: Francesca Borgonovi recognises support from the British Academy through its Global Professorship scheme. The authors would like to thank participants in the award-holder's work-in-progress semnar organised by the British Academy; Marie-Anne Deussing and Susan Bennet from the Labour Market and Skills Research Division, Employment and Social Development Canada for work on the Canadian longitudinal data.

[1] Contact details: Francesca Borgonovi (F.Borgonovi@ucl.ac.uk), UCL Social Research Institute, University College London, 20 Bedford Way London, WC1H 0AL, United Kingdom.
[2] European University Institute (Alessandro.Ferrara@eui.eu), Department of Political and Social Sciences, European University Institute, Via della Badia dei Roccettini, 9, 50014 Fiesole, FI, Italy
[3] Mario Piacentini (Mario.Piacentini@OECD.org), Directorate for Education and Skills, 2 Rue André Pascal, 75116 Paris, France.

1. **Introduction**

Research in economics, psychology, sociology and education indicates that socio-emotional and motivational skills, also referred to as personality traits, temperament, non-cognitive skills or character skills in the literature, play an important role (and one independent from cognitive skills) in shaping individuals' long term outcomes (Duckworth & Seligman, 2005; Gutman & Schoon, 2016; Kautz, Heckman, Diris, & Borghans, 2014; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Moreover, there is evidence that education policy and school-level practices have the potential to shape the acquisition of some of these skills (Heckman, Pinto, & Savelyev, 2013) although not all such skills may be amenable to the influence of external factors (Credé, Tynan, & Harms, 2017; Revelle, 2007).

The emphasis on socio-emotional and motivational skills - the term we adopt in this work - and the interest in how they can be promoted has prompted policy makers and researchers to consider how their measurement can best be integrated in benchmarking and accountability systems at international, regional and national levels. Standardised low-stakes large-scale tests such as the Programme for International Student Assessment (PISA), the Programme for International Assessment of Adult Competences (PIAAC), the IEA's Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS) as well as regional and national standardised tests have, in recent years, introduced the measurement of selected socio-emotional and motivational skills within questionnaires administered to participants, alongside the measurement of cognitive skills.

In fact, the increased attention for the role played by socio-emotional and motivational skills has also spurred a parallel and lively debate in the economics of education literature over what exactly low-stakes large-scale tests measure and the importance of factors such as test taking motivation to determine differences in measures of cognitive skills across countries and population subgroups (Balart & Oosterveen, 2019; Borghans & Schils, 2013; Borgonovi and Biecek, 2016; Brunello, Crema, & Rocco, 2018; Gneezy, et al., 2019; Zamarro, Hitt, & Mendez, 2019; Akyol, Krishna and

Wang, 2018). Such research builds upon the psychometrics literature on response styles and item position effects (Nagy et al., 2018) and is gaining momentum due to the shift to computer-based administration of many large-scale assessments. Computer-based testing in fact enables the collection of behavioural indicators, in particular timing indicators that can complement or augment what was already possible in paper-based administration conditions (Wise & Kong, 2005). Data on test-taking processes collected through computers are generally referred to in the literature as log-file data since they contain complete logs of the test-taking experience of test takers. Log-files have been used to identify test-taking effort and motivation and have been considered as indicators of socio-emotional and motivational skills (Azzolini et al., 2019; Goldhammer et al., 2014; Goldhammer, Martens and Lüdtke, 2017).

Questions remain on how best socio-emotional and motivational skills can be measured and how reliable different measurement frameworks are (Heckman & Kautz, 2012). Methodological considerations on the influence of measurement on substantive findings pertain both socio-emotional and motivational skills and cognitive skills. For example, although cognitive skills are typically measured using relatively homogeneous instruments (achievement tests), research has indicated that meaningful differences in outcomes can be found depending on the specific tests and assessments that are used (Rindermann, 2007; Wu, 2009). However, measurement concerns are generally more pronounced in the case of socio-emotional and motivational skills because of the multitude of approaches that are used to measure them and the fact that little is known about the properties of different measurement tools.

Examples of approaches to measure socio-emotional and motivational skills include: reports (by the self or people close to the self); laboratory experiments; behavioural observations; and think-aloud protocols. Likert-type scales administered to individuals through self-reported questionnaires are by far the most widely used instrument in large-scale settings. Self-reports have the advantage of being designed to reflect well-defined theoretical constructs and can be measured in a questionnaire with a relatively short time burden for respondents (Kyllonen & Kell, 2018). Self-reported measures of socio-emotional and motivational skills have been found to have substantial predictive validity for

important education, work, and life outcomes (Soto, 2019; Soto 2020; Wilmot & Ones, 2019). However, in applications of such assessments, practitioners often raise concerns about potential limitations of these methods, such as misinterpretation, lack of information and memory bias, social desirability bias, response style bias and reference-group bias (Kankaraš, 2017) even if some researchers argue that these biases do not have a large influence on the association between personality measures and individual outcomes (Ones, Visweran and Reiss, 1996). Concerns over the measurement properties of self-reports raise questions as to whether the reliability and validity of socio-emotional and motivational skills assessments can be improved by supplementing (or even replacing) self-report questionnaires with other measurement methods (Soland et al., 2019).

Laboratory and field experiments in which study participants are asked to complete tasks designed to reveal their socio-emotional and motivational skills can arguably be used to measure specific skills more reliably and allow for better comparisons across population groups and across countries than self-reports. Some studies have conducted global performance-task experiments (Cohn, Maréchal, Tannenbaum, & Zünd, 2019) but the high cost and response burden that would result from administering a wide battery of tests in controlled conditions prevents their use in large-scale assessment programmes. An alternative approach has been to use experiments to validate self-reports. For example, Falk & Hermle (2018) conducted a smallexperimental pilot study involving 402 participants to validate individuals' self-reports of willingness to take risks, patience, altruism, positive and negative reciprocity, as well as trust. The questionnaire-based and experimentally-validated self-reported measures were then used to conduct an extensive survey among 80,000 individuals in 76 representative country samples. However, the experimental selection and the validation of the survey items through laboratory experiments was limited to participants at a specific university and therefore the validation exercise did not consider potential cultural or linguistic factors shaping discrepancies in self-reports.

A new line of research has developed behavioural and opportunistic measures of socio-emotional and motivational skills derived from observing and coding the behaviours of individuals when they participate in standardised assessment programmes (see Soland et al., 2019 for a review).

Proponents of these indicators maintain that behavioural measures share many characteristics with laboratory or field experiments, i.e. participants are not asked to report perceptions of themselves or others but their behaviour as they complete a task is evaluated. Moreover, they do not involve additional costs or burden for respondents. Their main disadvantage is that they are theoretically ill-defined, because their construction is based on the behaviour of respondents while they perform a task that was designed with the intention of measuring a different set of constructs (Kroehne & Goldhammer, 2018). This is a typical challenge related to the use of behavioural data in social research (Salganik, 2019). The challenge is to clearly investigate and highlight how different measures were constructed, what they are likely to measure, and what types of measurement problems they suffer from.

Measures constructed using the behaviour of students while they respond to questionnaires include: item non-response rate (Zamarro, Cheng, Shakeel, & Hitt, 2018); response non-consistency (Hitt, 2015; Zamarro et al., 2018) and non-differentiation (Knowles, 1988; Krosnick, 1991; Vannette & Krosnick, 2014). Behavioural measures developed using students' responses during a cognitive test include: performance decline (Borghans & Schils, 2013; Zamarro et al., 2019) and, for computer-based tests, response time effort (Wise & Kong, 2005). We describe these measures in detail in the Data and Methods section.

The aim of this work is to consider behavioural measures that have been used in the economics of education literature as indicators of four socio-emotional and motivational constructs: perseverance, self-regulation, endurance and conscientiousness (PSEC from now on). The objective is to establish if behavioural measures can become useful complements or substitutes of other established self-reported indicators of socio-emotional and motivational skills in large-scale assessment and monitoring systems.

The use of task-based and performance-based measures has been promoted by researchers to avoid social desirability problems that may arise in the context of self-reports as well as by the fact they do not entail additional burden and administration costs. However, these measures might suffer from interpretive and construct validity challenges. So far relatively little attention has been devoted

to testing convergent validity (Fiske, 1971, p. 164), i.e. if there are substantial and significant correlations between instruments developed with the intention of assessing a common construct. Establishing construct validity is necessary to avoid jingle-jangle fallacies: situations in which two scales with similar labels measure different constructs (jingle fallacy) or two scales with dissimilar labels measure similar constructs (jangle fallacy) (Kelley, 1927; Block, 1995).

Prior work in psychology has been conducted to identify convergent validity of measures of important socio-emotional and motivational skills operationalised using measurement frameworks based on self-reports and task-based performance and such studies have generally found low correlations between task-based performance measures and self-reports as well as low correlations between task-based performance measures and self-reported measures (see for examples Duckworth & Kern, 2011; Sharma, Markon, & Clark, 2014; Cyders & Coskunpinar, 2011). No study to date has examined convergent validity of behavioural measures that are routinely used in the economics of education literature and that are developed using data from large-scale international assessments. This is the gap that we fill with our work.

The primary goal of researchers that developed behavioural measures using large-scale assessment data has been to conduct cross-country comparisons in socio-emotional and motivational skills and to compare cross-country differences in the distribution of socio-emotional and motivational skills among key population groups defined in terms of socio-economic background, immigration history and gender. We use PISA data to identify the convergent validity of these measures, i.e. how related these measures are with each other and with self-reported indicators both at the individual level within each country and at the country level. Convergent validity is also examined by studying how differences in constructs across population sub-groups (i.e. gender, socio-economic status) vary according to the selected measure. We also compare behavioural measures and self-report methods according to other measurement properties. In particular, we compare reliability by assessing whether measures based on different survey rounds provide consistent conclusions about cross-country differences in socio-emotional and motivational skills. We also conduct several tests to identify how sensitive different instruments are to the use of different

administration protocols or instruments. Finally, we evaluate extrinsic convergent validity by considering how strongly different indicators are related to contemporaneous achievement and, for a subset of countries, predictive validity, i.e. how related they are to educational attainment in young adulthood.

2. Data and Methods

2.1. Data

We use data from PISA. PISA is a triennial large-scale international assessment of 15-year-old students conducted since 2000 and targeting the schooled population of children between the ages of 15 years and three months and 16 years and two months at the time of administration (15-year-olds from now on). It covers large, representative samples of 15-year-old students and over 80 countries have participated at least once since the first edition was conducted in 2000[i]. PISA is a low-stake assessment at least at the individual level because no individual scores are released to students or schools. Therefore, intrinsic motivation plays a considerable role in guiding the behaviour of participants in the study. In the absence of external pressure and motivational drivers, variability in motivation is higher than in the presence of external incentives and the influence of PSEC on performance tends be stronger (Barry, Horst, Finney, Brown, & Kopp, 2010; Cole, Bergin, & Whittaker, 2008; Wise & DeMars, 2010).

The core instruments of PISA are a two-hours long low-stakes assessment developed by international experts to test students' proficiency in reading, mathematics and science, and a 30-minutes background questionnaire. Greater information on PISA assessment instruments is available in the Online Supplementary Annex A. We use data from the questionnaire to identify self-reported measures of PSEC and to derive three behavioural indicators and we use data from the assessment to derive two behavioural indicators. The PISA surveys are conducted on two-stage stratified representative samples of 15-year-old students enrolled in lower-secondary or upper secondary institutions [for details, see (OECD, 2009)]. We used sampling and replication weight in our analyses in line with recommended PISA procedures (OECD, 2009).

Most analyses are based on the PISA 2012 study. While more recent editions of PISA are available, 2012 is the only edition that contains design features that allow to validate alternative operational definitions of the behavioural measures and test the robustness of behavioural measures to different administration conditions. We use PISA 2003 data to examine the stability of measures over time and PISA 2000 for Canada, Denmark and Switzerland and PISA 2003 for Australia to examine how predictive measures are of educational attainment at age 25. In these countries PISA participants were followed in national longitudinal studies. The follow-up data for Denmark is derived from PISA and the OECD Survey of Adult Skills (PIAAC). For Switzerland, we used data from the Transitions from Education to Employment (TREE1) surveys. For Australia, we used data from the Longitudinal Study of Australian Youth (LSAY). For Canada we used data from the Youth in Transition Survey (YITS). Canadian data have not been released for public access, so all estimates were obtained through collaborations with national researchers at Statistics Canada, who conducted the analyses on the basis of statistical programmes prepared by the authors. While the frequency and timeline of follow-up surveys vary by country, all the datasets collect information on young people when they were approximately 25 years old (the only exception is Denmark, where the analysis refers to individuals who were either 26 or 27 years old). Table B1 in the Supplementary Online Annex provides additional details on the longitudinal studies.

2.2. PSEC Measures

2.2.1. Self-reported measures

The PISA index of lack of perseverance. The first measure of noncognitive skills based on students' self-reports is the index of perseverance. The index is derived using five questions from the core PISA student questionnaire that ask participants to report the extent to which they felt that the following statements described them: "When confronted with a problem, I give up easily"; "I put off difficult problems"; "I remain interested in the tasks that I start"; "I continue working on tasks until everything is perfect"; "When confronted with a problem, I do more than is

expected of me". Possible answers were "Very much like me", "Mostly like me", "Somewhat like me", "Not much like me" and "Not at all like me". The index was recoded so that higher values indicate lower student perseverance. The index was standardised such that it had a mean of zero across the pooled student population and a standard deviation of one. The median Chronbach alpha across OECD countries was 0.8.

Measures from the PISA effort thermometer. The second self-reported measure that we consider was derived from the PISA effort thermometer. In the last page of the PISA assessment booklet, students who sat the assessment were asked how much effort they put in the test and to consider how much effort they would have put in the test if their performance in PISA had counted towards their school marks. Responses could range from 1 to 10 with 10 being maximum effort. We include the answers to the question on the effort students reported having put in the PISA test and recoded it so that a higher value signals lower effort. We also consider a second index, which is the difference between the effort that would have been invested if the test were marked and the actual effort invested. Students with a greater score on such index should be seen as being less conscientious or intrinsically motivated, since they reported having invested less effort in the PISA test compared to how much effort they would have invested had the PISA test been associated with an extrinsic reward, i.e. a school grade. Both indices were standardised such that they had a mean of zero across the pooled student population and a standard deviation of one.

2.2.2. Questionnaire-based behavioural indicators

We consider the following questionnaire-based behavioural indicators based on the core student questionnaires[ii]: item non-response, non-differentiation and inconsistency.

Item non-response rate. Survey item non-response is considered to provide information on the level of effort and motivation of students (Hitt, Trivitt, & Cheng, 2016). The indicator is

constructed using multiple-choice questions for which an answer was possible and applicable to the responding student (we dropped items where the response is conditional on prior items). The item non-response rate indicator is defined as:

$$Item\ non-response = \frac{Number\ of\ applicable\ items\ left\ blank}{Number\ of\ applicable\ items}$$

The index was standardised such that it had a mean of zero across the pooled student population and a standard deviation of one. Higher values indicate that students left more questions unanswered.

Inconsistency. The inconsistency indicator (also called response non-consistency in the literature) was developed by (Hitt, 2015) and adopted in subsequent papers, such as (Zamarro et al., 2019). The measure is based on item-rest correlations. The measure relies on the assumption that, in an internally consistent scale, the answer to a specific item should be correlated with the answers to the rest of the items that make up that scale. A student who answers a specific item in a way that is unpredictable based on the answers to other items in the same scale is thus considered to provide a "careless answer". The response non-consistency indicator is constructed using Likert-type items belonging to a scale meant to measure some latent attitude or behaviour of respondents. First, we adjusted answers of reverse-coded items (such as negatively-phrased items in positively-phrased scales) and then for every item, we regressed the item response on the average of answers in the remaining items belonging to the same scale. We run regressions separately for each country in our sample to account for the fact that the internal consistency of scales might differ across countries and considered only respondents who had non-missing responses for at least 3 items in a scale. For each item, we fit the following linear model:

$$Y_{ijs} = \beta_0 + \beta_1 \bar{X}_{is,-j} + \eta_{ijs}$$

Where Y_{ijs} represents the answer of student *i* to item *s* in scale *j* and $\bar{X}_{is,-j}$ represents the average response on all remaining items in the scale, β_0 represents a constant and η_{ijs} is an item-specific, scale-specific and student-specific error term representing the degree to which student *i* gave an unpredictable answer to item *s* in scale *j*. These sets of regressions are equivalent to item-rest correlations common in psychometrics (Hitt, 2015). For each student, we averaged the absolute values of all residuals to obtain a measure of non-consistency in their answers across the entire PISA survey. Higher values imply that a student gave more inconsistent answers than another student. In PISA 2012, we examine 93 items across 17 scales in the main questionnaire. The index was standardised such that it had a mean of zero across the pooled student population and a standard deviation of one.

Non-differentiation. The non-differentiation indicator reflects the extent to which students tend to select the same response across a set of similar and related items. This can be the result of a careful analysis of questions, but also of satisficing behaviour. Respondents could realize that the items in a certain set are similar and, in order to minimize the effort exerted, give the same response to all. Several studies have found support for this claim: non-differentiation is more common among less educated individuals and towards the end of a questionnaire compared to the beginning (Knowles, 1988; Krosnick, 1991; Vannette & Krosnick, 2014). We focus only on Likert-type item sets in PISA, since for this type of questions the likelihood of mindful respondents not differentiating their answers is lower. We limit the analysis to item sets having at least 4 items (a total of 16 in PISA 2012) and to students having non-missing responses for at least 3 items. In the most stringent version of our measure, a student is considered to be non-differentiating within an item set if he or she gave the same answer to all items in the item set. To compute a non-differentiation metric for each student, we compute the percentage of item sets in which they did not differentiate their answers out of all valid items sets. The index was standardised

such that it had a mean of zero across the pooled student population and a standard deviation of one

$$Non - differentiation = \frac{Number\ of\ applicable\ items\ sets\ without\ differentiation}{Number\ of\ applicable\ item\ sets}$$

Since absolute non-differentiation can be too stringent, we decided to also adopt a more lenient version of the index (Barge & Gehlbach, 2012). We relaxed the definition such that a student is considered not to differentiate within an item set when he or she gives the same response in all but one item in the set. All analyses were replicated using this index and are available from the authors upon request.

2.2.3. Test-based behavioural indicators

Decline in performance. The core cognitive test in PISA 2012 was delivered as a paper and pencil test. This was designed to last two hours in total for each participant and was organised around a series of clusters of test questions. Each cluster was designed to take about 30 minutes to complete. Each student was randomly allocated a booklet containing four clusters of test questions and a total of 13 booklets were administered in 2012. Each booklet contained different clusters, which were rotated across the booklets so that each cluster was administered at least once with any other cluster and each cluster appeared at least once in one of the four potential positions within the booklet. Table B1 in the Supplementary Annex illustrates the design of the standard paper and pencil booklets in PISA 2012.

On average, in various PISA waves, the number of correct responses students give declines across the test: students tend to display higher performance in the first cluster than in the second and so on (Borghans & Schils, 2013; Zamarro et al., 2019). While different strategies have been used to develop PSEC indicators based on decline in performance in low-stakes standardised

assessments (Borghans & Schils, 2013; Brunello et al., 2018; Zamarro et al., 2019) we develop an individual level indicator following Zamarro et al. (2019). The assumption behind the measure is that students with high levels of socio-emotional and motivational skills are more likely to maintain a similar level of performance throughout the test while those with lower levels of socio-emotional and motivational skills are more likely to display steep declines in performance as a function of item position. We only consider performance in the first three clusters to deal with end of test non-response. Some students in fact fail to reach items at the end of the test, which could lead to biased estimates (Author, 2016; Debeer & Janssen, 2013).

To obtain our measure, we estimate the following linear random coefficient model:

$$y_{ij} = \delta_0 + \delta_0^i + \delta_1 Q_{ij} + \delta_1^i Q_{ij} + \gamma_j + \theta_j + \varepsilon_{ij}$$

Where y_{ij} is equals zero if respondent i answered incorrectly to question j and 1 if he or she answered correctly or if he or she got partial credit for his/her answers[iii]. Q_{ij} represents the position of question j rescaled for each student such that the first question takes value zero and the last item in the third cluster takes value 1. δ_0 represents the average student's performance on the first question in the test and δ_1 is the average performance drop from the first question to the last. γ_j are question fixed effects, which allow us to control for question difficulty and nature (such as multiple choice versus open-ended question). θ_j are booklet fixed effects to control for the sequence of clusters in the booklet (for example starting with a math or reading cluster). δ_0^i is a random intercept and δ_1^i is a random coefficient that allow for students to deviate from the average values.

The model was estimated separately for each country using Maximum Likelihood methods and allowing maximum flexibility in the covariance matrix for random effects (all variances and covariances could be distinctly estimated). Standard errors were clustered at the school level, to account for the clustered nature of the PISA samples. Fitting the model produced estimates of the

standard deviation of random effects (δ_0^i and δ_1^i) and the correlation between them. We used these to predict the best linear unbiased predictions (BLUPS) for the random effects.

The index of performance decline corresponds to BLUP estimate of the random slope parameter δ_1^i, which measures individuals' decline in performance between the first and the last question (in the third cluster), accounting for question difficulty and the booklet that the student was assigned.

The PISA 2000 test design was different from subsequent editions in that the major domain (reading) was assessed significantly more in depth than the other two. Seven booklets began with three reading clusters and ended with a different domain. To maintain a balanced sample, we decided to focus our analyses only on those booklets. Therefore, our measure in PISA 2000 is entirely based on performance decline in the reading assessment. The index was standardised such that it had a mean of zero across the pooled student population and a standard deviation of one. Higher values indicate that students had a steeper performance decline.

Response time effort. In 2012 some countries administered an optional computer-based assessment on top of the paper-based assessment instruments, generating data on participants' interactions with the testing platform, including a timestamp to mark each interaction. We follow Wise and Kong (2005) and develop a response-time effort indicator using the log files for the PISA 2012 computer-based assessment. The indicator constitutes the proportion of items, out of the total number of items in the test, on which respondents spent less than a threshold time T. We use a threshold of 5 seconds for items based on short texts but also provide robustness checks using more stringent or lenient thresholds (3 seconds and 8 seconds respectively). The index was standardised such that it had a mean of zero across the pooled student population and a standard deviation of one. Higher values indicate that students exerted less effort in their responses.

2.2.4. Validation of measure construction

We performed a number of robustness checks to assess the questionnaire-based and test-based behavioural indicators, exploring how they might vary according to certain features of the PISA assessment.

In 2012 the background questionnaire had a rotation design and therefore different students were randomly allocated different questionnaire material (scales with a different number of items or with a different length of the prompts or the same material but placed in different positions within the questionnaire). In Supplementary Annex C we illustrate the rotation design and provide estimates of how much the value of different questionnaire-based behavioural indicators depends on the questionnaire set students were administered. We find that non-differentiation and inconsistency appear to be sensitive while non-response is less dependent on the specific questions contained in the questionnaire. Given the structure of our data, we could separately assess the robustness of indicators to specific features of the questionnaires such as their length or the type and order of items they contained.

We performed other robustness checks for the test-based indicator of performance decline. We examined variations in the performance decline indicator depending on the content of the test (amount of reading and mathematics material and position of mathematics and reading items). Results in Supplementary Online Annex D suggest that the performance decline is not related to the structure of the assessment booklet. In order to identify the robustness of the indicator to mode of administration, in Supplementary Annex table E we estimated differences in performance decline across the PISA 2012 computer-based assessment and the PISA 2012 paper-based assessment. Results indicate that students' ranking in the performance decline measure based on the computer-based test is positively related to their ranking in paper-based performance decline.

2.3. Background variables

In a set of analyses, we analyse differences in PSEC indicators by students' gender, socio-economic background and immigrant background. These variables were constructed using information gathered in the PISA students questionnaire. Socio-economic background (SES) is defined using the PISA index of economic, social and cultural status The index is an aggregate indicator reflecting differences across students in parental educational attainment, parental occupational status and household resources (see Organisation for Economic Co-operation and Development (OECD), 2014) for a detailed description of the index). We provide results comparing students in the top and bottom quartile of their national distribution. Students are defined as having an immigrant background if their parents were born outside the country in which they sat the PISA test; in the analysis, these students are compared with those with two native-born parents.

In addition to these variables, in the longitudinal analyses, we also control for other variables gathered in the PISA background questionnaire: students' school grade and age when they sat the PISA assessment, and whether respondents reported most frequently speaking the language of the PISA assessment at home.We also control for for their achievement in the first cluster of the assessment (using an indicator of the average percentage of correct responses in the first booklet).

3. Results

3.1. Individual and country level correlations of PSEC indicators

In order to illustrate differences across countries in average levels of PSEC indicators we compare, for each indicator, if the country specific mean was in line, above or below the estimated

mean across all the countries that took part in the analysis[4]. Table F1 in the Supplementary Annex illustrates country specific results for means, Table F2 illustrates country specific results for standard deviations and Table F3 reports country level correlations between different measures. Results indicate that at the country level the correlation between different PSEC indicators is weak. Countries with comparatively high mean levels on some indicators have comparatively low levels on other indicators.

Table 1 indicates that individual level associations between different PSEC measures are generally weak although almost all associations are in the expected directions. Correlations are higher between the three self-reported measures: r = .21 between lack of perseverance and lack of effort in the PISA test; r = .69 between lack of effort in the PISA test and the difference between this measure and this measure in a hypothetical PISA test that counted towards their grade; and r = .10 between lack of perseverance and the difference in effort between the PISA low-stake condition and the hypothetical higher stake condition.

Meta-analytic reviews of cross-method convergence of several psychological constructs have found weak correlations between informant-report and task measures (Meyer, 2001). Our estimates are comparable to some of Meyer's result. However, correlations in Table 1 provide only weak evidence of convergent validity across measures and cast some doubt on the interpretation of behavioural measures as indicators of PSEC psychological traits.

[4] The average mean across countries is not precisely 0 because the z score standardization was conducted on the weighted pooled sample while the mean across countries represents the arithmetic mean of weighted country specific samples,i.e. while the country specific mean reflects the country specific population distribution, countries with large overall populations contribute to the same extent to the average as do countries with small populations. By contrast, in the pooled sample, these countries contribute more.

Table 1. Individual level correlations between PSEC indicators

	Self-reported lack of perseverance	Self-reported lack of effort in test	Difference in effort (PISA-marked)	Non-response	Non-differentiation	Non-differentiation (more lenient)	Inconsistency	Performance decline	Response time effort (3 seconds)	Response time effort (5 seconds)	Response time effort (10 seconds)
Self-reported lack of perseverance	1.00										
Self-reported lack of effort in test	0.21	1.00									
Difference in effort (PISA-marked)	0.10	0.69	1.00								
Non-response	0.07	0.08	0.02	1.00							
Non-differentiation	-0.07	-0.01	-0.05	0.03	1.00						
Non-differentiation (more lenient)	-0.07	-0.01	-0.05	0.01	1.00	1.00					
Inconsistency	0.14	0.10	0.06	0.12	-0.33	-0.33	1.00				
Performance decline	0.05	0.07	0.05	0.01	0.05	0.05	0.00	1.00			
Response time effort (3 seconds)	0.01	0.05	0.02	0.04	0.04	0.04	0.03	0.03	1.00		
Response time effort (5 seconds)	0.03	0.06	0.04	0.06	0.09	0.08	0.05	0.05	0.74	1.00	
Response time effort (10 seconds)	0.05	0.07	0.03	0.10	0.14	0.13	0.09	0.08	0.45	0.74	1.00

Note: PISA 2012 data. The number of individual level observations is 224 748 for results involving indicators obtained from the paper-based assessment and 59 190 for results involving indicators obtained from the computer-based assessment. All correlations are statistically significant at the 0.001 level, except for the one between response time effort (3 seconds) and self-reported lack of perseverance, and between self-reported lack of effort in test and non-differentiation at the 0.10 level.

In order to establish the stability of PSEC indicators over time, we examine country level correlations between each indicator measured in 2012 and in 2003, the only two waves including at least one self-reported instrument, the effort thermometer. We present comparisons for available PSEC measures as well as the PISA indexes of sense of belonging and instrumental motivation, which we used as a benchmark of changes over time in self-reported constructs in a non-PSEC measure. Although successive cohorts of students might have different levels of PSEC, we do not expect large differences over a nine-year period. We present Pearson correlation coefficients between the two survey waves in Table 3 and scatterplots indicating how countries fared in 2003 and 2012 in Supplementary Online Annex G.

Table 2 suggests that the questionnaire-based indicators are more stable over time than the performance decline indicator.

Table 2. Correlations between country level PSEC indicators measured in PISA 2003 and PISA 2012

	Absolute correlation	Rank correlation
Self-reported lack of effort in test	.842	.755
Difference in effort (PISA-marked)	.860	.704
Non-response	.282	.525
Non-differentiation	.845	.848
Inconsistency	.897	.880
Performance decline	.405	.526
Instrumental motivation	0.945	0.945
Sense of belonging at school	0.714	0.724

Note: PISA 2012 data.

3.2. Differences across student background characteristics

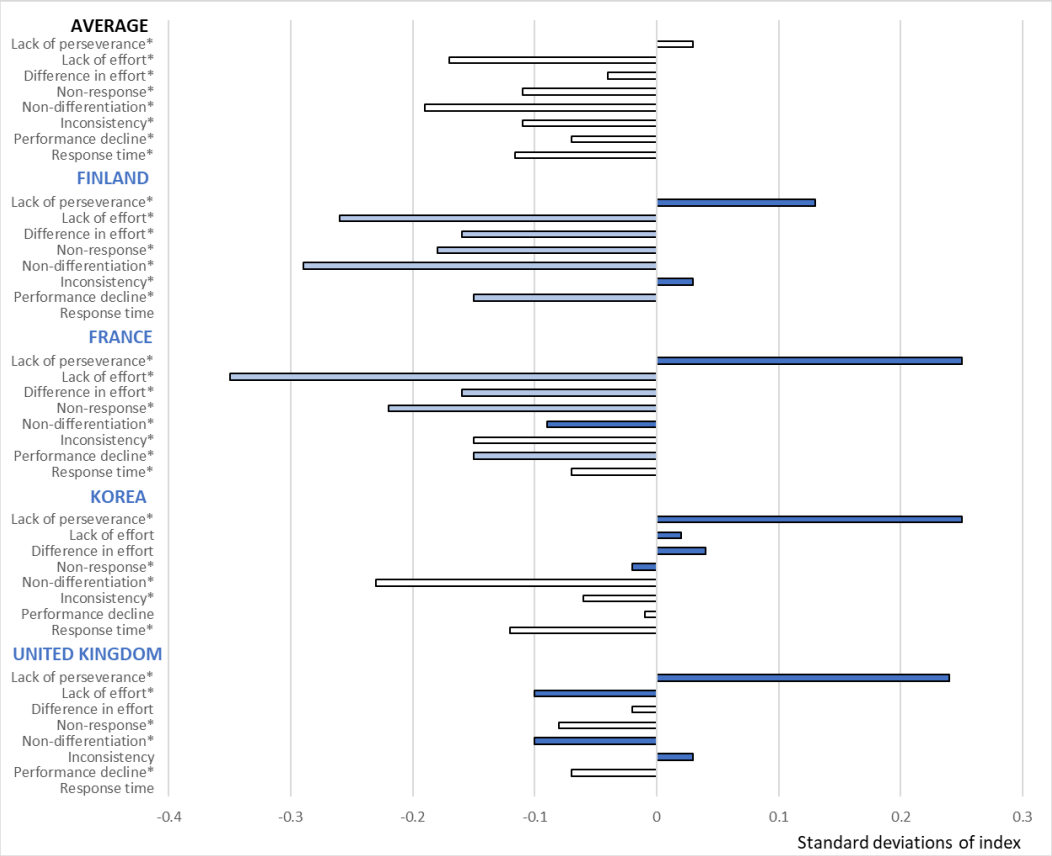
In Figure 1 we present gender differences in the PSEC indicator for a sample of countries. Positive values indicate that females have higher z-values than males (and thus worse PSEC outcomes), while negative values indicate that males have higher z-values than females (and thus worse PSEC outcomes). Figure 2 shows results by socio-economic status (positive values indicate worse PSEC outcomes for socio-economically advantaged students), and Figure 3 results by immigrant background (positive values indicate worse PSEC outcomes for students without an immigrant background) for a sample of countries. Results on differences across gender, socio-economic background and immigrant background for the full set of countries with available data can be found in Tables H2, I2 and, J2 in the Supplementary Online Annex.

On average, the gender gap in the standard self-reported lack of perseverance, although statistically significant, is quantitatively very small: 15-year-old female students report slightly more lack of perseverance than 15-year boys (d = 0.03). By contrast, on all other behavioural indicators, females displayed lower overall z scores than males (i.e. had better PSEC outcomes). In the questionnaire, they were less likely than males to provide inconsistent response patters (d = -0.11), to fail to provide answers (d = -0.11) and to provide undifferentiated responses (d = -0.19). Among

test based behavioural indicators, females displayed, on average, less steep performance decline than males (d = -0.07) and were less likely to skip rapidly or guess an answer in the computer-based assessment (d = -0.12).

Gender differences varied significantly across countries and indicators. Variability across countries in gender gaps was particularly pronounced for the self-reported lack of perseverance indicator and less pronounced for the performance decline indicator, although a number of outlier countries could be identified. These results suggest that on all behavioural dimensions males appear to have lower PSEC outcomes than females. However, when examining self-reports, males report similar levels of PSEC as females, except for effort expended on the PISA test, indicating an awareness of lower behavioural involvement in the survey. The gender gap in this indicator is similar in size to the behavioural indicators.

Figure 1. Gender differences in PSEC indicators

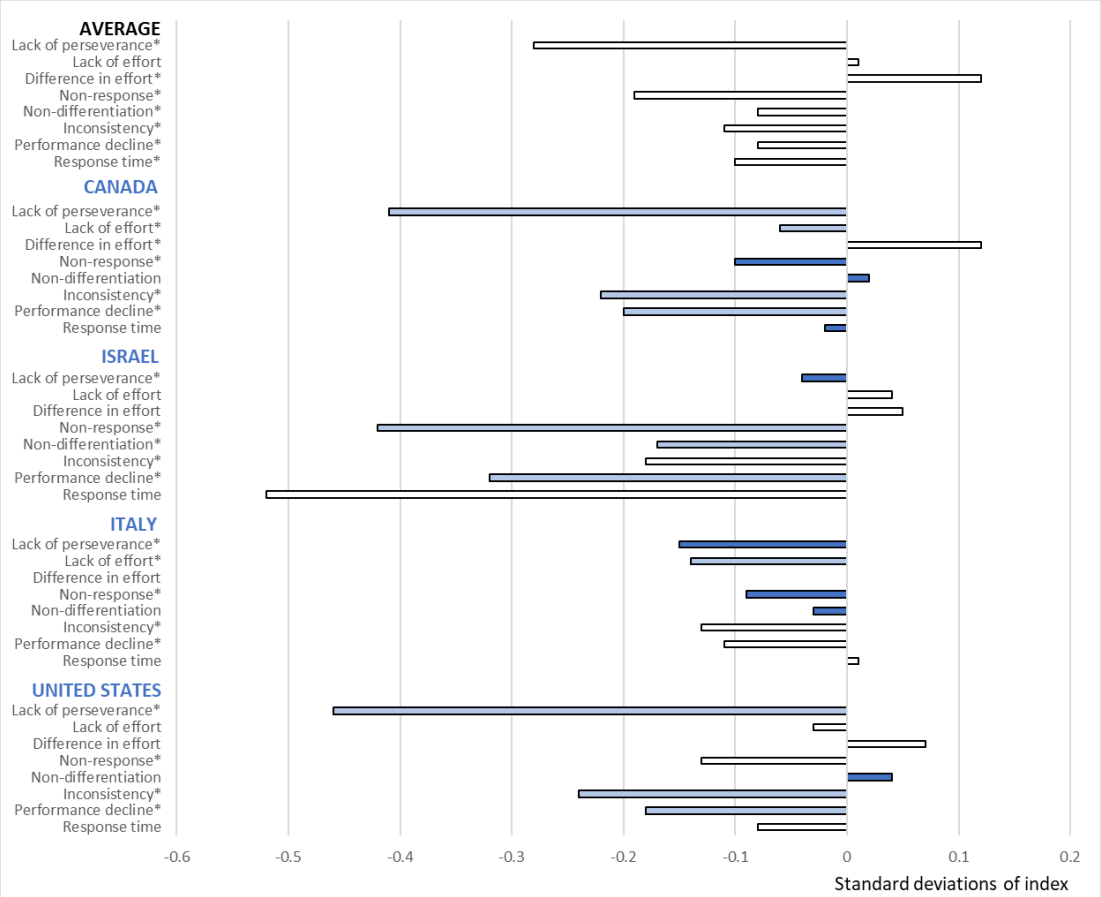


Notes: The gender gap refers to the difference between females and males. A negative number implies that females have lower mean values than males. A positive value implies that females have higher mean values than males. An asterisk (*) next to an indicator name indicates that the estimated d value is statistically significant at least at the 5% level. A dark shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null hypothesis of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Data for all countries in the sample can be consulted in Table I2 in the Online Supplementary Annex. Source: PISA 2012 data.

Figure 2 indicates that socio-economically disadvantaged students see themselves as less perseverant than socio-economically advantaged students. On average across countries in our sample the difference between the two groups in the self-reported lack of perseverance indicator corresponds to a medium size gap (d=-.28). However, both groups report having expended a similar amount of effort on the PISA test (d = 0.01; $p \geq 0.05$). On average, socio-economic differences in behavioural indicators of PSEC exist but are smaller than gaps in self-reported lack of perseverance (d = -0.12 for self-reported difference in effort between PISA and a graded PISA test; d = -.019 for item non-response; d = -0.07 for non-differentiation; d = -0.11 for inconsistency; d = -.09 for performance

decline and *d* = -0.012 for response time effort). Country rankings differ greatly depending on which indicator is considered.

Figure 2. Socio-economic differences in PSEC indicators



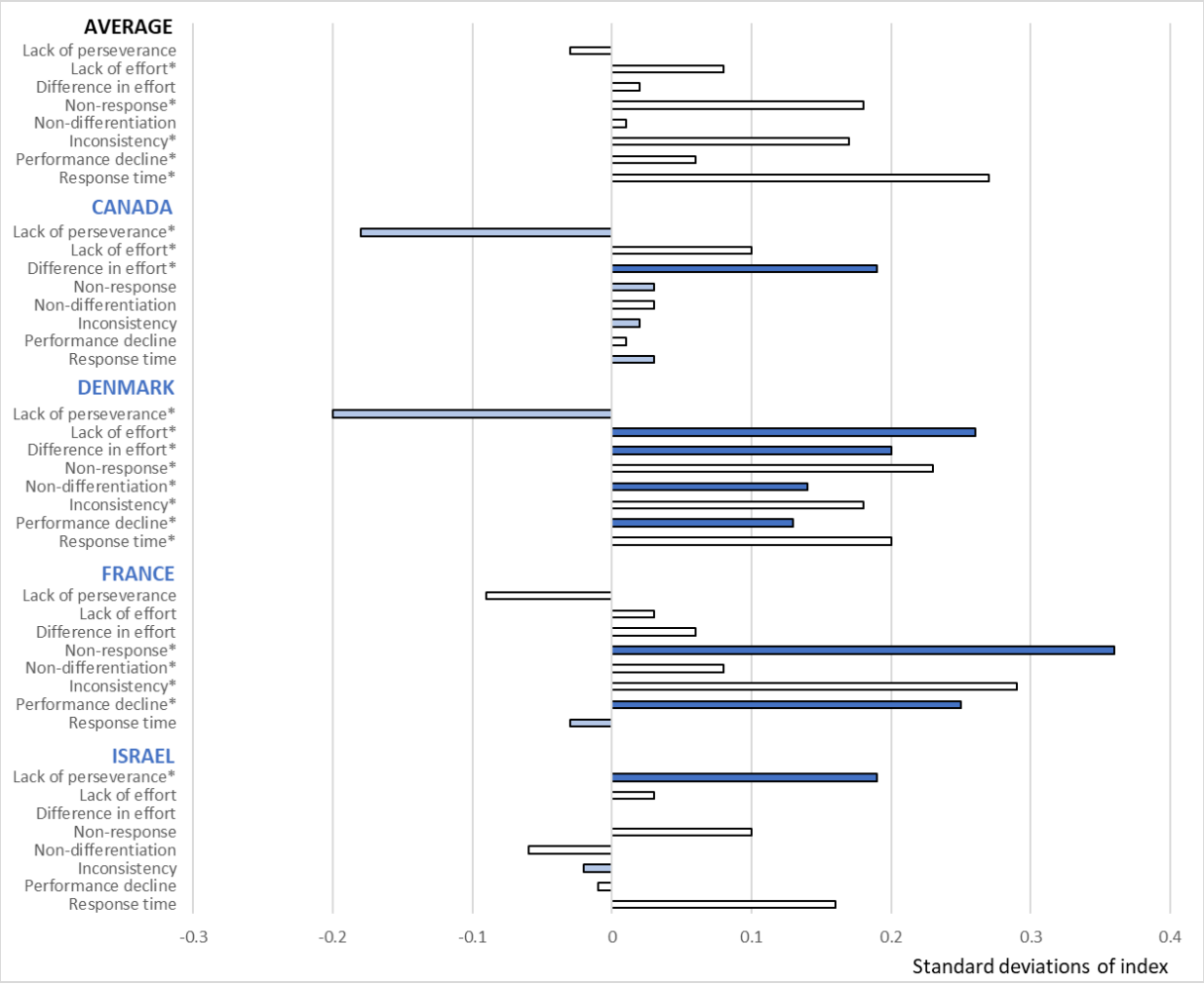
Notes: The socio-economic (SES) gap refers to the difference between socio-economically advantaged and socio-economically disadvantaged students. PISA contains an aggregate SES measure based on students' reports on their parents' educational attainment, occupational status and availability of economic and cultural resources in their home. Advantaged students are students in the top 25% of the country specific distribution of SES. Disadvantages students are students in the bottom 25% of the country specific distribution of SES. A negative number implies that advantaged have lower mean values than disadvantaged students. A positive value implies that advantaged have higher mean values than disadvantaged students. An asterisk (*) next to an indicator name indicates that the estimated d value is statistically significant at least at the 5% level. A dark shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null hypothesis of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Data for all countries in the sample can be consulted in Table J2 in the Online Supplementary Annex. Source: PISA 2012 data.

Figure 3 indicates that there are no differences between students with and those without an immigrant background in self-reported lack of perseverance (*d* = -0.03) and the questionnaire-based behavioural indicator of non-differentiation (*d* = 0.01). However, immigrant students have higher mean values on behavioural indicators based on questionnaire items (*d* = 0.18 for non-response and

d = 0.17 for inconsistency). The difference between immigrant and non-immigrant students in performance decline is small (d = 0.06) and highest for the response time effort indicator (d = 0.27).

In many OECD countries students without an immigrant background indicate that they are less perseverant than immigrant students, although on behavioural indicators students with an immigrant background have lower PSEC values than students without such background, a possible reflection of their lower language abilities. Figure 3 shows that gaps varied significantly across countries and indicators.

Figure 3. Differences by immigrant background in PSEC indicators



Notes: The immigrant gap refers to the difference between students with and students without an immigrant background. A negative number implies that immigrant students have lower mean values than students without an immigrant background. A positive value implies that immigrant students have higher mean values than students without an immigrant background. An asterisk (*) next to an indicator name indicates that the estimated d value is statistically significant at least at the 5% level. A dark shade indicates that the estimated d value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light shade indicates that the estimated d value for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null hypothesis of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Data for all countries in the sample can be consulted in Table K2 in the Online Supplementary Annex. Source: PISA 2012 data.

3.3. Relationship with contemporaneous academic achievement

We estimate the following models:

$$y_{ik} = \delta_0 + \delta_1 Q_{ik} + \gamma_k + \varepsilon_{ij} \quad \text{(A)}$$

$$y_{ik} = \delta_0 + \delta_1 Q_{ik} + \delta_2 V_{ik} + \gamma_k + \varepsilon_{ij} \quad \text{(B)}$$

$$y_{ik} = \delta_0 + \delta_1 Q_{ik} + \delta_2 V_{ik} + \delta_3 U_{ik} + \gamma_k + \varepsilon_{ij} \quad \text{(C)}$$

Baseline models described by equation (A) estimate the association δ_1 between each achievement domain *y* for student *i* in country *k* and each PSEC indicator *Q* without controls (except for country fixed effects γ). Models described by equation (B) estimate the association between each achievement domain and each PSEC indicator controlling for country fixed effects as well as a set of additional control variables *V* including grade level, age, gender, immigrant background, socio-economic status, and if the student was a native-speaker of the language in which the PISA assessment was delivered. For behavioural measures, Models described by equation (C) further control for the self-reported perseverance measure *U*, to assess whether the relationship between behavioural indicators of PSEC and performance is robust to inclusion of self-reported measures of PSEC. All models include students with non-missing values for all PSEC indicators and control variables to ensure comparability of results across specifications.

In Table 3 we report average findings across countries in our sample on the association between each achievement domain in PISA (reading, mathematics, science and problem solving) and each indicators of PSEC for the three sets of models.

Table 3 suggests that all indicators of PSEC are negatively associated with contemporaneous achievement in reading, mathematics, science and domain general problem solving. Introducing background characteristics generally reduces the association between PSEC indicators and contemporaneous achievement but the reduction is small. On average, across countries in our sample, a change in one SD in the self-reported lack of perseverance is associated with a change of 14% of a *SD* in reading, 17% *SD* in math, 16% *SD* in science and 14% *SD* in problem solving performance.

The questionnaire-based behavioural indicator of item non-response is the PSEC measure that is most strongly associated with contemporaneous academic achievement: a change of one SD in non-response is associated with a change of 37% of a *SD* in reading, 33% *SD* in math, 34% *SD* in science and 34% *SD* in problem solving performance. The association between the other questionnaire-based PSEC measures of non-differentiation and inconsistency is smaller: i.e. a change in one *SD* in these measures is associated with a change of between 8% of a *SD* and 19% of a *SD* depending on the achievement domain considered.

Table 3. The relationship between PSEC indicators and achievement

		Reading			Math			Science			Collaborative problem-solving		
Indicator	Model	b	Sig	Adjusted R-squared	b	sig	Adjusted R-squared	b	sig	Adjusted R-squared	b	sig	Adjusted R-squared
Self-reported lack of perseverance	A	-0.14	***	0.0069	-0.17	***	0.007	-0.16	***	0.0056	-0.14	***	.0075
	B	-0.11	***	0.3	-0.13	***	0.31	-0.11	***	0.3	-0.10	***	.19
Self-reported lack of effort in test	A	-0.13	***	0.0000024	-0.10	***	0.0014	-0.11	***	0.00072	-0.09	***	0.000015
	B	-0.11	***	0.26	-0.10	***	0.26	-0.10	***	0.26	-0.10	***	.18
Difference in effort (PISA-marked)	A	-0.02	***	0.0083	-0.02	***	0.0098	-0.02	***	0.0096	-0.03	***	.00095
	B	-0.04	***	0.26	-0.05	***	0.26	-0.05	***	0.25	-0.06	***	.17
Non-response	A	-0.37	***	0.07	-0.33	***	0.062	-0.34	***	0.066	-0.34	***	.049
	B	-0.27	***	0.33	-0.26	***	0.33	-0.27	***	0.33	-0.27	***	.21
	C	-0.32	***	0.33	-0.32	***	0.33	-0.32	***	0.33	-0.33	***	.22
Non-differentiation	A	-0.14	***	0.036	-0.09	***	0.023	-0.11	***	0.029	-0.12	***	.011
	B	-0.10	***	0.32	-0.08	***	0.31	-0.09	***	0.31	-0.10	***	.2
	C	-0.13	***	0.32	-0.09	***	0.32	-0.11	***	0.32	-0.12	***	.2
Inconsistency	A	-0.14	***	0.014	-0.13	***	0.013	-0.13	***	0.011	-0.13	***	.024
	B	-0.09	***	0.31	-0.10	***	0.31	-0.09	***	0.31	-0.10	***	.2
	C	-0.08	***	0.31	-0.09	***	0.31	-0.07	***	0.31	-0.09	***	.2
Performance decline	A	-0.15	***	0.0027	-0.14	***	0.00093	-0.14	***	0.0016	-0.18	***	.0076
	B	-0.13	***	0.31	-0.12	***	0.31	-0.12	***	0.3	-0.15	***	.2
	C	-0.12	***	0.31	-0.12	***	0.31	-0.12	***	0.31	-0.14	***	.2
Response time effort (5 seconds)	A	-0.19	***	0.021	-0.16	***	0.016	-0.18	***	0.017	-0.23	***	.032
	B	-0.15	***	0.21	-0.13	***	0.19	-0.15	***	0.19	-0.20	***	.18
	C	-0.13	***	0.2	-0.12	***	0.19	-0.14	***	0.2	-0.19	***	.18

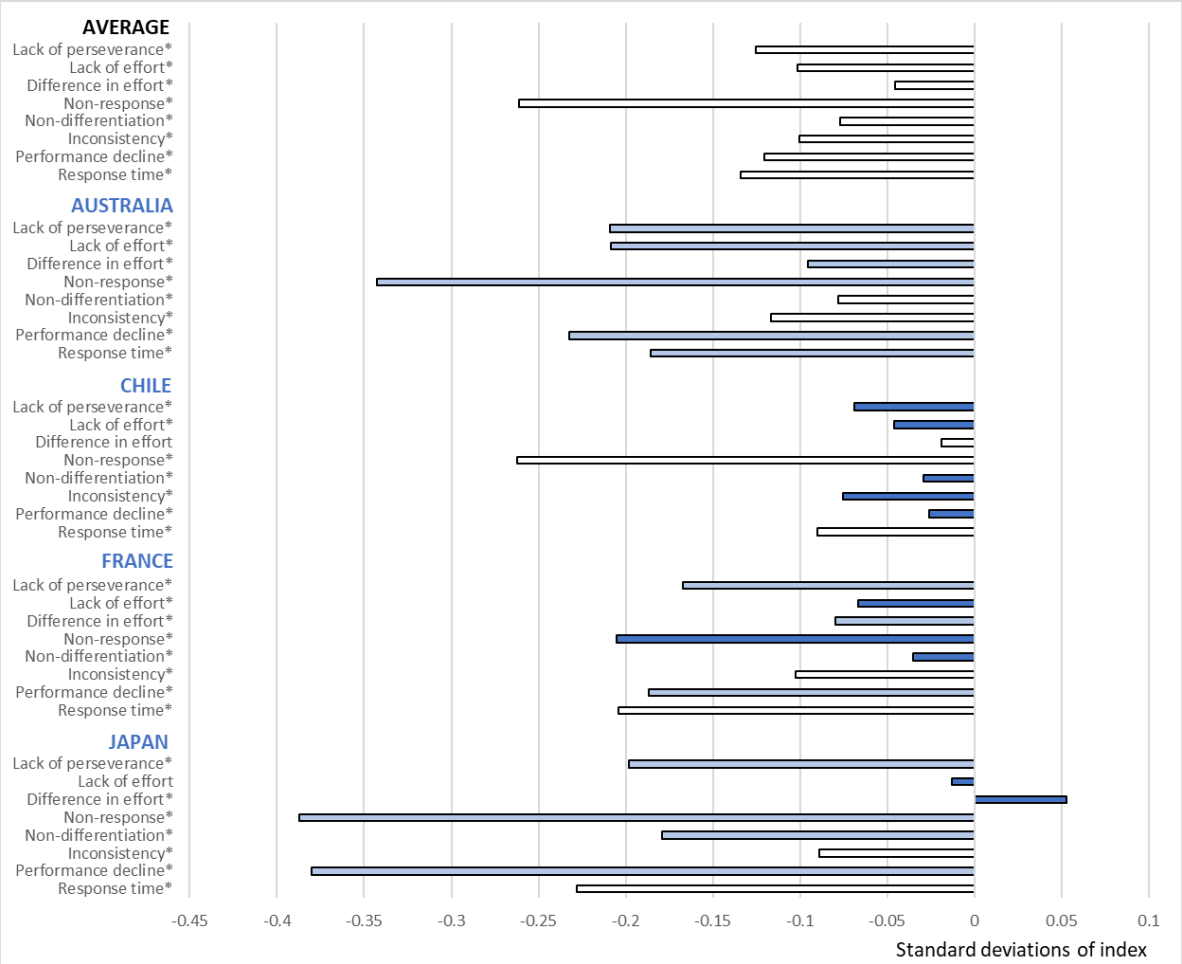
Notes: The number of observations in models A and B testing the significance of self-reported indicators for math, science and reading is: 247598 (self-reported lack of perseverance) 340096 (self-reported lack of effort) and 340096 (difference in effort (PISA-marked). For collaborative problem-solving it is: 150252 (self-reported lack of perseverance) 210212 (self-reported lack of effort) and 210212 (difference in effort (PISA-marked)).
Models B and C control for students' school grade, age, gender, socio-economic status, immigrant background and linguistic background. For indicators obtained from the paper-based assessment, the number of observations is 375657 in models A and B and 247598 in model C for math, science and reading (227316 in models A and B and 150252 in model C for collaborative problem-solving). For the indicator obtained from the computer-based assessment, the number of observations is 48502 in A and B and 31982 in model C for math, science and reading (48502 in A and B and 31982 in model C). * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Results in Model (C) of Table 3 suggests that estimated associations between behavioural indicators and achievement are robust to the introduction of self-reported measures (i.e. standardised regression coefficients remain similar in size and statistically significant).

In Figure 4 we illustrate variation across a selection of countries in the strength of the association between PSEC indicators and mathematics achievement accounting for control variables (as in model B above). Tables for the full sample of countries and the associations of PSEC indicators with all other PISA domains (mathematics, reading, science and problem solving) are available in Tables J1-J4 in the Supplementary Online Annex.

Figure 4 and the materials in the Supplementary Online annex suggest that associations vary considerably across countries. The strength of the association varies for the same country across different indicators. Moreover, relative country rankings depend, to an extent, on the specific indicator used and that the same country can be considered to be above, in line, or below the average depending on the indicator analysed.

Figure 4. Associations between PSEC indicators and mathematics achievement



Notes: An asterisk (*) next to an indicator name indicates that the estimated coefficient is statistically significant at least at the 5% level. A dark shade indicates that the estimated association for a specific country is statistically significantly above the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated association for a specific indicator and for a specific country is statistically significantly below the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Data for all countries in the sample can be consulted in Table K1 in the Online Supplementary Annex and data for other PISA outcomes can be consulted in Tables L2-L4. Source: PISA 2012 data.

3.4. Predictive validity of PSEC indicators of outcomes in young adulthood

We analyse the predictive validity of PSEC indicators measured in PISA for 15 years old students by examining their association with the likelihood of completing upper secondary education (academic or vocational) and of completing a university degree by the age of 25.

We estimate the following two linear probability models:

$$y_{it} = \delta_0 + \delta_1 Q_{it-10} + \delta_2 V_{it-10} + \varepsilon_{ijt} \qquad (1)$$

$$y_{it} = \delta_0 + \delta_1 \bar{Q}_{it-1} + \delta_2 V_{it-10} + \delta_3 u_{it-10} + \varepsilon_{ijt} \quad (2)$$

In model (1) we regressed the two binary outcome variables *y* (having completed high-school and having completed university by age 25) on one PSEC indicator (*Q*) at a time and a set of control variables *V* including students' school grade, age, gender, immigrant background, socio-economic status[iv], whether respondents were native-speakers at the time of the PISA assessment as well as for their achievement in the first cluster of the assessment (using an indicator of the average percentage of correct responses in the first booklet). The latter was taken as a proxy of their cognitive skills and was preferred to actual assessment scores because of the possibility of collinearity with the index of performance decline. In model (2), we estimate the relationship between the education completion at 25 and the behavioural indicators of PSEC $\bar{Q}$ (performance decline, inconsistency, non-response, non-differentiation), when accounting for self-reported measures of PSEC. For Denmark and Switzerland, model (2) includes a self-reported index perseverance (u), obtained from students responses to questions that were administered as part of an optional Cross-Curricular Competences questionnaire (CCC) in PISA 2000 (see Supplementary Online Annex L). For Australia, the only self-reported measure of PSEC available is the effort thermometer, so model 2 for Australia includes this measure as additional control. No self-reported measure of PSEC is available for Canada.

Table 4 illustrates results for model (1). Coefficients associated with the PSEC indices δ_1 measure the change in the probability of completing upper-secondary education and university by age 25 as a function of a change in one standard deviation in PSEC indicators[v]. The results indicate that performance decline is associated with university completion three out of four countries: in Australia and Switzerland, students with similar characteristics and performance who had a one SD greater decline in performance in the PISA test in 2000 had a 7point lower probability of completing university by age 25and in Canada they had a 13% points lower porbability. By contrast, a difference of one SD in the self-reported lack of perseverance index was associated with a change of 5% points

in the probability of completing university by age 25 in Denmark, 4% points in Switzerland. Among questionnaire-based behavioural indicators, the non-differentiation indicator is not associated with university completion in any of the countries analysed. Inconsistency is associated with university completion in Canada and Switzerland, while non-response is associated with university completion in Denmark and Switzreland and with upper secondary completion in Canada. Indicators of PSEC are less strongly associated with upper secondary school completion than university completion: performance decline is the only indicator that are associated with the probability that individuals will complete upper secondary school in at least two countries (Denmark and Australia).

Table 4. The predictive power of PSEC indicators with respect to university and upper secondary school completion

		Denmark (N=1192)		**Switzerland** (N=1948)		**Australia** (N=3196)		**Canada** (N=5863)	
Indicator	Outcome	b	Adjusted R squared	b	Adjusted R squared	b	Adjusted R squared	b	Adjusted R squared
Performance decline	Upper secondary	-0.045*	.0814	-0.047	.0611	-0.0093**	.0496	-0.027	0.039
	University	-0.066	.148	-0.074***	.125	-0.070*	.205	-0.132***	0.20
Inconsistency	Upper secondary	-0.062	.0949	-0.027	.054	-0.015	.051	-0.009	0.036
	University	-0.012	.139	-0.031*	.111	0.011	.187	-0.025*	0.18
Non-response	Upper secondary	-0.053	.0811	0.023	.0499	-0.028	.0535	-0.036*	0.042
	University	-0.050*	.143	-0.072***	.115	-0.014	.187	-0.034	0.18
Non-differentiation	Upper secondary	0.0049	.0729	0.0052	.0487	-0.012	.05	0.003	0.035
	University	0.015	.139	-0.022	.107	-0.0015	.187	0.014	0.18
Self-reported lack of perseverance	Upper secondary	-0.017	.0754	0.015	.0512				
	University	-0.049*	.149	0.037*	.116				
Self-reported of lack of self-efficacy	Upper secondary	-0.023	.077	0.017	.0507				
	University	-0.050	.149	0.017	.108				
Self-reported of lack of instrumental motivation	Upper secondary	0.0058	.0731	0.026	.0546				
	University	-0.0040	.139	0.033*	.112				
Self-reported of lack of control expectations	Upper secondary	-0.0052	.0731	0.024	.0539				
	University	-0.044	.147	0.044**	.117				
Self-reported lack of effort in text	Upper secondary					-0.016	.0534		
	University					0.0078	.187		

Notes: Denmark: PISA 2000-PIAAC 2012 link. Switzerland: PISA 2000-TREE 2010 follow-up. Australia: PISA 2003-LSAY 2013. Canada: PISA 2000-YITS 2010.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 5 presents results for model (2). Results indicate that the association between performance decline and educational outcomes at age 25 is statistically significant, quantitatively moderate and robust to the inclusion of self-reported measures of PSEC. Furthermore, models that include the performance decline measure are the ones with the highest explained variance overall and best model fit. We only present the outcome of regressions accounting for the PISA index of perseverance for Denmark and Switzerland and self-reported effort for Australia, but results were

similar when we control for any self-reported measure of PSEC and they are available from authors upon request.

Table 5. The incremental predictive power of behavioural PSEC indicators

		Denmark (N=1192)		**Switzerland** (N=1948)		**Australia** (N=3196)	
Indicator	Outcome	b	Adjusted R squared	b	Adjusted R squared	b	Adjusted R squared
Performance decline	Upper secondary	-0.043*	.0823	-0.046	.0627	-0.0074	.0545
	University	-0.059	.156	-0.071***	.133	-0.072*	.206
Inconsistency	Upper secondary	-0.061	.096	-0.025	.0553	-0.014	.0563
	University	-0.0088	.149	-0.025	.118	0.011	.187
Non-response	Upper secondary	-0.050	.0821	0.027	.0527	-0.026	.0587
	University	-0.042	.151	-0.063**	.122	-0.015	.187
Non-differentiation	Upper secondary	0.0047	.0747	0.0096	.051	-0.012	.056
	University	0.014	.149	-0.012	.116	-0.0013	.187

Notes: Denmark: PISA 2000-PIAAC 2012 link. Switzerland: PISA 2000-TREE 2010 follow-up. Australia: PISA 2003-LSAY 2013. Canada: PISA 2000-YITS 2010.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

4. Discussion

In recent years, low-stakes large-scale assessments conducted at the national, regional and international level have come to play a prominent role in policy making (Egelund, 2008; Ertl, 2006; Grek, 2009; Takayama, 2008). These assessments have been used to benchmark progress, set standards and foster policy learning across different education systems (Breakspear, 2012). However, just as assessments have gained in visibility and use, they have also started to attract an increasing level of criticism. A first set of critics question the narrow focus of assessments on information processing abilities, maintaining that, by equating education quality with performance measures in specific academic disciplines, assessments lead to poorer educational experiences for young children since different stakeholders divert resources to maximising only what is measured (Meyer et al. 2014). A second set of critics question the validity and reliability of comparisons of achievement measures obtained through low-stakes assessments across groups that may differ in level of test engagement and motivation (Gneezy et al., 2019; Brunello, Crema and Rocco, 2018). These critics essentially maintain that observed differences in achievement could be attributed not to underlying

differences in ability and the quality of learning children undergo but, rather, to the level of motivation they exerted during the test.

In response to these critics some assessment programmes have begun to develop indicators that go beyond achievement measures. In particular, because of mounting evidence on the importance of socio-emotional and motivational skills such as conscientiousness, perseverance and task persistence for individuals' long term outcomes and the malleability of such skills in response to educational interventions, some assessments have introduced self-reported indicators aimed at capturing these skills. However, because the validity of self-reports has been questioned by some, over the last decade researchers have developed indicators of perseverance, self-regulation, endurance and conscientiousness based on students' behaviour during standardised tests and questionnaires and used them to compare countries, groups of students or to infer relationships between cognitive and socio-emotional and motivational skills (Zamarro et al., 2019; Soland et al. 2019).

Little attention has been paid to evaluating the measurement properties of these indicators, and to investigating whether these measures of how students behave on a test can be interpreted as general measures of socio-emotional and motivational skills

Our results based on PISA data confirm previous findings on the existence of considerable differences in measures of socio-emotional and motivational measures both across and within countries, and on the robustness of the relation between these measures and achievement in low-stake, standardised tests. A novel finding based on longitudinal data from four countries is that some measures of students' behaviour on the test, such as performance decline, are also strongly associated with life outcomes ten years after the test.

Our work suggests that no measure is without limitations. For example, some behavioural measures that have strong predictive validity of future outcomes (such as performance decline) have relatively low contemporaneous correlations with achievement and are less stable than others across cycles. Furthermore, although the correlation between self-reports and behavioural indicators is low at both individual and country level, so is the correlation between behavioural indicators, suggesting

that different indicators measure different aspects of socio-emotional and motivational skills and/or suffer from different sources of bias or measurement error. Measurement issues pertain to comparisons across countries with different languages and cultural traditions as well as across groups of students within an individual country. For example, differences across gender are more marked across countries for some measures than for others.

The evidence from our studies suggests that the measurement quality of behavioural indicators needs to be improved before they can be reliably used to compare students or countries. Building indicators of complex constructs from assessment tasks that are meant to measure different constructs increases the risk of falling into jingle-jangle interpretation fallacies. One way to achieve more robust measures of socio-emotional and motivational skills within large-scale assessment programmes consists in designing tasks with the explicit objective of measuring these skills, in addition to the primary, cognitive construct that is targeted by the assessment. The process of design of interactive test units or tasks can incorporate experts' consideration on which behaviours on the tasks are indicative of specific socio-emotional and motivational skills. Such interpretative model increases the value of the process data, as traces or log-data become an integral part of the measurement model underlying the assessment units. Even when adopting such improved design process, indicators based on process data need be validated, because trace data can often be interpreted in different ways. For example, trace data signalling an extended period of inactivity on a task might indicate disengagement for students with higher cognitive skills, but might reflect confusion for those students with lower cognitive skills. Validation can occur through small-scale cognitive laboratories where the trace data are complemented either by observational data collected by experts following standardised protocols or by asking students to verbalise their thinking and feelings. Validation studies in which students' are assessed across different disciplines and with different typologies of problems are also needed to confirm the generalisability of findings about students' conscientiousness, persistence or other socio-emotional traits.

The most pressing problem with the measures of students' behaviour on the test considered in this study and the related literature is how they are being interpreted and used. Behavioural

indicators have been interpreted as indicators of socio-emotional and motivational skills without acknowledging that there are still considerable gaps in our understanding of the inferences that can be made from traces of students' behaviour on a test. These gaps lie both at the theoretical level – we lack evidence demonstrating that a certain behaviour constitutes a generalisable indicator of how students behave in other contexts – and at the measurement level - we still need to better understand the magnitude and sources of measurement bias associated with each indicator. The fact that empirically different measures designed to measure the same construct yield markedly different results points to the need for additional work to validate these measures. Until then, we suggest that these measures should be considered in purely operational terms: i.e. performance decline should be considered to indicate the decline in accuracy during a cognitive test rather than as an indicator of test takers' underlying conscientiousness or perseverance.

References

Akyol, S. P., Krishna, K., & Wang, J. (2018). Taking PISA seriously: How accurate are low stakes exams? NBER Working Paper No. 24930. Cambridge: NBER.

Azzolini, D., Bazoli, N., Lievore, I., Schizzerotto, A., & Vergolini, L. (2019). *Beyond achievement: A comparative look into 15- year-olds' school engagement, effort and perseverance in the European Union.* Brussels: Publication Office of the European Union.

Balart, P., & Oosterveen, M. (2019). Females show more sustained performance during test-taking than males. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11691-y

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200. https://doi.org/10.1007/s11162-011-9251-2

Barry, C. L., Horst, S. J., Finney, S. J., Brown, A. R., & Kopp, J. P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing*, *10*(4), 342–363. https://doi.org/10.1080/15305058.2010.508569

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Borghans, L., & Schils, T. (2013). *The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components.* Unpublished manuscript. Draft version: July 22, 2013.

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.

Breakspear, S. (2012). The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers n. 71.* Paris: OECD Publishing. https://doi.org/10.1787/19939019

Brunello, G., Crema, A., & Rocco, L. (2018). Testing at length if it is cognitive or non-cognitive. *IZA Discussion Papers, N. 11603.*Bonn: Institute of Labor Economics (IZA).

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*(6448), 70–73. https://doi.org/10.1126/science.aau8712

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, *33*(4), 609–624. https://doi.org/10.1016/j.cedpsych.2007.10.002

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, *10*(JULY).

https://doi.org/10.3389/fpsyg.2019.01522

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? Clinical Psychology Review 31(6), 965-982.

Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement*, *50*(2), 164–185. https://doi.org/10.1111/jedm.12009

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45,* 259-268.

Egelund, N. (2008). The value of international comparative studies of achievement–a Danish perspective. *Assessment in Education: Principles, Policy and Practice*, *15*(3), 245–251.

Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education, 32*, 619–634. https://doi.org/10.1080/03054980600976320

Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412).

Fiske, D. W. (1971). *Measuring the concepts of personality.* Chicago, Illinois: Aldline Publishing Co.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & 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.

Goldhammer, F., Martens, T., & Lüdtke, O. (2017). Conditioning factors of test-taking engagement in PIAAC: an exploratory IRT modelling approach considering person and item characteristics. *Large Scale Assessments in Education, 18*(5).

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review, 1*(3), 291-308.

Grek, S. (2009). Governing by numbers: The PISA "effect" in Europe. *Journal of Education Policy*, *24*(1), 23–37. https://doi.org/10.1080/02680930802412669

Gutman, L. M., & Schoon, I. (2016). A synthesis of causal evidence linking non-cognitive skills to later outcomes for children and adolescents. In Khine, M. S., & Areepattamannil, S. (Eds.) *Non-cognitive Skills and Factors in Educational Attainment* (pp. 171–198). https://doi.org/10.1007/978-94-6300-591-3_9

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

Heckman, J. J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*,

103(6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052

Hitt, C. (2015). *Just filling in the bubbles: Using careless answer patterns on surveys as a proxy measure of noncognitive skills* EDRE Working Paper 2015-06.

Hitt, C., Trivitt, J., & Cheng, A. (2016). When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*, *52*, 105–119.

Kankaraš, M. (2017). Personality matters: Relevance and assessment of personality characteristics. *OECD Education Working Papers 157.* Paris: OECD Publishing.

Kautz, T., Heckman, J. J., Diris, R., & Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *OECD Education Working Papers 110.* Paris: OECD Publishing.

Kelley, T. L. (1927). *Interpretation of educational measurements.* Oxford: World Book Co.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*(2), 312–320. https://doi.org/10.1037/0022-3514.55.2.312

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, *45*(2), 527–563. https://doi.org/10.1007/s41237-018-0063-y

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

Kyllonen, P., & Kell, H. (2018). Ability tests measure personality, personality tests measure ability: Disentangling construct and method in evaluating the relationship between personality and ability. *Journal of Intelligence*, *6*(3), 32. https://doi.org/10.3390/jintelligence6030032

Levin, H. M. (2013). The utility and need for incorporating noncognitive skills into large-scale educational assessments. In von Davier, M., Gonzales, E., Kirsch, I., Yamamoto, K. (Eds.) *The role of international large-scale assessments: perspectives from technology, economy, and educational research* (pp. 67–86). Springer. https://doi.org/10.1007/978-94-007-4629-9_5

Nagy, G., Nagengast, B., Becker, M., Rose, N., & Frey, A. Item position effects in a reading comprehension test: An IRT study of individual differences and individual correlates. *Psychological Test and Assessment Modeling, 60*(2), 165-187.

Organisation for Economic Co-operation and Development OECD. (2009). *PISA Data Analysis Manual: SAS, Second Edition* (2nd ed.). Paris: OECD Publishing.

OECD. (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660–679.

Revelle, W. (2007). Experimental approaches to the study of personality. In R. W. Robins, R. C.

Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 37–61). The Guilford Press.

Rindermann, H. (2007). Theg-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, *21*(5), 667–706. https://doi.org/10.1002/per.634

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Salganik, M. (2019). *Bit by Bit: Social research in the digital age*. Princeton: Princeton University Press.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin, 140*(2), 374–408.

Soland, J., Zamarro, G., Cheng, A, & Hitt, C. (2019). Identifying naturally occurring direct assessments of social-emotional competencies: The promise and limitations of survey and assessment disengagement metadata. *Educational Researcher, 48*(7), 466-478.

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science, 30*(5), 711-727.

Soto, C. J. (2020). Do links between personality and life outcomes generalize? Testing the robustness of trait–outcome associations across gender, age, ethnicity, and analytic approaches. *Social Psychological and Personality Science*.

Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, *44*(4), 387–407.

Vannette, D. L., & Krosnick, J. A. (2014). Answering questions: A Comparison of Survey Satisficing and Mindlessness. In Ie, A., Ngnoumen, C. T., & Langer, E. J. (Eds.) *The Wiley Blackwell Handbook of Mindfulness* (pp. 312–327). John Wiley & Sons.

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *PNAS, 116*(46), 23004-23010.

Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, *15*(1), 27–41. https://doi.org/10.1080/10627191003673216

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 163–183. https://doi.org/10.1207/s15324818ame1802_2

Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects*, *39*(1), 33–46. https://doi.org/10.1007/s11125-009-9109-y

Zamarro, G., Cheng, A., Shakeel, M. D., & Hitt, C. (2018). Comparing and validating measures of non-cognitive traits: Performance task measures and self-reports from a nationally representative internet panel. *Journal of Behavioral and Experimental Economics*, *72*, 51–60. https://doi.org/10.1016/j.socec.2017.11.005

Zamarro, G., Hitt, C., & Mendez, I. (2019). When students don't care: Reexamining international differences in achievement and student effort. *Journal of Human Capital, 13*(4), 519-552. https://doi.org/10.1086/705799

Annex A The PISA assessment instruments

PISA is based on a dynamic and forward-looking model of lifelong learning, exploring the

knowledge and skills students need to adapt successfully to a changing world and apply their knowledge and experience to real issues. International experts defined each of the competency domains that are examined in PISA science, reading, mathematics, problem solving and drafted the frameworks for assessing them. Competency is not something that an individual either does or does not have, but is measured on a continuum. There is no precise dividing line between a person who is fully competent and one who is not. However, it is necessary for measurement purposes to define at which level of competencies students are able to participate effectively in society (in PISA, international experts set the baseline at Level 2 on the PISA proficiency scales) (OECD, 2013).

Reading: In PISA reading literacy in defined as "understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society" (OECD, 2013, p. 61).

Mathematics: In PISA, mathematical literacy is defined as "an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens" (OECD, 2013, p. 25).

Science: PISA defines scientific literacy as "the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2017, p. 20) A scientifically competent person is willing to engage in reasoned discourse about science and technology which requires the competencies to: explain phenomena scientifically – recognising, offering and evaluating explanations for a range of natural and technological phenomena; evaluate and design scientific enquiry – describing and appraising scientific investigations, and proposing ways to address questions scientifically; and interpret data and evidence scientifically – analysing and evaluating data, claims and arguments in a variety of representations, and drawing appropriate scientific conclusions.

Problem solving: in PISA problem solving is defined as "an individual's capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential as a constructive and reflective citizen" (OECD, 2013, p. 122).

Examples of the PISA assessment tasks can be found at: http://www.oecd.org/pisa/test/

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

OECD (2017), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264281820-en.

Annex B A description of the longitudinal follow-ups of PISA

The datasets from Australia, Canada, Denmark and Switzerland followed the transition of early PISA cohorts into adulthood. While the frequency and timeline of follow-up surveys vary by country, all the datasets collect information on young people when they are approximately 25 years old (the only exception is Denmark, where the analysis refers to individuals who are either 26 or 27 years old).

The follow-up data for Denmark is derived from PISA and the Survey of Adult Skills (a product of the OECD Programme for the International Assessment of Adult Competencies T[PIAAC]). Participants in the PISA 2000 cycle were tested and interviewed again in the 2012 Survey of Adult Skills (PIAAC).

Transitions from Education to Employment (TREE1) surveys the post-secondary education and labour market pathways of students in Switzerland, and is the country's first longitudinal study of this type at the national level. TREE1, the project's first cohort, is based on a sample of students who participated in PISA 2000. The sample was tracked for follow-up surveys annually from 2001 to 2007, and twice more (in 2010 and 2014).

The Longitudinal Study of Australian Youth (LSAY) tracks students annually for a period of 10 years as they move from school into further study, work and other destinations. While the first surveys began in 1995, participants have been recruited from Australian schools that have taken part in PISA since 2003 (Y03).

Canada designed the Youth in Transition Survey (YITS) to examine the patterns of, and influences on, major transitions in young people's lives, particularly with respect to education, training and work. PISA/YITS is a special project that aligned the two survey programmes. The PISA 2000 cohort was selected to participate in YITS, and was surveyed in follow-up interviews every two years through 2010. These data have not been released for public access, so all estimates were obtained through collaborations with national researchers at Statistics Canada, who conducted the analyses on the basis of statistical programmes prepared by the authors.

Table B1

Description of the longitudinal studies examined in the study

Data source (surveys)	Australia	Canada	Denmark	Switzerland
Study name	Longitudinal Surveys of Australian Youth (LSAY)	Youth in Transition Survey (YITS)	PISA - PIAAC	Transitions from Education to Employment (TREE1)
Baseline data collection	PISA 2003	PISA 2000	PISA 2000	PISA 2000
Age of students at last follow-up survey	25	25	26-27	25[1]
Baseline sample	10370	25190	4235	6343
Sample at last follow-up	3741	9183	1881	3423

Annex C Testing the dependence of questionnaire-based indicators on questionnaires

All our questionnaire-based behavioural indicators are standardised and can be compared across students. However, as a robustness check, we test the dependence of these indicators on questionnaire content and structure. We make use of the fact that PISA 2012 adopted a rotated student questionnaire design. All students were administered a short common questionnaire containing questions on socio-economic and demographic information. The core questionnaire was complemented by a rotated part. Three sets of questions were developed and each was randomly administered to two-thirds of the overall student sample, so each student was administered the common questionnaire part and two out of three rotated parts. Table A1 illustrates the PISA 2012 rotation design.

Table C1

Standard Booklet Design in PISA 2012

Form A	**Form B**	**Form C**
	Common part	
Question set 1	Question set 3	Question set 2
Question set 2	Question set 1	Question set 3

The contents of rotating question sets differed along various dimensions, which could affect students' effort depletion. Moreover, not all items in each question set could be used to compute our PSEC measures. Items from Question set 2 included some questions that were not multiple-choice and were dropped in our analyses. It also did not include any Likert-type questions that could be used for the calculation of the indices of non-differentiation and non-consistency. Question sets 1 and 3 had the same number of valid Likert-type item sets. This implies that students who took form B had more data for the calculation of our questionnaire-based measures. Students who took forms A and C also took more open-ended questions in set 2 (although in different positions within their form) which could affect effort expended. We ran regressions to measure differences in questionnaire-based behavioral measures of PSEC across the three forms.

Estimates presented in Table A2 indicate that non-differentiation, non-response and inconsistency vary across questionnaire forms. Non-response is highest in form C, followed by form A and then form B. However, differences are not substantively large. Non-differentiation is smallest in form A and largest in form C. The difference between questionnaire forms is large: it is around 40% of a standard deviation larger in form B than in form A, while form C is 25% of a standard deviation larger than form B. Inconsistency is smallest in form C and largest in form A. Differences across questionnaire forms are also large: compared to form B, the inconsistency indicator is 40% of a standard deviation larger in form A than in form B while form C is 25% of a standard deviation smaller than form B.

Table C2

Questionnaire-based Behavioral Measures by Questionnaire Form

	Non-response				Non-differentiation				Inconsistency			
	b	se	min95	max95	b	se	min95	max95	b	se	min95	max95
Questionnaire form A	0.041***	(0.0087)	0.024	0.058	-0.41***	(0.0092)	-0.43	-0.39	0.39***	(0.0081)	0.37	0.41
Questionnaire form C	0.083***	(0.010)	0.063	0.10	0.25***	(0.0078)	0.23	0.26	-0.25***	(0.0084)	-0.27	-0.24
Constant	-0.041***	(0.0095)	-0.060	-0.022	0.055***	(0.0084)	0.038	0.071	-0.046***	(0.0063)	-0.058	-0.033
Observations	404764				402701				402551			
Adjusted R-squared	0.0011				0.072				0.070			

* p<0.05, ** p<0.01, *** p<0.001

Analyses presented in Table C2 suggest that differences in the indices of non-differentiation and inconsistency across forms are particularly large. Given this pattern, it is likely that the extent of non-differentiation is related to the items contained in each booklet. Intuitively, inconsistency is to a certain extent inversely related to non-differentiation, so it also varies with booklet items. Specifically, some questions contained in set 3, which were administered in questionnaire forms B and C, appear to be especially prone to non-differentiation, compared to questions contained in set 1, which was included in forms A and B.

Table C3 displays the non-differentiation measure by item set and confirms these results. Non-differentiation was about 8 percentage points more prevalent for questions in set 3 compared to those in set 1. Over 30% of students assigned to forms B and C gave non-differentiated responses to items in questions 83[vi], 86[vii] and 89[viii]. By contrast non-differentiation was never above 15% in question set 1. Item specific analyses suggest that differences may be driven by the specific item pool presented in each booklet and that the non-differentiation measure is particularly sensitive to the framing and content of questionnaire items.

Table C3

Non-differentiation by Questionnaire Item Set

Question(s)	Percentage of non-differentiation	Question(s)	Percentage of non-differentiation
Set 1: Questions contained in questionnaire forms A and B	10.60	Set 3: Questions contained in questionnaire forms B and C	18.30
st29	13.80	st42	2.25
st35	9.20	st83	37.50
st37	13.20	st85	13.20
st43	5.53	st86	31.30
st44	7.04	st87	1.74
st46	14.80	st88	4.46
st93	5.78	st89	51.90
st94	14.70	st91	4.48

Annex D The association between paper-based decline in performance and assessment contents

The core cognitive test in PISA 2012 adopted an incomplete matrix design. The test was designed to last two hours in total for each participant and was organised around a series of clusters of test questions. Each cluster was designed to take about 30 minutes to complete. Each student was randomly allocated a booklet containing four clusters of test questions and a total of 13 booklets were administered in 2012. Each booklet contained different clusters, which were rotated across the booklets so that each cluster was administered at least once with any other cluster and each cluster appeared at least once in one of the four potential position within the booklet. Table B1 below shows the design of standard booklets in PISA 2012.

Table D1

Standard Booklet design in PISA 2012

Booklet ID	**Cluster**			
1	PM5	PS3	PM6A	PS2
2	PS3	PR3	PM7A	PR2
3	PR3	PM6A	PS1	PM3
4	PM6A	PM7A	PR1	PM4
5	PM7A	PS1	PM1	PM5
6	PM1	PM2	PR2	PM6A
7	PM2	PS2	PM3	PM7A
8	PS2	PR2	PM4	PS1
9	PR2	PM3	PM5	PR1
10	PM3	PM4	PS3	PM1
11	PM4	PM5	PR3	PM2
12	PS1	PR1	PM2	PS3
13	PR1	PM1	PS2	PR3

Performance decline could be associated with the contents and the order of material in the assessment, since some subject domains could require more effort depletion than others. The random allocation of booklets in the PISA assessment gives a unique possibility to test for variability in the index of performance decline across different test forms. Although our index is measured net of question and booklet fixed effects, ideally, in order to draw meaningful comparisons across students, the index should not be very sensitive to the specific characteristics of the assessment.

Table D2 reports differences in the performance decline measure between different sets of cognitive booklets in the PISA 2012 assessment that vary depending on their mathematics and reading content. Results show that the measure is not significantly different depending on the amount of mathematics content they received or depending on whether the booklet started with reading questions. Having reading questions at the very beginning or towards the end of the assessment was associated with the performance decline measure, but the effect is quantitatively very small. Overall, these analyses show that the performance decline measure is not related to the structure of the assessment booklet.

Table D2

Index of Performance Decline by Test Contents and Order

	Performance decline			
	b	se	min95	max95
Booklet with 1hr of math	0.0088	(0.0068)	-0.0046	0.022
Booklet with a reading cluster	-0.025***	(0.0068)	-0.038	-0.012
Booklet starting with a reading cluster	-0.017	(0.0088)	-0.034	0.00045
Booklet with reading as third cluster	0.038***	(0.0090)	0.021	0.056
Observations	407110			

* p<0.05, ** p<0.01, *** p<0.001

Annex E The association between computer- and paper-based assessment indicators

The unique features of PISA 2012 allow us to test the extent to which test-based indicators of PSEC differed depending on the mode of administration. In 32 participating countries students sat both the core paper-based assessment and an extra assessment of computer-based literacies. The computer-based test differed from the paper-based test because it included mathematics, reading and problem-solving rather than mathematics, reading and science content, and because it was composed of only two forms of 20 minutes each (rather than 4 booklets of 30 minutes each in the paper-based with a 15 minutes break after the first hour of testing time). Moreover, while it was possible for test takers to move freely within the paper-based-booklet, test takers could go back only within an individual unit in the computer-based test.

For students who sat both forms of assessment, we regressed performance decline in the computer-based assessment on performance decline in the paper-based one. We compared their absolute scores in the indicators as well as their ranking within their national distribution. We also tested for differences between students who sat the computer-based assessment on the same day as their paper-based assessment or on a different day.

Estimates from the first model in table E1 below show that students with a large performance decline in the paper-based assessment tend to also have a large performance decline in the computer-based assessment. However, the correlation between the two is low: a one standard deviation increase in paper-based decline is associated with an increase of 0.04 of a standard deviation in computer-based decline. The second model accounts for whether students sat the computer-based assessment on the same day as the paper-based test, as well as an interaction term to examine differences in the relationship between performance decline in the paper-based test and computer-based test as a function of when the computer-based test was held. Results show that students who sat the two assessments on the same day had a higher performance decline in the computer-based assessment. Results from the third model show that students' ranking in the performance decline measure based on the computer-based test within their country is positively related to their ranking in paper-based performance decline.

Table E1

The Association Between Computer- and Paper-based Assessment Indicators

	Model 1: Computer-based decline				Model 2: Computer-based decline				Model 3: Ranking in computer-based decline			
	b	se	min95	max95	b	se	min95	max95	b	se	min95	max95
Paper-based decline	0.039***	(0.0069)	0.026	0.053	0.073*	(0.034)	0.0074	0.14				
Tests taken in the same day					0.030***	(0.0023)	0.025	0.035				
Paper decline * test same day					-0.023	(0.035)	-0.093	0.046				
Ranking in paper-based decline									0.045***	(0.0069)	0.032	0.059
Observations	106834				106834				106834			
Adjusted R-squared	0.0014				0.0084				0.0020			

* p<0.05, ** p<0.01,
***p<0.001

Annex F Country-specific means and standard deviations in PSEC indicators

Table F1 – Country-specific means in PSEC indicators

	Non-response		Non-differentiation		Inconsistency		Performance decline		Response time effort (5 seconds)		Self-reported lack of perseverance		Self-reported lack of effort in test		Difference in effort (PISA-marked)	
	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.
Australia	-0.12	(0.01)	0.06	(0.01)	-0.31	(0.01)	0.06	(0.01)	-0.02	(0.01)	-0.01	(0.01)	0.05	(0.01)	0.10	(0.01)
Austria	-0.14	(0.02)	-0.38	(0.01)	0.39	(0.02)	-0.13	(0.01)	0.05	(0.04)	0.12	(0.02)	0.14	(0.02)	0.08	(0.02)
Belgium	-0.03	(0.02)	-0.14	(0.01)	0.11	(0.01)	0.01	(0.01)	0.00	(0.03)	0.44	(0.02)	0.22	(0.01)	0.14	(0.01)
Canada	-0.15	(0.01)	0.00	(0.01)	-0.12	(0.01)	0.05	(0.01)	-0.07	(0.02)	-0.13	(0.01)	0.49	(0.01)	0.47	(0.01)
Chile	-0.20	(0.01)	-0.15	(0.02)	0.14	(0.02)	-0.08	(0.02)	-0.03	(0.04)	-0.19	(0.03)	-0.28	(0.02)	-0.14	(0.03)
Czech Republic	-0.18	(0.01)	-0.22	(0.02)	-0.03	(0.02)	-0.11	(0.01)	m	m	0.20	(0.02)	0.09	(0.02)	0.03	(0.02)
Denmark	-0.14	(0.01)	-0.10	(0.01)	-0.22	(0.01)	-0.05	(0.01)	-0.03	(0.03)	0.18	(0.02)	0.01	(0.02)	0.21	(0.02)
Estonia	-0.16	(0.01)	-0.19	(0.01)	-0.18	(0.01)	-0.17	(0.01)	-0.06	(0.02)	-0.22	(0.02)	0.18	(0.02)	0.26	(0.02)
Finland	-0.12	(0.01)	-0.12	(0.01)	-0.21	(0.01)	-0.12	(0.01)	m	m	0.09	(0.02)	-0.19	(0.02)	-0.08	(0.01)
France	0.11	(0.03)	-0.23	(0.02)	0.27	(0.02)	0.29	(0.02)	-0.05	(0.03)	0.55	(0.02)	0.51	(0.02)	0.34	(0.02)
Germany	0.10	(0.02)	-0.34	(0.01)	0.29	(0.02)	-0.05	(0.01)	0.06	(0.04)	0.10	(0.02)	0.32	(0.02)	0.41	(0.02)
Greece	-0.08	(0.02)	-0.15	(0.02)	0.35	(0.02)	0.73	(0.02)	m	m	0.19	(0.02)	0.16	(0.02)	0.12	(0.02)
Hungary	-0.23	(0.02)	-0.08	(0.02)	-0.02	(0.02)	-0.01	(0.01)	0.00	(0.04)	0.11	(0.02)	0.01	(0.02)	0.05	(0.02)
Iceland	0.00	(0.02)	0.30	(0.02)	-0.30	(0.02)	0.37	(0.02)	m	m	0.20	(0.02)	0.05	(0.02)	0.23	(0.02)
Ireland	-0.18	(0.01)	-0.13	(0.01)	-0.13	(0.01)	-0.20	(0.01)	-0.08	(0.03)	-0.05	(0.02)	-0.14	(0.02)	-0.01	(0.02)
Israel	0.61	(0.06)	-0.09	(0.02)	0.23	(0.02)	0.48	(0.03)	0.50	(0.13)	-0.27	(0.02)	0.02	(0.02)	0.07	(0.02)
Italy	-0.15	(0.01)	-0.25	(0.01)	0.18	(0.01)	0.23	(0.01)	-0.08	(0.02)	0.05	(0.01)	-0.13	(0.01)	-0.05	(0.01)
Japan	0.06	(0.01)	-0.06	(0.02)	0.07	(0.01)	0.33	(0.02)	-0.10	(0.02)	0.70	(0.02)	0.72	(0.03)	0.15	(0.02)
Korea	-0.16	(0.01)	0.08	(0.03)	-0.17	(0.01)	0.12	(0.03)	-0.09	(0.02)	0.19	(0.02)	0.10	(0.02)	0.04	(0.02)
Latvia	-0.10	(0.01)	-0.13	(0.02)	-0.16	(0.02)	-0.03	(0.01)	m	m	-0.06	(0.02)	0.00	(0.02)	-0.19	(0.02)
Luxembourg	0.07	(0.01)	-0.11	(0.02)	0.38	(0.01)	0.23	(0.01)	m	m	0.16	(0.02)	0.28	(0.01)	0.26	(0.02)
Mexico	0.04	(0.01)	-0.03	(0.01)	0.05	(0.01)	0.17	(0.01)	m	m	-0.23	(0.01)	-0.56	(0.01)	-0.38	(0.01)
Netherlands	-0.13	(0.02)	-0.02	(0.02)	-0.22	(0.02)	0.09	(0.01)	m	m	0.23	(0.02)	0.23	(0.02)	0.09	(0.02)
New Zealand	-0.11	(0.03)	0.09	(0.02)	-0.25	(0.02)	0.09	(0.02)	m	m	0.10	(0.02)	0.11	(0.02)	0.15	(0.02)
Norway	0.96	(0.02)	-0.02	(0.01)	-0.11	(0.01)	0.27	(0.02)	0.17	(0.08)	0.44	(0.02)	0.25	(0.03)	0.34	(0.03)
Poland	-0.29	(0.01)	-0.10	(0.02)	-0.03	(0.02)	0.05	(0.02)	0.05	(0.04)	0.06	(0.02)	0.08	(0.02)	0.04	(0.02)
Portugal	-0.06	(0.02)	0.20	(0.03)	-0.32	(0.01)	0.24	(0.02)	-0.10	(0.01)	-0.27	(0.03)	-0.01	(0.02)	0.08	(0.02)
Slovak Republic	-0.09	(0.03)	-0.06	(0.02)	0.15	(0.02)	0.06	(0.01)	-0.09	(0.02)	0.59	(0.02)	0.26	(0.02)	0.18	(0.02)
Slovenia	-0.14	(0.01)	-0.07	(0.01)	0.00	(0.01)	0.21	(0.01)	0.19	(0.04)	0.00	(0.02)	0.16	(0.01)	0.27	(0.02)
Spain	-0.05	(0.01)	-0.17	(0.01)	0.15	(0.01)	0.13	(0.01)	0.04	(0.03)	-0.01	(0.01)	0.15	(0.01)	0.18	(0.01)
Sweden	0.13	(0.02)	0.06	(0.02)	-0.12	(0.02)	0.37	(0.02)	0.02	(0.04)	0.35	(0.02)	0.32	(0.02)	0.50	(0.03)
Switzerland	-0.15	(0.02)	-0.30	(0.01)	0.21	(0.01)	-0.04	(0.01)	m	m	0.24	(0.02)	0.16	(0.02)	0.19	(0.02)
Turkey	0.05	(0.02)	0.03	(0.02)	0.29	(0.02)	-0.11	(0.01)	m	m	-0.36	(0.02)	-0.61	(0.02)	-0.47	(0.01)
United Kingdom	-0.09	(0.02)	-0.02	(0.01)	-0.22	(0.02)	0.11	(0.01)	m	m	-0.01	(0.02)	0.06	(0.02)	0.25	(0.02)
United States	-0.05	(0.02)	0.16	(0.02)	-0.25	(0.02)	0.01	(0.01)	-0.07	(0.02)	-0.29	(0.02)	-0.43	(0.02)	-0.23	(0.02)
Argentina	0.52	(0.05)	0.04	(0.03)	0.47	(0.02)	-0.41	(0.02)	m	m	0.10	(0.03)	-0.28	(0.03)	-0.42	(0.03)
Albania	0.61	(0.06)	0.14	(0.02)	0.07	(0.02)	0.36	(0.02)	m	m	-0.56	(0.02)	m	m	m	m
Brazil	0.30	(0.03)	0.04	(0.02)	0.22	(0.01)	-0.72	(0.02)	-0.10	(0.02)	-0.05	(0.02)	-0.34	(0.02)	-0.26	(0.02)
Bulgaria	0.28	(0.05)	0.36	(0.04)	0.21	(0.03)	0.23	(0.02)	m	m	-0.50	(0.03)	-0.18	(0.03)	-0.18	(0.03)
Colombia	0.94	(0.08)	-0.03	(0.03)	0.09	(0.03)	-1.54	(0.02)	-0.07	(0.02)	-0.32	(0.04)	-0.41	(0.02)	-0.30	(0.02)
Costa Rica	0.08	(0.03)	0.01	(0.02)	0.11	(0.03)	-1.79	(0.03)	m	m	-0.38	(0.04)	-0.34	(0.03)	-0.11	(0.03)
Croatia	-0.20	(0.01)	-0.06	(0.02)	-0.05	(0.02)	-0.10	(0.01)	m	m	0.00	(0.02)	0.10	(0.02)	0.06	(0.02)
Hong Kong (China)	-0.38	(0.01)	0.21	(0.02)	-0.40	(0.02)	-0.15	(0.02)	0.08	(0.05)	-0.03	(0.02)	0.15	(0.02)	0.14	(0.02)
Indonesia	0.16	(0.04)	0.39	(0.03)	-0.28	(0.02)	0.02	(0.01)	m	m	-0.17	(0.03)	-0.38	(0.04)	-0.68	(0.04)
Jordan	0.34	(0.05)	0.31	(0.03)	0.61	(0.03)	0.44	(0.02)	m	m	-0.24	(0.03)	-0.16	(0.03)	-0.10	(0.03)
Kazakhstan	-0.13	(0.03)	0.32	(0.03)	-0.37	(0.03)	-0.60	(0.02)	m	m	-0.71	(0.03)	-0.72	(0.02)	-0.66	(0.02)
Liechtenstein	-0.17	(0.03)	-0.30	(0.04)	0.19	(0.06)	0.28	(0.04)	m	m	0.20	(0.06)	0.09	(0.05)	0.11	(0.05)
Lithuania	-0.03	(0.02)	-0.15	(0.02)	0.32	(0.02)	0.00	(0.01)	m	m	-0.05	(0.02)	-0.24	(0.02)	-0.13	(0.02)
Macao (China)	-0.34	(0.01)	0.03	(0.01)	-0.21	(0.01)	-0.08	(0.01)	-0.07	(0.02)	-0.06	(0.01)	-0.16	(0.01)	-0.33	(0.01)
Malaysia	0.09	(0.03)	0.15	(0.02)	-0.03	(0.02)	-0.02	(0.01)	m	m	-0.13	(0.02)	-0.08	(0.02)	-0.66	(0.02)
Montenegro	0.21	(0.02)	0.16	(0.02)	0.20	(0.02)	0.25	(0.01)	m	m	-0.26	(0.02)	0.08	(0.02)	0.09	(0.02)
Peru	0.57	(0.06)	-0.02	(0.02)	0.02	(0.02)	-1.11	(0.02)	m	m	-0.28	(0.02)	-0.32	(0.02)	-0.35	(0.02)
Qatar	0.60	(0.02)	0.49	(0.02)	0.48	(0.01)	0.52	(0.01)	m	m	-0.17	(0.01)	0.04	(0.01)	-0.15	(0.02)
Romania	-0.12	(0.04)	0.03	(0.03)	0.50	(0.04)	-1.56	(0.02)	m	m	0.05	(0.03)	-0.27	(0.02)	-0.25	(0.02)
Russia	-0.05	(0.02)	-0.02	(0.02)	-0.09	(0.02)	0.45	(0.03)	-0.04	(0.03)	-0.41	(0.02)	-0.21	(0.02)	-0.32	(0.02)
Serbia	0.11	(0.03)	0.00	(0.03)	0.20	(0.03)	-1.47	(0.03)	m	m	-0.07	(0.03)	0.16	(0.04)	0.21	(0.03)
Shanghai	-0.37	(0.00)	0.18	(0.02)	-0.41	(0.01)	-0.11	(0.01)	-0.09	(0.02)	-0.15	(0.02)	-0.40	(0.02)	-0.26	(0.01)

Singapore	-0.33	(0.01)	0.18	(0.02)	-0.27	(0.01)	-0.05	(0.01)	-0.10	(0.01)	-0.20	(0.02)	0.05	(0.01)	0.10	(0.01)
Chinese Taipei	-0.21	(0.00)	0.23	(0.02)	-0.30	(0.01)	-0.20	(0.01)	-0.03	(0.04)	0.18	(0.02)	-0.36	(0.02)	-0.48	(0.01)
Thailand	-0.13	(0.01)	0.51	(0.03)	-0.30	(0.02)	0.23	(0.01)	m	m	-0.13	(0.01)	-0.61	(0.02)	-0.62	(0.02)
Tunisia	0.26	(0.04)	-0.06	(0.03)	0.75	(0.03)	0.01	(0.03)	m	m	-0.09	(0.04)	0.08	(0.03)	-0.06	(0.04)
United Arab Emirates	0.15	(0.03)	0.18	(0.03)	0.30	(0.02)	-0.80	(0.01)	0.14	(0.05)	-0.33	(0.02)	-0.12	(0.02)	-0.07	(0.02)
Uruguay	0.62	(0.05)	0.01	(0.02)	0.20	(0.02)	-1.01	(0.03)	m	m	-0.18	(0.03)	-0.06	(0.02)	-0.03	(0.02)
Viet Nam	-0.17	(0.02)	-0.24	(0.02)	-0.29	(0.02)	-0.68	(0.01)	m	m	-0.38	(0.03)	-0.59	(0.03)	-0.69	(0.02)
Average	0.03	(0.00)	0.01	(0.00)	0.03	(0.00)	-0.09	(0.00)	0.00	(0.01)	-0.03	(0.00)	-0.04	(0.00)	-0.04	(0.00)

Note: the index of response time effort has a missing value for countries that did not administer the computer-based assessment

Table F2 – Country-specific standard deviations in PSEC indicators

	Non-response		Non-differentiation		Inconsistency		Performance decline		Response time effort (5 seconds)		Self-reported lack of perseverance		Self-reported lack of effort in test		Difference in effort (PISA-marked)	
	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.	Value	s.e.
Australia	0.77	(0.04)	0.91	(0.01)	0.87	(0.01)	0.87	(0.01)	0.90	(0.11)	0.95	(0.01)	0.96	(0.01)	0.94	(0.01)
Austria	0.69	(0.08)	0.71	(0.03)	0.89	(0.01)	0.78	(0.01)	1.06	(0.15)	0.91	(0.01)	1.00	(0.02)	1.02	(0.02)
Belgium	0.90	(0.05)	0.88	(0.02)	1.03	(0.01)	0.79	(0.01)	1.03	(0.19)	0.97	(0.02)	0.86	(0.01)	0.91	(0.01)
Canada	0.68	(0.06)	0.91	(0.01)	0.92	(0.01)	1.02	(0.01)	0.63	(0.12)	1.01	(0.01)	0.89	(0.01)	0.99	(0.01)
Chile	0.46	(0.03)	0.82	(0.02)	0.91	(0.02)	0.86	(0.01)	0.79	(0.18)	1.00	(0.03)	0.89	(0.02)	0.94	(0.03)
Czech Republic	0.57	(0.04)	0.77	(0.02)	0.94	(0.01)	0.76	(0.01)	m	m	0.88	(0.02)	0.93	(0.02)	0.90	(0.02)
Denmark	0.72	(0.03)	0.82	(0.02)	0.87	(0.01)	0.87	(0.01)	0.99	(0.29)	0.94	(0.02)	0.91	(0.02)	0.93	(0.02)
Estonia	0.47	(0.03)	0.77	(0.03)	0.86	(0.02)	0.74	(0.01)	0.60	(0.09)	0.93	(0.01)	0.96	(0.01)	0.91	(0.02)
Finland	0.70	(0.04)	0.84	(0.02)	0.85	(0.01)	0.89	(0.01)	m	m	0.90	(0.02)	0.87	(0.02)	0.80	(0.02)
France	1.08	(0.06)	0.78	(0.03)	1.00	(0.01)	0.91	(0.01)	0.74	(0.17)	1.06	(0.02)	1.00	(0.01)	1.01	(0.02)
Germany	0.92	(0.07)	0.75	(0.03)	0.88	(0.01)	0.79	(0.01)	1.09	(0.18)	0.90	(0.02)	1.04	(0.01)	1.02	(0.02)
Greece	0.77	(0.05)	0.96	(0.03)	1.07	(0.02)	1.26	(0.01)	m	m	1.01	(0.02)	1.00	(0.01)	1.06	(0.03)
Hungary	0.69	(0.07)	0.91	(0.03)	0.99	(0.02)	0.60	(0.01)	0.90	(0.17)	0.85	(0.02)	0.94	(0.02)	0.89	(0.02)
Iceland	1.06	(0.06)	1.12	(0.02)	0.94	(0.01)	1.02	(0.01)	m	m	0.98	(0.03)	1.11	(0.02)	1.11	(0.02)
Ireland	0.53	(0.05)	0.75	(0.01)	0.90	(0.01)	0.61	(0.01)	0.60	(0.20)	1.03	(0.02)	0.83	(0.01)	0.82	(0.02)
Israel	1.39	(0.12)	0.95	(0.03)	1.03	(0.02)	1.21	(0.02)	2.56	(0.37)	1.19	(0.02)	1.09	(0.02)	1.10	(0.02)
Italy	0.70	(0.03)	0.80	(0.01)	1.02	(0.01)	1.01	(0.01)	0.53	(0.15)	1.04	(0.01)	0.85	(0.01)	0.85	(0.01)
Japan	0.35	(0.04)	0.96	(0.03)	0.97	(0.01)	1.00	(0.01)	0.49	(0.09)	0.87	(0.02)	1.24	(0.01)	1.17	(0.02)
Korea	0.35	(0.03)	1.01	(0.04)	0.88	(0.01)	0.86	(0.01)	0.64	(0.16)	0.76	(0.02)	1.19	(0.02)	1.14	(0.02)
Latvia	0.55	(0.05)	0.82	(0.02)	0.90	(0.02)	0.74	(0.01)	m	m	0.87	(0.02)	0.83	(0.02)	0.88	(0.03)
Luxembourg	1.08	(0.04)	0.97	(0.03)	1.01	(0.01)	0.94	(0.01)	m	m	0.97	(0.02)	1.02	(0.01)	1.05	(0.02)
Mexico	0.85	(0.04)	0.96	(0.01)	0.96	(0.01)	0.98	(0.01)	m	m	1.02	(0.01)	0.67	(0.01)	0.65	(0.01)
Netherlands	0.81	(0.05)	0.89	(0.02)	0.97	(0.02)	0.65	(0.01)	m	m	0.83	(0.02)	0.88	(0.02)	0.86	(0.02)
New Zealand	0.91	(0.13)	0.98	(0.03)	0.88	(0.01)	1.00	(0.01)	m	m	0.92	(0.02)	0.97	(0.02)	0.96	(0.01)
Norway	1.04	(0.05)	0.96	(0.02)	0.93	(0.02)	0.99	(0.01)	1.58	(0.38)	1.10	(0.02)	1.05	(0.02)	1.03	(0.03)
Poland	0.48	(0.05)	0.88	(0.03)	0.95	(0.02)	0.84	(0.01)	0.96	(0.18)	1.04	(0.02)	1.05	(0.02)	0.99	(0.02)
Portugal	0.93	(0.06)	1.09	(0.04)	0.85	(0.01)	0.95	(0.01)	0.49	(0.09)	1.07	(0.02)	0.94	(0.01)	0.95	(0.02)
Slovak Republic	0.96	(0.07)	0.96	(0.03)	1.07	(0.02)	0.82	(0.01)	0.48	(0.11)	1.02	(0.02)	1.06	(0.02)	1.10	(0.02)
Slovenia	0.61	(0.03)	0.95	(0.01)	0.95	(0.01)	0.90	(0.01)	1.43	(0.15)	0.96	(0.02)	1.00	(0.01)	1.05	(0.02)
Spain	0.75	(0.03)	0.82	(0.01)	0.97	(0.01)	0.81	(0.01)	1.11	(0.25)	0.99	(0.01)	0.94	(0.01)	0.95	(0.01)
Sweden	0.97	(0.06)	1.02	(0.03)	0.95	(0.02)	1.10	(0.01)	1.07	(0.18)	1.02	(0.02)	1.09	(0.02)	1.15	(0.02)
Switzerland	0.71	(0.04)	0.77	(0.02)	0.90	(0.01)	0.84	(0.01)	m	m	0.89	(0.02)	0.93	(0.02)	0.95	(0.02)
Turkey	0.73	(0.06)	1.03	(0.03)	1.00	(0.02)	0.71	(0.01)	m	m	1.09	(0.02)	0.84	(0.02)	0.88	(0.02)
United Kingdom	0.82	(0.04)	0.85	(0.01)	0.92	(0.02)	0.79	(0.01)	m	m	1.01	(0.02)	0.95	(0.02)	0.97	(0.03)
United States	1.00	(0.05)	1.03	(0.03)	0.96	(0.02)	0.63	(0.01)	0.57	(0.08)	1.07	(0.02)	0.76	(0.01)	0.76	(0.02)
Argentina	1.63	(0.09)	1.13	(0.04)	1.13	(0.03)	1.00	(0.02)	m	m	0.96	(0.03)	0.83	(0.02)	0.96	(0.04)
Albania	2.41	(0.10)	1.10	(0.02)	0.96	(0.01)	1.46	(0.02)	m	m	1.10	(0.02)	m	m	m	m
Brazil	1.60	(0.05)	1.09	(0.02)	1.05	(0.01)	1.07	(0.01)	0.34	(0.07)	0.95	(0.02)	0.90	(0.02)	0.88	(0.02)
Bulgaria	1.70	(0.11)	1.43	(0.05)	1.18	(0.03)	1.11	(0.02)	m	m	1.19	(0.03)	1.06	(0.03)	1.04	(0.03)
Colombia	1.91	(0.09)	1.03	(0.04)	1.07	(0.03)	1.05	(0.02)	0.50	(0.07)	1.01	(0.03)	0.81	(0.02)	0.83	(0.03)
Costa Rica	0.93	(0.07)	0.93	(0.03)	1.00	(0.02)	1.13	(0.02)	m	m	1.01	(0.03)	0.83	(0.03)	0.83	(0.03)
Croatia	0.55	(0.05)	0.90	(0.02)	0.97	(0.01)	0.82	(0.01)	m	m	1.04	(0.02)	0.99	(0.01)	1.06	(0.02)

Hong Kong (China)	0.39	(0.04)	1.04	(0.02)	0.91	(0.02)	0.86	(0.01)	1.20	(0.17)	0.81	(0.02)	1.04	(0.02)	1.01	(0.02)
Indonesia	1.14	(0.10)	1.29	(0.03)	0.95	(0.02)	0.85	(0.01)	m	m	0.88	(0.03)	0.83	(0.04)	0.86	(0.03)
Jordan	1.41	(0.14)	1.41	(0.04)	1.15	(0.03)	1.02	(0.01)	m	m	1.02	(0.03)	1.08	(0.03)	1.14	(0.04)
Kazakhstan	0.92	(0.08)	1.09	(0.03)	0.94	(0.02)	1.06	(0.02)	m	m	1.14	(0.02)	0.82	(0.02)	0.80	(0.03)
Liechtenstein	0.54	(0.09)	0.75	(0.08)	0.95	(0.08)	0.67	(0.03)	m	m	0.79	(0.07)	0.90	(0.05)	0.84	(0.06)
Lithuania	0.67	(0.07)	0.90	(0.02)	0.95	(0.01)	0.76	(0.01)	m	m	0.85	(0.02)	0.82	(0.02)	0.82	(0.02)
Macao (China)	0.39	(0.04)	0.93	(0.02)	0.93	(0.01)	0.70	(0.01)	0.60	(0.11)	0.81	(0.02)	0.93	(0.01)	0.89	(0.02)
Malaysia	1.03	(0.08)	1.13	(0.03)	1.00	(0.02)	0.75	(0.01)	m	m	0.85	(0.02)	1.00	(0.02)	0.90	(0.02)
Montenegro	1.13	(0.04)	1.14	(0.02)	1.08	(0.02)	0.95	(0.01)	m	m	1.14	(0.02)	1.16	(0.02)	1.19	(0.02)
Peru	1.80	(0.07)	1.02	(0.03)	1.01	(0.03)	1.15	(0.02)	m	m	0.89	(0.03)	0.84	(0.02)	0.84	(0.02)
Qatar	2.19	(0.03)	1.57	(0.02)	1.24	(0.01)	1.16	(0.01)	m	m	1.00	(0.01)	1.06	(0.01)	1.28	(0.01)
Romania	0.97	(0.19)	1.12	(0.04)	1.21	(0.02)	1.06	(0.02)	m	m	0.95	(0.03)	0.97	(0.02)	0.91	(0.03)
Russia	0.72	(0.07)	0.95	(0.02)	0.91	(0.01)	1.22	(0.02)	0.91	(0.18)	1.05	(0.02)	1.01	(0.02)	0.97	(0.02)
Serbia	1.05	(0.10)	1.05	(0.04)	1.04	(0.03)	1.13	(0.02)	m	m	1.10	(0.04)	1.08	(0.03)	1.05	(0.03)
Shanghai	0.30	(0.05)	0.93	(0.02)	0.83	(0.01)	0.66	(0.01)	0.52	(0.09)	0.84	(0.02)	0.84	(0.01)	0.75	(0.02)
Singapore	0.44	(0.04)	0.97	(0.02)	0.92	(0.01)	0.76	(0.01)	0.44	(0.10)	0.84	(0.02)	0.98	(0.01)	0.99	(0.02)
Chinese Taipei	0.40	(0.04)	1.00	(0.02)	0.95	(0.01)	0.67	(0.01)	1.09	(0.24)	0.90	(0.02)	0.92	(0.02)	0.84	(0.02)
Thailand	0.53	(0.06)	1.37	(0.03)	0.95	(0.02)	0.85	(0.01)	m	m	0.74	(0.02)	0.74	(0.02)	0.66	(0.02)
Tunisia	1.30	(0.08)	1.03	(0.04)	1.14	(0.02)	1.24	(0.02)	m	m	1.11	(0.03)	1.03	(0.02)	1.13	(0.04)
United Arab Emirates	1.35	(0.06)	1.20	(0.03)	1.08	(0.02)	0.89	(0.01)	1.21	(0.16)	0.98	(0.02)	0.98	(0.02)	1.06	(0.03)
Uruguay	1.58	(0.10)	1.11	(0.04)	1.03	(0.02)	1.25	(0.02)	m	m	1.00	(0.03)	0.89	(0.02)	0.93	(0.02)
Viet Nam	0.65	(0.10)	0.79	(0.02)	0.88	(0.02)	0.25	(0.00)	m	m	0.89	(0.02)	0.72	(0.04)	0.60	(0.02)
Average	0.92	(0.01)	0.98	(0.00)	0.97	(0.00)	0.91	(0.00)	0.88	(0.03)	0.97	(0.00)	0.95	(0.00)	0.95	(0.00)

Note: the index of response time effort has a missing value for countries that did not administer the computer-based assessment

Table F3

Country level correlations between PSEC indicators

	Self-reported lack of perseverance	Self-reported lack of effort in test	Difference in effort (PISA-marked)	Non-response	Non-differentiation	Non-differentiation (more lenient)	Inconsistency	Performance decline	Response time effort (3 seconds)	Response time effort (5 seconds)	Response time effort (10 seconds)
Self-reported lack of perseverance	1.00										
Self-reported lack of effort in test	0.69***	1.00									
Difference in effort (PISA-marked)	0.55***	0.86***	1.00								
Non-response	-0.17	-0.04	-0.06	1.00							
Non-differentiation	-0.40**	-0.36**	-0.40**	0.18	1.00						
Non-differentiation (more lenient)	-0.40**	-0.37**	-0.41***	0.15	1.00***	1.00					
Inconsistency	0.07	0.15	0.10	0.43***	-0.25*	-0.26*	1.00				
Performance decline	0.28*	0.38**	0.29*	-0.15	0.05	0.06	-0.07	1.00			
Response time effort (3 seconds)	-0.04	0.18	0.28	0.44*	-0.17	-0.18	0.32	0.22	1.00		
Response time effort (5 seconds)	-0.09	0.12	0.23	0.40*	-0.14	-0.14	0.31	0.22	0.98***	1.00	
Response time effort (10 seconds)	-0.14	0.07	0.17	0.45*	-0.11	-0.12	0.33	0.14	0.96***	0.99***	1.00

Note: PISA 2012 data. The number of country level observations is 64 for results involving indicators obtained from the paper-based assessment and 32 for results involving indicators obtained from the computer-based assessment.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Annex G Stability of PSEC indicators over time

Figure G1 – Correlation in index of self-reported effort between PISA 2003 and PISA 2012

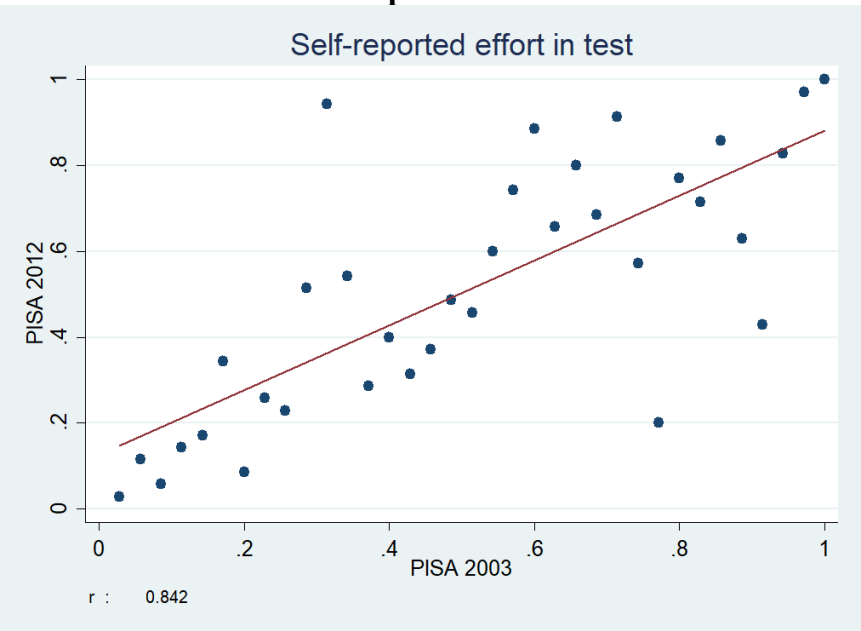


Figure G2 – Correlation in index of self-reported difference in effort (marked-PISA) between PISA 2003 and PISA 2012

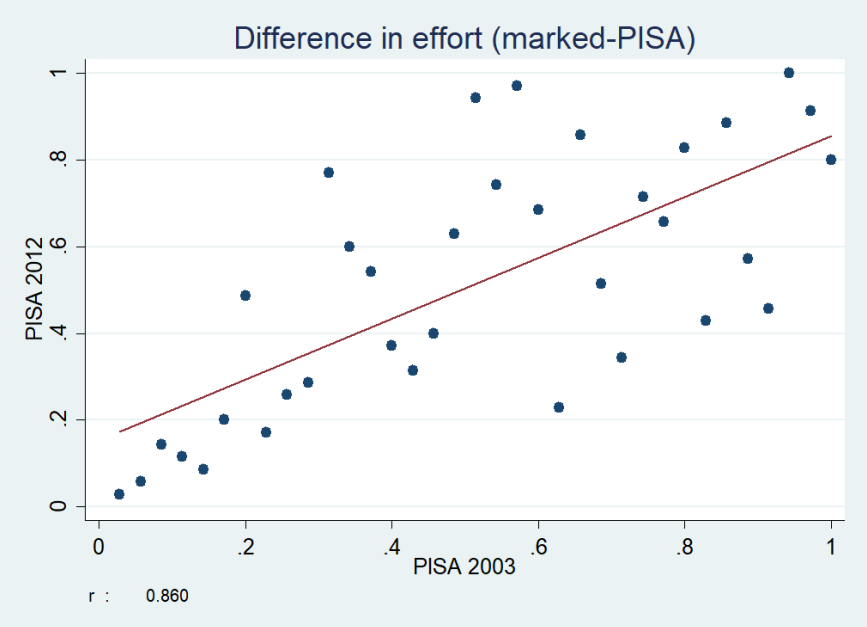


Figure G3 – Correlation in index of non-response between PISA 2003 and PISA 2012

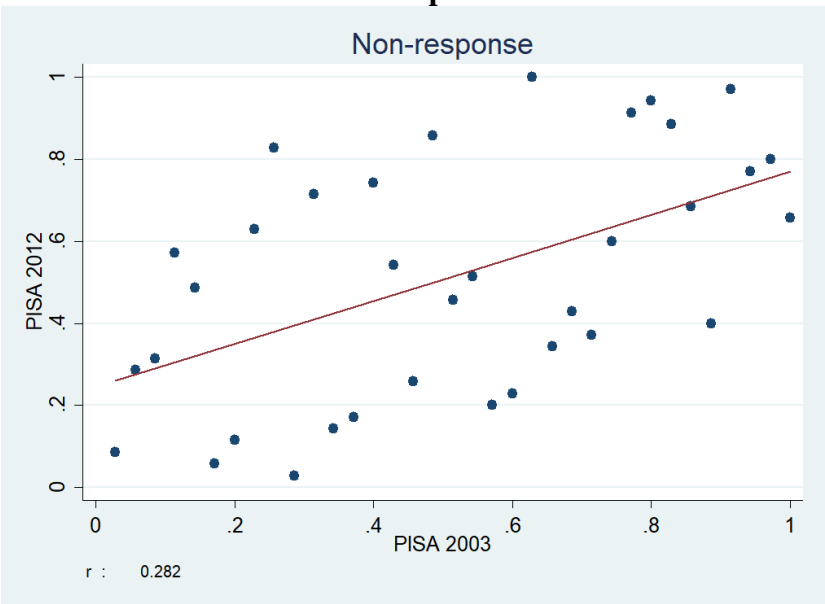


Figure G4 – Correlation in index of non-differentiation between PISA 2003 and PISA 2012

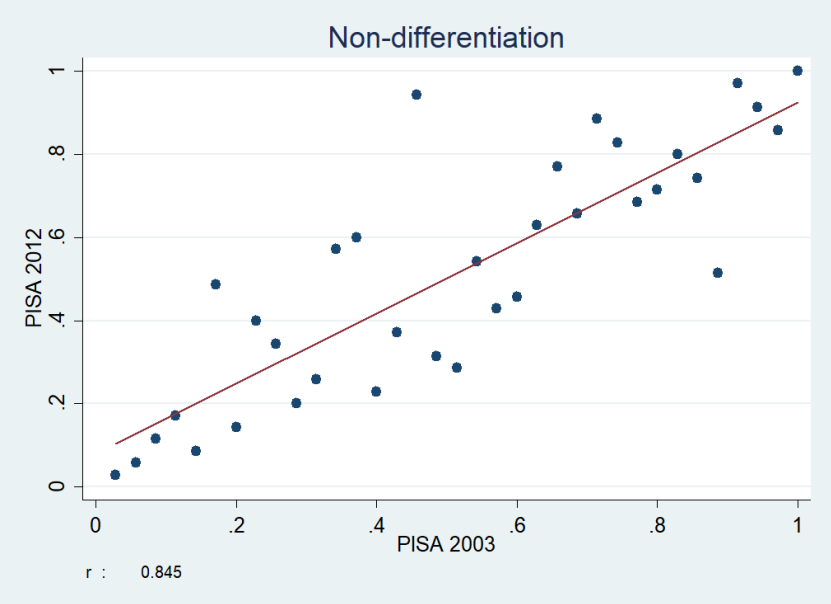


Figure G5 – Correlation in index of inconsistency between PISA 2003 and PISA 2012

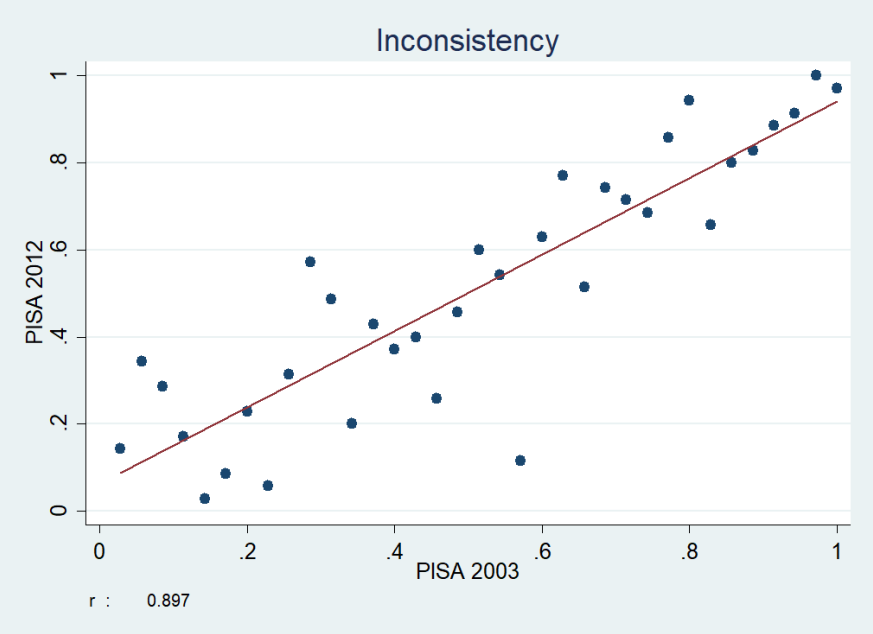


Figure G6 – Correlation in index of performance decline between PISA 2003 and PISA 2012

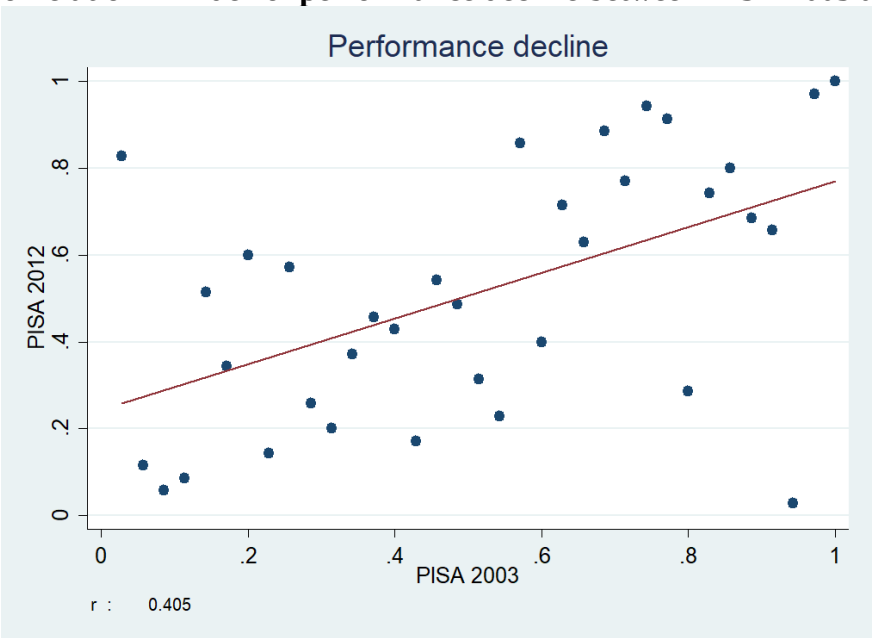


Figure G7 – Correlation of rank in index of self-reported effort between PISA 2003 and PISA 2012

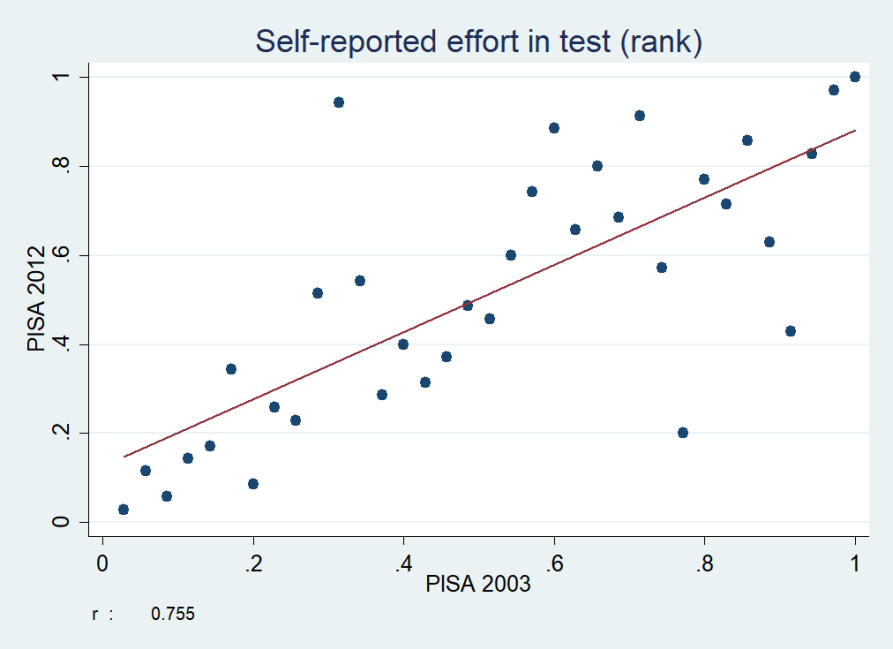


Figure G8 – Correlation of rank in index of self-reported difference in effort (marked-PISA) between PISA 2003 and PISA 2012

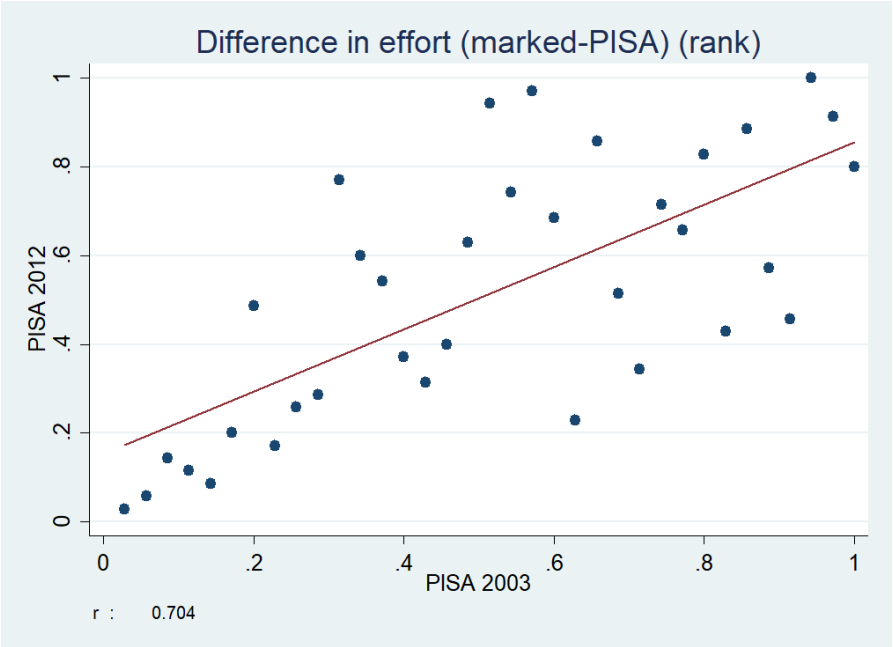


Figure G9 – Correlation of rank in index of non-response between PISA 2003 and PISA 2012

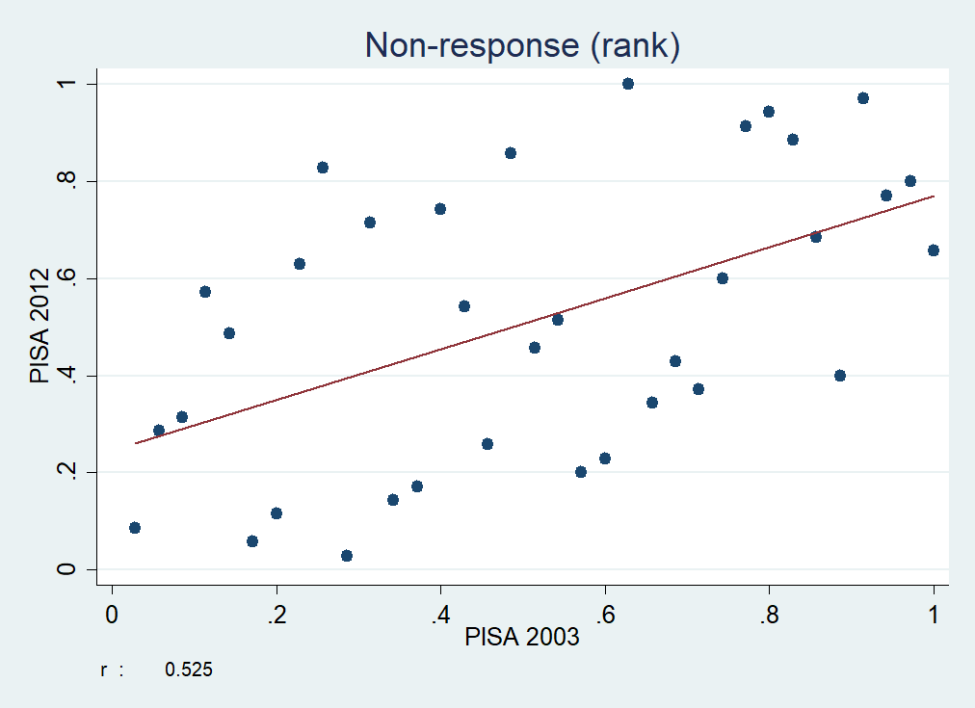


Figure G10 – Correlation of rank in index of non-differentiation between PISA 2003 and PISA 2012

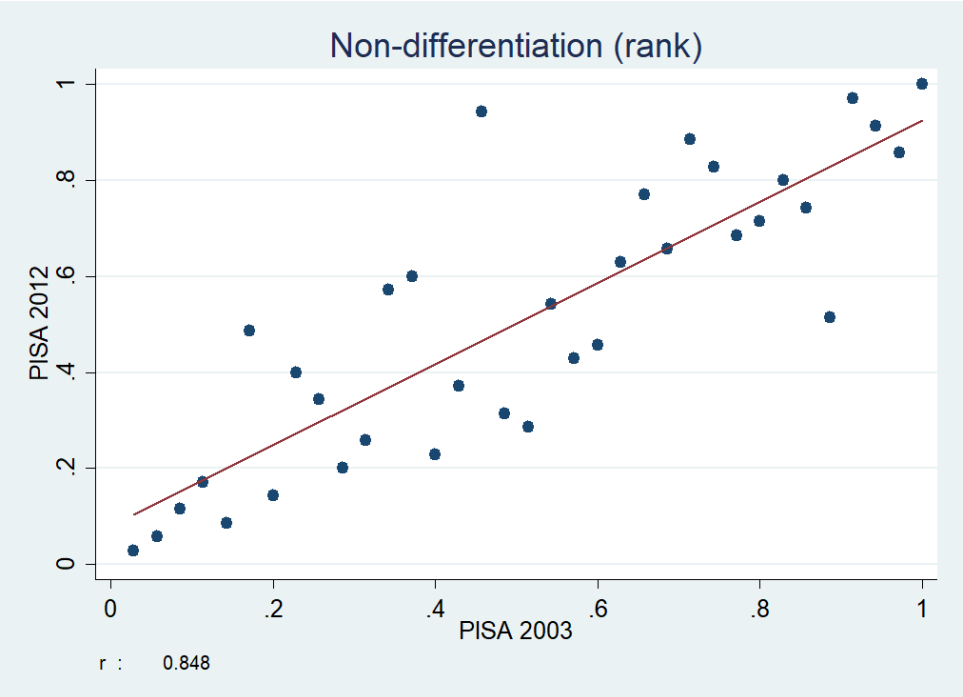


Figure G11 – Correlation of rank in index of inconsistency between PISA 2003 and PISA 2012

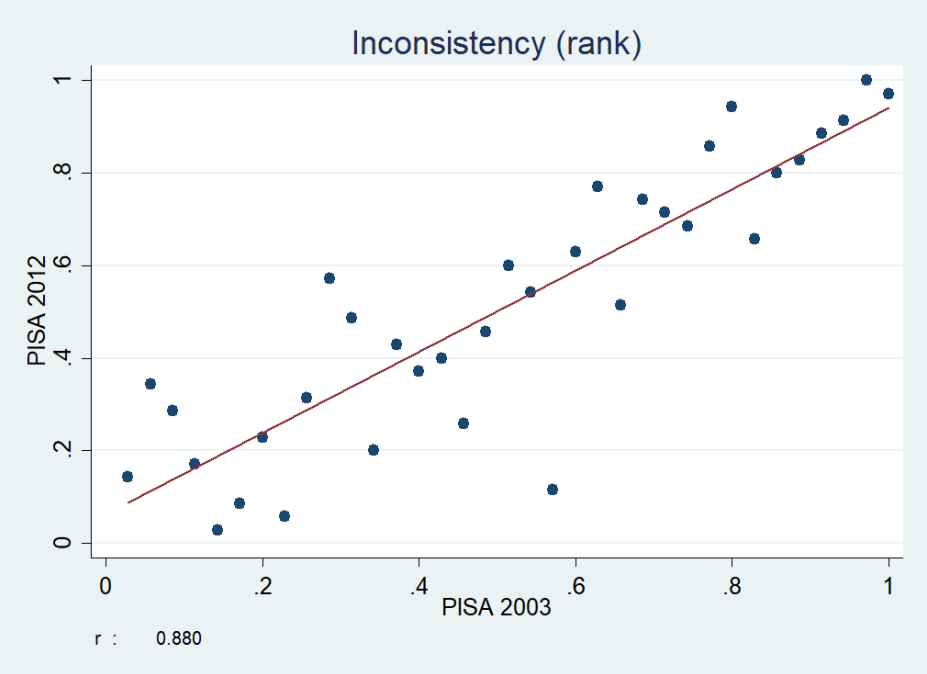
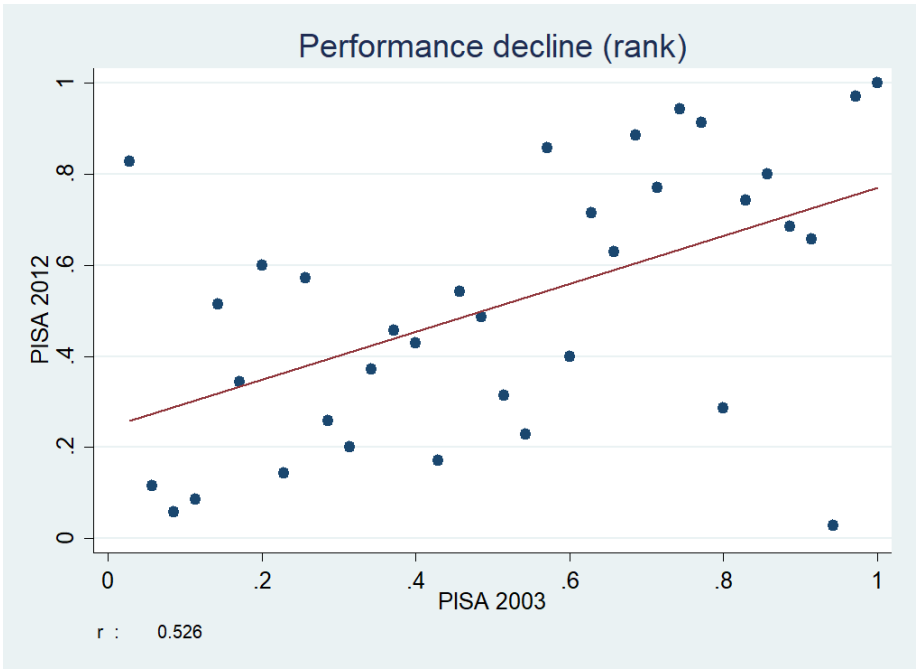


Figure G12 – Correlation of rank in in index of performance decline between PISA 2003 and PISA 2012



Annex H Measuring gender differences in PSEC indicators

Table H1 – Gender differences in PSEC indicators by country

	Self-reported lack of perseverance		Self-reported lack of effort in Test		Difference in effort (PISA-test)		Nonesponse		Non-differentiation		Inconsistency		Performance decline		Response-time-effort (5 seconds)	
Australia	0.19	***	-0.11	***	0.00		-0.11	***	-0.18	***	0.03		-0.09	***	-0.09	**
Austria	0.23	***	-0.22	***	-0.09	*	-0.11	***	-0.13	***	-0.06		-0.06	*	-0.04	
Belgium	0.19	***	-0.11	***	0.01		-0.13	***	-0.11	***	-0.09	**	-0.11	***	-0.07	
Canada	0.06	*	-0.17	***	0.00		-0.08	***	-0.17	***	-0.05	**	-0.13	***	-0.08	*
Chile	0.00		-0.16	***	-0.03		-0.02		-0.09	**	0.01		-0.03		-0.11	
Czech Republic	-0.08	*	-0.13	***	0.07		-0.10	***	-0.12	***	-0.14	***	-0.10	***	m	
Denmark	0.22	***	-0.21	***	-0.15	***	-0.05	**	-0.15	***	0.04		-0.11	***	-0.11	
Estonia	-0.05		-0.36	***	-0.20	***	-0.07	***	-0.12	***	-0.09	**	-0.08	***	-0.07	
Finland	0.13	***	-0.26	***	-0.16	***	-0.18	***	-0.29	***	0.03		-0.15	***	m	
France	0.25	***	-0.35	***	-0.16	***	-0.22	***	-0.09	***	-0.15	***	-0.15	***	-0.07	
Germany	0.22	***	-0.23	***	-0.05		-0.16	***	-0.21	***	-0.09	**	-0.12	***	-0.06	
Greece	0.07	*	-0.20	***	-0.07		-0.09	***	-0.26	***	-0.22	***	-0.10	*	m	
Hungary	0.00		-0.23	***	-0.08	**	-0.12	***	-0.15	***	-0.20	***	-0.09	***	-0.07	
Iceland	0.14	**	-0.26	***	-0.14	***	-0.19	***	-0.38	***	-0.03		-0.08	*	m	
Ireland	0.15	***	-0.23	***	-0.13	***	-0.03		-0.07	**	-0.07	*	-0.04		-0.07	
Israel	-0.07		-0.22	***	-0.05		-0.21	*	-0.21	***	-0.05		-0.14	*	-0.55	**
Italy	0.05		-0.14	***	-0.06	***	-0.11	***	-0.10	***	-0.22	***	-0.12	***	-0.06	
Japan	0.10	***	-0.05		0.03		-0.02	*	-0.25	***	-0.15	***	-0.03		-0.05	*
Korea	0.25	***	0.02		0.04		-0.02	*	-0.23	***	-0.06	*	-0.01		-0.12	**
Latvia	-0.07		-0.24	***	-0.11	**	-0.12	***	-0.14	***	-0.18	***	-0.12	***	m	
Luxembourg	0.19	***	-0.23	***	-0.12	***	-0.10	***	-0.29	***	-0.12	***	-0.08	**	m	
Mexico	-0.08	**	-0.09	***	-0.04	*	-0.04	*	-0.16	***	-0.07	**	-0.05	*	m	
Netherlands	0.06	*	-0.06		0.01		0.03		-0.05		-0.13	***	0.00		m	
New Zealand	0.16	***	-0.05		0.02		-0.12	***	-0.16	***	0.02		-0.12	***	m	
Norway	0.24	***	-0.28	***	-0.14	***	-0.17	***	-0.30	***	0.02		-0.15	***	-0.29	**
Poland	0.03		-0.41	***	-0.11	**	-0.07	***	-0.22	***	-0.15	***	-0.09	***	-0.15	*
Portugal	-0.12	**	-0.26	***	-0.13	***	-0.04		-0.27	***	-0.08	***	-0.13	***	-0.05	
Slovak Republic	0.16	***	-0.30	***	-0.06		-0.07		-0.22	***	-0.15	***	-0.07	*	-0.06	
Slovenia	-0.03		-0.12	***	0.00		-0.12	***	-0.24	***	-0.18	***	-0.15	***	-0.34	***
Spain	0.05		-0.18	***	-0.10	***	-0.06	***	-0.09	***	-0.11	***	-0.09	***	-0.11	
Sweden	0.25	***	-0.25	***	-0.13	***	-0.20	***	-0.36	***	0.01		-0.18	***	-0.12	*
Switzerland	0.21	***	-0.20	***	-0.07	*	-0.11	***	-0.17	***	-0.08	**	-0.13	***	m	
Turkey	-0.12	***	-0.12	***	0.01		-0.07	*	-0.18	***	-0.11	***	-0.08	***	m	
United Kingdom	0.24	***	-0.10	**	-0.02		-0.08	*	-0.10	***	0.03		-0.07	***	m	
United States	0.03		-0.08	**	-0.01		-0.20	***	-0.22	***	0.01		-0.07	***	-0.05	
Argentina	0.05		-0.08		0.05		-0.15		-0.07		-0.21	***	-0.07		m	
Albania	0.03			***		***	0.13		0.00		-0.01		-0.06		m	
Brazil	-0.15	***	-0.18	***	-0.04		-0.14	**	-0.17	***	-0.10	**	-0.06		-0.03	
Bulgaria	-0.19	**	-0.34	***	-0.13	**	-0.16		-0.28	***	-0.32	***	0.05		m	
Colombia	-0.18	**	-0.15	***	-0.05		0.03		-0.04		-0.08		-0.05		-0.05	
Costa Rica	-0.06		-0.16	***	-0.07		-0.08		-0.07		-0.10		0.00		m	
Croatia	0.00		-0.14	***	0.05		-0.09	***	-0.17	***	-0.16	***	-0.10	***	m	
Hong Kong (China)	0.13	***	-0.17	***	-0.05		-0.04	**	-0.20	***	-0.17	***	0.02		-0.21	***
Indonesia	0.02		-0.02		0.11	*	-0.06		-0.08	*	-0.10	**	-0.05	*	m	
Jordan	-0.21	***	-0.20	**	-0.03		-0.51	***	-0.45	***	-0.43	***	0.14	**	m	
Kazakhstan	-0.08		-0.14	***	0.01		-0.05		-0.21	***	-0.09		0.07		m	
Liechtenstein	0.17		-0.08		0.12		0.05		-0.26	***	-0.13		-0.07		m	
Lithuania	-0.14	***	-0.33	***	-0.17	***	-0.13	***	-0.24	***	-0.14	***	-0.07	**	m	
Macao (China)	0.09	***	-0.10	***	0.12	***	-0.06	***	-0.14	***	-0.15	***	-0.05	**	-0.05	
Malaysia	-0.03		-0.04		-0.01		-0.09	*	-0.20	***	-0.11	***	0.00		m	
Montenegro	-0.27	***	-0.28	***	-0.07		-0.22	***	-0.33	***	-0.26	***	-0.06	*	m	
Peru	-0.21	***	-0.18	***	-0.04		0.08		-0.09	*	-0.05		0.03		m	
Qatar	-0.10	***	-0.01		0.04		-0.82	***	-0.40	***	-0.15	***	-0.06	**	m	
Romania	-0.04		-0.23	***	-0.05		0.04		-0.06		-0.19	***	-0.05		m	
Russia	-0.06	*	-0.34	***	-0.15	***	-0.12	***	-0.18	***	-0.06	*	-0.08	*	-0.05	
Serbia	-0.05		-0.22	***	-0.04		-0.18	**	-0.25	***	-0.20	***	-0.04		m	
Shanghai	0.16	***	-0.01		0.07	**	-0.03	***	-0.15	***	-0.14	***	-0.04	*	-0.03	
Singapore	0.13	***	-0.16	***	-0.06	*	-0.06	***	-0.14	***	-0.18	***	-0.07	***	-0.03	
Chinese Taipei	0.09	**	-0.03		0.10	***	-0.07	***	-0.25	***	-0.15	***	-0.07	*	-0.14	
Thailand	-0.14	***	-0.18	***	0.04		-0.07	***	-0.35	***	-0.13	***	-0.10	***	m	
Tunisia	0.04		-0.14	*	-0.01		-0.10		-0.15	**	-0.30	***	-0.07		m	
United Arab Emirates	-0.10	*	-0.19	***	-0.02		-0.21	**	-0.33	***	-0.23	***	-0.03		-0.37	***
Uruguay	0.02		-0.29	***	-0.17	***	-0.25	***	-0.22	***	-0.12	**	0.00		m	

Viet Nam	0.03		-0.03		0.03		-0.06		-0.09	*	-0.07	*	0.02		m	
Average	0.03	***	-0.17	***	-0.04	***	-0.11	***	-0.19	***	-0.11	***	-0.07	***	-0.12	***

Notes: The gender gap refers to the difference between females and males. A negative number implies that females have lower mean values than males. A positive value implies that females have higher mean values than males.

* p<0.05, ** p<0.01, *** p<0.001

Table H2 – Gender differences in PSEC indicators by country compared to average

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**0.19**	**-0.11**	0	**-0.11**	**-0.18**	0.03	**-0.09**	**-0.09**
Austria	**0.23**	**-0.22**	**-0.09**	**-0.11**	**-0.13**	-0.06	**-0.06**	-0.04
Belgium	**0.19**	**-0.11**	0.01	**-0.13**	**-0.11**	**-0.09**	**-0.11**	**-0.07**
Canada	**0.06**	**-0.17**	0	**-0.08**	**-0.17**	**-0.05**	**-0.13**	-0.08
Chile	0	**-0.16**	-0.03	-0.02	**-0.09**	0.01	-0.03	-0.11
Czech Republic	**-0.08**	**-0.13**	**0.07**	**-0.1**	**-0.12**	**-0.14**	**-0.1**	
Denmark	**0.22**	**-0.21**	**-0.15**	**-0.05**	**-0.15**	0.04	**-0.11**	-0.11
Estonia	-0.05	**-0.36**	**-0.2**	**-0.07**	**-0.12**	**-0.09**	**-0.08**	-0.07
Finland	**0.13**	**-0.26**	**-0.16**	**-0.18**	**-0.29**	0.03	**-0.15**	
France	**0.25**	**-0.35**	**-0.16**	**-0.22**	**-0.09**	**-0.15**	**-0.15**	-0.07
Germany	**0.22**	**-0.23**	-0.05	**-0.16**	**-0.21**	**-0.09**	**-0.12**	-0.06
Greece	0.07	**-0.2**	**-0.07**	**-0.09**	**-0.26**	**-0.22**	**-0.1**	
Hungary	0	**-0.23**	**-0.08**	**-0.12**	**-0.15**	**-0.2**	**-0.09**	**-0.07**
Iceland	**0.14**	**-0.26**	**-0.14**	**-0.19**	**-0.38**	-0.03	**-0.08**	
Ireland	**0.15**	**-0.23**	**-0.13**	-0.03	**-0.07**	**-0.07**	**-0.04**	-0.07
Israel	-0.07	**-0.22**	-0.05	**-0.21**	**-0.21**	-0.05	**-0.14**	**-0.55**
Italy	**0.05**	**-0.14**	**-0.06**	**-0.11**	**-0.1**	**-0.22**	**-0.12**	-0.06
Japan	**0.1**	-0.05	0.03	**-0.02**	**-0.25**	**-0.15**	-0.03	**-0.05**
Korea	**0.25**	0.02	0.04	**-0.02**	**-0.23**	**-0.06**	-0.01	-0.12
Latvia	-0.07	**-0.24**	**-0.11**	**-0.12**	**-0.14**	**-0.18**	**-0.12**	
Luxembourg	**0.19**	**-0.23**	**-0.12**	**-0.1**	**-0.29**	**-0.12**	**-0.08**	
Mexico	**-0.08**	**-0.09**	**-0.04**	**-0.04**	**-0.16**	**-0.07**	**-0.05**	
Netherlands	**0.06**	-0.06	0.01	0.03	-0.05	**-0.13**	0	
New Zealand	**0.16**	-0.05	0.02	**-0.12**	**-0.16**	0.02	**-0.12**	
Norway	**0.24**	**-0.28**	**-0.14**	**-0.17**	**-0.3**	0.02	**-0.15**	-0.29
Poland	0.03	**-0.41**	**-0.11**	**-0.07**	**-0.22**	**-0.15**	**-0.09**	-0.15
Portugal	**-0.12**	**-0.26**	**-0.13**	-0.04	**-0.27**	**-0.08**	**-0.13**	-0.05
Slovak Republic	**0.16**	**-0.3**	-0.06	-0.07	**-0.22**	**-0.15**	**-0.07**	**-0.06**
Slovenia	-0.03	**-0.12**	0	**-0.12**	**-0.24**	**-0.18**	**-0.15**	-0.34
Spain	0.05	**-0.18**	**-0.1**	**-0.06**	**-0.09**	**-0.11**	**-0.09**	-0.11
Sweden	**0.25**	**-0.25**	**-0.13**	**-0.2**	**-0.36**	0.01	**-0.18**	-0.12
Switzerland	**0.21**	**-0.2**	**-0.07**	**-0.11**	**-0.17**	**-0.08**	**-0.13**	
Turkey	**-0.12**	**-0.12**	0.01	**-0.07**	**-0.18**	**-0.11**	**-0.08**	
United Kingdom	**0.24**	**-0.1**	-0.02	**-0.08**	**-0.1**	0.03	**-0.07**	
United States	0.03	**-0.08**	-0.01	**-0.2**	**-0.22**	0.01	**-0.07**	-0.05
Argentina	0.05	-0.08	0.05	-0.15	-0.07	**-0.21**	-0.07	
Albania	0.03			0.13	0	-0.01	-0.06	
Brazil	**-0.15**	**-0.18**	**-0.04**	**-0.14**	**-0.17**	**-0.1**	**-0.06**	-0.03
Bulgaria	**-0.19**	**-0.34**	**-0.13**	**-0.16**	**-0.28**	**-0.32**	0.05	
Colombia	**-0.18**	**-0.15**	-0.05	0.03	-0.04	-0.08	-0.05	-0.05
Costa Rica	-0.06	**-0.16**	-0.07	-0.08	-0.07	**-0.1**	0	
Croatia	0	**-0.14**	0.05	**-0.09**	**-0.17**	**-0.16**	**-0.1**	
Hong Kong (China)	**0.13**	**-0.17**	-0.05	**-0.04**	**-0.2**	**-0.17**	0.02	-0.21
Indonesia	0.02	-0.02	**0.11**	**-0.06**	**-0.08**	**-0.1**	**-0.05**	
Jordan	**-0.21**	**-0.2**	-0.03	**-0.51**	**-0.45**	**-0.43**	**0.14**	
Kazakhstan	-0.08	**-0.14**	0.01	-0.05	**-0.21**	-0.09	0.07	
Liechtenstein	0.17	-0.08	0.12	0.05	**-0.26**	-0.13	-0.07	
Lithuania	**-0.14**	**-0.33**	**-0.17**	**-0.13**	**-0.24**	**-0.14**	**-0.07**	
Macao (China)	**0.09**	**-0.1**	**0.12**	**-0.06**	**-0.14**	**-0.15**	**-0.05**	-0.05
Malaysia	-0.03	-0.04	-0.01	**-0.09**	**-0.2**	**-0.11**	0	
Montenegro	**-0.27**	**-0.28**	-0.07	**-0.22**	**-0.33**	**-0.26**	**-0.06**	
Peru	**-0.21**	**-0.18**	-0.04	0.08	**-0.09**	-0.05	0.03	
Qatar	**-0.1**	-0.01	0.04	**-0.82**	**-0.4**	**-0.15**	**-0.06**	
Romania	-0.04	**-0.23**	-0.05	0.04	-0.06	**-0.19**	-0.05	
Russia	**-0.06**	**-0.34**	**-0.15**	**-0.12**	**-0.18**	**-0.06**	**-0.08**	-0.05
Serbia	-0.05	**-0.22**	-0.04	**-0.18**	**-0.25**	**-0.2**	-0.04	
Shanghai	**0.16**	-0.01	**0.07**	**-0.03**	**-0.15**	**-0.14**	**-0.04**	-0.03
Singapore	**0.13**	**-0.16**	**-0.06**	**-0.06**	**-0.14**	**-0.18**	**-0.07**	**-0.03**
Chinese Taipei	**0.09**	-0.03	**0.1**	**-0.07**	**-0.25**	**-0.15**	**-0.07**	-0.14
Thailand	**-0.14**	**-0.18**	**0.04**	**-0.07**	**-0.35**	**-0.13**	**-0.1**	
Tunisia	0.04	**-0.14**	-0.01	-0.1	**-0.15**	**-0.3**	-0.07	
United Arab Emirates	**-0.1**	**-0.19**	-0.02	**-0.21**	**-0.33**	**-0.23**	-0.03	**-0.37**
Uruguay	0.02	**-0.29**	**-0.17**	**-0.25**	**-0.22**	**-0.12**	0	
Viet Nam	0.03	-0.03	0.03	**-0.06**	**-0.09**	**-0.07**	**0.02**	
Average	0.03	-0.17	-0.04	-0.11	-0.19	-0.11	-0.07	-0.12

Notes: The gender gap refers to the difference between females and males. A negative number implies that females have lower mean values than males. A positive value implies that females have higher mean values than males. Bolded values indicate that the estimated d value is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null hypothesis of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Annex I Measuring differences in PSEC indicators across socio-economic status
Table I1 – Differences in PSEC indicators across socio-economic status by country

	Self-reported lack of perseverance	Self-reported lack of effort in Test	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	-0.41 ***	-0.10 ***	0.10 ***	-0.19 ***	-0.03	-0.13 ***	-0.18 ***	-0.16 ***
Austria	-0.11 *	-0.05	-0.07	-0.10 **	-0.07 *	-0.12 **	-0.07	-0.13
Belgium	-0.07	-0.05	0.03	-0.19 ***	-0.13 ***	-0.22 ***	-0.32 ***	-0.18 *
Canada	-0.41 ***	-0.06 *	0.12 ***	-0.10 ***	0.02	-0.22 ***	-0.20 ***	-0.02
Chile	-0.15 *	0.20 **	0.26 ***	-0.15 ***	-0.16 ***	-0.03	0.08	0.05
Czech Republic	-0.22 ***	-0.06	-0.05	-0.06 **	-0.05	-0.04	-0.12 **	m
Denmark	-0.43 ***	-0.22 ***	-0.06	-0.18 ***	-0.02	-0.11 **	-0.19 ***	-0.06
Estonia	-0.11 **	0.06	0.03	-0.07 ***	0.01	0.05	-0.06	-0.05
Finland	-0.46 ***	-0.22 ***	-0.07 *	-0.11 ***	-0.01	-0.12 ***	-0.11 **	m
France	-0.40 ***	0.07	0.07	-0.40 ***	-0.04	-0.30 ***	-0.37 ***	-0.07
Germany	-0.18 ***	-0.08	-0.08	-0.26 ***	-0.02	-0.20 ***	-0.18 ***	-0.14
Greece	-0.42 ***	0.08	0.19 ***	-0.03	-0.06	-0.24 ***	-0.04	m
Hungary	-0.23 ***	0.02	0.09	-0.20 **	-0.09 *	-0.07	-0.12 ***	-0.17 *
Iceland	-0.41 ***	-0.20 ***	0.02	-0.22 ***	0.10	-0.19 ***	-0.11 *	m
Ireland	-0.38 ***	-0.10 *	-0.02	-0.10 ***	-0.01	-0.10 **	-0.16 ***	-0.09
Israel	-0.04	0.04	0.05	-0.42 ***	-0.17 ***	-0.18 ***	-0.32 ***	-0.52 *
Italy	-0.15 ***	-0.14 ***	0.00	-0.09 ***	-0.03	-0.13 ***	-0.11 ***	0.01
Japan	-0.19 ***	-0.03	0.09	-0.07 ***	-0.16 **	-0.03	-0.33 ***	-0.13 *
Korea	-0.30 ***	-0.09	0.18 ***	-0.02	-0.19 ***	0.06	-0.34 ***	-0.06
Latvia	-0.41 ***	-0.23 ***	0.11 *	-0.06	0.03	-0.02	-0.03	m
Luxembourg	-0.26 ***	-0.09	-0.01	-0.29 ***	-0.02	-0.14 ***	-0.16 ***	m
Mexico	-0.25 ***	0.13 ***	0.23 ***	-0.23 ***	-0.09 *	0.00	-0.01	m
Netherlands	0.00	-0.10 *	-0.06	-0.06	-0.08 *	-0.11 *	-0.17 ***	m
New Zealand	-0.39 ***	-0.24 ***	0.07	-0.30 ***	-0.20 ***	-0.19 ***	-0.28 ***	m
Norway	-0.44 ***	-0.01	0.15 ***	-0.07	0.08 *	-0.21 ***	-0.16 ***	-0.11
Poland	-0.43 ***	0.22 ***	0.25 ***	-0.09 ***	-0.03	-0.05	-0.12 **	-0.13
Portugal	-0.42 ***	-0.05	0.13 **	-0.07	-0.31 ***	0.00	-0.11 *	-0.02
Slovak Republic	-0.29 ***	0.02	0.21 ***	-0.35 ***	-0.03	-0.26 ***	-0.05	-0.06
Slovenia	-0.16 **	-0.09	0.03	0.00	-0.04	-0.02	-0.14 **	-0.21
Spain	-0.30 ***	0.08 **	0.22 ***	-0.07 *	-0.04 *	-0.16 ***	-0.07 **	-0.10
Sweden	-0.36 ***	-0.15 **	-0.06	-0.21 ***	-0.02	-0.15 **	-0.16 **	-0.10
Switzerland	-0.05	-0.03	-0.04	-0.12 ***	-0.01	-0.05	-0.18 ***	m
Turkey	-0.21 ***	0.19 ***	0.20 ***	-0.10 *	-0.10 *	-0.14 ***	-0.03	m
United Kingdom	-0.33 ***	-0.06	0.12 *	-0.18 ***	0.00	-0.18 ***	-0.21 ***	m
United States	-0.46 ***	-0.03	0.07	-0.13 ***	0.04	-0.24 ***	-0.18 ***	-0.08
Argentina	-0.20 *	0.04	0.17 *	-0.55 ***	-0.28 ***	-0.10	-0.02	m
Albania								m
Brazil	-0.09	0.21 ***	0.26 ***	-0.58 ***	-0.15 **	0.00	0.12 **	0.03
Bulgaria	-0.62 ***	0.13	0.21 **	-0.39 ***	-0.49 ***	-0.18 *	0.05	m
Colombia	-0.23 **	0.21 **	0.29 ***	-0.82 ***	-0.19 **	-0.07	0.18 *	0.02
Costa Rica	-0.19 *	0.24 **	0.23 **	-0.32 ***	-0.19 **	0.09	0.00	m
Croatia	-0.06	0.11 *	0.19 ***	-0.10 ***	-0.03	-0.05	-0.05	m
Hong Kong (China)	-0.30 ***	0.06	0.17 ***	0.02	0.02	-0.06	-0.26 ***	-0.03
Indonesia	-0.21 ***	0.15	0.27 ***	-0.08	-0.16 **	0.09 *	0.03	m
Jordan	-0.35 ***	0.20 *	0.27 ***	-0.12	0.08	-0.23 ***	0.36 ***	m
Kazakhstan	-0.69 ***	0.00	0.12 *	-0.23 ***	0.04	-0.27 ***	0.04	m
Liechtenstein	-0.46 **	-0.06	0.07	-0.03	0.08	-0.27	-0.23 *	m
Lithuania	-0.24 ***	-0.11 **	0.02	-0.09 **	-0.05	-0.23 ***	0.03	m
Macao (China)	-0.24 ***	-0.03	0.13 ***	-0.01	0.01	0.09 **	-0.06 **	0.02
Malaysia	-0.24 ***	-0.08	0.15 ***	-0.12 **	-0.13 **	-0.02	0.26 ***	m
Montenegro	-0.23 ***	0.20 ***	0.19 ***	-0.28 ***	-0.15 **	0.05	0.18 ***	m
Peru	-0.28 ***	0.03	0.18 ***	-0.90 ***	-0.12	-0.19 ***	0.12 *	m
Qatar	-0.29 ***	0.15 ***	0.35 ***	-0.08	0.02	-0.20 ***	0.16 ***	m
Romania	-0.40 ***	0.13	0.21 **	-0.27 ***	-0.27 **	-0.17	0.21 **	m
Russia	-0.34 ***	0.14 **	0.16 ***	-0.20 ***	-0.07	-0.04	0.02	0.02
Serbia	-0.29 **	0.09	0.10	-0.12 *	-0.07	-0.08	-0.08	m
Shanghai	-0.30 ***	0.12 ***	0.09 **	0.05 ***	0.15 ***	-0.09 *	-0.26 ***	-0.07
Singapore	-0.19 ***	-0.04	0.09 *	0.00	0.00	-0.23 ***	-0.38 ***	-0.05 *
Chinese Taipei	-0.34 ***	-0.27 ***	0.03	-0.01	-0.24 ***	0.01	-0.32 ***	-0.19 *
Thailand	-0.22 ***	0.00	0.08 *	-0.01	-0.29 ***	0.21 ***	0.04	m
Tunisia	-0.22 *	0.15 *	0.26 **	-0.22 *	0.12	-0.16	0.11	m
United Arab Emirates	-0.35 ***	0.16 **	0.30 ***	-0.27 ***	-0.05	-0.28 ***	0.06	-0.25 *
Uruguay	-0.23 **	0.12 *	0.28 ***	-0.64 ***	-0.32 ***	-0.19 **	-0.27 **	m
Viet Nam	-0.14	-0.08	0.19 ***	-0.08 *	-0.07	0.08	0.05 **	m
Average	-0.28 ***	0.01	0.12 ***	-0.19 ***	-0.08 ***	-0.11 ***	-0.08 ***	-0.10 ***

s: The socio-economic (SES) gap refers to the difference between socio-economically advantaged and socio-economically disadvantaged students. PISA contains an aggregate SES measure l on students' reports on their parents' educational attainment, occupational status and availability of economic and cultural resources in their home. Advantaged students are students in the top of the country specific distribution of SES. Disadvantages students are students in the bottom 25% of the country specific distribution of SES. A negative number implies that advantaged have r mean values than disadvantaged students. A positive value implies that advantaged have higher mean values than disadvantaged students. * p<0.05, ** p<0.01, *** p<0.001

Table I2 – Differences in PSEC indicators across socio-economic status by country compared to average

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.41**	**-0.1**	**0.1**	**-0.19**	-0.03	**-0.13**	**-0.18**	**-0.16**
Austria	**-0.11**	-0.05	-0.07	**-0.1**	**-0.07**	**-0.12**	-0.07	-0.13
Belgium	-0.07	-0.05	0.03	**-0.19**	**-0.13**	**-0.22**	**-0.32**	**-0.18**
Canada	**-0.41**	**-0.06**	**0.12**	**-0.1**	0.02	**-0.22**	**-0.2**	-0.02
Chile	**-0.15**	**0.2**	**0.26**	**-0.15**	**-0.16**	-0.03	0.08	0.05
Czech Republic	**-0.22**	-0.06	-0.05	**-0.06**	-0.05	-0.04	**-0.12**	
Denmark	**-0.43**	**-0.22**	-0.06	**-0.18**	-0.02	**-0.11**	**-0.19**	-0.06
Estonia	**-0.11**	0.06	0.03	**-0.07**	0.01	0.05	**-0.06**	-0.05
Finland	**-0.46**	**-0.22**	**-0.07**	**-0.11**	-0.01	**-0.12**	**-0.11**	
France	**-0.4**	0.07	0.07	**-0.4**	-0.04	**-0.3**	**-0.37**	-0.07
Germany	**-0.18**	-0.08	-0.08	**-0.26**	-0.02	**-0.2**	**-0.18**	-0.14
Greece	**-0.42**	**0.08**	**0.19**	-0.03	-0.06	**-0.24**	-0.04	
Hungary	**-0.23**	0.02	**0.09**	**-0.2**	**-0.09**	-0.07	**-0.12**	**-0.17**
Iceland	**-0.41**	**-0.2**	0.02	**-0.22**	**0.1**	**-0.19**	**-0.11**	
Ireland	**-0.38**	**-0.1**	-0.02	**-0.1**	-0.01	**-0.1**	**-0.16**	-0.09
Israel	-0.04	0.04	0.05	**-0.42**	**-0.17**	**-0.18**	**-0.32**	**-0.52**
Italy	**-0.15**	**-0.14**	0	**-0.09**	-0.03	**-0.13**	**-0.11**	0.01
Japan	**-0.19**	-0.03	0.09	**-0.07**	**-0.16**	-0.03	**-0.33**	**-0.13**
Korea	**-0.3**	-0.09	**0.18**	-0.02	**-0.19**	0.06	**-0.34**	-0.06
Latvia	**-0.41**	**-0.23**	**0.11**	**-0.06**	0.03	-0.02	-0.03	
Luxembourg	**-0.26**	-0.09	-0.01	**-0.29**	-0.02	**-0.14**	**-0.16**	
Mexico	**-0.25**	**0.13**	**0.23**	**-0.23**	**-0.09**	0	-0.01	
Netherlands	0	**-0.1**	**-0.06**	-0.06	**-0.08**	**-0.11**	**-0.17**	
New Zealand	**-0.39**	**-0.24**	0.07	**-0.3**	**-0.2**	**-0.19**	**-0.28**	
Norway	**-0.44**	-0.01	**0.15**	-0.07	**0.08**	**-0.21**	**-0.16**	-0.11
Poland	**-0.43**	**0.22**	**0.25**	**-0.09**	-0.03	-0.05	**-0.12**	-0.13
Portugal	**-0.42**	-0.05	**0.13**	-0.07	**-0.31**	0	**-0.11**	-0.02
Slovak Republic	**-0.29**	0.02	**0.21**	**-0.35**	-0.03	**-0.26**	-0.05	**-0.06**
Slovenia	**-0.16**	-0.09	0.03	0	-0.04	-0.02	**-0.14**	-0.21
Spain	**-0.3**	**0.08**	**0.22**	**-0.07**	**-0.04**	**-0.16**	**-0.07**	-0.1
Sweden	**-0.36**	**-0.15**	-0.06	**-0.21**	-0.02	**-0.15**	**-0.16**	-0.1
Switzerland	-0.05	-0.03	-0.04	**-0.12**	-0.01	-0.05	**-0.18**	
Turkey	**-0.21**	**0.19**	**0.2**	**-0.1**	**-0.1**	**-0.14**	-0.03	
United Kingdom	**-0.33**	-0.06	**0.12**	**-0.18**	0	**-0.18**	**-0.21**	
United States	**-0.46**	-0.03	0.07	**-0.13**	0.04	**-0.24**	**-0.18**	-0.08
Argentina	**-0.2**	0.04	**0.17**	**-0.55**	**-0.28**	-0.1	-0.02	
Albania								
Brazil	-0.09	**0.21**	**0.26**	**-0.58**	**-0.15**	0	**0.12**	0.03
Bulgaria	**-0.62**	0.13	**0.21**	**-0.39**	**-0.49**	**-0.18**	0.05	
Colombia	**-0.23**	**0.21**	**0.29**	**-0.82**	**-0.19**	-0.07	**0.18**	0.02
Costa Rica	**-0.19**	**0.24**	**0.23**	**-0.32**	**-0.19**	0.09	0	
Croatia	-0.06	**0.11**	**0.19**	**-0.1**	-0.03	-0.05	-0.05	
Hong Kong (China)	**-0.3**	0.06	**0.17**	0.02	0.02	-0.06	**-0.26**	-0.03
Indonesia	**-0.21**	0.15	**0.27**	-0.08	**-0.16**	**0.09**	0.03	
Jordan	**-0.35**	**0.2**	**0.27**	-0.12	0.08	**-0.23**	**0.36**	
Kazakhstan	**-0.69**	0	**0.12**	**-0.23**	0.04	**-0.27**	0.04	
Liechtenstein	**-0.46**	-0.06	0.07	-0.03	0.08	-0.27	**-0.23**	
Lithuania	**-0.24**	**-0.11**	0.02	**-0.09**	-0.05	**-0.23**	0.03	
Macao (China)	**-0.24**	-0.03	**0.13**	-0.01	0.01	**0.09**	**-0.06**	0.02
Malaysia	**-0.24**	-0.08	**0.15**	**-0.12**	**-0.13**	-0.02	**0.26**	
Montenegro	**-0.23**	**0.2**	**0.19**	**-0.28**	**-0.15**	0.05	**0.18**	
Peru	**-0.28**	0.03	**0.18**	**-0.9**	-0.12	**-0.19**	**0.12**	
Qatar	**-0.29**	**0.15**	**0.35**	-0.08	0.02	**-0.2**	**0.16**	
Romania	**-0.4**	0.13	**0.21**	**-0.27**	**-0.27**	-0.17	**0.21**	
Russia	**-0.34**	**0.14**	**0.16**	**-0.2**	-0.07	-0.04	0.02	0.02
Serbia	**-0.29**	0.09	0.1	**-0.12**	-0.07	-0.08	-0.08	
Shanghai	**-0.3**	**0.12**	**0.09**	**0.05**	**0.15**	**-0.09**	**-0.26**	-0.07
Singapore	**-0.19**	-0.04	**0.09**	0	0	**-0.23**	**-0.38**	**-0.05**
Chinese Taipei	**-0.34**	**-0.27**	0.03	-0.01	**-0.24**	0.01	**-0.32**	-0.19
Thailand	**-0.22**	0	**0.08**	-0.01	**-0.29**	**0.21**	0.04	
Tunisia	**-0.22**	**0.15**	**0.26**	**-0.22**	0.12	**-0.16**	0.11	
United Arab Emirates	**-0.35**	**0.16**	**0.3**	**-0.27**	-0.05	**-0.28**	0.06	**-0.25**
Uruguay	**-0.23**	**0.12**	**0.28**	**-0.64**	**-0.32**	**-0.19**	**-0.27**	
Viet Nam	**-0.14**	-0.08	**0.19**	**-0.08**	-0.07	0.08	**0.05**	
AVG	**-0.28**	0.01	**0.12**	**-0.19**	**-0.08**	**-0.11**	**-0.08**	**-0.1**

Notes: The socio-economic (SES) gap refers to the difference between socio-economically advantaged and socio-economically disadvantaged students. PISA contains an aggregate SES measure based on students' reports on their parents' educational attainment, occupational status and availability of economic and cultural resources in their home. Advantaged students are students in the top 25% of the country specific distribution of SES. Disadvantages students are students in the bottom 25% of the country specific distribution of SES. A negative number implies that advantaged have lower mean values than disadvantaged students. A positive value implies that advantaged have higher mean values than disadvantaged students. Bolded values indicate that the estimated d value is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is

statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Annex J Measuring differences in PSEC indicators across immigrant background
Table J1 – Differences in PSEC indicators across immigrant background by country

	Self-reported lack of perseverance		Self-reported lack of effort in Test		Difference in effort (PISA-test)		Non-response		Non-differentiation		Inconsistency		Performance decline		Response-time-effort (5 seconds)	
Australia	-0.17	***	-0.01		0.05		-0.04	*	-0.01		0.06	**	-0.13	***	-0.02	
Austria	-0.09		0.09		0.17	***	0.13	**	0.07		0.14	**	0.03		0.09	
Belgium	-0.27	***	0.13	**	0.09	*	0.26	***	0.01		0.34	***	0.18	***	0.21	***
Canada	-0.18	***	0.10	***	0.19	***	0.03		0.03		0.02		0.01		0.03	
Chile	-0.36		0.63	*	0.65		-0.06		0.19		-0.03		-0.13		1.68	
Czech Republic	-0.03		0.11		0.04		0.02		0.08		0.10		0.01		m	
Denmark	-0.20	***	0.26	***	0.20	***	0.23	***	0.14	***	0.18	***	0.13	***	0.20	***
Estonia	-0.28	***	0.03		0.05		0.13	*	0.05		0.31	***	0.05		0.01	
Finland	-0.12	**	0.05		-0.04		0.32	***	0.06		0.30	***	0.24	***	m	
France	-0.09		0.03		0.06		0.36	***	0.08	*	0.29	***	0.25	***	-0.03	
Germany	-0.07		0.06		0.11	*	0.23	***	0.07		0.19	***	0.17	***	0.09	
Greece	0.07		-0.07		-0.05		0.10	*	0.08		0.05		-0.01		m	
Hungary	-0.12		0.03		0.06		-0.02		0.02		-0.08		0.00		-0.10	
Iceland	-0.09		0.17		0.01		0.59	***	0.00		0.09		0.06		m	
Ireland	-0.09		-0.07		-0.08		0.10	*	0.07		0.07		0.00		0.03	
Israel	0.19	**	0.03		0.00		0.10		-0.06		-0.02		-0.01		0.16	
Italy	0.04		0.04		-0.07	*	0.19	***	0.02		0.25	***	0.02		0.01	
Japan	-0.24		0.16		-0.35		0.27		0.39		0.66	*	0.42	*	-0.06	***
Korea	0.64	***	-0.32		-1.27		-0.08	***	-0.39		-0.63		1.17	***	2.76	***
Latvia	-0.32	**	0.12		0.10		0.01		0.02		0.18	**	-0.03		m	
Luxembourg	0.09	**	0.03		-0.02		0.15	***	0.02		0.05		0.09	**	m	
Mexico	0.21		-0.07		-0.15	**	0.53	*	0.09		0.32	*	-0.22	*	m	
Netherlands	-0.25	***	0.20	**	0.19	***	0.15	**	0.05		0.33	***	0.18	***	m	
New Zealand	-0.33	***	-0.04		-0.06		0.02		0.04		0.06		-0.01		m	
Norway	-0.31	***	0.10		0.09		0.09		0.18	***	0.16	***	0.15	*	0.01	
Poland	0.10		0.34		0.43		0.03		-0.26		1.09	**	-0.20		2.48	
Portugal	0.03		0.10		0.02		0.22	***	-0.07		0.24	***	-0.05		0.03	
Slovak Republic	-0.15		-0.21		-0.22		0.22		-0.11		0.19		-0.04		-0.05	***
Slovenia	-0.07		0.03		0.10		0.07	*	0.03		0.00		-0.01		0.02	
Spain	0.10	**	0.06		-0.02		0.14	***	-0.04		0.12	***	0.03		-0.02	
Sweden	-0.28	***	0.07		0.02		0.28	***	0.15	***	0.17	***	0.26	***	-0.02	
Switzerland	-0.13	***	0.12	***	0.17	***	0.13	***	0.04		0.16	***	0.15	***	m	
Turkey	0.09		0.09		0.19		0.04		-0.02		-0.13		0.01		m	
United Kingdom	-0.27	***	0.01		0.00		0.17	***	0.11	*	0.09	*	0.02		m	
United States	0.00		0.05		0.07	**	0.02		-0.04		-0.06		0.05		-0.01	
Argentina	0.16		0.11		0.30	*	0.45		0.01		0.01		-0.13		m	
Albania	0.73	***		***		***	0.04		0.68		0.21		0.16		m	***
Brazil	0.13		-0.24	**	-0.22		0.02		-0.39	**	0.84	*	-0.32		1.15	*
Bulgaria	-0.27		-0.01		-0.40	***	0.52		-0.44		1.16		0.20		m	
Colombia	0.46	*	0.46		-0.13		3.16	*	-0.26		1.14	*	0.15		-0.08	***
Costa Rica	0.11		-0.06		-0.06		0.20		-0.03		-0.07		-0.06		m	
Croatia	0.09		0.08		0.09		0.03		0.04		0.06		0.03		m	
Hong Kong (China)	0.00		-0.05		-0.03		-0.03	**	-0.06		-0.05	*	0.01		-0.09	
Indonesia	-0.18		-0.03		0.22		-0.32	***	0.09		0.21		-0.08		m	
Jordan	-0.07		-0.04		-0.05		-0.10		-0.17	*	-0.07		0.12	*	m	
Kazakhstan	0.22	**	0.01		-0.04		0.02		-0.05		0.06		0.09		m	
Liechtenstein	0.07		0.20		0.10		0.16		0.01		0.31	*	0.18		m	
Lithuania	-0.03		0.10		-0.05		0.31	*	0.39	*	-0.06		-0.14		m	
Macao (China)	-0.09	**	-0.11	***	-0.06	*	-0.02		-0.04		-0.02		-0.08	***	-0.04	
Malaysia	0.04		0.12		-0.04		-0.18	***	-0.15		0.14		-0.10		m	
Montenegro	0.09		0.08		-0.01		0.02		0.05		0.19	*	0.09		m	
Peru	0.05		0.37		0.16		2.38	*	0.08		0.80	*	-0.11		m	
Qatar	-0.31	***	-0.08	**	0.31	***	-0.35	***	-0.43	***	-0.23	***	0.19	***	m	
Romania	0.77		-0.28		-0.05		0.16		0.13		0.04		-0.23		m	
Russia	0.02		0.00		0.04		0.10	*	0.11	*	0.06		0.00		0.05	
Serbia	0.04		-0.01		-0.06		0.06		0.03		0.01		-0.03		m	
Shanghai	-0.04		-0.03		-0.11		0.13	*	-0.09		0.21		0.49	***	0.08	
Singapore	-0.06		-0.01		0.01		0.04	*	0.08	*	-0.07	*	-0.08	**	-0.05	**
Chinese Taipei	0.08		0.06		-0.14		-0.01		-0.06		0.47	**	0.29		0.10	
Thailand	0.32	**	0.71		0.12		0.00		0.51	*	-0.28	**	-0.04		m	
Tunisia	0.47		0.33		-0.04		-0.48	***	-0.11		0.77		0.61	*	m	
United Arab Emirates	-0.16	***	0.13	***	0.23	***	0.02		-0.22	***	-0.22	***	0.03		-0.09	
Uruguay	-0.16		0.23		0.32		-0.12		0.01		0.07		-0.08		m	
Viet Nam	-1.45		0.22		0.20		-0.06		0.03		-0.09		0.08		m	
Average	-0.03		0.08	**	0.02		0.18	***	0.01		0.17	***	0.06	***	0.27	**

Notes: The immigrant gap refers to the difference between students with and students without an immigrant background. A negative number implies that immigrant students have lower mean values than students without an immigrant background. A positive value implies that immigrant students have higher mean values than students without an immigrant background.
* p<0.05, ** p<0.01, *** p<0.001

Table J1 – Differences in PSEC indicators across immigrant background by country compared to average

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.17**	-0.01	0.05	**-0.04**	-0.01	**0.06**	**-0.13**	-0.02
Austria	-0.09	0.09	**0.17**	**0.13**	0.07	**0.14**	0.03	0.09
Belgium	**-0.27**	**0.13**	**0.09**	**0.26**	0.01	**0.34**	**0.18**	**0.21**
Canada	**-0.18**	**0.1**	**0.19**	0.03	0.03	0.02	0.01	0.03
Chile	-0.36	**0.63**	0.65	-0.06	0.19	-0.03	-0.13	1.68
Czech Republic	-0.03	0.11	0.04	0.02	0.08	0.1	0.01	
Denmark	**-0.2**	**0.26**	**0.2**	**0.23**	**0.14**	**0.18**	**0.13**	**0.2**
Estonia	**-0.28**	0.03	0.05	**0.13**	0.05	**0.31**	0.05	0.01
Finland	**-0.12**	0.05	-0.04	**0.32**	0.06	**0.3**	**0.24**	
France	-0.09	0.03	0.06	**0.36**	**0.08**	**0.29**	**0.25**	-0.03
Germany	-0.07	0.06	**0.11**	**0.23**	0.07	**0.19**	**0.17**	0.09
Greece	0.07	-0.07	-0.05	**0.1**	0.08	0.05	-0.01	
Hungary	-0.12	0.03	0.06	-0.02	0.02	-0.08	0	**-0.1**
Iceland	-0.09	0.17	0.01	**0.59**	0	0.09	0.06	
Ireland	-0.09	-0.07	-0.08	**0.1**	0.07	0.07	0	0.03
Israel	**0.19**	0.03	0	0.1	-0.06	-0.02	-0.01	0.16
Italy	0.04	0.04	-0.07	**0.19**	0.02	**0.25**	0.02	0.01
Japan	-0.24	0.16	-0.35	0.27	0.39	**0.66**	**0.42**	**-0.06**
Korea	**0.64**	-0.32	-1.27	**-0.08**	-0.39	-0.63	**1.17**	**2.76**
Latvia	**-0.32**	0.12	0.1	0.01	0.02	**0.18**	-0.03	
Luxembourg	**0.09**	0.03	-0.02	**0.15**	0.02	0.05	**0.09**	
Mexico	0.21	-0.07	**-0.15**	**0.53**	0.09	**0.32**	**-0.22**	
Netherlands	**-0.25**	**0.2**	**0.19**	**0.15**	0.05	**0.33**	**0.18**	
New Zealand	**-0.33**	-0.04	-0.06	0.02	0.04	**0.06**	-0.01	
Norway	**-0.31**	0.1	0.09	0.09	**0.18**	**0.16**	**0.15**	0.01
Poland	0.1	0.34	0.43	0.03	-0.26	**1.09**	-0.2	2.48
Portugal	0.03	0.1	0.02	**0.22**	-0.07	**0.24**	-0.05	0.03
Slovak Republic	-0.15	-0.21	-0.22	0.22	-0.11	0.19	-0.04	**-0.05**
Slovenia	-0.07	0.03	0.1	**0.07**	0.03	0	-0.01	0.02
Spain	**0.1**	**0.06**	-0.02	**0.14**	**-0.04**	**0.12**	0.03	-0.02
Sweden	**-0.28**	0.07	0.02	**0.28**	**0.15**	**0.17**	**0.26**	-0.02
Switzerland	**-0.13**	**0.12**	**0.17**	**0.13**	**0.04**	**0.16**	**0.15**	
Turkey	0.09	0.09	0.19	0.04	-0.02	-0.13	0.01	
United Kingdom	**-0.27**	0.01	0	**0.17**	**0.11**	**0.09**	0.02	
United States	0	0.05	**0.07**	0.02	-0.04	-0.06	0.05	-0.01
Argentina	0.16	0.11	**0.3**	0.45	0.01	0.01	-0.13	
Albania	**0.73**			0.04	0.68	0.21	0.16	
Brazil	0.13	**-0.24**	-0.22	0.02	**-0.39**	**0.84**	-0.32	**1.15**
Bulgaria	-0.27	-0.01	**-0.4**	0.52	-0.44	1.16	0.2	
Colombia	**0.46**	0.46	-0.13	**3.16**	-0.26	**1.14**	0.15	**-0.08**
Costa Rica	0.11	-0.06	-0.06	**0.2**	-0.03	-0.07	-0.06	
Croatia	**0.09**	0.08	0.09	0.03	0.04	0.06	0.03	
Hong Kong (China)	0	-0.05	-0.03	**-0.03**	-0.06	-0.05	0.01	-0.09
Indonesia	-0.18	-0.03	0.22	**-0.32**	0.09	0.21	-0.08	
Jordan	-0.07	-0.04	-0.05	**-0.1**	**-0.17**	-0.07	**0.12**	
Kazakhstan	**0.22**	0.01	-0.04	0.02	-0.05	0.06	0.09	
Liechtenstein	0.07	0.2	0.1	**0.16**	0.01	**0.31**	0.18	
Lithuania	-0.03	0.1	-0.05	**0.31**	**0.39**	-0.06	-0.14	
Macao (China)	**-0.09**	**-0.11**	**-0.06**	**-0.02**	-0.04	-0.02	**-0.08**	-0.04
Malaysia	0.04	0.12	-0.04	**-0.18**	-0.15	0.14	-0.1	
Montenegro	0.09	0.08	-0.01	0.02	0.05	**0.19**	0.09	
Peru	0.05	0.37	0.16	**2.38**	0.08	**0.8**	-0.11	
Qatar	**-0.31**	**-0.08**	**0.31**	**-0.35**	**-0.43**	**-0.23**	**0.19**	
Romania	0.77	-0.28	-0.05	0.16	0.13	0.04	-0.23	
Russia	0.02	0	0.04	**0.1**	0.11	0.06	0	0.05
Serbia	0.04	-0.01	-0.06	0.06	0.03	0.01	-0.03	
Shanghai	-0.04	-0.03	-0.11	0.13	-0.09	0.21	**0.49**	0.08
Singapore	-0.06	-0.01	0.01	**0.04**	**0.08**	**-0.07**	**-0.08**	**-0.05**
Chinese Taipei	0.08	0.06	-0.14	-0.01	-0.06	**0.47**	0.29	0.1
Thailand	**0.32**	0.71	0.12	0	**0.51**	**-0.28**	-0.04	
Tunisia	0.47	0.33	-0.04	**-0.48**	-0.11	0.77	**0.61**	
United Arab Emirates	**-0.16**	**0.13**	**0.23**	0.02	**-0.22**	**-0.22**	0.03	-0.09
Uruguay	-0.16	0.23	0.32	-0.12	0.01	0.07	-0.08	
Viet Nam	-1.45	0.22	0.2	-0.06	0.03	-0.09	0.08	
AVG	-0.03	**0.08**	0.02	**0.18**	0.01	**0.17**	**0.06**	**0.27**

Notes: The immigrant gap refers to the difference between students with and students without an immigrant background. A negative number implies that immigrant students have lower mean values than students without an immigrant background. A positive value implies that immigrant students have higher mean values than students without an immigrant background. Bolded values indicate that the estimated d value is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated *d* value for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data

Annex K The association between PSEC indicators and academic achievement

Table K1 – Country-specific associations between PSEC indicators and math achievement

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.21**	**-0.21**	**-0.10**	**-0.34**	**-0.08**	**-0.12**	**-0.23**	**-0.19**
Austria	**-0.13**	**-0.06**	**-0.06**	**-0.30**	**-0.09**	**-0.12**	**-0.07**	**-0.08**
Belgium	**-0.12**	**-0.12**	**-0.09**	**-0.23**	**-0.08**	**-0.10**	**-0.22**	**-0.14**
Canada	**-0.15**	**-0.18**	**-0.11**	**-0.27**	**-0.07**	**-0.08**	**-0.19**	**-0.20**
Chile	**-0.07**	**-0.05**	-0.02	**-0.26**	-0.03	**-0.08**	-0.03	**-0.09**
Czech Republic	**-0.08**	**-0.14**	**-0.12**	**-0.28**	**-0.07**	**-0.13**	**-0.15**	
Denmark	**-0.22**	**-0.20**	**-0.15**	**-0.21**	0.00	**-0.15**	**-0.12**	**-0.15**
Estonia	-0.03	-0.01	**-0.10**	**-0.32**	**-0.05**	**-0.10**	**-0.11**	**-0.18**
Finland	**-0.29**	**-0.29**	**-0.22**	**-0.31**	**-0.10**	**-0.10**	**-0.18**	
France	**-0.17**	**-0.07**	**-0.08**	**-0.21**	-0.04	**-0.10**	**-0.19**	**-0.20**
Germany	**-0.14**	**-0.10**	**-0.14**	**-0.24**	**-0.09**	**-0.13**	**-0.20**	**-0.08**
Greece	**-0.19**	**-0.06**	-0.01	**-0.31**	**-0.10**	**-0.15**	**-0.02**	
Hungary	**-0.09**	**-0.09**	**-0.07**	**-0.29**	**-0.06**	**-0.12**	**-0.17**	**-0.13**
Iceland	**-0.26**	**-0.26**	**-0.16**	**-0.26**	**-0.10**	**-0.14**	**-0.18**	
Ireland	**-0.15**	**-0.15**	**-0.09**	**-0.49**	**-0.03**	**-0.11**	**-0.29**	**-0.21**
Israel	-0.01	**-0.07**	**-0.06**	**-0.18**	**-0.13**	**-0.18**	**-0.16**	**-0.09**
Italy	**-0.09**	**-0.22**	**-0.11**	**-0.31**	**-0.07**	**-0.16**	**-0.10**	-0.13
Japan	**-0.20**	-0.01	**0.05**	**-0.39**	**-0.18**	**-0.09**	**-0.38**	**-0.23**
Korea	**-0.27**	**-0.11**	0.02	**-0.76**	**-0.24**	-0.02	**-0.53**	**-0.26**
Latvia	**-0.11**	**-0.21**	-0.03	**-0.29**	-0.01	**-0.11**	**-0.04**	
Luxembourg	**-0.11**	**-0.09**	**-0.08**	**-0.18**	**-0.07**	**-0.13**	**-0.08**	
Mexico	**-0.10**	**-0.06**	0.02	**-0.21**	**-0.04**	**-0.09**	**-0.02**	
Netherlands	**-0.07**	**-0.13**	**-0.14**	**-0.21**	**-0.13**	**-0.15**	**-0.39**	
New Zealand	**-0.26**	**-0.23**	**-0.11**	**-0.30**	**-0.11**	**-0.10**	**-0.23**	
Norway	**-0.30**	**-0.24**	**-0.16**	**-0.19**	**-0.09**	**-0.16**	**-0.20**	**-0.09**
Poland	**-0.19**	-0.02	-0.03	**-0.25**	**-0.06**	**-0.12**	**-0.13**	**-0.13**
Portugal	**-0.13**	**-0.08**	**-0.06**	**-0.13**	**-0.06**	**-0.06**	**-0.07**	-0.06
Slovak Republic	**-0.17**	**-0.07**	**-0.03**	**-0.27**	**-0.09**	**-0.11**	**-0.06**	**-0.10**
Slovenia	**-0.08**	**-0.18**	**-0.10**	**-0.28**	**-0.14**	**-0.12**	**-0.18**	**-0.10**
Spain	**-0.13**	**-0.07**	**-0.04**	**-0.18**	**-0.04**	**-0.10**	**-0.10**	**-0.06**
Sweden	**-0.28**	**-0.21**	**-0.16**	**-0.23**	**-0.07**	**-0.14**	**-0.16**	**-0.12**
Switzerland	**-0.14**	**-0.14**	**-0.12**	**-0.28**	**-0.10**	**-0.13**	**-0.16**	
Turkey	**-0.09**	0.01	-0.01	**-0.24**	**-0.09**	**-0.11**	-0.03	
United Kingdom	**-0.20**	**-0.19**	**-0.11**	**-0.40**	**-0.06**	**-0.16**	**-0.28**	
United States	**-0.12**	**-0.16**	**-0.11**	**-0.20**	**-0.07**	**-0.12**	**-0.28**	**-0.20**
Argentina	**-0.09**	**-0.08**	0.02	**-0.11**	**-0.04**	**-0.05**	0.02	
Brazil	**-0.07**	-0.01	0.00	**-0.09**	**-0.05**	**-0.06**	**0.03**	-0.14
Bulgaria	**-0.07**	-0.03	-0.01	**-0.15**	**-0.11**	**-0.11**	**-0.03**	
Colombia	**-0.04**	-0.01	0.00	**-0.12**	**-0.05**	**-0.06**	-0.01	0.01
Costa Rica	**-0.04**	0.03	0.02	**-0.12**	**-0.07**	**-0.03**	**-0.03**	
Croatia	**-0.05**	**-0.10**	**-0.03**	**-0.50**	**-0.12**	**-0.11**	**-0.06**	
Hong Kong (China)	**-0.11**	**-0.16**	**-0.08**	**-0.23**	**-0.15**	**-0.05**	**-0.43**	**-0.19**
Indonesia	**-0.05**	-0.03	**0.09**	**-0.16**	**-0.06**	**-0.04**	**0.12**	
Jordan	**-0.13**	0.00	**0.05**	**-0.21**	**-0.04**	**-0.10**	**0.18**	
Kazakhstan	**-0.06**	-0.02	**0.07**	**-0.10**	**-0.05**	**-0.08**	0.02	
Liechtenstein	-0.03	-0.05	-0.07	-0.20	0.00	**-0.14**	**-0.37**	
Lithuania	**-0.09**	**-0.16**	**-0.07**	**-0.34**	**-0.09**	**-0.15**	0.01	
Macao (China)	**-0.13**	**-0.15**	**-0.07**	**-0.34**	**-0.05**	**-0.04**	**-0.19**	-0.08
Malaysia	**-0.09**	**-0.04**	**0.09**	**-0.15**	**-0.06**	**-0.10**	**0.26**	
Montenegro	**-0.13**	0.01	**0.03**	**-0.24**	**-0.12**	**-0.08**	**0.12**	
Peru	**-0.08**	-0.02	**0.05**	**-0.14**	**-0.05**	**-0.04**	0.00	
Qatar	**-0.18**	**-0.07**	**0.06**	**-0.10**	**-0.07**	**-0.09**	**0.03**	
Romania	**-0.08**	0.00	**0.05**	**-0.22**	**-0.09**	-0.03	0.03	
Russia	**-0.05**	-0.02	-0.01	**-0.39**	**-0.08**	**-0.12**	**-0.03**	**-0.06**
Serbia	**-0.07**	**-0.12**	**-0.10**	**-0.27**	**-0.13**	**-0.08**	**-0.14**	
Shanghai	**-0.10**	**-0.08**	**-0.06**	**-0.69**	-0.02	**-0.11**	**-0.55**	**-0.27**
Singapore	**-0.09**	**-0.13**	**-0.04**	**-0.44**	**-0.04**	**-0.18**	**-0.52**	**-0.17**
Chinese Taipei	**-0.22**	**-0.31**	**-0.06**	**-0.49**	**-0.21**	**-0.08**	**-0.59**	**-0.11**
Thailand	**-0.17**	**-0.18**	-0.04	**-0.35**	**-0.09**	-0.02	**0.04**	
Tunisia	**-0.09**	0.00	0.01	**-0.07**	**-0.03**	**-0.03**	-0.01	
United Arab Emirates	**-0.16**	**-0.06**	0.02	**-0.10**	**-0.05**	**-0.12**	-0.03	**-0.06**
Uruguay	**-0.07**	-0.03	0.00	**-0.15**	**-0.05**	**-0.08**	**-0.05**	
Viet Nam	**-0.07**	**-0.09**	0.06	**-0.14**	-0.03	**-0.05**	**0.53**	
Average	**-0.13**	**-0.10**	**-0.05**	**-0.26**	**-0.08**	**-0.10**	**-0.12**	**-0.13**

Notes: Bolded values indicate that the estimated coefficient is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated association for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated association for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Table K2 – Country-specific associations between PSEC indicators and reading achievement

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.20**	**-0.22**	**-0.10**	**-0.36**	**-0.12**	**-0.09**	**-0.24**	**-0.21**
Austria	**-0.12**	**-0.07**	**-0.06**	**-0.32**	**-0.12**	**-0.12**	**-0.09**	**-0.09**
Belgium	**-0.06**	**-0.12**	**-0.07**	**-0.25**	**-0.11**	**-0.08**	**-0.25**	**-0.16**
Canada	**-0.14**	**-0.18**	**-0.10**	**-0.29**	**-0.10**	**-0.08**	**-0.21**	**-0.22**
Chile	**-0.08**	**-0.05**	-0.01	**-0.26**	**-0.07**	**-0.07**	-0.01	-0.05
Czech Republic	**-0.08**	**-0.13**	**-0.10**	**-0.26**	**-0.08**	**-0.12**	**-0.16**	
Denmark	**-0.17**	**-0.20**	**-0.14**	**-0.25**	**-0.03**	**-0.13**	**-0.14**	**-0.16**
Estonia	-0.03	-0.01	**-0.08**	**-0.32**	**-0.06**	**-0.09**	**-0.14**	**-0.20**
Finland	**-0.24**	**-0.30**	**-0.23**	**-0.35**	**-0.12**	**-0.10**	**-0.19**	
France	**-0.08**	**-0.09**	**-0.08**	**-0.22**	**-0.08**	**-0.08**	**-0.21**	**-0.24**
Germany	**-0.10**	**-0.11**	**-0.13**	**-0.23**	**-0.11**	**-0.11**	**-0.20**	**-0.08**
Greece	**-0.12**	**-0.06**	0.00	**-0.36**	**-0.14**	**-0.16**	**-0.06**	
Hungary	**-0.06**	**-0.09**	**-0.05**	**-0.32**	**-0.09**	**-0.13**	**-0.20**	**-0.15**
Iceland	**-0.20**	**-0.28**	**-0.16**	**-0.28**	**-0.14**	**-0.13**	**-0.20**	
Ireland	**-0.15**	**-0.15**	**-0.08**	**-0.49**	**-0.06**	**-0.11**	**-0.29**	**-0.25**
Israel	-0.01	**-0.09**	**-0.06**	**-0.23**	**-0.17**	**-0.19**	**-0.21**	**-0.10**
Italy	**-0.06**	**-0.21**	**-0.10**	**-0.35**	**-0.10**	**-0.15**	**-0.11**	-0.14
Japan	**-0.19**	-0.02	**0.06**	**-0.40**	**-0.23**	**-0.09**	**-0.41**	**-0.28**
Korea	**-0.19**	**-0.11**	0.01	**-0.65**	**-0.23**	-0.02	**-0.47**	**-0.29**
Latvia	**-0.11**	**-0.19**	-0.01	**-0.24**	**-0.05**	**-0.08**	**-0.04**	
Luxembourg	**-0.06**	**-0.10**	**-0.08**	**-0.25**	**-0.09**	**-0.13**	**-0.11**	
Mexico	**-0.11**	**-0.08**	0.02	**-0.23**	**-0.06**	**-0.09**	**-0.03**	
Netherlands	**-0.05**	**-0.15**	**-0.14**	**-0.17**	**-0.15**	**-0.15**	**-0.41**	
New Zealand	**-0.20**	**-0.25**	**-0.10**	**-0.36**	**-0.15**	**-0.08**	**-0.26**	
Norway	**-0.26**	**-0.25**	**-0.17**	**-0.21**	**-0.14**	**-0.16**	**-0.23**	**-0.11**
Poland	**-0.13**	-0.02	-0.02	**-0.29**	**-0.10**	**-0.10**	**-0.13**	**-0.13**
Portugal	**-0.10**	**-0.07**	**-0.04**	**-0.08**	**-0.08**	**-0.04**	**-0.08**	-0.08
Slovak Republic	**-0.12**	**-0.08**	-0.01	**-0.31**	**-0.11**	**-0.12**	**-0.07**	**-0.15**
Slovenia	**-0.05**	**-0.18**	**-0.09**	**-0.32**	**-0.17**	**-0.12**	**-0.19**	**-0.11**
Spain	**-0.10**	**-0.09**	**-0.05**	**-0.21**	**-0.07**	**-0.10**	**-0.13**	**-0.06**
Sweden	**-0.23**	**-0.25**	**-0.19**	**-0.30**	**-0.10**	**-0.15**	**-0.19**	**-0.13**
Switzerland	**-0.09**	**-0.14**	**-0.11**	**-0.27**	**-0.11**	**-0.11**	**-0.15**	
Turkey	**-0.11**	-0.01	-0.01	**-0.25**	**-0.08**	**-0.09**	**-0.05**	
United Kingdom	**-0.16**	**-0.21**	**-0.11**	**-0.44**	**-0.09**	**-0.15**	**-0.30**	
United States	**-0.12**	**-0.16**	**-0.09**	**-0.23**	**-0.12**	**-0.09**	**-0.29**	**-0.22**
Argentina	**-0.05**	**-0.10**	0.01	**-0.16**	**-0.07**	**-0.08**	0.00	
Albania								
Brazil	**-0.06**	-0.02	0.01	**-0.11**	**-0.07**	**-0.07**	**0.03**	-0.22
Bulgaria	**-0.11**	-0.04	-0.01	**-0.19**	**-0.17**	**-0.14**	-0.02	
Colombia	**-0.08**	-0.04	0.00	**-0.15**	**-0.07**	**-0.07**	0.00	0.02
Costa Rica	-0.03	0.01	0.01	**-0.14**	**-0.11**	-0.03	**-0.04**	
Croatia	**-0.03**	**-0.09**	-0.02	**-0.50**	**-0.14**	**-0.10**	**-0.07**	
Hong Kong (China)	**-0.10**	**-0.15**	**-0.06**	**-0.22**	**-0.17**	-0.03	**-0.39**	**-0.17**
Indonesia	**-0.08**	-0.02	**0.08**	**-0.16**	**-0.07**	**-0.05**	**0.10**	
Jordan	**-0.14**	-0.02	**0.05**	**-0.25**	**-0.07**	**-0.13**	**0.20**	
Kazakhstan	**-0.07**	-0.01	**0.08**	**-0.12**	**-0.09**	**-0.08**	**0.03**	
Liechtenstein	0.00	-0.04	-0.05	-0.28	-0.01	**-0.16**	**-0.38**	
Lithuania	**-0.09**	**-0.16**	**-0.07**	**-0.34**	**-0.09**	**-0.14**	0.00	
Macao (China)	**-0.09**	**-0.13**	**-0.04**	**-0.27**	**-0.06**	-0.02	**-0.17**	**-0.11**
Malaysia	**-0.13**	**-0.06**	**0.08**	**-0.16**	**-0.07**	**-0.10**	**0.25**	
Montenegro	**-0.14**	0.00	**0.03**	**-0.29**	**-0.15**	**-0.06**	**0.09**	
Peru	**-0.07**	-0.02	0.04	**-0.17**	**-0.08**	**-0.05**	0.01	
Qatar	**-0.19**	**-0.08**	**0.06**	**-0.11**	**-0.09**	**-0.09**	**0.02**	
Romania	**-0.07**	-0.04	0.04	**-0.29**	**-0.12**	-0.03	**0.07**	
Russia	**-0.04**	-0.02	0.00	**-0.38**	**-0.13**	**-0.08**	**-0.06**	-0.05
Serbia	**-0.07**	**-0.14**	**-0.11**	**-0.26**	**-0.16**	**-0.08**	**-0.12**	
Shanghai	**-0.08**	**-0.06**	**-0.05**	**-0.47**	**-0.04**	**-0.08**	**-0.43**	**-0.24**
Singapore	**-0.05**	**-0.12**	**-0.03**	**-0.43**	**-0.07**	**-0.15**	**-0.49**	**-0.16**
Chinese Taipei	**-0.14**	**-0.24**	**-0.04**	**-0.39**	**-0.18**	**-0.06**	**-0.47**	**-0.12**
Thailand	**-0.20**	**-0.15**	0.00	**-0.25**	**-0.10**	-0.02	0.01	
Tunisia	-0.04	-0.02	0.01	**-0.09**	**-0.05**	**-0.04**	-0.01	
United Arab Emirates	**-0.18**	**-0.07**	0.03	**-0.13**	**-0.08**	**-0.12**	-0.01	**-0.06**
Uruguay	**-0.05**	**-0.05**	0.01	**-0.17**	**-0.07**	**-0.08**	**-0.09**	
Viet Nam	-0.04	-0.08	0.05	-0.16	-0.04	-0.03	0.50	
Average	**-0.11**	**-0.11**	**-0.04**	**-0.27**	**-0.10**	**-0.09**	**-0.13**	**-0.15**

Notes: Bolded values indicate that the estimated coefficient is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated association for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated association for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Table K3 – Country-specific associations between PSEC indicators and science achievement

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.22**	**-0.23**	**-0.10**	**-0.34**	**-0.12**	**-0.10**	**-0.25**	**-0.21**
Austria	**-0.08**	**-0.06**	**-0.06**	**-0.30**	**-0.09**	**-0.11**	**-0.09**	**-0.09**
Belgium	**-0.08**	**-0.12**	**-0.09**	**-0.28**	**-0.10**	**-0.09**	**-0.24**	**-0.15**
Canada	**-0.15**	**-0.19**	**-0.12**	**-0.28**	**-0.08**	**-0.11**	**-0.21**	**-0.24**
Chile	**-0.06**	**-0.05**	-0.01	**-0.25**	**-0.06**	**-0.08**	**-0.04**	**-0.08**
Czech Republic	**-0.08**	**-0.14**	**-0.12**	**-0.29**	**-0.10**	**-0.10**	**-0.14**	
Denmark	**-0.20**	**-0.23**	**-0.16**	**-0.27**	-0.03	**-0.15**	**-0.16**	**-0.16**
Estonia	**-0.04**	-0.01	**-0.09**	**-0.30**	**-0.06**	**-0.10**	**-0.11**	**-0.19**
Finland	**-0.29**	**-0.33**	**-0.26**	**-0.36**	**-0.13**	**-0.10**	**-0.20**	
France	**-0.11**	**-0.08**	**-0.08**	**-0.26**	**-0.06**	**-0.08**	**-0.20**	**-0.25**
Germany	**-0.11**	**-0.10**	**-0.14**	**-0.26**	**-0.11**	**-0.10**	**-0.20**	**-0.11**
Greece	**-0.15**	**-0.06**	-0.01	**-0.32**	**-0.13**	**-0.14**	**-0.04**	
Hungary	**-0.09**	**-0.08**	**-0.05**	**-0.31**	**-0.07**	**-0.11**	**-0.18**	**-0.16**
Iceland	**-0.24**	**-0.27**	**-0.17**	**-0.28**	**-0.13**	**-0.13**	**-0.21**	
Ireland	**-0.16**	**-0.17**	**-0.10**	**-0.54**	**-0.06**	**-0.11**	**-0.31**	**-0.25**
Israel	**-0.03**	**-0.08**	**-0.07**	**-0.20**	**-0.15**	**-0.18**	**-0.20**	**-0.09**
Italy	**-0.08**	**-0.21**	**-0.10**	**-0.31**	**-0.09**	**-0.14**	**-0.11**	-0.11
Japan	**-0.16**	-0.03	**0.05**	**-0.41**	**-0.21**	**-0.09**	**-0.41**	**-0.26**
Korea	**-0.22**	**-0.10**	0.01	**-0.66**	**-0.23**	0.00	**-0.46**	**-0.26**
Latvia	**-0.11**	**-0.19**	-0.02	**-0.30**	**-0.05**	**-0.09**	**-0.06**	
Luxembourg	**-0.09**	**-0.11**	**-0.09**	**-0.23**	**-0.08**	**-0.14**	**-0.10**	
Mexico	**-0.08**	**-0.05**	**0.04**	**-0.20**	**-0.05**	**-0.08**	**-0.04**	
Netherlands	**-0.08**	**-0.14**	**-0.14**	**-0.18**	**-0.15**	**-0.13**	**-0.41**	
New Zealand	**-0.25**	**-0.25**	**-0.11**	**-0.35**	**-0.14**	**-0.10**	**-0.25**	
Norway	**-0.27**	**-0.25**	**-0.17**	**-0.19**	**-0.14**	**-0.14**	**-0.23**	**-0.10**
Poland	**-0.13**	-0.02	-0.03	**-0.23**	**-0.09**	**-0.08**	**-0.12**	**-0.15**
Portugal	**-0.11**	**-0.07**	**-0.04**	**-0.10**	**-0.06**	**-0.07**	**-0.08**	-0.09
Slovak Republic	**-0.14**	**-0.08**	-0.02	**-0.27**	**-0.11**	**-0.10**	**-0.07**	**-0.12**
Slovenia	**-0.08**	**-0.19**	**-0.10**	**-0.30**	**-0.16**	**-0.12**	**-0.20**	**-0.11**
Spain	**-0.11**	**-0.08**	**-0.05**	**-0.19**	**-0.06**	**-0.09**	**-0.11**	**-0.06**
Sweden	**-0.24**	**-0.24**	**-0.18**	**-0.26**	**-0.11**	**-0.14**	**-0.18**	**-0.14**
Switzerland	**-0.08**	**-0.13**	**-0.12**	**-0.29**	**-0.11**	**-0.10**	**-0.15**	
Turkey	**-0.07**	0.00	-0.01	**-0.22**	**-0.09**	**-0.09**	**-0.04**	
United Kingdom	**-0.19**	**-0.21**	**-0.11**	**-0.45**	**-0.09**	**-0.16**	**-0.31**	
United States	**-0.10**	**-0.17**	**-0.11**	**-0.24**	**-0.10**	**-0.12**	**-0.30**	**-0.20**
Argentina	**-0.09**	**-0.11**	0.00	**-0.15**	**-0.06**	**-0.06**	0.01	
Albania								
Brazil	**-0.07**	0.00	0.01	**-0.11**	**-0.05**	**-0.06**	**0.03**	-0.21
Bulgaria	**-0.07**	-0.04	-0.01	**-0.18**	**-0.15**	**-0.11**	**-0.04**	
Colombia	**-0.05**	-0.01	0.01	**-0.13**	**-0.04**	**-0.05**	-0.01	0.01
Costa Rica	-0.03	0.02	0.01	**-0.17**	**-0.11**	0.00	**-0.05**	
Croatia	**-0.06**	**-0.10**	**-0.03**	**-0.49**	**-0.13**	**-0.10**	**-0.06**	
Hong Kong (China)	**-0.12**	**-0.15**	**-0.07**	**-0.15**	**-0.16**	**-0.03**	**-0.38**	**-0.17**
Indonesia	**-0.09**	-0.03	**0.09**	**-0.14**	**-0.07**	**-0.04**	**0.10**	
Jordan	**-0.15**	-0.02	**0.05**	**-0.23**	**-0.05**	**-0.10**	**0.18**	
Kazakhstan	**-0.10**	-0.01	**0.08**	**-0.16**	**-0.08**	**-0.07**	0.01	
Liechtenstein	0.00	-0.02	-0.08	-0.18	-0.01	**-0.11**	**-0.34**	
Lithuania	**-0.11**	**-0.17**	**-0.07**	**-0.36**	**-0.10**	**-0.13**	0.01	
Macao (China)	**-0.10**	**-0.13**	**-0.05**	**-0.25**	**-0.06**	-0.02	**-0.17**	-0.09
Malaysia	**-0.12**	**-0.06**	**0.08**	**-0.19**	**-0.07**	**-0.09**	**0.24**	
Montenegro	**-0.12**	0.00	0.02	**-0.26**	**-0.14**	**-0.07**	**0.11**	
Peru	**-0.05**	-0.01	**0.05**	**-0.17**	**-0.06**	-0.02	-0.02	
Qatar	**-0.18**	**-0.08**	**0.07**	**-0.12**	**-0.09**	**-0.09**	**0.02**	
Romania	**-0.05**	-0.01	0.03	**-0.23**	**-0.11**	-0.02	**0.04**	
Russia	-0.03	-0.02	-0.01	**-0.35**	**-0.11**	**-0.09**	**-0.04**	**-0.05**
Serbia	**-0.07**	**-0.12**	**-0.10**	**-0.26**	**-0.13**	**-0.07**	**-0.13**	
Shanghai	**-0.10**	**-0.07**	**-0.05**	**-0.56**	**-0.04**	**-0.09**	**-0.45**	**-0.24**
Singapore	**-0.09**	**-0.12**	-0.03	**-0.45**	**-0.07**	**-0.15**	**-0.52**	**-0.17**
Chinese Taipei	**-0.14**	**-0.23**	**-0.04**	**-0.39**	**-0.18**	**-0.05**	**-0.44**	**-0.09**
Thailand	**-0.18**	**-0.16**	-0.01	**-0.29**	**-0.10**	-0.01	0.02	
Tunisia	**-0.08**	-0.01	0.01	**-0.11**	**-0.04**	**-0.05**	-0.02	
United Arab Emirates	**-0.16**	**-0.07**	**0.03**	**-0.15**	**-0.07**	**-0.12**	-0.02	**-0.08**
Uruguay	**-0.08**	**-0.05**	0.00	**-0.20**	**-0.06**	**-0.07**	**-0.10**	
Viet Nam	-0.03	-0.08	0.05	-0.12	-0.04	-0.03	0.52	
Average	**-0.11**	**-0.10**	**-0.05**	**-0.27**	**-0.09**	**-0.09**	**-0.12**	**-0.15**

Notes: Bolded values indicate that the estimated coefficient is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated association for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated association for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Table K4 – Country-specific associations between PSEC indicators and problem-solving

	Self-reported lack of perseverance	Self-reported lack of effort	Difference in effort (PISA-test)	Non-response	Non-differentiation	Inconsistency	Performance decline	Response-time-effort (5 seconds)
Australia	**-0.18**	**-0.21**	**-0.10**	**-0.33**	**-0.11**	**-0.10**	**-0.22**	**-0.27**
Austria	**-0.08**	**-0.07**	**-0.06**	**-0.29**	**-0.12**	**-0.12**	**-0.08**	**-0.15**
Belgium	**-0.11**	**-0.13**	**-0.08**	**-0.16**	**-0.06**	**-0.13**	**-0.20**	**-0.18**
Canada	**-0.16**	**-0.18**	**-0.11**	**-0.27**	**-0.08**	**-0.09**	**-0.18**	**-0.31**
Chile	**-0.08**	**-0.04**	0.00	**-0.25**	**-0.07**	**-0.11**	-0.01	-0.09
Czech Republic	**-0.05**	**-0.14**	**-0.11**	**-0.30**	**-0.08**	**-0.12**	**-0.13**	
Denmark	**-0.14**	**-0.19**	**-0.15**	**-0.22**	-0.02	**-0.15**	**-0.12**	**-0.26**
Estonia	0.00	-0.03	**-0.10**	**-0.35**	**-0.09**	**-0.13**	**-0.11**	**-0.27**
Finland	**-0.26**	**-0.30**	**-0.23**	**-0.33**	**-0.12**	**-0.10**	**-0.18**	
France	**-0.12**	**-0.06**	**-0.09**	**-0.21**	-0.03	**-0.10**	**-0.15**	**-0.25**
Germany	**-0.08**	**-0.10**	**-0.12**	**-0.24**	**-0.12**	**-0.13**	**-0.18**	**-0.14**
Greece								
Hungary	**-0.07**	**-0.10**	**-0.04**	**-0.27**	**-0.11**	**-0.13**	**-0.17**	**-0.20**
Iceland								
Ireland	**-0.19**	**-0.13**	**-0.07**	**-0.48**	**-0.06**	**-0.09**	**-0.28**	**-0.29**
Israel	0.00	**-0.08**	**-0.08**	**-0.21**	**-0.21**	**-0.21**	**-0.19**	**-0.15**
Italy	**0.05**	**-0.15**	-0.06	**-0.36**	**-0.10**	**-0.09**	**-0.06**	**-0.29**
Japan	**-0.11**	0.01	**0.05**	**-0.32**	**-0.16**	**-0.08**	**-0.29**	**-0.30**
Korea	**-0.15**	**-0.10**	0.02	**-0.64**	**-0.20**	-0.01	**-0.44**	**-0.28**
Latvia								
Luxembourg								
Mexico								
Netherlands	**-0.07**	**-0.13**	**-0.16**	**-0.19**	**-0.16**	**-0.16**	**-0.39**	
New Zealand								
Norway	**-0.21**	**-0.23**	**-0.16**	**-0.11**	**-0.13**	**-0.15**	**-0.20**	**-0.18**
Poland	**-0.15**	-0.01	-0.01	**-0.35**	**-0.08**	**-0.13**	**-0.13**	**-0.35**
Portugal	**-0.10**	**-0.05**	**-0.04**	**-0.07**	**-0.05**	**-0.04**	**-0.05**	-0.08
Slovak Republic	**-0.07**	**-0.05**	-0.01	**-0.26**	**-0.09**	**-0.09**	**-0.05**	**-0.17**
Slovenia	**-0.06**	**-0.18**	**-0.10**	**-0.30**	**-0.15**	**-0.10**	**-0.18**	**-0.16**
Spain	**-0.11**	**-0.06**	**-0.04**	**-0.19**	**-0.07**	**-0.11**	**-0.10**	**-0.12**
Sweden	**-0.25**	**-0.22**	**-0.17**	**-0.19**	**-0.12**	**-0.13**	**-0.15**	**-0.19**
Switzerland								
Turkey	**-0.07**	0.01	0.00	**-0.18**	**-0.06**	**-0.09**	**-0.04**	
United Kingdom	**-0.18**	**-0.19**	**-0.10**	**-0.39**	**-0.09**	**-0.17**	**-0.28**	
United States	**-0.15**	**-0.17**	**-0.12**	**-0.23**	**-0.08**	**-0.13**	**-0.28**	**-0.31**
Argentina								
Albania								
Brazil	**-0.13**	0.01	0.02	**-0.07**	**-0.05**	**-0.06**	0.04	**-0.25**
Bulgaria	**-0.07**	-0.04	-0.02	**-0.18**	**-0.16**	**-0.14**	-0.03	
Colombia	-0.04	-0.03	-0.02	**-0.08**	**-0.06**	**-0.07**	**0.03**	-0.07
Costa Rica								
Croatia	**-0.05**	**-0.10**	**-0.03**	**-0.56**	**-0.16**	**-0.09**	**-0.05**	
Hong Kong (China)	-0.02	**-0.11**	**-0.06**	-0.12	**-0.16**	0.00	**-0.35**	**-0.19**
Indonesia								
Jordan								
Kazakhstan								
Liechtenstein								
Lithuania								
Macao (China)	**-0.08**	**-0.11**	**-0.05**	**-0.25**	**-0.05**	**-0.04**	**-0.13**	-0.09
Malaysia	**-0.09**	-0.03	**0.10**	**-0.17**	**-0.06**	**-0.10**	**0.25**	
Montenegro	**-0.10**	0.02	**0.04**	**-0.21**	**-0.15**	**-0.07**	**0.13**	
Peru								
Qatar								
Romania								
Russia	-0.03	-0.03	-0.01	**-0.29**	**-0.08**	**-0.08**	**-0.04**	**-0.12**
Serbia	**-0.05**	**-0.08**	**-0.06**	**-0.28**	**-0.15**	**-0.05**	**-0.10**	
Shanghai	-0.03	**-0.04**	-0.03	**-0.56**	**-0.04**	**-0.07**	**-0.42**	**-0.24**
Singapore	**-0.09**	**-0.10**	-0.02	**-0.36**	**-0.08**	**-0.13**	**-0.39**	**-0.21**
Chinese Taipei	**-0.07**	**-0.23**	**-0.05**	**-0.39**	**-0.19**	**-0.04**	**-0.43**	**-0.13**
Thailand								
Tunisia								
United Arab Emirates	**-0.16**	**-0.05**	0.03	**-0.09**	**-0.08**	**-0.15**	**-0.04**	**-0.13**
Uruguay	**-0.06**	-0.01	0.01	**-0.15**	**-0.09**	**-0.08**	**-0.05**	
Viet Nam								
Average	**-0.10**	**-0.10**	**-0.06**	**-0.27**	**-0.10**	**-0.10**	**-0.15**	**-0.20**

Notes: Bolded values indicate that the estimated coefficient is statistically significant at least at the 5% level. A dark grey shade indicates that the estimated association for a specific country is statistically significantly above than the average for that indicator across the countries with available information ($p \leq 0.05$). A light grey shade indicates that the estimated association for a specific indicator and for a specific country is statistically significantly below than the average for that indicator across the countries with available information ($p \leq 0.05$). No shading indicates that the null of no difference between the average indicator across countries in the sample and a specific country mean cannot be rejected at the 5% level. Source: PISA 2012 data.

Annex L – Description of the CCC questionnaire

Countries participating in PISA 2000 had the option of administering a Cross-Curricular Competences questionnaire (CCC). The questionnaire included two large sets of items, although sub-groups of items within the sets were intended to capture specific constructs. The first set asked students how often certain statements applied to them. Possible answers were "almost never", "sometimes", "often" and "almost always". The statements were:

1) When I study, I try to memorise everything that might be covered.
2) I'm certain I can understand the most difficult material presented in texts.
3) When I study, I start by figuring out exactly what I need to learn.
4) When I sit myself down to learn something really difficult, I can learn it.
5) When I study, I memorise as much as possible.
6) I study to increase my job opportunities.
7) When studying, I work as hard as possible.
8) I'm confident I can understand the most complex material presented by the teacher.
9) When I study, I try to relate new material to things I have learned in other subjects.
10) When I study, I memorise all new material so that I can recite it.
11) If I decide not to get any bad grades, I can really do it.
12) When studying, I keep working even if the material is difficult.
13) When I study, I force myself to check to see if I remember what I have learned.
14) I study to ensure that my future will be financially secure.
15) When I study, I practise by saying the material to myself over and over.
16) If I decide not to get any problems wrong, I can really do it.
17) When I study, I figure out how the information might be useful in the real world.
18) I'm confident I can do an excellent job on
19) When I study, I try to figure out which concepts I still haven't really understood.
20) When studying, I try to do my best to acquire the knowledge and skills taught
21) When I study, I try to understand the material better by relating it to things I already know.
22) I study to get a good job.
23) When I study, I make sure that I remember the most important things.
24) If I want to learn something well, I can.
25) When I study, I figure out how the material fits in with what I have already learned
26) I'm certain I can master the skills being taught.
27) When I study, and I don't understand something I look for additional information to clarify this.
28) When studying, I put forth my best effort.

The second item set asked students whether they "strongly disagreed", "disagreed", "agreed" or "strongly agreed" with the following statements:

1) When I do mathematics, I sometimes get totally absorbed.
2) I like to work with other students.
3) I learn things quickly in most school subjects.
4) I like to try to be better than other students.
5) I'm hopeless in <test language> classes.
6) Because reading is fun, I wouldn't want to give it up.
7) I'm good at most school subjects.
8) I learn most when I work with other students.

[i] PISA 2012 had a total of 64 participating countries: Albania, Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Chinese Taipei, Colombia, Costa Rica, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao (China), Malaysia, Mexico, Montenegro, Netherlands, New

Zealand, Norway, Peru, Poland, Portugal, Qatar, Romania, Russia, Serbia, Shanghai, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Thailand, Tunisia, Turkey, United Arab Emirates, United Kingdom, United States, Uruguay and Vietnam.

[ii] Each PISA round includes a core student questionnaire, as well as a series of optional questionnaires designed to gather for a restricted number of interested countries, additional information. We construct two sets of questionnaire-based behavioural indicators. The first is based solely on responses to the core questionnaires so that we have valid measures for all PISA-participating countries in each round. We present these results in the manscript. In addition, we developed all indicators using both the core questionnaire as well as the optional Information and Computer Technology (ICT) questionnaires in order to maximize the observations used to construct the indices for each individual (for the restricted set of countries that administered also the optional questionnaires). Results are in line with those presented and are available from the authors upon request.

[iii] In some test items respondents could obtain a partial credit, for example when the final response provided was incorrect (because of a typo or small calculation mistake) but the respondent correctly followed the procedure to solve an item. For simplicity and in line with most research in this area we coded these answers as 1 (correct).

[iv] In PISA 2000, the PISA index of Economic Social and Cultural Status (ESCS) had not yet been developed, so instead, we use three indicators that were available at the time – the one of parental highest educational attainment, highest occupational status and cultural possessions at home. These three are almost identical to the three indicators that are used to construct the ESCS index (only the indicator of home possessions differs slightly).

[v] The standardization of PSEC indicators is based on the pooled SD of each indicator based on 2012 data.

[vi] Question 83 of the questionnaire asked students whether they "strongly agree", "agree", "disagree" or "strongly disagree" with the following statements: "My teacher lets us know we need to work hard."; "My teacher provides extra help when needed."; "My teacher helps students with their learning."; "My teacher gives students the opportunity to express opinions."

[vii] Question 86 of the questionnaire asked students whether they "strongly agree", "agree", "disagree" or "strongly disagree" with the following statements: "Students get along well with most teachers."; "Most teachers are interested in students' well-being."; "Most of my teachers really listen to what I have to say."; "If I need extra help, I will receive it from my teachers."; "Most of my teachers treat me fairly.".

[viii] Question 89 of the questionnaire asked students whether they "strongly agree", "agree", "disagree" or "strongly disagree" with the following statements: "Trying hard at school will help me get a good job."; "Trying hard at school will help me get into a good college."; "I enjoy receiving good grades."; "Trying hard at school is important.".