# Performance Decline in a Low-Stakes Test at Age 15 and Educational Attainment at Age 25: Cross-Country Longitudinal Evidence

F. Borgonovi[1,3*]

A. Ferrara[2]

M. Piacentini[3]

1. Social Research Institute, Institute of Education, University College London

f.borgonovi@ucl.ac.uk

2. European University Institute

Alessandro.Ferrara@eui.eu

3. Directorate for Education and Skills

Organisation for Economic Cooperation and Development

Mario.Piacentini@oecd.org

*Correspondence concerning this article should be addressed to:

**Francesca Borgonovi**

Social Research Institute, University College London,

55-59 Gordon Square

London WC1H 0NU

UNITED KINGDOM

Phone: +44 – (0)20 7612 6453

f.borgonovi@ucl.ac.uk

Performance Decline in a Low-Stakes Test at Age 15 and Educational Attainment at Age 25: Cross-Country Longitudinal Evidence


## Abstract

### Introduction

Educational attainment is associated with important life outcomes including labour market performance, health status, well-being, civic and political participation. An important question is whether it is possible to identify early those students who lack the achievement motivation that is needed to complete a higher education degree.

### Methods

Longitudinal follow-ups of representative samples of participants in the 2000 and 2003 Programme for International Student Assessment (PISA) from Australia, Denmark and Switzerland (N=3,110; 1,130; and 1,962; age = 15 to 27; % females 51%, 51%, 49%; ethnicity/race unknown) were used to identify the association between a measure of effort on a cognitively demanding low-stake task at age 15 – performance decline during the test – and educational attainment at age 25-27.

### Results

A one SD difference in performance decline was associated with a 5-6 percentage point difference in the probability of obtaining tertiary-level qualifications ($r$ = -0.15 in Australia; -0.11 in Denmark and -0.11 in Switzerland). We find no evidence of differences in this

relationship across genders, socio-economic status and baseline levels of ability in the three countries. The association between performance decline and educational attainment is homogeneous across these groups. Self-reported measures of achievement motivation were not predictive of educational attainment in the three countries.

**Conclusions**

Our work contributes new longitudinal evidence to the body of research in education employing behavioural measures of motivation and engagement. It can be used to understand the potential long-term consequences of disparities in students' preparation to sustain effort over cognitively demanding tasks.

## Introduction

A growing body of empirical research documents that psychological constructs measured in childhood predict later-life outcomes (Almlund, Duckworth, Heckman, and Kautz, 2011; Gutman and Schoon, 2016; Hitt, 2015; Hitt et al., 2016; Roberts, Kuncel, Shiner, Caspi, and Goldberg, 2007, Soland et al., 2019). This paper investigates whether adolescents' capacity to sustain effort in academic tasks – a capacity that reflects not only cognitive skills but also motivation - can predict their odds of completing higher education. The hypothesis is that students' level of engagement on low-stake assignments at school is indicative of stable motivational traits that influence life outcomes beyond school.

In order to test this hypothesis, we exploit longitudinal follow-ups of participants in the Programme for International Student Assessment (PISA) conducted in Australia, Denmark and Switzerland. We investigate the association between the decline in performance over the course

of the PISA test at age 15 and completion of upper-secondary school and tertiary level education. Declines in performance during the PISA test are related to either motivation or the cognitive capacity to sustain concentration over an extended period (Borgonovi and Biecek, 2016; Zamarro et al., 2019). We find that performance decline at age 15 predicts completion of a college degree by age 25, while the association with upper secondary school completion is weaker and less precisely estimated. We do not find evidence of performance decline mediating or moderating the associations between students' gender, parental education and achievement in the first part of the PISA test and their later life outcomes.

## Background

### Determinants of students' effort on academic tasks

Individuals' educational attainment depends not only on their academic knowledge and cognitive skills, but also on their capacity to sustain effort and attention in cognitively demanding tasks for sustained periods of time and with low levels of external supervision (Kankaraš, 2017).

The capacity to keep a stable level of performance over long and demanding academic tasks is determined by cognitive, socio-emotional and motivational skills. At the cognitive level, students with higher general intelligence and more opportunities to learn in a domain will find completing a task less challenging and tiring (Kuncel et al. 2004, Borgonovi and Biecek, 2016). Differences in the capacity to sustain attention and to endure fatigue are other cognitive characteristics that influence the capacity to keep a stable level of performance over a long cognitive task. Individuals differ in their capacity to prevent and manage the mental fatigue that arises during extended cognitive tasks (Mezzacappa, 2004; Slimani et al., 2018). For example, Sievertsen et al. (2016) showed that students who took a standardized test in Denmark later in

the day performed worse than students who completed the test earlier and were thus less fatigued.

Diverse motivational constructs interact with cognitive functions in determining how likely individuals are to avoid declines in performance during academic work. Many studies have found that students' motivational beliefs are significantly related to their academic achievement (e.g., Steinmayr et al., 2019; Linnenbrink-Garcia et al., 2018; Huang, 2011; Ryan and Deci, 2000; Vecchione, Alessandri and Marsicano, 2014). The social-cognitive approach to motivation (Pintrich et al., 1993; Wigfield and Cambria, 2010) and expectancy value theory (Wigfield and Eccles, 2000) emphasize the role of performance expectancies as well as task values and goals for achievement. Self-determination theory further proposes that sustained performance on a task results from actions motivated by intrinsic interests or by extrinsic values that have become integrated and internalized (Deci and Ryan, 1985; Ryan and Deci, 2000). Empirically, after controlling for differences in students' intelligence and prior academic achievements, expectancy components (ability self-concept, self-efficacy) were the best predictors of achievement in different domains followed by task values (i.e., enjoyment and perceived utility), need for achievement, and learning goals (Steinmayr and Spinath, 2009; Kriegbaum et al., 2015). According to the social cognitive perspective, students' motivation to do well on a task is largely context specific (Pintrich et al., 1993). Context specificity implies that students' effort on a test in a given domain does not necessarily predict their behavior on a test in another domain, or how much effort these students will spend later in their studies. By contrast, personality theories, such as McClelland's motivational theory (McClelland et al., 1953), conceptualize students' motivation as a relatively stable and general trait.

Conscientiousness is the personality traits that is more closely related to individuals' motivation to engage effort on academic work, and implies a desire to do a task well, and to take obligations seriously. Conscientiousness predicts success more strongly than the other traits in the Big Five model (Barrick and Mount, 1991; Chamorro-Premuzic and Furnham, 2003; Egan et al, 2017, Xu et al., 2021). Persistence is one of the main facets of conscientiousness. It represents the ability to maintain an action or complete a task regardless of a person's inclination towards the task (Duckworth et al., 2007). Individuals who are capable of completing demanding tasks are characterized by both distinct personality traits, such as persistence, and more task-specific motivational drivers, such as self-efficacy and valence (Schunk and DiBenedetto, 2019). Personality traits and task-specific drivers of motivation are both strongly related to the level of effort students spend on academic work (Usher et al., 2019).

**Measuring capacity to sustain effort: from self-reports to performance decline**

Typically, students' capacity to exert effort has been measured through self-reports. Existing self-reported measures of motivational constructs are associated with later life outcomes (Soto, 2019; Soto 2021; Wilmot and Ones, 2019). However, there are concerns about possible issues surrounding such measures (Duckworth and Jaeger, 2015) such as reference group bias (Kankaraš, 2017) and social desirability bias (Krosnick, et al., 1996).

Concerns over the measurement properties of self-reports have led to the development of measurements that can supplement (or even replace) self-reports (Soland et al., 2019). Alternate instruments have been employed to assess students' achievement motives (Eisenberger and Leonard, 1980; Ventura and Shute, 2013). However, administering ad hoc instruments in large-

scale settings is both impractical and costly. By contrast, recent work has pioneered the use of measures derived from observing and coding the behaviours of individuals when they participate in low-stakes standardised assessments (Soland et al., 2019). Proponents of these indicators maintain that completing long, cognitively challenging tests that are low stakes requires the capacity to stay focused and avoid distractors and, crucially, that the low-stakes nature requires high levels of intrinsic motivation (Kyllonen and Kell, 2018). The question explored in this paper is whether tracking students' level of effort on a low stake task can provide useful information to predict their capacity to succeed in their studies.

There have been some efforts devoted to investigating the predictive validity of behavioural measures of socio-emotional and motivational factors in the United States (Borghans and Schils, 2012; Zamarro et al., 2019). In particular, prior research indicates that measures constructed using careless responding to questionnaires, for example the extent to which individuals skip responding to questionnaire items (Hitt et al., 2016) or the extent to which they provide inconsistent and unpredictable responses (Hitt, 2015; Soland et al., 2018) at a young age predict educational attainment and labour market outcomes net of background characteristics, self-reported measures and cognitive abilities (Cheng, 2015; Hitt et al., 2016, Soland et al. 2019; Zamarro et al. 2018).

## The current study

We explore the association between performance decline in the PISA test at age 15 and the likelihood of completing upper-secondary and tertiary education by age 25 using ex-post standardized longitudinal follow-ups of PISA participants in Australia, Denmark, and Switzerland. Furthermore, we examine potential differences in the extent to which performance

decline is associated with educational attainment across gender, socio-economic status, and baseline levels of ability.

This study is the first to examine the predictive validity of performance decline as an indicator of individuals' capacity to sustain effort. Another contribution of the study consists in inspecting whether performance decline is more predictive of educational outcomes for some groups than for others. During the teenage years, females tend to report higher levels of motivation to do their best in low-stake tests (DeMars, Bashkov, and Socha, 2013) and to have lower performance decline than male students (Balart and Oosterveen, 2019; Borgonovi and Biecek, 2016). Females' higher conscientiousness and intrinsic motivation are considered to play a role in explaining why females have higher educational attainment than males (De Bolle et al., 2015; Duckworth and Seligman, 2006; Matthews, Ponitz, and Morrison, 2009). However, there is no prior evidence suggesting that socio-emotional and motivational constructs measured at school age shape long-term outcomes differently among males and females. Consequently, we hypothesise that gender differences in performance decline at age 15 explain gender differences in educational attainment in the three countries but do not have prior hypotheses on the extent to which the association between performance decline and educational attainment depends on gender.

Socio-economic differences in educational attainment are large (Blossfeld and Shavit, 1993; Barone and Ruggera, 2018). Moreover, socio-economically advantaged students have higher levels of conscientiousness than their disadvantaged peers (Hanushek, Welch, Machin and Woessmann, 2011; Liu, 2020). We hypothesise that socio-economic differences in performance decline at age 15 explain socio-economic differences in educational attainment in the three countries. According to the theory of compensatory advantage, upper-class families are able to

compensate for obstacles in their children's educational career, so that their educational outcomes dependent on cognitive or socio-emotional and motivational skills (Bernardi, 2014). Therefore, we also hypothesise that socio-economic status is a moderator of the association between performance decline and educational attainment. We also hypothesise that the capacity to sustain effort during a task is especially important among individuals with low baseline levels of achievement, because these individuals have lower self-efficacy and feelings of mastery, and generally experience coursework and exams as more cognitively challenging (Bandura, 1977; Binder, 1996; Montanello and Martens, 2005).

Our sample allows us to test hypotheses in three systems characterised by different types of upper-secondary education and rates of participation in tertiary education. Vocational education and training (VET) is highly developed in Switzerland and Denmark although the two systems differ: Switzerland adopts a traditional apprenticeship system, Denmark relies more on school-based vocational training (Buchman and Kriesi, 2011). In systems with strong VET, few students fail to complete compulsory schooling (lower-secondary education) and many complete their education in vocational upper-secondary rather than in university (Brzinsky-Fay, 2007; Gangl, Müler and Raffe, 2003). By contrast, in Australia upper secondary education is more academically oriented and less diversified (Buchman and Kriesi, 2011). Table 1 indicates that many students in Australia continue their studies beyond upper secondary school and graduate with tertiary-level qualifications (by young adulthood, 42% of 15-year-old students did while in Denmark and Switzerland the corresponding figures were 29% and 18%).

PISA is a low-stake test for students, but it is high-stakes for policymakers in several countries (Zamarro et al., 2019). Countries like Singapore attach a great importance to PISA scores, and consider them as indicators of the effectiveness of government policies and a factor

of national pride (Akyol et al., 2021). This political dimension might influence the attitudes students have towards PISA, and their motivation to do their best. More generally, cultural differences in the value assigned to academic performance might induce between-country differences in the effort students spend on low-stake tests. There is evidence of differences across world regions, and in particular between Asian students and students from other regions: Gneezy and colleagues (2019) found that increasing the stakes associated with the test increased performance in the United States but not in Shanghai. Nonetheless, differences in cultural norms about academic work between Switzerland, Denmark and Australia are considerably smaller than those between the United States and Shanghai. Moreover, cultural differences might influence the level of performance decline in the three countries, but there are no theoretical reasons as to why they should affect the relationship between performance decline and educational attainment within each country.

## Data and methods

### Data

Our baseline data come from the public use 2000 PISA survey for Denmark and Switzerland and 2003 for Australia. PISA is a triennial large-scale international assessment surveys 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)].

. At the individual level, PISA is a low-stake assessment because no individual scores are released to students or schools. As a result, although students are supervised by teachers and take the test in a classroom context, conscientiousness and intrinsic motivation play a role in guiding the behaviour of participants in the study (Borgonovi and Biecek, 2016). The core instruments of PISA are a two-hours assessment developed by international experts and a 30-minutes

background questionnaire. We use data from the background questionnaire to identify students' background characteristics and data from the assessment to evaluate achievement at the start of the test and performance decline during the test.

We complemented PISA with country specific longitudinal follow-ups, containing information up to young adulthood (age 25 in Switzerland and Australia, 26/27 in Denmark). The Danish follow-up comes from the Survey of Adult Skills (PIAAC) (Maehler and Konradt, 2020). The Swiss follow-up comes from the Transitions from Education to Employment (TREE1) study. Finally, the Australian follow-up comes from the Longitudinal Study of Australian Youth (LSAY).

**Measures**

Our dependent variables are upper secondary education (academic or vocational) completion and university completion by the age of 25 (26/27 in Denmark).

Our key explanatory variable is performance decline. Performance decline measures the extent to which students are able to sustain a consistent level of performance throughout the test. While different strategies have been used to operationalise decline in performance in low-stakes standardised assessments (Borghans and Schils, 2012; Soland et al., 2019; Zamarro et al., 2019) we develop an individual level indicator following Zamarro et al. (2019). This is computed by fitting linear regressions predicting the likelihood of a question being answered correctly based on its position in a test booklet. Thanks to the random allocation of PISA booklets and the different order of the same items in different booklets, we can fit models with item fixed effects to estimate the item order effects independent of question difficulty and booklet composition.

PISA 2000 and 2003 were paper-based tests designed to last two hours and organised around a series of clusters of test questions. Each cluster was designed to take about 30 minutes to complete. Students were randomly allocated booklets containing four clusters of test questions each and 13 booklets in total were administered. 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 appeared at least once in one of the four potential positions within booklets.

We only consider performance in the first three clusters to deal with end of test non-response. Some students fail to reach items at the end of the test, which could lead to biased estimates since item non-response due to lack of time varies across groups and countries (Borgonovi and Biecek, 2016; Debeer and Janssen, 2013; Adams and Wu, 2002). Some of the weakest students fail to complete tasks in the fourth cluster because PISA is a timed assessment and students who require a lot of time to read question stimuli fail to reach the end. As a result, considering performance decline using the fourth cluster would lead to bias because it would mean assigning a wrong code to questions for which, in fact, no information is available. In practical terms, our approach means that our estimates represent a lower bound of the association between performance decline and educational attainment, since the weakest students are less likely to attempt questions in cluster 4 and to complete upper secondary and tertiary education. Figure 1 illustrates the booklets in PISA 2000 and 2003.

FIGURE 1

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 $y_{ij}$ equals zero if respondent $i$ answered incorrectly to question $j$ and 1 if he/she answered correctly or got partial credit for his/her answers[i]. $Q_{ij}$ 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. $\delta_0$ represents the average student's performance on the first question in the test and $\delta_1$ is the average performance drop from the first question to the last in cluster 3. $\gamma_j$ are question fixed effects, to control for question difficulty and the nature of the question (for example if the question is multiple choice versus open-ended question). $\theta_j$ are booklet fixed effects to control for the sequence of clusters in the booklet. $\delta_0^i$ is a random intercept and $\delta_1^i$ 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, 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 PISA samples. Fitting the model produced estimates of the standard deviation of random effects ($\delta_0^i$ and $\delta_1^i$) and their correlation (coefficients presented in Appendix Table A1). We used these to predict the best linear unbiased predictions (BLUPS) for the random slope parameter $\delta_1^i$, which represents our measure of performance decline.

The PISA 2000 and 2003 test designs differed in that the major domain (reading in 2000, mathematics in 2003) was assessed significantly more in depth than the other two (Figure 1). Seven booklets began with three reading clusters and ended with a different domain. To maintain a balanced sample in terms of booklet composition and to ensure greater consistency in the measure of performance decline for students who participated in the PISA 2000 assessment, 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.

In the PISA 2003 assessment, mathematics was the core domain, but it was tested less in depth than reading was in PISA 2000: only one third of the booklets tested contained only mathematics in the first three clusters (see Figure 1). To maintain high statistical power in our analyses, we decided to consider all test booklets rather than focus only on the four math intensive ones. Differences across booklets are captured by booklet fixed effects, so our indicator should measure performance decline net of these. Results are robust to the use of the subset of booklets, although are less precise because of the smaller sample.

There is evidence that students' drop in performance may be greater in reading than mathematics (Borgonovi and Biecek, 2016). If true, this would lead to a larger performance decline in PISA 2000. However, since all analyses are conducted within countries (both the computation of the index and its relationship with later life outcomes) and since the index is standardized within-country (so within PISA wave), our estimates are comparable in spite of differences across the two assessments.

In all analyses, we control for students' background characteristics. These include: gender; immigrant background (defined as students reporting having two foreign-born parents or one foreign-born parent if the student lives in a single-parent household); being a non-native speaker (defined using reports by students on whether they most often speak a language at home that is different from the one used in the PISA test); age; grade attended; socio-economic status (using an indicator of whether students had at least one tertiary-educated parent); the prestige of the occupation of the parent with the most prestigious occupation (defined according to the ISEI index of occupational prestige [Ganzeboom and Treiman, 1996; 2003]). We control for baseline

achievement using students' share of correct answers in the first cluster instead of actual PISA scores to avoid problems of collinearity with the measure of performance decline. The balanced nature of the PISA test means that different assessment clusters are equally easy or hard, since they contain a sample of questions which, on average, have similar levels of difficulty (see Borgonovi and Biecek, 2016).

In some models, we also account for a self-reported measure of effort and persistence on tasks to identify the additional predictive value of performance decline. These included self-reported lack of perseverance in PISA 2000 and lack of effort in PISA 2003. A detailed description of the construction of the self-reported measures is presented in Appendix A.

**Methods**

For each country and for each of our two outcomes of interest, we run different sets of linear probability models (LPM) which we prefer to logistic models following Mood (2010). Logistic regressions suffer from unobserved heterogeneity even when omitted variables are unrelated to the independent variables included in the model. As a result, it is problematic to compare estimates from logistic regressions across samples, groups within samples and over time, as is the case in our work, because unobserved heterogeneity can vary across the samples, groups, and points in time. LPM do not suffer from the same shortcomings and are therefore preferable when modelling non-extreme probabilities which is the case in our work. In the first set of models, we regress outcomes of interest on individual-level controls and on the performance decline indicator. Then, we additionally control for self-reported indicators of effort and persistence to investigate associations between the behavioural measure and educational outcomes net of the self-reported measures.

In the second set of models, we test the moderating and mediating effects of the performance decline index. We investigate how the association between outcome variables and three predictors (gender, parental education and performance in the first cluster) changes after including the performance decline index in the regression, and after additionally controlling for self-reported measures. Finally, we run a model that includes interactions between the index of performance decline and gender, parental education, and performance in the first cluster.

To aid comparisons between variables of interest, we standardised performance decline, performance in the first cluster and self-reported measures of effort and persistence within each country. We used sampling and replication weight in our analyses in line with recommendations (OECD, 2009). Response rates varied across the longitudinal studies considered, ranging from 36% in the latest wave of the 2003 LSAY, 44% for the Danish PIAAC follow-up and 86% in the latest wave of the TREE1 study (LSAY, 2010; Sacchi, 2011; Rosdahl, 2014). We combined PISA weights with the recommended weights for the relevant wave of each longitudinal survey as recommended in technical reports (LSAY, 2017; Rosdahl, 2014; Sacchi, 2011).

## Results

### Descriptive Results

Table 1 indicates that the vast majority of 15-year-old students completed upper secondary education by age 25 with few differences across countries. Differences in university completion across countries are larger: completion rates are lowest in Switzerland and highest in Australia. Performance at the start of the PISA test differed little across countries, while performance decline was larger in Australia and Switzerland than in Denmark.

TABLE 1

**Performance Decline at age 15 and Educational Attainment at age 25**

Figure 2 displays the association between performance decline and later life outcomes, net of the control variables listed above and before and after controlling for self-reported indicators of effort and persistence (Model 1 and Model 2). Figure 2 shows that students with a one SD higher decline in performance in the PISA test at age 15 were three percentage points less likely to complete upper secondary school than students with a stable performance in the test in Denmark. The association was quantitatively small and not statistically significant in Australia and was larger but imprecisely estimated in Switzerland. Controlling for self-reported measures in Model 2 did not change the strength of the association between performance decline and the likelihood of completing upper secondary education. By contrast, we identify a quantitatively meaningful association between performance decline at age 15 and tertiary completion by age 25. The association appears to be quantitatively similar across the three countries examined despite large differences in overall rates of tertiary level completion and is unaffected by the inclusion of self-reported perseverance. The partial correlation coefficient for university completion corresponds to ($r$ = -0.15 $p<0.05$ for Australia; $r$ = -0.11 $p=0.11$ for Denmark and $r$ = -0.11 $p<0.001$ for Switzerland). Interestingly, results presented in Appendix Tables B1 and B2 indicate that self-reported measures of effort and persistence are not associated with neither upper secondary level nor tertiary level completion. According to results presented in Figure 2 a difference of one SD in performance decline was associated with a difference of 6-7 percentage points in tertiary level completion in the three countries, although in Denmark the association was imprecisely estimated.

FIGURE 2

**Performance Decline and Educational Attainment: Mediating and Moderating Effects**

We examine the mediating role of performance decline at age 15 with respect to gender, socio-economic condition, and achievement differences in educational attainment by comparing the strength of the association between background factors before and after accounting for performance decline in Models 1 and 2 in Tables 2, 3 and 4. In the three tables, Panel A reports results for upper secondary-school completion and Panel B reports results for tertiary completion.

Results indicate that neither gender nor parental educational attainment were associated with the likelihood of completing upper secondary education in the three countries in our sample, neither before nor after controlling for performance decline. By contrast, performance in the first cluster was associated with an increased likelihood of completing upper secondary education in Switzerland and Denmark, the two countries in which vocationally oriented secondary education is more prevalent and upper secondary qualifications more often lead to entrance in the labour market or vocationally oriented non-tertiary qualifications. However, performance decline was not an important mediator of the association between baseline achievement and upper secondary school completion.

Socio-economic differences and differences related to baseline levels of ability are more pronounced in the case of tertiary completion. For example, females were 16 percentage points more likely than males to complete tertiary education in Australia. Parental educational attainment was associated with an increased likelihood of obtaining tertiary qualifications by age 25 in all three countries: compared to individuals without tertiary educated parents, individuals

with at least one tertiary educated parent were 15 percentage points more likely of having obtained tertiary level qualifications in Australia, 13 percentage points more likely in Denmark and 8 percentage points more likely in Switzerland. Associations were not affected by the inclusion of performance decline.

We also investigated whether the strength of the association between performance decline at age 15 and completion of upper-secondary school or completion of a tertiary degree differed by gender, socio-economic condition, or achievement in the three countries considered adding interaction effects in our models. We find no evidence of heterogeneous effects of performance decline across groups: all interaction effects are quantitatively small and not statistically significantly different from zero.

TABLE 2

TABLE 3

TABLE 4

## Discussion

Our work contributes new longitudinal evidence to research employing behavioural measures of socio-emotional and motivational skills (see Soland et al., 2019 for a review). Despite the increasing use of the performance decline indicator as a measure of capacity to sustain effort on the test, and as a proxy for general conscientiousness and intrinsic motivation, there was no evidence documenting the extent to which performance decline during long low-stakes assessments such as PISA is associated with individuals' outcome.

Our study fills this gap using ex-post harmonised data from national longitudinal studies developed in Australia, Denmark and Switzerland, three countries with very different education systems that follow samples of individuals who took part in the PISA study. We relate differences in performance decline in the PISA test at age 15 to differences in the likelihood of having completed upper secondary and tertiary education by age 25. As such, our study can be used to understand the potential long-term consequences for individual outcomes of disparities in the capacity to sustain effort over cognitively demanding tasks, as measured by performance decline in the PISA test. The effect size is quantitatively small at conventional levels. However, even effects that are small according to conventional levels (Cohen, 1988), when reliably estimated, can be indicative of consequential effects (Funder and Ozer, 2019). A difference of 6 percentage points in the probability of completing college is consequential from a policy standpoint.

We find evidence that, net of achievement differences and other individual-level characteristics, the decline in accuracy over a long test is strongly associated with completion of a tertiary degree in all three countries, although results are imprecisely estimated in Denmark. The association between performance decline and upper-secondary completion is quantitatively smaller and less precisely estimated than the association observed for tertiary degree completion. The difference in tertiary level graduation that is associated with a standard deviation difference in performance decline corresponds to between half and a third of the difference in tertiary level graduation that is associated with a standard deviation difference in achievement, a key driver of attainment differentials. Among students of equal achievement potential and similar background characteristics, their ability to sustain attention to solve cognitively demanding problems at age 15 significantly contributes to their odds of completing a tertiary degree. Interestingly, self-reported measures of perseverance and effort are not associated with the likelihood of completing upper secondary or tertiary education in the three countries in our sample and all our estimates are robust to the inclusion of self-reported effort and persistence. Crucially, our estimates on tertiary level attainment are in line with estimates for the United States obtained using behavioural indicators of conscientiousness and intrinsic motivation derived from response behaviour to questionnaires, rather than cognitive tests (Hitt et al., 2016).

Contrary to our hypotheses, we do not find evidence that differences in performance decline at age 15 explain differences across socio-economic groups in tertiary level graduation rates. Furthermore, in all three countries individuals with higher performance decline appeared less likely to obtain tertiary level qualifications, and the association did not differ across genders, socio-economic statues and baseline levels of achievement.

Succeeding in tertiary education is a highly cognitively demanding endeavour that requires individuals to process large amounts of information over years. The correlation between performance decline in PISA and completion of tertiary education suggests that measures of sustained effort in extended performance tasks might capture motivational traits that matter also outside the context of a test and that are relatively stable.

Our study suffers from a number of limitations that should be addressed in future research. First, although our study provides longitudinal evidence on the outcomes associated with performance decline in three countries that differ in how education systems are organised, the three countries come from rather homogeneous levels of economic development and baseline levels of achievement. PISA contains information on over 80 education systems worldwide and it is therefore not possible to infer from the evidence presented the likely impact of performance decline on student outcomes in countries that differ in meaningful ways from those analysed here. Moreover, the cohorts examined refer to the outcomes at age 25-27 of individuals who were 15-year-old students in 2000 and 2003. It is possible that social, economic and technological transformations changed the association between individuals' capacity to sustain baseline levels of accuracy in the PISA test and their educational attainment.

Second, our analyses are based on performance decline indicators calculated using the PISA 2000 and 2003 assessments. Since then, the assessment has undergone several changes which may influence the measurement of performance decline by influencing participants' level of motivation to put effort in the test. The test moved to computer-based administration in 2015 and in 2018 it introduced an adaptive design for the reading part of the assessment. Third, further work is also needed to understand what are the main cognitive and motivational drivers of performance decline during a demanding, but relatively low-stake task. Performance decline

might in fact reflect low levels of motivation, but might also be due to low cognitive ability to sustain attention and manage fatigue. Finally, results reflect the association between a specific indicator of academic effort and educational attainment measured through completion of upper secondary school and graduation from a tertiary level institution. It is not possible to extrapolate from these findings to associations between effort on cognitively demanding tasks in adolescence and other outcomes such as labour market participation and success.

## References


Adams, R. and Wu, M. (2002). *PISA 2000 Technical Report.* Paris: OECD Publishing. http://www.oecd.org/pisa/data/33688233.pdf

Akyol, P., Krishna, K. and Wang, J. (2021). Taking PISA Seriously: How Accurate are Low-Stakes Exams?. *Journal of Labor Research* **42,** 184–243 (2021). https://doi.org/10.1007/s12122-021-09317-8

Almlund, M., Duckworth, A. L., Heckman, J. J., and Kautz, T. D. (2011). Personality psychology and economics. Working Paper 16822. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w16822.pdf

Duckworth, A. L. and Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–25. https://doi.org/10.3102%2F0013189X15584327

Balart, P., and 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

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Barone, C. and Ruggera, L. (2018). Educational equalization Stalled? Trends in inequality of educational opportunity between 1930 and 1980 across 26 European Nations. *European Societies, 20*(1), 1-25. https://doi.org/10.1080/14616696.2017.1290265

Barrick, M. R., and Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bernardi, F. (2014). Compensatory advantage as a mechanism of educational inequality: A regression discontinuity based on month of birth. *Sociology of Education, 87*(2), 74-88. https://doi.org/10.1177%2F0038040714524258

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163–197. https://doi.org/10.1007/BF03393163

Blossfeld, H. P. and Shavit, Y. (1993). Persisting barriers. Changes in educational opportunities in thirteen countries. in: Blossfeld, H. P. and Shavit, Y. (Eds.). *Persisting inequality. Changing educational attainment in thirteen countries* (pp. 1-23). Westview Press.

Borghans, L. and Schils, T. (2012). The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components. Unpublished manuscript. Draft version: July 22, 2012.

Borgonovi, F. and 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. https://doi.org/10.1016/j.lindif.2016.06.001

Brzinsky-Fay, C. (2007). Lost in transition? Labour market entry sequences of school leavers in Europe. *European Sociological Review, 23*(4), 409-422. https://doi-org.eui.idm.oclc.org/10.1093/esr/jcm011

Buchmann, C. and Kriesi, I. (2011). Transition to adulthood in Europe. *Annual Review of Sociology, 37,* 481–503. https://doi.org/10.1146/annurev-soc-081309-150212

Chamorro-Premuzic, T., and Furnham, A. (2003). Personality traits and academic examination performance. European Journal of Personality, 17, 237–250

Cheng, A. (2015). Like teacher, like student: Teachers and the development of student noncognitive skills. EDRE Working Paper No. 2015-02. Fayetteville, AR: University of Arkansas.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa Jr, P. T., Aguilar-Vafaie, M. E.,... and Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology, 108*(1), 171 – 185. https://doi.org/10.1037/a0038497

Debeer, D., and 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

DeMars, C.A., Bashkov, B.M., and Socha, A.B. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Research and Practice in Assessment, 8,* 69-82.

Deci, E.L. and Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press, New York.

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087-1101.

Duckworth, A. L. and Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198–208. https://doi.org/10.1037/0022-0663.98.1.198

Egan, M., Daly, M., Delaney, L., Boyce, C. J., and Wood, A. M. (2017). Adolescent conscientiousness predicts lower lifetime unemployment. *Journal of Applied Psychology, 102*(4), 700–709. https://doi.org/10.1037/apl0000167

Eisenberger, R., and Leonard, J. M. (1980). Effects of conceptual task difficulty on generalized persistence. *American Journal of Psychology, 93,* 285–298.

Funder, D.C., and Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. Advances in Methods and Practices in Psychological Science, 2(2), 156-168. https://doi.org/10.1177/2515245919847202

Gangl, M., Müler, W., and Raffe, D. (2003). Conclusions: Explaining cross-national differences in school- to-work transitions. in: W., Müller, and M., Gangl (Eds) *Transitions from Education to Work in Europe. The Integration of Youth into EU Labour Markets* (pp.277-305. Oxford: Oxford University Press.

Ganzeboom, H. B.G and Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. in: J. H. P., Hoffmeyer-Zlotnik and C. Wolf (Eds) *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (pp. 159-193). New York: Kluwer Academic Press.

Ganzeboom, H. B. G., and Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25,* 201-239. https://doi.org/10.1006/ssre.1996.0010

Garcia, E. (2014). The need to address noncognitive skills in the education policy agenda. EPI Briefing Paper 386. Economic Policy Institute, Washington DC.

Gneezy, U., J. A. List, J. A. Livingston, X. Qin, S. Sadoff, and Y. Xu. (2019). "Measuring Success in Education: The Role of Effort on the Test Itself." *American Economic Review: Insights*, 1 (3): 291-308.

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

Hanushek, E. A., Welch, F., Machin, S. and Woessmann, L. (2011). *Handbook of the Economics of Education (Vol. 4)*. North-Holland: Elsevier.

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. http://www.uaedreform.org/downloads/2015/07/edre-working-paper-2015-06.pdf

Hitt, C., Trivitt, J. and Cheng, A. (2016) When you say nothing at all: The predictive power of student effort on surveys, *Economics of Education Review*, *52,* 105-119. https://doi.org/10.1016/j.econedurev.2016.02.001

Huang, C. (2011). Self-concept and academic achievement: a meta-analysis of longitudinal relations. J. School Psychol. 49, 505–528. https://doi: 10.1016/j.jsp.2011. 07.001

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

Kriegbaum, K., Jansen, M., and Spinath, B. (2015). Motivation: a predictor of PISA's mathematical competence beyond intelligence and prior test achievement. Learning and Individual Differences 43, 140–148. doi: 10.1016/j.lindif.2015.08.026

Kuncel, N. R., Hezlett, S. A., and Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: can one construct predict them all? J. Person. Soc. Psychol. 86, 148-161. doi: 10.1037/0022-3514.86.1.148

Kyllonen, P. C., and 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), 1–26. https://doi.org/10.3390/jintelligence6030032

Krosnick, J. A., Narayan, S. S., and Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In: M. T. Braveman and J. K. Slater (Eds) *Advances in survey research* (pp. 29–44). San Francisco, CA: Jossey- Bass.

Linnenbrink-Garcia, L., Wormington, S. V., Snyder, K. E., Riggsbee, J., Perez, T., Ben-Eliyahu, A., et al. (2018). Multiple pathways to success: an examination of integrative motivational profiles among upper elementary and college students. Journal of Educational Psychology 110, 1026–1048 https://doi: 10.1037/edu0000245.

Liu, A. (2020). Non-Cognitive skills and the growing achievement gap. *Research in Social Stratification and Mobility*, 69, 100546. https://doi.org/10.1016/j.rssm.2020.100546

LSAY (2017). Longitudinal Surveys of Australian Youth, 2003 cohort (Version 7.1). Australian Government Department of Education, Skills and Employment, doi:10.4225/87/5IOBPG, ADA Dataverse, V4.

Maehler D.B. and Konradt I. (2020). Adult cognitive and non-cognitive skills: An overview of existing PIAAC data. in: D., Maehler and B., Rammstedt (Eds) *Large-Scale Cognitive Assessment. Methodology of Educational Measurement and Assessment* (pp.49-91). Springer, Cham. https://doi.org/10.1007/978-3-030-47515-4_4

Matthews, J. S., Ponitz, C. C. and Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689–704. https://doi.org/10.1037/a0014240

McClelland, D. C., Atkinson, J., Clark, R., and Lowell, E. (1953). The Achievement Motive. New York, NY: Appleton-Century-Crofts.

Mezzacappa E. (2004) Alerting, orienting and executive attention. Developmental properties and sociodemographic correlated in and epidemiological sample of young, urban children. Child Development 75: 1373–1386.

Montanello, S. and Martens, B. K. (2005). Effects of interspersed brief problems on students' endurance at completing math work. *Journal of Behavioral Education, 14,* 249–266. https://doi-org.eui.idm.oclc.org/10.1007/s10864-005-86493

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82. https://doi-org.eui.idm.oclc.org/10.1093/esr/jcp006

OECD (2009). *PISA data analysis manual: SAS.* Second Edition. Paris: OECD Publishing.

Pintrich, P. R., Marx, R. W., and Boyle, R. A. (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. Review of Education Research 63, 167–199. https://doi:10.3102/00346543063002167

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

Rosdahl, A. (2014). From 15 to 27 years old. The PISA 2000 students in 2011/12. The Danish Centre for Social Science Research. SFI - Det Nationale Forskningscenter for Velfærd. https://www.vive.dk/en/publications/from-15-to-27-years-old-5070/

Ryan, R.M. and Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Sacchi, Stefan (2011). Construction of Tree Panel Weights: Documentation for the panel waves from 2000 to 2010. TREE and cue sozialforschung, Basel/Zurich https://doi.org/10.7892/boris.131036

Schunk, D. H., and DiBenedetto, M. K. (2020). Motivation and social cognitive theory. Contemporary Educational Psychology, 61, 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Sievertsen H.H., Gino F., Piovesan M. Cognitive fatigue influences students' performance on standardized tests. Proc Natl Acad Sci U S A. 2016 Mar 8;113(10):2621-4. doi: 10.1073/pnas.1516947113. Epub 2016 Feb 16. PMID: 26884183; PMCID: PMC4790980.

Slimani, M., Znazen, H., Bragazzi, N. L., Zguira, M. S., and Tod, D. (2018). The Effect of Mental Fatigue on Cognitive and Aerobic Performance in Adolescent Active Endurance Athletes: Insights from a Randomized Counterbalanced, Cross-Over Trial. *Journal of clinical medicine, 7*(12), 510. https://doi.org/10.3390/jcm7120510

Soland, J., Zamarro, G., Cheng, A and 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. https://doi.org/10.3102/0013189X19861356

Soto, C. J. (2021). 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,* 12(1), 118-130. https://doi.org/10.1177/1948550619900572

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. https://doi.org/10.1177/0956797619831612

Steinmayr, R., and Spinath, B. (2009). The importance of motivation as a predictor of school achievement. Learning and Individual Differences 19, 80–90. doi: 10.1016/j.lindif. 2008.05.004

Steinmayr R, Weidinger AF, Schwinger M, Spinath B. (2019) The Importance of Students' Motivation for Their Academic Achievement - Replicating and Extending Previous Findings. Frontiers in Psychology. 2019 Jul 31;10:1730. https://doi: 10.3389/fpsyg.2019.01730

Vecchione, M., Alessandri, G. and Marsicano, G. (2014). Academic motivation predicts educational attainment: Does gender make a difference? *Learning and Individual Differences, 32,* 124-13. https://doi.org/10.1016/j.lindif.2014.01.003

Ventura, M., and Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior, 29*(6), 2568-2572.

Wigfield, A., and Cambria, J. (2010). Students' achievement values, goal orientations, and interest: definitions, development, and relations to achievement outcomes. Developmental Review 30, 1–35. https://doi: 10.1016/j.dr.2009.12.001

Wigfield, A., and Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 6881. https://doi.org/10.1006/ceps.1999.1015

Wilmot, M. P. and Ones, D. S. (2019). A century of research on conscientiousness at work. *PNAS, 116*(46), 23004-23010. https://doi.org/10.1073/pnas.1908430116

Xu K.M., A. R. Cunha-Harvey, R. B. King, B. B. de Koning, F. Paas, M. Baars, J. Zhang and Renate de Groot (2021) A cross-cultural investigation on perseverance, self-regulated learning, motivation, and achievement, Compare: *A Journal of Comparative and International Education*, https://doi.org/10.1080/03057925.2021.1922270

Zamarro, G., Cheng, A., Shakeel, M. D. and 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., and 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

Figure 1

*Booklet design in PISA 2000 and 2003*

| PISA 2000 | | | | |
|---|---|---|---|---|
| Booklet | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| 1 | R1 | R2 | R4 | M1/M2 |
| 2 | R2 | R3 | R5 | S1/S2 |
| 3 | R3 | R4 | R6 | M3/M4 |
| 4 | R4 | R5 | R7 | S3/S4 |
| 5 | R5 | R6 | R1 | M2/M3 |
| 6 | R6 | R7 | R2 | S2/S3 |
| 7 | R7 | R1 | R3 | R8 |
| 8 | M4/M2 | S1/S3 | R8 | R9 |
| 9 | S4/S2 | M1/M3 | R9 | R8 |

| PISA 2003 | | | | |
|---|---|---|---|---|
| Booklet | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| 1 | M1 | M2 | M4 | R1 |
| 2 | M2 | M3 | M5 | R2 |
| 3 | M3 | M4 | M6 | PS1 |
| 4 | M4 | M5 | M7 | PS2 |
| 5 | M5 | M6 | S1 | M1 |
| 6 | M6 | M7 | S2 | M2 |
| 7 | M7 | S1 | R1 | M3 |
| 8 | S1 | S2 | R2 | M4 |
| 9 | S2 | R1 | PS1 | M5 |
| 10 | R1 | R2 | PS2 | M6 |
| 11 | R2 | PS1 | M1 | M7 |
| 12 | PS1 | PS2 | M2 | S1 |

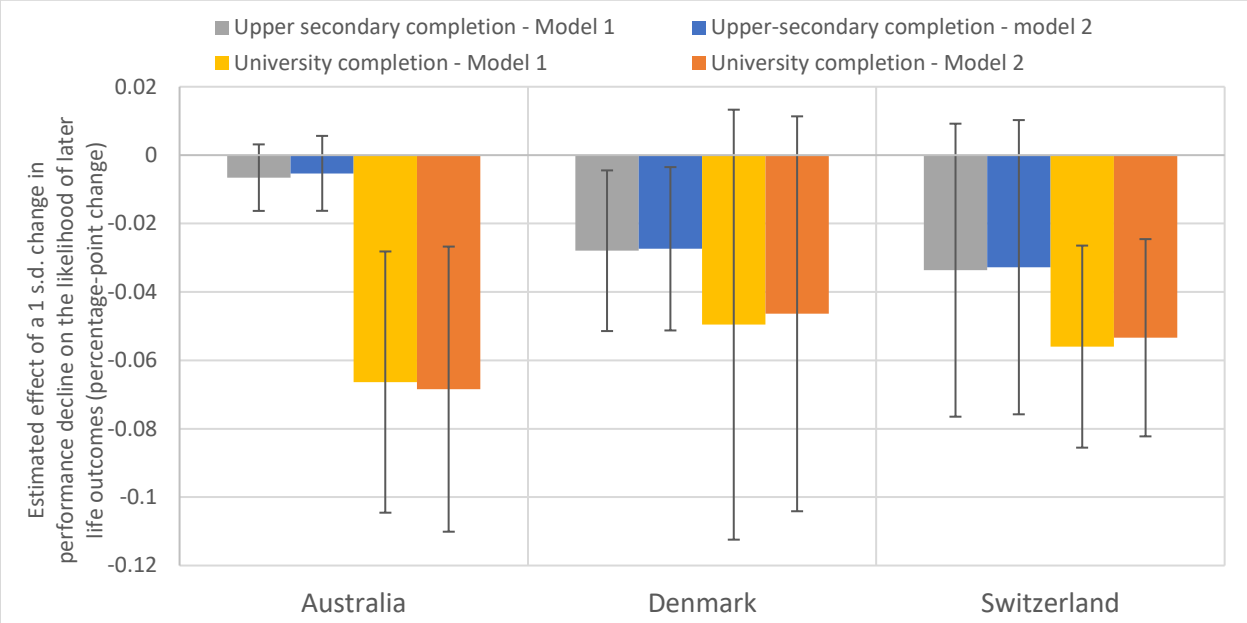
Table 1

*Descriptive statistics*

| | Australia | | Denmark | | Switzerland | |
|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd |
| Completed upper secondary | 0.964 | 0.185 | 0.895 | 0.307 | 0.893 | 0.309 |
| Completed university | 0.416 | 0.493 | 0.291 | 0.454 | 0.184 | 0.387 |
| Female | 0.512 | 0.500 | 0.514 | 0.500 | 0.493 | 0.500 |
| Age at test | 15.781 | 0.289 | 15.708 | 0.281 | 15.704 | 0.637 |
| School grade at test | 10.133 | 0.511 | 8.941 | 0.310 | 9.000 | 0.012 |
| Parent with tertiary education | 0.413 | 0.492 | 0.522 | 0.500 | 0.386 | 0.487 |
| Parental ISEI | 52.991 | 16.306 | 49.549 | 15.848 | 48.948 | 16.611 |
| Non-native speaker | 0.092 | 0.289 | 0.030 | 0.172 | 0.268 | 0.613 |
| Immigrant student | 0.224 | 0.417 | 0.041 | 0.200 | 0.221 | 0.415 |
| Correct 1st cluster | 0.632 | 0.234 | 0.709 | 0.181 | 0.685 | 0.189 |
| Performance decline | 0.066 | 0.078 | 0.040 | 0.052 | 0.061 | 0.057 |
| Self-reported perseverance | . | . | 0.018 | 0.926 | -0.157 | 1.040 |
| Self-reported effort | -7.694 | 1.697 | . | . | . | . |
| Observations | 3110 | | 1130 | | 1962 | |

Figure 2

*The associations between performance decline and later life outcomes*




**Notes:** The figure shows estimates of the effect of a 1 standard deviation change in the performance decline index. Estimates are presented with 95 percent confidence intervals. Model 1 controls for students' gender, immigrant background, age at test, school grade at test, whether they spoke the language of instruction at home, parental occupational class (isei score) and whether they had a parent with tertiary education. Model 2 also controls for self-reported perseverance for Denmark and Switzerland and for self-reported effort in the test for Australia. Regression Tables B1 and B2 in the Appendix display full model coefficients for explanatory and control variables.

Table 2 *Likelihood of completing upper-secondary and tertiary education in Australia*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A: Completing upper-secondary education** | | | | | | |
| Female | 0.022 | 0.021 | 0.020 | 0.016 | 0.020 | 0.020 |
| | (1.49) | (1.49) | (1.45) | (1.07) | (1.47) | (1.43) |
| Parent tertiary ed. | 0.012 | 0.012 | 0.012 | 0.012 | 0.007 | 0.012 |
| | (1.11) | (1.09) | (1.14) | (1.17) | (0.57) | (1.15) |
| Correct 1st cluster | 0.024 | 0.024* | 0.022* | 0.022* | 0.022* | 0.028 |
| | (1.89) | (2.04) | (2.07) | (2.04) | (2.06) | (1.44) |
| Performance decline | | -0.007 | -0.005 | -0.007 | -0.008 | 0.012 |
| | | (-1.33) | (-0.95) | (-0.90) | (-0.87) | (0.38) |
| Self-reported effort | | | -0.010 | -0.010 | -0.010 | -0.010 |
| | | | (-1.87) | (-1.82) | (-1.82) | (-1.84) |
| Performance decline*Female | | | | 0.004 | | |
| | | | | (0.62) | | |
| Performance decline*Parent tertiary ed. | | | | | 0.006 | |
| | | | | | (0.88) | |
| Performance decline* correct 1st cluster | | | | | | -0.007 |
| | | | | | | (-0.63) |
| Observations | 3110 | 3110 | 3110 | 3110 | 3110 | 3110 |
| R-squared | 0.034 | 0.036 | 0.038 | 0.038 | 0.039 | 0.040 |
| **Panel B: Completing tertiary education** | | | | | | |
| Female | 0.162*** | 0.159*** | 0.161*** | 0.192*** | 0.161*** | 0.162*** |
| | (6.49) | (5.84) | (5.54) | (7.36) | (5.33) | (5.55) |
| Parent tertiary ed. | 0.150*** | 0.144*** | 0.144*** | 0.143*** | 0.113*** | 0.144*** |
| | (6.71) | (6.37) | (6.53) | (6.44) | (3.91) | (6.51) |
| Correct 1st cluster | 0.150*** | 0.154*** | 0.158*** | 0.158*** | 0.158*** | 0.172*** |
| | (15.06) | (13.78) | (12.44) | (12.47) | (12.84) | (13.86) |
| Performance decline | | -0.066*** | -0.068** | -0.051 | -0.083** | -0.024 |
| | | (-3.41) | (-3.22) | (-1.31) | (-2.79) | (-0.78) |
| Self-reported effort | | | 0.016 | 0.015 | 0.017 | 0.016 |
| | | | (0.78) | (0.67) | (0.79) | (0.76) |
| Performance decline*Female | | | | -0.036 | | |
| | | | | (-0.91) | | |
| Performance decline*Parent tertiary ed. | | | | | 0.037 | |
| | | | | | (1.83) | |
| Performance decline* correct 1st cluster | | | | | | -0.017 |
| | | | | | | (-1.76) |
| Observations | 3110 | 3110 | 3110 | 3110 | 3110 | 3110 |
| R-squared | 0.197 | 0.215 | 0.216 | 0.217 | 0.217 | 0.217 |

Notes: t statistics in parentheses. * p < 0:05, ** p < 0:01, *** p < 0.001. All estimates control for students' gender, immigrant background, age at test, school grade at test, whether they spoke the language of instruction at home, parental occupational class (isei score) and whether they had a parent with tertiary education. The coefficients for performance decline, the share of correct answers in the first cluster, and the index of self-reported effort measure the effect of a change of 1 standard deviation.

Table 3 *Likelihood of completing upper-secondary and tertiary education in Denmark*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A: Completing upper-secondary education** | | | | | | |
| Female | -0.042 | -0.042 | -0.043 | -0.026 | -0.043 | -0.043 |
| | (-1.70) | (-1.65) | (-1.67) | (-1.28) | (-1.65) | (-1.66) |
| Parent tertiary ed. | -0.006 | -0.012 | -0.012 | -0.012 | -0.003 | -0.011 |
| | (-0.35) | (-0.70) | (-0.70) | (-0.73) | (-0.12) | (-0.65) |
| Correct 1st cluster | 0.038*** | 0.035*** | 0.034*** | 0.034** | 0.034** | 0.019 |
| | (3.65) | (3.47) | (3.30) | (3.27) | (3.25) | (1.95) |
| Performance decline | | -0.028* | -0.027* | -0.016 | -0.022 | -0.091* |
| | | (-2.33) | (-2.25) | (-1.29) | (-1.14) | (-1.96) |
| Self-reported effort | | | -0.006 | -0.006 | -0.006 | -0.006 |
| | | | (-0.74) | (-0.76) | (-0.75) | (-0.75) |
| Performance decline*Female | | | | -0.022 | | |
| | | | | (-0.98) | | |
| Performance decline*Parent tertiary ed. | | | | | -0.012 | |
| | | | | | (-0.40) | |
| Performance decline* correct 1st cluster | | | | | | 0.017 |
| | | | | | | (1.46) |
| Observations | 1110 | 1110 | 1110 | 1110 | 1110 | 1110 |
| R-squared | 0.068 | 0.076 | 0.076 | 0.078 | 0.077 | 0.080 |
| **Panel B: Completing tertiary education** | | | | | | |
| Female | 0.026 | 0.026 | 0.022 | 0.020 | 0.020 | 0.022 |
| | (0.86) | (0.85) | (0.71) | (0.48) | (0.67) | (0.73) |
| Parent tertiary ed. | 0.133*** | 0.123*** | 0.123*** | 0.123*** | 0.168*** | 0.122*** |
| | (3.96) | (3.69) | (3.80) | (3.80) | (4.14) | (3.72) |
| Correct 1st cluster | 0.071 | 0.064 | 0.060* | 0.060* | 0.062* | 0.088* |
| | (1.92) | (1.94) | (1.96) | (1.99) | (1.96) | (2.38) |
| Performance decline | | -0.050 | -0.046 | -0.047 | -0.018 | 0.072 |
| | | (-1.55) | (-1.57) | (-1.02) | (-0.60) | (1.21) |
| Self-reported effort | | | -0.035 | -0.035 | -0.035 | -0.035 |
| | | | (-1.39) | (-1.40) | (-1.45) | (-1.40) |
| Performance decline*Female | | | | 0.002 | | |
| | | | | (0.04) | | |
| Performance decline*Parent tertiary ed. | | | | | -0.060 | |
| | | | | | (-1.79) | |
| Performance decline* correct 1st cluster | | | | | | -0.032 |
| | | | | | | (-1.67) |
| Observations | 1110 | 1110 | 1110 | 1110 | 1110 | 1110 |
| R-squared | 0.145 | 0.156 | 0.162 | 0.162 | 0.166 | 0.167 |

Notes: t statistics in parentheses. * p < 0:05, ** p < 0:01, *** p < 0.001. All estimates control for students' gender, immigrant background, age at test, school grade at test, whether they spoke the language of instruction at home, parental occupational class (isei score) and whether they had a parent with tertiary education. The coefficients for performance decline, the share of correct answers in the first cluster, and the index of self-reported perseverance measure the effect of a change of 1 standard deviation.

Table 4 *Likelihood of completing upper-secondary and tertiary education in Switzerland*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A: Completing upper-secondary education** | | | | | | |
| Female | -0.021 | -0.024 | -0.026 | -0.029 | -0.025 | -0.025 |
| | (-0.57) | (-0.67) | (-0.72) | (-0.72) | (-0.70) | (-0.71) |
| Parent tertiary ed. | 0.052 | 0.049 | 0.050 | 0.050 | 0.020 | 0.049 |
| | (1.64) | (1.61) | (1.65) | (1.64) | (0.61) | (1.63) |
| Correct 1st cluster | 0.058* | 0.054* | 0.054* | 0.054* | 0.052* | 0.043 |
| | (2.36) | (2.32) | (2.32) | (2.32) | (2.30) | (1.42) |
| Performance decline | | -0.034 | -0.033 | -0.034 | -0.045 | -0.064 |
| | | (-1.54) | (-1.49) | (-0.92) | (-1.45) | (-0.62) |
| Self-reported effort | | | 0.013 | 0.013 | 0.014 | 0.013 |
| | | | (0.77) | (0.76) | (0.81) | (0.76) |
| Performance decline*Female | | | | 0.003 | | |
| | | | | (0.07) | | |
| Performance decline*Parent tertiary ed. | | | | | 0.030 | |
| | | | | | (0.85) | |
| Performance decline* correct 1st cluster | | | | | | 0.009 |
| | | | | | | (0.33) |
| Observations | 1953 | 1953 | 1953 | 1953 | 1953 | 1953 |
| R-squared | 0.058 | 0.069 | 0.071 | 0.071 | 0.074 | 0.072 |
| **Panel B: Completing tertiary education** | | | | | | |
| Female | 0.036 | 0.031 | 0.025 | 0.041 | 0.026 | 0.024 |
| | (1.38) | (1.18) | (0.98) | (1.03) | (1.00) | (0.91) |
| Parent tertiary ed. | 0.076** | 0.071* | 0.076** | 0.077** | 0.048 | 0.078** |
| | (2.64) | (2.54) | (2.72) | (2.74) | (1.13) | (2.88) |
| Correct 1st cluster | 0.073*** | 0.067*** | 0.067*** | 0.067*** | 0.065*** | 0.083** |
| | (5.06) | (4.38) | (4.49) | (4.49) | (4.58) | (2.79) |
| Performance decline | | -0.056*** | -0.053*** | -0.047* | -0.065*** | -0.010 |
| | | (-3.71) | (-3.63) | (-2.31) | (-4.30) | (-0.16) |
| Self-reported effort | | | 0.039* | 0.039* | 0.040* | 0.040* |
| | | | (2.25) | (2.24) | (2.33) | (2.23) |
| Performance decline*Female | | | | -0.015 | | |
| | | | | (-0.55) | | |
| Performance decline*Parent tertiary ed. | | | | | 0.028 | |
| | | | | | (0.91) | |
| Performance decline* correct 1st cluster | | | | | | -0.012 |
| | | | | | | (-0.67) |
| Observations | 1953 | 1953 | 1953 | 1953 | 1953 | 1953 |
| R-squared | 0.106 | 0.125 | 0.135 | 0.136 | 0.136 | 0.136 |

Notes: t statistics in parentheses. * p < 0:05, ** p < 0:01, *** p < 0.001. All estimates control for students' gender, immigrant background, age at test, school grade at test, whether they spoke the language of instruction at home, parental occupational class (isei score) and whether they had a parent with tertiary education. The coefficients for performance decline, the share of correct answers in the first cluster, and the index of self-reported perseverance measure the effect of a change of 1 standard deviation.

[i] 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).