Education Economics

ISSN: 0964-5292 (Print) 1469-5782 (Online) Journal homepage: www.tandfonline.com/journals/cede20

Locus of control, educational attainment, and college aspirations: the relative role of effort and expectations

Ágnes Szabó-Morvai & Hubert János Kiss

To cite this article: Ágnes Szabó-Morvai & Hubert János Kiss (2024) Locus of control, educational attainment, and college aspirations: the relative role of effort and expectations, Education Economics, 32:6, 862-881, DOI: 10.1080/09645292.2023.2273220

To link to this article: https://doi.org/10.1080/09645292.2023.2273220

EDUCATION ECONOMICS
2024, VOL. 32, NO. 6, 862–881
https://doi.org/10.1080/09645292.2023.2273220

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Locus of control, educational attainment, and college aspirations: the relative role of effort and expectations

Ágnes Szabó-Morvai[a,b†] and Hubert János Kiss[a,c†]

[a]KRTK KTI, Budapest, Hungary; [b]University of Debrecen, Debrecen, Hungary; [c]Corvinus University of Budapest, Budapest, Hungary

ABSTRACT

We study the relationship between locus of control and both educational attainment and college aspirations in adolescence, focusing on the potential channels through which locus of control may influence these outcomes. Even after controlling for a comprehensive set of socio-demographic factors and cognitive skills, there remains a strong correlation between locus of control, college aspirations, and attendance. We show that effort is an essential channel through which locus of control operates. Mediation analysis further reveals that this mechanism is more influential than future expectations in determining high-school graduation, college aspirations, and attendance.

ARTICLE HISTORY
Received 19 July 2022
Accepted 13 October 2023

KEYWORDS
Educational attainment; college aspirations; effort; expectations; locus of control; machine learning; PDS lasso

JEL CLASSIFICATION
D91; I21; I23; I24; I26

1. Introduction

A growing literature indicates that besides cognitive abilities, non-cognitive skills also play a crucial role in educational attainment (Almlund et al. 2011; Borghans et al. 2008). Locus of control (LoC), the subjective belief about the extent to which one's actions determine life outcomes, is one of the most widely studied non-cognitive skills. Individuals who believe that life outcomes are due to their decisions and behavior have an internal locus of control. In contrast, those who attribute those outcomes to external factors, like luck, or fate, have an external locus of control (Rotter 1966).[1] According to J. S. Coleman (1966), LoC is one of the most strongly related factors in a student's background to achievement.[2] A large body of evidence corroborates this statement (Almlund et al. 2011; Wang et al. 1999).

While the significance of LoC in educational attainment is well-documented, the specific mechanisms through which it operates remain less explored. M. Coleman and DeLeire (2003) and Cebi (2007) delved into the association between future expectations and LoC. Specifically, M. Coleman and DeLeire (2003) show that more internal LoC correlates with teenagers' more positive expectations about expected income and occupation at the age of 30, pointing to expectations as a pivotal channel. This study shows that effort is another channel through which LoC operates. The underlying premise is that when individuals believe that their actions do not significantly affect future outcomes, they tend to exert lower effort, leading to less favorable future results. This hypothesis is in line with Delaney, Harmon, and Ryan (2013), who document that conscientiousness [a Big Five personality trait that correlates positively with LoC (Judge et al. 2002)] is associated

CONTACT Ágnes Szabó-Morvai szabomorvai.agnes@krtk.hu, szabomorvai.agnes@krtk.hun-ren.hu KRTK KTI, Tóth Kálmán u. 4, Budapest 1097, Hungary; University of Debrecen, Egyetem tér 1, Debrecen 4032, Hungary
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/09645292.2023.2273220.
†Both the authors contributed equally.

with increased lecture attendance and study hours in college. To our knowledge, no existing study has directly examined the association between LoC, effort in academic activities, and the subsequent consequences on educational attainment.

Studying exogenous variability in LoC is challenging due to its observed stability (Cobb-Clark and Schurer 2013; Elkins, Kassenboehmer, and Schurer 2017). Significant shifts in LoC typically arise from major life events, such as prolonged health issues or the loss of a close family member. However, these events not only impact LoC but can also influence broader familial contexts and an individual's capabilities, potentially affecting educational outcomes. The related literature does its best to control for all possible confounding factors and thus get close to the causal effect. We add to the literature by utilizing machine learning to select the most appropriate controls.

We utilize data from the Hungarian Life Course Survey, which encompasses 10,000 adolescents. This survey offers information about educational attainment, detailing school-leaving age, high-school graduation, and college attendance. Additionally, it captures college aspirations and incorporates a section dedicated to LoC. The dataset also provides information on future expectations and variables associated with effort, such as diligence grades in school and the number of hours dedicated to studying, enabling us to explore these potential channels.

This study contributes to the literature on the relationship between LoC and educational attainment in two main ways. First, we investigate the association between LoC and four distinct educational outcomes. We find that LoC is an important determinant of school-leaving age, college aspirations, and college attendance even after controlling for socioeconomic background and cognitive abilities.[3] However, the landscape of educational outcomes is complex, with a wide array of factors potentially affecting those outcomes. Influences can range from individual attributes, such as cognitive abilities and non-cognitive skills (including LoC), to external factors like family background, peers, and school environment. Moreover, these factors often exhibit a strong correlation (as in our case, see Table 4), making it hard to establish causal relationships.[4] Thus, it is difficult to untangle the effect of LoC on educational outcomes. Possibly, any relationship between LoC and educational outcomes is merely correlational, with confounding factors (e.g. family background, cognitive ability) lurking behind the observed association. To ameliorate this problem of omitted variables, we consider more and more controls in the analysis, taking into account many potential confounders.

Second, we show that effort is a potential channel linking LoC to educational attainment. Previous studies, such as those by M. Coleman and DeLeire (2003) and Cebi (2007) proposed future expectations (viewed as a proxy for the expected rate of return of educational attainment) as a pathway through which LoC works. Consistent with findings in M. Coleman and DeLeire (2003), we observe that students with a higher internal LoC harbor brighter future expectations. Building on studies that explore the connection between LoC and effort in job searches (e.g. Caliendo, Cobb-Clark, and Uhlendorff 2015, or McGee 2015) or maintaining a healthy lifestyle (e.g. Cobb-Clark, Kassenboehmer, and Schurer 2014), we introduce effort as a previously undocumented mechanism in this context. We report that students with a higher internal LoC exert more effort in studying, which leads to higher achievement and aspirations. Moreover, we demonstrate that the effort channel is different from the expectation channel. Mediation analysis indicates that while expectations play a pivotal role in influencing school-leaving age, effort accounts for a larger portion of LoC's effect on other educational outcomes. While the existence of an association between LoC and the proposed mediating mechanisms related to educational outcomes does not by any means demonstrate that LoC affects educational outcomes, the absence of such associations would indeed be cause for concern.

Overall, our findings reveal a significant association between LoC and educational outcomes that remains robust in most instances even after introducing a comprehensive set of controls. Importantly, both future expectations and effort emerge as potential mediating mechanisms through which LoC may influence educational outcomes, with effort frequently standing out as the more dominant channel.

In Section 2, we review the literature. Section 3 introduces the data, followed by a description of our methodology in Section 4. The results are presented in Section 5. Section 6 concludes.

2. Related literature

In this section, we begin by reviewing the literature on the relationship between LoC, educational attainment, and college aspirations. We then explore the potential mediators of LoC's effects: expectations and effort.

Table 1 summarizes the main results of the literature.[5] As illustrated in column 3 (labeled 'Outcomes'), prior research has investigated the relationship between LoC and human capital investment decisions related to finishing high school (dropping out, choosing subjects required to go to university, graduating from high school), college aspirations, and university entrance and attendance.

Column 4 (labeled 'Controls') of Table 1 lists the controls employed in previous studies.[6] The first specifications in these studies incorporate exogenous variables, such as ethnicity/race and gender. Subsequent specifications include variables related to family background and cognitive abilities.

Column 5 (labeled 'Findings') in Table 1 highlights the main findings from prior studies.[7] Previous literature consistently finds that individuals with a higher internal LoC experience more positive high-school outcomes, such as later dropping out of school (Coneus, Gernandt, and Saam 2011), a higher probability of graduation (Barón and Cobb-Clark 2010; Cebi 2007; M. Coleman and DeLeire 2003), and the probability of following subject tracks required to enter university (Mendolia and Walker 2014). The results of Coneus, Gernandt, and Saam (2011) suggest that LoC plays a more significant role than GPA for dropout probability in specific periods of high school. A similarly important role of LoC is demonstrated for higher education outcomes. Kay, Shane, and Heckhausen (2016) find that internal LoC exhibits a positive association with university aspirations, even when family background and cognitive abilities are considered. Several previous studies find that LoC is associated with performance in university entrance exams and the probability of attending college (Barón and Cobb-Clark 2010; Cebi 2007; M. Coleman and DeLeire 2003).

In Figure 1, we visualize the findings of Table 1, and compare them to our results by indicating the 95% confidence interval of the LoC coefficients.[8] For clarity, 'M. Coleman and DeLeire (2003) – 1' indicates that we consider the coefficient in the first specification of the corresponding study. As more controls are added in subsequent specifications, the coefficient of LoC typically diminishes, and sometimes turns insignificant. The coefficient of our study, always presented last, is derived from our most exhaustive specification. Similarly to Cebi (2007), the positive association between internal LoC and high-school graduation turns insignificant once we consider a host of controls. Yet, the relationship between college attendance (or related outcomes) and internal LoC remains significant at 5% even after taking into account a wide array of factors, while in the other studies, the significance vanishes. Regarding university aspiration, our finding is in line with Kay, Shane, and Heckhausen (2016).

Almlund et al. (2011) emphasize the significance of the timing when measuring a trait (in our context, LoC) in relation to the outcome variable under examination. Since most studies are interested in the predictive power of LoC, most of them (except Barón and Cobb-Clark 2010) use LoC measures that were assessed before the measurement of the outcome variables. We follow this practice as well.

The body of literature exploring the channels through which LoC operates is relatively limited. M. Coleman and DeLeire (2003) show that expectations about the future serve as a pivotal channel for LoC. In their interpretation, more positive expectations reflect a higher subjective rate of return on human capital investment. As a consequence, students with a more internalized LoC are more likely to graduate from high school and attend college. On the other hand, Cebi (2007) also investigates the role of expectations, but she does not find a connection between LoC and future expectations.

Effort presents itself as another potential conduit through which LoC can influence outcomes. Psychologists often link LoC to motivation. Atkinson (1964) posits that motivation has two key elements: motive, and expectancy. The latter refers to an individual's judgment as to what extent their efforts and actions are causally related to desired results.[9] Therefore, motivation can be seen

Table 1. Summary of the main findings in the literature.

Study	Methods	Outcome	Controls	Finding
M. Coleman and DeLeire (2003)	Probit	Graduates from high school	(1) Hispanic, Black, Female, Urban, Region, (2) Math, Reading, GPA, Parents' education, (3) Parenting controls, (4) Family structure	A sd increase in LoC results in a 0.09 / 0.02** / 0.02** / 0.02** pp higher probability of outcome variable.
		Attends college	Same as above	A sd increase in LoC results in a 0.11** / 0.01** / 0.01 / 0.01 pp higher probability of outcome variable.
Cebi (2007)	Probit	Graduates from high school	(1) Black, Hispanic, Female, Urban, Age, Residence, (2) Parental education, (3) Family structure, (4) Home life, (5) AFQT	A sd increase in LoC results in 0.09*** / 0.08*** / 0.07*** / 0.07*** / 0.03 pp higher probability of outcome variable.
		Attends college	Same as above	A sd increase in LoC results in a 0.13*** / 0.1*** / 0.11*** / 0.1*** / 0.04* pp higher probability of outcome variable.
Barón and Cobb-Clark (2010)	Probit	Graduates from high school	Social disadvantage, Family structure, Male, Indigenous, Home environment, Parental education, Parent immigrant, Early born	A sd increase in LoC results in a 4.5* pp higher probability of outcome variable.
		Passes college entry exam	Same as above	A sd increase in LoC results in a 2.9** pp higher probability of outcome variable.
		College entrance rank	Same as above	A sd increase in LoC is associated with an increase of less than one (0.95*) percentile in one's college ranking.
Coneus, Gernandt, and Saam (2011)	Probit	Drops out at age 18 /19 /20	GPA, Mother LoC, Female, Family structure, Migration background, Mother education, Mother occupation, West	A sd increase in LoC results in a 1.9* / 2.8*** / 3.7*** pp higher probability of outcome variable at age 18 / 19 / 20.
Mendolia and Walker (2014)	OLS, Probit with propensity score matching	GCSE performance (Has 5+GCSE with A*-C, Has GCSEA*-C in English, Has GCSEA*-C in Maths)	(1) at-birth characteristics (birth weight, premature, ethnicity, gender, family characteristics), (2) other family's characteristics (child's or parent's disability, maternal education and employment status, single parent family, grandparents' education, family income and older siblings)	Being external decreases GCSE performance. Very significant (***) effect in both specifications for all the elements.
		Has A levels (overall, in Maths, Science, English)	Same as above	Being external decreases probability to have A levels. Very significant (***) effect in both specifications overall, ** for Maths and Science, not consistent for English.
		Points in A levels (overall, Maths, Science, English)	Same as above	Being external decreases test scores in A levels. Very significant (***) effect in both specifications overall, ** for Maths and English, * for Science.
		No. of facilitating subjects	Same as above	Being external decreases number of facilitating subjects. Very significant (***) effect in both specifications for all the elements.
Kay, Shane, and Heckhausen (2016)	Maximum likelihood structural equation modeling (SEM)	Education aspirations	(1) Female, High-school type, Grades, Living away from home, Parental education (2) Parental warmth, Parental interest, Parental involvement	Internal LoC positively associated with educational aspirations, β of 0.1** / 0.09**.
		University completion	(1) Female, High-school type, Grades, Living away from home, Parental education, Educational aspirations, Parental warmth, Parental interest, Parental involvement	External LoC not associated with educational aspirations, β of 0.01.

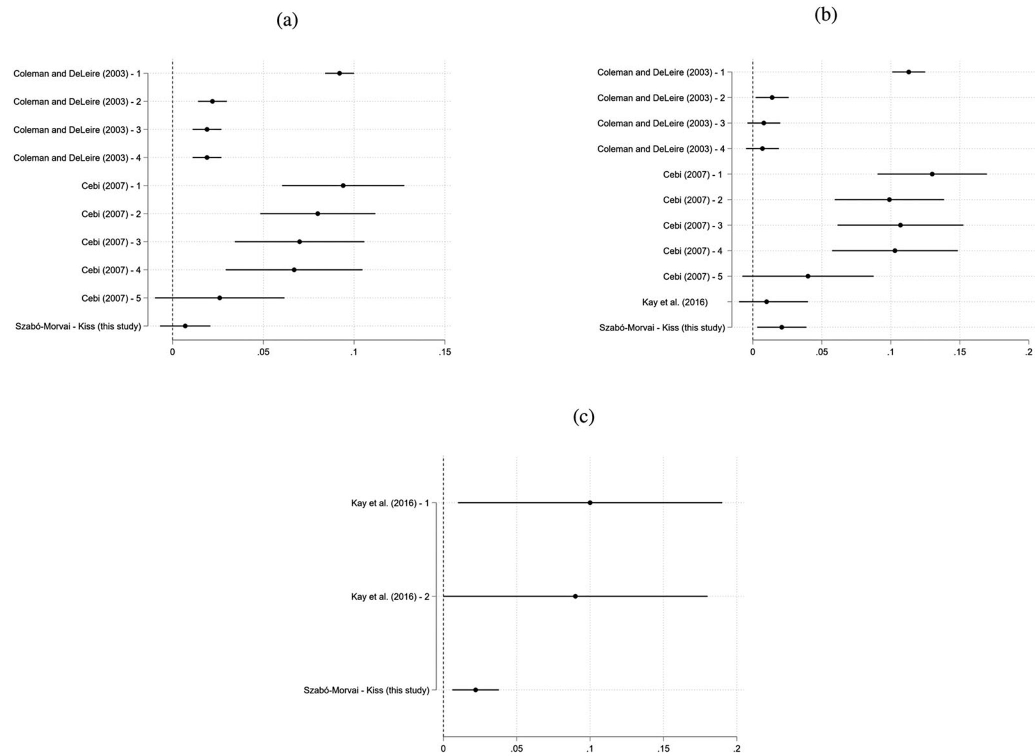

Figure 1. Main findings of the literature – the standardized coefficients of LoC with 95% confidence interval. (a) Graduates from high school. (b) Attends college / Passes college entry exam / university completion and (c) Education aspirations.

Note: The graphs include the most important findings of the previous literature that make a meaningful comparison to the results of the present study.

as a prerequisite for exerting effort. In a lab experiment, Borghans, Meijers, and Ter Weel (2008) show that internal LoC is associated with a higher motivation, which subsequently manifests as increased effort. Other studies also report effort as a potential link between LoC and academic endeavor. For instance, Mendolia and Walker (2014) hypothesize that the adverse effects of external LoC on educational outcomes may stem from the diminished effort displayed by such individuals.

3. Data

We use the six waves of a highly detailed longitudinal database, the Life Course Survey (Életpálya) conducted by the TÁRKI Social Research Institute in Hungary. A representative sample of approximately 10,000 adolescents was selected from the students who completed the 8th grade Hungarian National Assessment of Basic Competences (NABC) in May 2006.[10] The sample comprised students born in 1990 (9.8%), 1991 (64%), and 1992 (24%). However, due to sample attrition by the 6*th* wave in 2012, our panel shrank to 7638 students.[11] To preserve the representativeness of the sample despite the attrition, we use appropriate weighting. Furthermore, due to missing values of the dependent and the LoC-related variables, the sample size was reduced to 5101 participants. Our analysis consistently uses this reduced sample. The database contains answers to 4910 distinct questions, some of which were asked in each wave, while others were not.[12]

The Life Course Survey, in its 1st and 4th waves conducted in 2006 and 2009, integrated a LoC section. This section utilized a condensed version of the renowned Rotter test, similar to what the National Longitudinal Surveys of Youth (NLSY) employs. Participants were presented with the following pairs of statements and were asked to select the one that best describes their judgment about their own lives, the statements in italics indicating external LoC:

- 1A – What happens to me is first of all my own doing.
- *1B – Sometimes I feel that I don't have enough control over the direction my life is taking.*
- 2A – When I make plans, I am almost certain that I can make them work.
- *2B – It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.*
- 3A – In my case getting what I want has little or nothing to do with luck.
- *3B – Many times we might just as well decide what to do by flipping a coin.*
- 4A – Many times I feel that I have little influence over the things that happen to me.
- *4B – It is impossible for me to believe that chance or luck plays an important role in my life.*

During the 2006 test, respondents were predominantly 15 years old, and 18 years old during the 2009 test.[13] Between 60% to 80% of the respondents chose the answer indicating internal LoC tendencies for each item in both years. To construct our LoC measure, we use factor analysis like Mendolia and Walker (2014), Piatek and Pinger (2016), and Caliendo et al. (2020). We utilize the first factor and standardize the variable to zero mean and unit standard deviation.[14] Larger values indicate a more pronounced internal LoC.[15]

We study four outcome variables: school-leaving age (measured in years), high-school graduation (dummy variable, = 1 if graduated), college aspirations (dummy variable, = 1 if planned to apply to university), and college attendance (dummy variable, = 1 if attended college in any wave of the survey).[16] Our outcome measures align with existing literature. Specifically, high-school graduation and college attendance are also employed by M. Coleman and DeLeire (2003), Cebi (2007), and Barón and Cobb-Clark (2010). School-leaving age is tightly associated with the probability of dropping out at a specific age, as used by Coneus, Gernandt, and Saam (2011). Our binary measure of college aspirations is a simple version of the educational aspiration measure applied in Kay, Shane, and Heckhausen (2016) that offers four possible answers (university, technical college, apprenticeship, no aspirations).

The two main potential channels that we investigate in this study are expectations and effort. While the role of expectations in education has been explored (Cebi 2007; M. Coleman and DeLeire 2003), we are not aware of any studies on the importance of effort in academic outcomes.[17] We measure expectations with five questions from 2008. Respondents were asked to rate the probability that at the age of 35, (i) they would earn more money than the average, (ii) they would be in the top 10% of earners, (iii) they would secure a permanent job after finishing school, (iv) they would earn over HUF 100,000 (EUR 278) per month, and (v) they would earn above HUF 200,000 (EUR 556) per month.[18] We use factor analysis to derive a factor that captures expectations. The eigenvalue of the first factor is 3.81, and the Cronbach's alpha is 0.8.[19]

We assess effort in different ways. First, effort is measured with teacher-given grades on diligence in 2007, 2008, and 2009.[20] Additionally, we rely on responses to questions concerning the number of hours spent studying in a week and whether the individual studied after 8 PM on weekdays or during weekends, questions posed in 2007 and 2008. To further evaluate effort, we use factor analysis to derive a dedicated measure (with an eigenvalue of 2.52 and a Cronbach's alpha of 0.63). This variable is then standardized to have a zero mean and unit standard deviation.[21]

Our dataset includes variables on parental education, occupation, and household income. Additionally, it also contains information on the level of education of the grandparents, health, nationality, and languages spoken by the parents. There is also a rich collection of variables related to current and past individual characteristics. This includes questions related to babyhood and childhood (e.g. birth weight; breastfeeding duration; whether parents read fairy tales; and if they played board games), health (major diseases), self-evaluation, employment, future expectations, friends, lifestyle habits (e.g. exercise, smoking, alcohol or drug use), prejudices and political orientation.

The database offers comprehensive details about the home environment.[22] A widely used measure in empirical research is the HOME (Home Observation for Measurement of the Environment) scale, which is integrated into our dataset. This scale incorporates various metrics related to objects, activities, circumstances, and events at home that may play a role in adolescent development. In this survey, a tailored condensed version for young adolescents was administered, based on the National Longitudinal Survey of Youth (NLSY 2004).[23]

The database also contains data related to the school environment. This includes questions about the respondent's current school, academic performance, and schooling history – such as the age they began school and previous schools attended. Additional details cover class composition in terms of socioeconomic status, participation in extracurricular activities, parental involvement in schooling, attachment to the school, and patterns of absenteeism or dropping out.

Given the extensive range of variables in the database, there is a significant likelihood that at least one variable is missing for an individual. To avoid sample selection on missing variables, we impute missing values of control variables. For continuous variables, we use the mean of the non-missing observations, and for categorical variables, we employ the mode.[24] Furthermore, we introduce a binary variable to denote which observations were missing for each specific variable. Notably, we abstain from imputing values for the outcome variables and LoC.

3.1. Descriptive statistics

Table 2 presents the summary statistics for the main variables used in the analysis. Regarding the outcome variables, about 73% of our sample graduates from high school, 44% plan to pursue university studies, and approximately one-third of the sample attend university. Turning to LoC, our sample contains a high share of students with internal tendencies, two-thirds of the respondents scoring 3 or 4 points on the Rotter scale ranging from 0 to 4.[25] Previous literature indicates that family background variables strongly predict LoC. These variables have the expected sign in our dataset: a more favorable home environment and higher education level of the mother are positively associated with the level of internal tendencies in 2009 (see Table 4). Our findings related to the stability and determinants of LoC are in line with the literature.[26]

The middle panel in Table 2 displays the descriptive statistics of the pre-determined variables that were selected by the lasso method (see section 4 for details on lasso). These variables are pre-determined as they may affect LoC, but are not affected by LoC.[27] Most of the variables are related to aspects of family background. Thus, variables related to parental education and employment status are included. We have variability in parental education. While both education level and employment status are important determinants of income, the lasso method selects several variables that offer more direct insights into a household's financial status. Hence, household income, social disadvantage (such as eligibility for free books or free meals in school), and financial distress signals (like the inability to cover bills, food, or rent) represent facets that are chosen as relevant independent variables. The cognitive and emotional aspects of the home environment (HOME indicators, derived from the Early Adolescent version of the Home Observation for Measurement of the Environment of the NLSY, ranging from 0 to 130) are also selected by the lasso method. The selection underscores the importance of both emotional and cognitive dimensions. A very negative facet of the home environment (mental, physical, or sexual abuse before age 14) is also relevant for understanding the outcomes. The lasso technique also selects three items that are related to parental aspirations and (minimum) expectations regarding the educational level of their children, and the ability and capacity to finance the studies. About 66% state that ideally their child should go to university. While the previous question expresses aspirations, the database also contains information on the minimum expectations of the parents. Around 31% of parents expect their child to go to university. About 72% of the parents are able and/or willing to pay for their children's university studies.

Cognitive aspects are included in the HOME cognitive scale, but the lasso method also selects direct measures of cognitive abilities, namely, the respondents' reading and mathematics scores as measured by the National Assessment of Basic Competencies. One variable related to the school environment, the number of students in the class is chosen by the lasso method. Finally, being Roma (around 4% of our sample) is related to almost all previous variables, as Roma people tend to have lower education, worse employment status, lower income, disadvantaged home environment, modest educational aspirations/expectations for their children, lower cognitive scores, and worse school environment.

The bottom panel in Table 2 shows variables of future expectations and effort. The lasso method selected four expectations-related variables. Three of them express beliefs about relative and absolute earnings, while the last concerns future expectations about securing a permanent job. Regarding effort, five variables have been selected by the lasso method. Two of them measure the amount of time allocated to studying across different years. Another two variables are teacher-given grades that reflect the respondent's diligence and perseverance over various years. The last variable expresses extra dedication as most students do not typically study after 8 pm.[28][29]

Table 2. Descriptive statistics.

Outcome variables and LoC	Mean	Std. err.
School-leaving age	21.471	0.005
Graduates from high school	0.727	0.002
College aspiration	0.439	0.002
Attends college (2011, 2012)	0.349	0.002
Predetermined variables (included by lasso)		
LoC score in 2009	−0.004	0.004
Mother less than high school	0.457	0.002
Mother high school	0.337	0.002
Mother university	0.206	0.002
Father less than high school	0.673	0.002
Father high school	0.195	0.002
Father university	0.132	0.001
HOME cognitive scale	85.902	0.104
HOME emotional scale	99.610	0.092
Mental, physical or sexual abuse before age 14	1.418	0.010
Ideal education for the child: university (2006)	0.657	0.002
Minimum wanted education for the child: university (2006)	0.314	0.002
Able to pay for child's university (2006)	0.718	0.002
Reading score	−0.019	0.004
Mathematics score	0.000	0.004
# of students in the class	22.910	0.025
Household income (2006)	216836	644.798
Social disadvantage (2006)	0.401	0.002
Mother works (2007)	0.786	0.002
Mother works (2008)	0.788	0.002
Mother works (2009)	0.788	0.002
Father works (2006)	1.030	0.002
Financial distress (2009)	0.303	0.002
How often did the parents read tales from a book	17.817	0.036
Roma	0.042	0.001
Channel variables		
Exp: earn more than avg (2008)	0.554	0.001
Exp: earn more than net HUF100.000 (2008)	0.628	0.001
Exp: earn more than net HUF200.000 (2008)	0.347	0.001
Hours a week spent studying (2007)	4.675	0.012
Hours a week spent studying (2008)	4.488	0.013
Diligence grade (2007)	3.859	0.004
Diligence grade (2008)	3.810	0.004
Exp: permanent employment (2008)	0.698	0.001
Study after 8pm on weekdays (2008)	0.597	0.002
No. of observations	5101	

Table 3. Differences in means of the main outcomes for low- and high-LoC students.

	Mean (Low LoC)	Mean (High LoC)	T-test (*p*-value)
School-leaving age	21.285	21.450	0.000
Graduates from high school	0.614	0.701	0.000
College aspiration	0.335	0.421	0.000
Attends college	0.248	0.338	0.000

Note: Number of observations for each outcome: 5101.

Table 3 showcases stark differences in main outcomes based on students' LoC. Low-LoC students are those with a below-median LoC, and the high-LoC group is formed along the same lines. Students in the high-LoC group tend to leave school later, are more likely to graduate from high school and attend college, and have higher college aspirations. Moreover, these differences in the means are significant.

Table 4 reports the pairwise correlations between the main explanatory variables. A better family background that we proxy here with the mother's education and the cognitive and emotional aspects of the home environment (HOME cognitive and emotional scale) is positively associated with cognitive abilities (captured by the reading score in the NABC test), expectations, efforts, and also LoC. Cognitive abilities have a positive correlation with expectations, effort, and LoC. Moreover, expectations and efforts also exhibit a positive relationship, and both of them are positively correlated with LoC.[30] All the correlations are statistically significant at the 0.1% level. The positive association between LoC and the outcome variables may be due to correlation with the other factors. To isolate better the role of these factors we proceed with a regression analysis.

4. Empirical method

Economists often seek causal effects. However, it is challenging, if not impossible, to identify exogenous variability in the LoC suitable for causal analysis. LoC has been found to be relatively stable (Cobb-Clark and Schurer 2013; Elkins, Kassenboehmer, and Schurer 2017), changes occurring only after major life shocks like long-term health problems or loss of a close relative. However, such shocks do not only affect LoC but the broader family background or abilities of the individual which may itself influence educational outcomes. A second-best solution is to include the best possible explanatory variables to mitigate omitted variable bias. In the previous literature, several variables were used as controls to measure the association of LoC with educational attainment. Unless restricted by data availability, the specifications chosen reflect a professional judgment. Consequently, various studies use different sets of control variables (as shown in Table 1 in Section 2), indicating that there is no scientific consensus on this question. We circumvent this issue by making use of machine learning.

Table 4. Correlations between the main explanatory variables of interest.

	Mother education	HOME cognitive scale	HOME emotional scale	Reading test score	Expectations (2008)	Effort (2007-8-9)
HOME cogn.	0.522	1.000				
HOME emot.	0.107	0.262	1.000			
Reading test score	0.423	0.448	0.080	1.000		
Exp. (2008)	0.271	0.296	0.086	0.259	1.000	
Effort (2007-8-9)	0.250	0.261	0.096	0.349	0.152	1.000
LoC (2009)	0.061	0.095	0.038	0.091	0.138	0.078

Note: (a) Number of observations: 5101.
(b) Each correlation is significant at the 0.1% significance level.
(b) Cognitive abilities are proxied by the reading score in the NABC test.
(c) Expectations (2008) and Effort (2008) represent the factors that we constructed from our expectations and effort variables, respectively.

We employ the Post Double Selection (PDS) lasso method of Belloni et al. (2012) designed to estimate causal effects after a lasso variable selection procedure, using high-dimensional data. Lasso shares similarities with ordinary least squares (OLS), except that the minimand function of the optimization does not only include the residual sum of squares (RSS) but also a penalty term that increases with larger absolute values of the regression coefficients (see Equation (1)). In practice, this optimization method finds the curve that fits the data best, while maximizing parsimony, favoring a model with as few non-zero coefficients ($\beta \neq 0$) as possible.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j| \tag{1}$$

The lasso technique uses shrinkage, providing an efficient means to select a model with a limited number of variables, which performs the best out-of-sample prediction of the dependent variable (see James et al. 2013). Recently, the adoption of machine learning techniques for control variable selection has gained traction, as seen in works by Angrist and Frandsen (2019), Böheim and Stöllinger (2020), and Fluchtmann et al. (2020), among others. Although we do not claim our results to be causal, we aim to get as close to the causal population parameter as possible.

During the double selection, PDS lasso selects control variables that make the best out-of-sample prediction for the actual dependent variable (Y_i) in the first step, and for the LoC variable (LoC_i) in the second step. In the third step (see equation 2), a simple OLS regression is estimated using LoC and the union of the control variables selected in the first two steps (denoted as X_i).

$$Y_i = \alpha \mathrm{LoC}_i + X_i'\gamma + \xi_i \tag{2}$$

In the regressions, the timing of the variables is selected such that the control variables (X_i) refer to years 2006 to 2008, the LoC values are from 2009, and the outcome variables ($Y_{i,t+n}$) (graduating from high school, college aspirations, attending college) refer to 2009, 2011 or 2012. The school-leaving age is calculated using each year, from 2006 to 2012. The robust standard errors are clustered at the school level.

Ideally, our database would include all determinants of LoC as well as all the factors influencing the outcome variables. Additionally, the ideal timing of these data would be such that the explanatory variables predate the outcome variables. In our regressions, we use LoC as measured in 2009, so that we can include a wide range of factors that determine LoC, measured before 2009. If we were to use LoC measured in 2006, we could include only a limited number of such factors, measured in May 2006 or determined before 2006, such as gender or parental education.

In this setup, graduating from high school and college attendance are observed after the measurement of LoC in October 2009, as students were graduating in May 2010 and were observed as attending college even later. School-leaving age is measured throughout the six waves, including the years preceding 2009. College application plans were measured in 2009 on the same day as LoC. In all four cases, we claim to be measuring associations; however, our estimates are likely to be closer to the causal effects in the two outcomes (graduating from high school and college attendance) where the LoC measurement date precedes the outcome (Almlund et al. 2011; Piatek and Pinger 2016).

We study the association between LoC and the outcome variables, incorporating various control variables in our regressions. To identify the relationship between LoC and the outcome variables we rely on cross-sectional variation. In line with the literature, the first set of control variables (which we call pre-determined controls) are not affected by the student's LoC. These encompass parental education level and labor market status, home environment, birth weight, financial status, and the living circumstances of the family. Such variables are the predominant controls in the related literature (see Table 1). Second, we control for cognitive ability by including NABC mathematics and reading test scores. In specifications 4 through 6, we introduce expectations and effort –first individually, and then in tandem. These serve as potential channel variables that might mediate the association between LoC and the outcome variables.

We also carry out a formal mediation analysis, building upon the approach by Tubeuf, Jusot, and Bricard (2012). The total effect of LoC on the outcome variable can be divided: a portion is exerted directly (direct effect), while another part is channeled through mediator variables, namely effort or expectations (mediating effect). The total effect is measured by regression 2, where Y represents one of the educational outcome variables, α_1 is the total effect of LoC, and X embodies the vector of the explanatory variables of our main regression specifications. Subsequently, we estimate the following regression:

$$Y_i = \alpha_2 LoC_i + \beta E_i + X_i'\gamma + \xi_{2,i} \tag{3}$$

where E is either effort or expectations and α_2 is the direct effect of LoC on Y.[31]

5. Results

5.1. Main results

Table 5 presents the main regression results. For each outcome variable, we have six specifications. In each column, we extend the dictionary of variables and run a PDS lasso model to select the most relevant variables from each dictionary.[32]

5.1.1. Predetermined controls

Column 1 indicates the raw association between LoC and the outcome variable. This specification reveals that adolescents exhibiting stronger internal tendencies leave school later and are more likely to graduate from high school, have college aspirations, and attend college.

In Column 2, we add predetermined controls. A detailed look at the comprehensive regression tables (Tables H.15–H.18 in the Online Appendix H) reveals that most variables selected by the PDS lasso procedure are related to family background. Notably, HOME cognitive scale appears in the regressions for each outcome variable and is significant at 1%. Similarly, the variable about parental preferences on the ideal education level is selected for each outcome variable and is significant at 1%. However, other control variables are chosen only for some of the outcomes. For instance, the family's financial distress exhibits a strong negative association with school-leaving age but does not prove to be a relevant factor for the other outcomes.

As the row of selected controls in Table 5 indicates in each panel, the PDS lasso procedure selects 12 to 21 controls from the set of 119 predetermined variables in Column 2. The LoC coefficient exhibits a substantial reduction, ranging from 54% to 71% in all cases. Focusing on high-school graduation as the dependent variable, we observe that LoC remains significant at 5%. The coefficient reveals that a one standard deviation increase in LoC is associated with a 1.2 percentage point (corresponding to a 1.9% surge) higher probability of graduating from high school.[33] Similarly, a standard deviation increase in LoC correlates with a 2.4 percentage point (that is, a 6.6% increase) higher propensity for college aspirations.[34] Finally, a standard deviation increase in LoC is associated with a 2.6 percentage point (an 8.8% increase) higher probability of college attendance.[35]

While LoC turns out to be insignificant in determining the school-leaving age, it retains a 1% significance level for both college aspirations and attendance after accounting for these control variables. It is noteworthy that the inclusion of variables that largely predict LoC does not eliminate the association between LoC and three out of our four outcome variables.

5.1.2. Cognitive abilities

In Column 3, we extend the dictionary of variables with cognitive measures. Cebi (2007) points out that controlling appropriately for cognitive abilities may weaken or remove the significance of LoC. Our cognitive measures include reading and mathematics test scores in the NABC test, in addition to

Table 5. PDS Lasso models of LoC and educational attainment.

	Basic controls		Cognitive ability, Expectations and Effort			
	None (1)	Pre (2)	Pre + Cogn (3)	Pre + Cogn + Exp (4)	Pre + Cogn + Eff (5)	Pre + Cogn + Exp + Eff (6)
				A. School-leaving age		
LoC score in 2009	0.089***	0.026	0.016	0.003	0.019	0.017
	[0.020]	[0.016]	[0.016]	[0.015]	[0.013]	[0.013]
Observations	5101	5101	5101	5101	5101	5101
Clusters	1492	1492	1492	1492	1492	1492
Selected controls	0	21	18	19	23	26
Dictionary size	0	119	121	131	137	147
R-squared (in sample)	0.007	0.281	0.295	0.333	0.495	0.496
R-squared (out of sample)	0.018	0.343	0.335	0.396	0.564	0.574
				B. Graduates from high school		
LoC score in 2009	0.042***	0.012**	0.007	0.007	0.003	0.004
	[0.007]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Observations	5101	5101	5101	5101	5101	5101
Clusters	1492	1492	1492	1492	1492	1492
Selected controls	0	16	14	17	22	24
Dictionary size	0	119	121	131	137	147
R-squared (in sample)	0.009	0.412	0.447	0.448	0.472	0.473
R-squared (out of sample)	0.026	0.510	0.523	0.525	0.540	0.540
				C. College aspiration		
LoC score in 2009	0.057***	0.024***	0.018***	0.017***	0.011*	0.010*
	[0.008]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Observations	5101	5101	5101	5101	5101	5101
Clusters	1492	1492	1492	1492	1492	1492
Selected controls	0	15	14	16	18	20
Dictionary size	0	119	121	131	137	147
R-squared (in sample)	0.013	0.387	0.427	0.429	0.466	0.466
R-squared (out of sample)	0.011	0.421	0.474	0.483	0.496	0.504
				D. Attends college		
LoC score in 2009	0.057***	0.026***	0.020***	0.018***	0.015***	0.013**
	[0.008]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Observations	5101	5101	5101	5101	5101	5101
Clusters	1492	1492	1492	1492	1492	1492
Selected controls	0	12	11	12	15	17
Dictionary size	0	119	121	131	137	147
R-squared (in sample)	0.014	0.345	0.406	0.407	0.442	0.444
R-squared (out of sample)	0.005	0.409	0.462	0.464	0.505	0.507

Note: (a) Robust standard errors in parentheses. */**/*** denotes significance at 1 / 5 / 10 % level.
(b) The full regression tables are reported in Tables H.15–H.18 of the Online Appendix.
(c) The list of control variables are listed in Online Appendix D.
(d) We use appropriate population weights to preserve representativeness of the sample.

the cognitive aspects of the home environment already introduced in the previous specification. For each outcome variable, both scores are selected by the lasso method, and in each model, they are significant at least at the 5% level. Note that the number of controls included in the model drops in three cases even though we added two new items to the dictionary of variables. The coefficient of LoC decreases moderately in all cases. The significance level changes only in one case, as LoC ceases to be significant when predicting high-school graduation.[36] Even after the inclusion of cognitive abilities, LoC maintains a significant association with both college aspirations and college attendance at the 1% significance level.[37]

5.1.3. Controlling for expectations and effort

In Columns 4, 5, and 6, we explore the role of expectations and effort. In specification 4, we add expectations variables to the variable dictionary. In specification 5, we expand the variable dictionary of model 3 with the effort-related variables, but excluding expectations variables. In model 6, we include variables related to both channels in the variable dictionary of model 3.

In Column 4, three expectation variables are consistently chosen for each outcome. Two of these variables refer to the probability that the respondent's salary in their first job will be over HUF 100,000 / 200,000. The third variable reflects the perceived likelihood of earning a higher-than-average wage at the age of 35. Existing literature suggests that such positive expectations could indicate that the respondents have a more favorable view on the returns of human capital investments. Considering the school-leaving age, we observe a substantial drop in the coefficient of LoC after the inclusion of expectations, and it remains insignificant as in specification (3). For the other outcomes, the coefficient of LoC remains largely unchanged or decreases slightly and retains its significance at least at 5% for the college-related outcomes.

In Column 5, we specifically examine the influence of effort, distinguishing it from the expectations channel. Each of the effort variables is selected in at least one specification, and they are significant in most cases. Looking at the change in the LoC coefficients, expectations seem to matter more for school-leaving age, and less for the rest of outcomes, compared to effort. Effort is clearly a more relevant factor in understanding college aspirations as the significance of the LoC coefficient drops when we consider effort, relative to including expectations.[38]

When we consider both channels at the same time (see Column 6), the PDS lasso method consistently selects the future expectations variables that were chosen also in Column 4. From the effort variables, diligence grades in the years 2007 and 2008 and weekly study time are selected for each outcome variable. These effort variables tend to be significant more often than the future expectations variables. Compared to Column 5, we see only minor changes in the coefficient of LoC. The inclusion of both channels decreases the significance of LoC as a predictor of college attendance, suggesting that LoC exerts its impact by affecting both future expectations and effort.

Remarkably, even after accounting for both channels, the significance of LoC at 5% persists for college attendance. This indicates that LoC's influence on attendance is not solely through future expectations or effort.

Our hypotheses about the association between LoC and the outcomes are well-established in existing literature. Therefore, adjusting for multiple hypotheses testing seems unnecessary. Nevertheless, in Table J.20 of the Online Appendix J, we apply the Bonferroni correction to examine how it affects our findings. Without further controls, LoC continues to be significantly associated with all outcome variables. The correction increases *p*-values, and if pre-determined controls are added, then relative to the findings reported in Table 5 LoC ceases to be significant for school-leaving age. However, college aspirations and attendance remain significantly related to LoC even if cognitive abilities are accounted for. Moreover, the LoC coefficient remains at least marginal even after including the channels.

If we use bootstrapped standard errors, the results are very similar and even more significant, see Table K.21 in the Online Appendix K.

5.2. Channels: expectations and effort

In this section, we first investigate the association between LoC and the two proposed channels (expectations and effort) to see if these channels mediate between LoC and educational outcomes. Note that while we refer to expectations and effort as potential mediators between LoC and educational outcomes, LoC is measured in a later year. Thus, the causality could possibly go the other way around. However, previous research indicates that LoC is remarkably stable over time, and shifts in LoC are often associated with traumatic life events, such as the death of a close family member.[39] As a consequence, it is likely that LoC measured in 2009 represents a stable trait, that in turn shapes expectations and effort, possibly through affecting motivation (see Borghans, Meijers, and Ter Weel (2008)).

Table 6. LoC, expectations and effort.

	Dependent variable: Expectations (2008)			
	None (1)	+Pre-determined (2)	+Cognitive (3)	+Effort (4)
LoC score in 2006	0.119***	0.065***	0.062***	0.061***
	[0.015]	[0.014]	[0.015]	[0.014]
Observations	5101	5101	5101	5101
Clusters	1492	1492	1492	1492
Selected controls	0	14	13	19
Dictionary size	0	119	121	137
R-squared (in-sample)	0.017	0.164	0.170	0.237
R-squared (out-of-sample)	0.031	0.178	0.178	0.275
	Dependent variable: Effort (2007-8-9)			
	None (1)	+Pre-determined (2)	+Cognitive (3)	+Expectations (4)
LoC score in 2006	0.117***	0.056***	0.046***	0.042***
	[0.015]	[0.013]	[0.012]	[0.012]
Observations	5101	5101	5101	5101
Clusters	1492	1492	1492	1492
Selected controls	0	14	13	16
Dictionary size	0	119	121	131
R-squared (in-sample)	0.015	0.330	0.359	0.363
R-squared (out-of-sample)	0.026	0.350	0.358	0.359

Note: (a) Robust standard errors in parentheses. */**/*** denotes significance at 1/5/10 % level. Appropriate population weights used.
(b) The full regression tables are reported in Online Appendix L.
(c) The list of control variables are listed in Online Appendix D.
(d) Expectations (2008): Exp: earn more than avg (2008), Exp: earn best 10% (2008), Exp: permanent employment (2008), Exp: earn more than net HUF100.000 (2008), Exp: earn more than net HUF200.000 (2008)
(e) Effort (2007-8-9): Study time (2007-8), Night study (2007-8), Weekend study (2007-8), Diligence grade (2007-8-9)

To highlight the relationship between LoC and the channels, in Table 6 we report regression results, using either expectations or effort as a dependent variable. Following the logic of Table 5, we deploy more and more control variables in the different specifications.

Column 1 of Table 6 presents a univariate regression with LoC as the sole explanatory variable. In Column 2, we add pre-determined controls; in Column 3, cognitive controls are also included. In Column 4, we introduce the channel not being used as the dependent variable. Thus, in the regression explaining expectations, we include effort as an independent variable and vice versa. This approach helps us determine whether the two proposed channels mirror the same underlying mechanism.

We observe a strong association between LoC and future expectations, even after controlling for pre-determined variables, cognitive abilities, and effort. This finding is in line with M. Coleman and DeLeire (2003), who argue that an important channel through which LoC operates is future expectations.[40] The regressions on effort show a very similar picture. LoC correlates strongly with both effort factor variables, even after including pre-determined variables, cognitive variables, and expectations. Overall, there is strong evidence that individuals with a more internal LoC are more likely to exert effort. The point estimate for the effect of LoC on effort (from 2007–2009) in Column 3 indicates that, after adjusting for pre-determined variables and cognitive ability, a one standard deviation surge in LoC corresponds to an increase in effort by 0.047. This represents a 5.3% increase in the standard deviation of effort.

The significant association of LoC with effort (or with expectations) remains even when controlling for expectations (or effort, respectively). It is a strong indication that the two channels are different. It suggests that those who exert more effort do not do so mainly because they have more positive future expectations and vice versa. To confirm that these are two separate channels, we compute correlations between effort and expectations (see Table L.22 in the Online Appendix L).

Table 7. Formal mediation analysis.

	A. Effort			
	Total effect (1)	Direct effect (2)	Mediating effect (3) =(2)-(1)	% of Mediating effect (4) =(3)/(2)
School-leaving age	0.016	0.011	0.005	31.3%
	(0.016)	(0.016)		
Graduates from high school	0.007	0.003	0.004	57.1%
	(0.006)	(0.006)		
College aspiration	0.018***	0.011*	0.007	38.9%
	(0.006)	(0.006)		
Attends college	0.020***	0.014**	0.006	30.0%
	(0.006)	(0.006)		
	B. Expectations			
	Total effect	Direct effect	Mediating effect	% of Mediating effect
School-leaving age	0.016	0.000	0.016	100.0%
	(0.016)	(0.015)		
Graduates from high school	0.007	0.006	0.001	14.3%
	(0.006)	(0.006)		
College aspiration	0.018***	0.016***	0.002	11.1%
	(0.006)	(0.006)		
Attends college	0.020***	0.018***	0.002	10.0%
	(0.006)	(0.006)		

Note: (a) In each case, the number of observations is 5101.
(b) Robust standard errors in parentheses. */**/*** denotes significance at 1 / 5 / 10 % level.
(c) The explanatory variables of Model (1) are identical to those of Model (3) in Table 5. The differences of the point estimates are the result of the difference of pdslasso and reg commands of Stata.

The correlation among expectations variables is mostly high (up to 0.89), similar to effort variables (up to 0.85). In contrast, the correlation between effort and expectations variables is relatively low (ranging from 0.00 to 0.17 in absolute value). Moreover, the correlation between the effort factor and the expectations factor variables is only 0.15. In sum, this correlation pattern indicates a weak link between the two channels.

Finally, we turn to the mediation analysis. As explained in Section 4, the mediating effect can be calculated as total effect - direct effect. The resulting estimates are reported in Table 7. The effort channel mediates 30.0 to 57.1% of the total effect, whereas 10.0 to 100% is mediated by the expectations channel. While the mediating effect of expectations is relatively higher compared to effort when considering school-leaving age, for the remaining three outcomes, the effort channel seems to be more important.

Our findings are consistent when we simply test for the difference of the LoC parameter estimates in Models 3, 4, 5, and 6 in Table 5. Specifically, when the dependent variable is school-leaving age and expectations are added in Model 3, we observe a significant shift in the LoC point estimates. However, for all other outcomes, it is incorporating effort into Model 3 that significantly alters the LoC point estimates.

6. Conclusion

This study explores a detailed database from Hungary and studies how locus of control correlates with educational attainment and college aspirations with a special focus on the potential channels. We are the first to show in the realm of human capital investment decisions that LoC may operate through effort: students with a more internal LoC tend to exert more effort in studying. We also find that after controlling for the pre-determined factors and cognitive ability, a one standard deviation rise in LoC is associated with an increase of 1.8 percentage points (or 5.4%) in the probability of harboring college aspirations. Similarly, there is a 2.0 percentage points (or 8.3%) surge in the probability of attending college. To provide context for these findings, the magnitude of these associations are comparable to an increase of math test scores by 27% and 23% of a standard deviation, respectively.[41] At the same time, LoC does not correlate significantly with the probability of graduating from high school and school-leaving age, once pre-determined factors and cognitive abilities are taken into account. The point estimates become marginally significant after including expectations and effort in the regressions of college application plans. Wald tests confirm that the coefficients between these specifications are different. This result suggests that these are important channels between LoC and this outcome. Remarkably, the coefficient of LoC remains significant even after considering the expectations and effort channel in the regressions of college attendance, indicating that LoC's influence is beyond these two channels.

While we do not assert that our results provide unbiased estimates of causal effects, it is plausible that they closely approximate these effects, especially regarding high-school graduation and college attendance. This is because these outcomes are not affected by a potential contemporaneity issue (Almlund et al. 2011; Piatek and Pinger 2016).[42] Even if we accept that there is a causal relationship between LoC and educational outcomes, leveraging this relationship for practical interventions remains ambiguous. Existing literature suggests that LoC is a stable trait, so interventions are unlikely to be able to modify it.[43] However, understanding the mechanisms through which LoC operates can provide insights for potential intervention strategies. For instance, affecting future expectations may lead to better educational outcomes. Alan and Ertac (2018) is a good example of how to encourage forward-looking behavior. There is also some evidence that effort, the other mechanism that we study, is malleable in childhood. Alan, Boneva, and Ertac (2019) show that interventions on grit (defined as perseverance in a productive task) can be nurtured among elementary school students. Notably, these interventions emphasized the pivotal role of effort in attaining goals. Consequently, while LoC's core might be stable, the channels through which it impacts outcomes appear malleable. It remains to be seen if such interventions correlate with LoC and educational outcomes.

Notes

1. Locus of control does not capture the role of other people as a determinant of what happens to somebody, as it is clear from the questions used to measure it, see Table C.10 in Online Appendix C. We are grateful to an anonymous reviewer for pointing out this issue.
2. Coleman did not call it LoC but referred to it as 'an attitude which indicated the degree to which the student felt in control of his fate'.
3. In univariate regressions LoC associates positively with all outcome variables, including high-school graduation, at the 1% significance level.
4. While the family background is exogenous to the individual, cognitive and non-cognitive abilities are intricately intertwined and affected by family background.
5. Online Appendix A summarizes the data sets used for the studies presented in Table 1.
6. The numbers in parentheses indicate which specification utilized which variables. For instance, in M. Coleman and DeLeire (2003) the variables following (1) and before (2) are used in specification 1. The variables following (2) (and before (3)) are the ones that the authors use in specification 2 in addition to the previous variables, and so on.
7. The different coefficients correspond to different specifications. For instance, in Cebi (2007), the first coefficient (5.4***) corresponds to specification 1, the next one (4.6***) to specification 2, and so on.
8. We could not include Barón and Cobb-Clark (2010) in Figure 1, as we did not find standard errors or confidence interval in their study. Coneus, Gernandt, and Saam (2011) study whether students drop out at a certain age, while we investigate at which age they leave school, so we could not prepare a meaningful comparison for dropout.
9. Bandura (1989) also links motivation to the effort individuals are willing to make.
10. The NABC is a nationwide low-stake test in reading and maths, similar to the PISA test, see Sinka (2010) for details.
11. Those dropping out of the sample have a bit lower socioeconomic status compared to those who stay in the sample until the 6th wave. 13.3% of them have a high-educated mother (vs. 17.4% of the stayers) and 9.2% (vs. 10.8%) have a high-educated father. Table B.9 in Online Appendix B shows the structure of the data collection with information on when the key variables that we use in our analysis were measured.
12. Table D.11 in Online Appendix D presents the dictionary of the variables used in the analysis, indicating that we use 147 variables.

13. Table C.10 in Online Appendix C contains the LoC questions that we use, the 4-question version of the Rotter-test with the valuation of the answers in brackets. Table A.8 in Online Appendix A shows that this test is often used in studies with large samples.
14. Factor analysis extracts maximum common variance from all variables and puts them into a common score, this is the first factor (UCLA 2023).
15. The eigenvalue of factor 1 is 0.64, and Cronbach's alpha is 0.48. It is not uncommon in the literature that in the case of short scales (scales with less than 5 items) Cronbach's alpha is around 0.5. Nevertheless, Cronbach (1951) suggests that a high alpha is desirable if a score is assigned to an individual, otherwise the key point is that the instrument should be interpretable. See Taber (2018) for a thorough discussion on this point.
16. In each wave, we know if the student still attends school or quit studying. From this information, we can calculate the school-leaving age. If a student in the sample started university after 2012 (aged 20 years or older), then we do not see it.
17. As noted earlier, the association between effort and LoC has been investigated in other domains (Caliendo, Cobb-Clark, and Uhlendorff 2015; Cobb-Clark, Kassenboehmer, and Schurer 2014; McGee 2015).
18. The corresponding current amounts in USD are 338 and 676. In 2008, HUF 200,000 was considered a high salary.
19. We show the graphical representation of the factor loadings for expectations in Figure F.3 of the Online Appendix F.1.
20. In Hungary, students receive grades on two general criteria: behavior (reflecting their conduct in school) and diligence (indicating the effort they put forth in school). The latter is closely tied to the concept of effort. Given that this is a grade assigned by the teacher, it is susceptible to subjectivity. A burgeoning body of literature addresses biases in teachers' assessments of their students. Such biases are observed in relation to ethnicity (Burgess and Greaves 2013; Kisfalusi, Janky, and Takács 2021; Pigott and Cowen 2000), gender (Hinnant, O'Brien, and Ghazarian 2009; Holder and Kessels 2017; Lazarides and Watt 2015), and socioeconomic status (Auwarter and Aruguete 2008; Ready and Chu 2015; Ready and Wright 2011), among other factors. Several recent studies specifically investigate these biases in relation to diligence, effort, and learning motivation (Brandmiller, Dumont, and Becker 2020; Lorenz 2021; Tobisch and Dresel 2017). This emerging literature consistently reports biases. Consequently, we acknowledge that teacher-assigned diligence grades might contain subjective elements influenced by ethnicity, gender, and socioeconomic status, and could be biased.
21. We show the graphical representation of the factor loadings for expectations in Figures F.4 and F.5 of the Online Appendix F.1.
22. According to the developmental psychology literature, psychological and physiological development of the children is strongly related to stimuli in the home environment (Davis-Kean 2005; Sarsour et al. 2011; Strauss and Knight 1999).
23. The elements of the scale are described in Table F.13 of the Online Appendix F.2.
24. Our results are robust to imputing conditional means based on basic characteristics.
25. See Table C.10 in Online Appendix C for details.
26. LoC is a stable trait, as the distribution of the change between 2006 and 2009 is bell-shaped, peaking at zero. There is no change in LoC for 15% of the sample, and 56% of the changes are at most 0.25 standard deviation. Only 0.78% (0.75%) of the individuals experience the maximum downward (upward) change, see Figure C.2 in Online Appendix C. These findings are in line with Cobb-Clark and Schurer (2013) and Elkins, Kassenboehmer, and Schurer (2017).
27. Moreover, most of these variables are measured before LoC.
28. Studying past 8 pm can be interpreted in various ways. For one, it might indicate poor time management if the student is disorganized. Alternatively, the student might be engaged in numerous activities, necessitating evening study sessions. A third interpretation could be that the student, due to lower cognitive abilities, requires more hours to study. We cannot definitively determine the exact reason behind studying past 8 pm. In our regressions, this variable consistently displays a significant positive coefficient, suggesting it may represent a form of effort. We are grateful to an anonymous reviewer for pointing out this issue.
29. Table E.12 in Online Appendix E indicates how the outcome variables are associated with the mother's education, cognitive abilities, channels, and LoC. Having more educated mothers / better cognitive skills / more positive future expectations / more effort / more internal LoC correlate clearly with better outcomes.
30. It is noteworthy that the associations with LoC are generally the weakest.
31. In these models, we use the Effort(2007-8-9) and Expectations(2008) variables generated with principal component analysis.
32. In Table G.14 in the Online Appendix G we show an alternative set of regressions, where the model specifications are nested in the sense that consecutive models incorporate each variable of the previous models. In this case, we perform PDS lasso estimation with the variable dictionary of Model 6 for each outcome. Then, going backward, we include relevant subsets of the variables selected in Model 6. Our results do not change significantly.
33. This is very similar to the 1.4 percentage points finding in the most closely related specification in M. Coleman and DeLeire (2003) and lower than the 3.8 and 4.5 percentage points increase documented by Cebi (2007) and Barón and Cobb-Clark (2010), respectively.

34. Kay, Shane, and Heckhausen (2016) also find that internal/external LoC associates positively/negatively with college aspirations. Their non-standardized coefficients do not allow a direct comparison between their model and ours.
35. That is higher than the non-significant 0.5 percentage points reported by M. Coleman and DeLeire (2003). Still, lower than the significant 6 percentage points documented by Cebi (2007).
36. In Cebi (2007) it is also at this stage when LoC becomes insignificant. Moreover, the size of the coefficients is similar (1 vs 1.5 percentage points).
37. Our results are akin to Cebi (2007)'s findings if we consider college attendance, as she reports a significant coefficient of 2.3 percentage points, similar in magnitude to our 1.9 percentage points.
38. In Table I.19 of the Online Appendix I we report the *p*-values obtained from Wald-tests, after testing for the difference of the coefficient on LoC in the different specifications. The test results suggest that adding expectations to the model does not matter in the case of high-school graduation. Incorporating effort does not seem to be influential for school-leaving age. The comparisons of Models 3 vs. 4 and Models 5 vs. 6 inform us about how the coefficient estimate of LoC changes after including expectations variables. Whereas the comparisons of Models 3 vs. 5 and Models 4 vs. 6 show the impact of adding effort variables similarly.
39. On the stability of LoC in general, see Cobb-Clark and Schurer (2013), and with a focus on the adolescence, consider Elkins, Kassenboehmer, and Schurer (2017). In Subsection 3.1 we indicated that LoC is stable in our data as well, see footnote 26.
40. Caliendo et al. (2020) show the relevance of the expectations channel in labor economics.
41. The coefficients of the math test score in the same specifications are 0.066***, 0.086***, as can be seen in the full regression Tables H.15–H.18 of the Online Appendix H.
42. Piatek and Pinger (2016) find that if contemporaneous measurements of LoC are used along with outcome measures, then the coefficients are larger and more precisely estimated, suggesting that LoC is an important determinant of the outcome (in their case, wages). However, when previously measured LoC is used, the significant association vanishes. Almlund et al. (2011) state that contemporaneity is problematic because we do not know which way the influences go, that is, does the trait affect the outcome, the other way around, or do they mutually influence each other.
43. There are some studies claiming that interventions may alter LoC (Huang and Ford 2012; Wolinsky et al. 2010). Clearly, more research is needed to understand the malleability of LoC.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Hungarian National Research, Development and Innovation Office NKFIH (K119683 and FK131422) and the Lendület programme of the Hungarian Academy of Sciences (grant number: LP2018-2/2018, LP2021-2). The funding source had no involvement in the study design; data handling; writing and submission decision.

Credit author statement

Ágnes Szabó-Morvai: Methodology; Software; Validation; Formal analysis; Investigation; Writing – Review & Editing; Visualization; Funding acquisition **Hubert János Kiss:** Conceptualization; Writing - Original Draft; Writing – Review & Editing; Funding acquisition

Data availability statement

The data is available upon request from the data owner, TÁRKI (https://tarki.hu/). The Stata codes are available on Github (https://github.com/szabomorvai).

References

Alan, S., and S. Ertac. 2018. "Fostering Patience in the Classroom: Results From Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911. https://doi.org/10.1086/699007.

Alan, S., T. Boneva, and S. Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results From a Randomized Educational Intervention on Grit." *The Quarterly Journal of Economics* 134 (3): 1121–1162.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." In Vol. 4 of *Handbook of the Economics of Education*, 1–181. Amsterdam: Elsevier.

Angrist, J., and B. Frandsen. 2019. *Machine labor*. Tech. rep., National Bureau of Economic Research.

Atkinson, J. W. 1964. *An Introduction to Motivation*. Princeton, NJ: van Nostrand.

Auwarter, A. E., and M. S. Aruguete. 2008. "Effects of Student Gender and Socioeconomic Status on Teacher Perceptions." *The Journal of Educational Research* 101 (4): 242–246. https://doi.org/10.3200/JOER.101.4.243-246.

Bandura, A. 1989. "Human Agency in Social Cognitive Theory." *American Psychologist* 44 (9): 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175.

Barón, J. D., and D. A. Cobb-Clark. 2010. "Are Young People's Educational Outcomes Linked to Their Sense of Control?" SSRN 1594792.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With An Application to Eminent Domain." *Econometrica* 80 (6): 2369–2429. https://doi.org/10.3982/ECTA9626.

Böheim, R., and P. Stöllinger. 2020. "Decomposition of the Gender Wage Gap Using the Lasso Estimator." *Applied Economics Letters* 28 (10): 1–12.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4): 972–1059. https://doi.org/10.1353/jhr.2008.0017.

Borghans, L., H. Meijers, and B. Ter Weel. 2008. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46 (1): 2–12. https://doi.org/10.1111/ecin.2008.46.issue-1.

Brandmiller, C., H. Dumont, and M. Becker. 2020. "Teacher Perceptions of Learning Motivation and Classroom Behavior: The Role of Student Characteristics." *Contemporary Educational Psychology*63:101893. https://doi.org/10.1016/j.cedpsych.2020.101893.

Burgess, S., and E. Greaves. 2013. "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities." *Journal of Labor Economics* 31 (3): 535–576. https://doi.org/10.1086/669340.

Caliendo, M., D. A. Cobb-Clark, C. Obst, H. Seitz, and A. Uhlendorff. 2020. "Locus of Control and Investment in Training." *Journal of Human Resources* 57 (4): 1311–1349.

Caliendo, M., D. A. Cobb-Clark, and A. Uhlendorff. 2015. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics* 97 (1): 88–103. https://doi.org/10.1162/REST_a_00459.

Cebi, M. 2007. "Locus of Control and Human Capital Investment Revisited." *Journal of Human Resources*42 (4): 919–932. https://doi.org/10.3368/jhr.XLII.4.919.

Cobb-Clark, D. A., S. C. Kassenboehmer, and S. Schurer. 2014. "Healthy Habits: The Connection Between Diet, Exercise, and Locus of Control." *Journal of Economic Behavior & Organization* 98:1–28. https://doi.org/10.1016/j.jebo.2013.10.011.

Cobb-Clark, D. A., and S. Schurer. 2013. "Two Economists' Musings on the Stability of Locus of Control." *The Economic Journal* 123 (570): F358–F400. https://doi.org/10.1111/ecoj.12069.

Coleman, J. S. 1966. "Equal Schools Or Equal Students?" *The Public Interest* 4:70–75.

Coleman, M., and T. DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Investment Decision." *Journal of Human Resources* 38 (3): 701–721. https://doi.org/10.2307/1558773.

Coneus, K., J. Gernandt, and M. Saam. 2011. "Noncognitive Skills, School Achievements and Educational Dropout." *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift Für Wirtschafts- Und Sozialwissenschaften* 131 (4): 547–568. https://doi.org/10.3790/schm.131.4.547.

Cronbach, L. J. 1951, September. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika*16 (3): 297–334. https://doi.org/10.1007/BF02310555.

Davis-Kean, P. E. 2005. "The Influence of Parent Education and Family Income on Child Achievement: the Indirect Role of Parental Expectations and the Home Environment." *Journal of Family Psychology*19 (2): 294–304. https://doi.org/10.1037/0893-3200.19.2.294.

Delaney, L., C. Harmon, and M. Ryan. 2013. "The Role of Noncognitive Traits in Undergraduate Study Behaviours." *Economics of Education Review* 32:181–195. https://doi.org/10.1016/j.econedurev.2012.07.009.

Elkins, R. K., S. C. Kassenboehmer, and S. Schurer. 2017. "The Stability of Personality Traits in Adolescence and Young Adulthood." *Journal of Economic Psychology* 60:37–52. https://doi.org/10.1016/j.joep.2016.12.005.

Fluchtmann, J., A. M. Glenny, N. Harmon, and J. Maibom. 2020. *The Gender Application Gap: Do Men and Women Apply for the Same Jobs?* Tech. rep., Mimeo University of Copenhagen.

Hinnant, J. B., M. O'Brien, and S. R. Ghazarian. 2009. "The Longitudinal Relations of Teacher Expectations to Achievement in the Early School Years." *Journal of Educational Psychology* 101 (3): 662–670. https://doi.org/10.1037/a0014306.

Holder, K., and U. Kessels. 2017. "Gender and Ethnic Stereotypes in Student Teachers' Judgments: A New Look From a Shifting Standards Perspective." *Social Psychology of Education* 20 (3): 471–490. https://doi.org/10.1007/s11218-017-9384-z.

Huang, J. L., and J. K. Ford. 2012. "Driving Locus of Control and Driving Behaviors: Inducing Change Through Driver Training." *Transportation Research Part F: Traffic Psychology and Behaviour* 15 (3): 358–368. https://doi.org/10.1016/j.trf.2011.09.002.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. New York: Springer.

Judge, T. A., A. Erez, J. E. Bono, and C. J. Thoresen. 2002. "Are Measures of Self-esteem, Neuroticism, Locus of Control, and Generalized Self-efficacy Indicators of a Common Core Construct?" *Journal of Personality and Social Psychology* 83 (3): 693–710. https://doi.org/10.1037/0022-3514.83.3.693.

Kay, J. S., J. Shane, and J. Heckhausen. 2016. "High-school Predictors of University Achievement: Youths' Self-reported Relationships with Parents, Beliefs About Success, and University Aspirations." *Journal of Adolescence* 53 (1): 95–106. https://doi.org/10.1016/j.adolescence.2016.08.014.

Kisfalusi, D., B. Janky, and K. Takács. 2021. "Grading in Hungarian Primary Schools: Mechanisms of Ethnic Discrimination Against Roma Students." *European Sociological Review* 37 (6): 899–917. https://doi.org/10.1093/esr/jcab023.

Lazarides, R., and H. M. Watt. 2015. "Girls' and Boys' Perceived Mathematics Teacher Beliefs, Classroom Learning Environments and Mathematical Career Intentions." *Contemporary Educational Psychology*41:51–61. https://doi.org/10.1016/j.cedpsych.2014.11.005.

Lorenz, G. 2021. "Subtle Discrimination: Do Stereotypes Among Teachers Trigger Bias in Their Expectations and Widen Ethnic Achievement Gaps?" *Social Psychology of Education* 24 (2): 537–571. https://doi.org/10.1007/s11218-021-09615-0.

McGee, A. D. 2015. "How the Perception of Control Influences Unemployed Job Search." *ILR Review*68 (1): 184–211. https://doi.org/10.1177/0019793914556245.

Mendolia, S., and I. Walker. 2014. "The Effect of Personality Traits on Subject Choice and Performance in High School: Evidence From An English Cohort." *Economics of Education Review* 43:47–65. https://doi.org/10.1016/j.econedurev.2014.09.004.

NLSY. 2004. *Nlsy79 Child and Young Adult Data Users Guide*. https://www.nlsinfo.org/content/cohorts/NLSY79.

Piatek, R., and P. Pinger. 2016. "Maintaining (locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models." *Journal of Applied Econometrics* 31 (4): 734–755. https://doi.org/10.1002/jae.v31.4.

Pigott, R. L., and E. L. Cowen. 2000. "Teacher Race, Child Race, Racial Congruence, and Teacher Ratings of Children's School Adjustment." *Journal of School Psychology* 38 (2): 177–195. https://doi.org/10.1016/S0022-4405(99)00041-2.

Ready, D. D., and E. M. Chu. 2015. "Sociodemographic Inequality in Early Literacy Development: The Role of Teacher Perceptual Accuracy." *Early Education and Development* 26 (7): 970–987. https://doi.org/10.1080/10409289.2015.1004516.

Ready, D. D., and D. L. Wright. 2011. "Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities: The Role of Child Background and Classroom Context." *American Educational Research Journal* 48 (2): 335–360. https://doi.org/10.3102/0002831210374874.

Rotter, J. B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs: General and Applied* 80 (1): 1–28. https://doi.org/10.1037/h0092976.

Sarsour, K., M. Sheridan, D. Jutte, A. Nuru-Jeter, S. Hinshaw, and W. T. Boyce. 2011. "Family Socioeconomic Status and Child Executive Functions: The Roles of Language, Home Environment, and Single Parenthood." *Journal of the International Neuropsychological Society: JINS* 17 (1): 120–132. https://doi.org/10.1017/S1355617710001335.

Sinka, E. 2010. *Hungary – Country Background Report*. Tech. rep., OECD.

Strauss, R. S., and J. Knight. 1999. "Influence of the Home Environment on the Development of Obesity in Children." *Pediatrics* 103 (6): e85–e85. https://doi.org/10.1542/peds.103.6.e85.

Taber, K. S. 2018, December. "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education." *Research in Science Education* 48 (6): 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Tobisch, A., and M. Dresel. 2017. "Negatively Or Positively Biased? Dependencies of Teachers' Judgments and Expectations Based on Students' Ethnic and Social Backgrounds." *Social Psychology of Education*20 (4): 731–752. https://doi.org/10.1007/s11218-017-9392-z.

Tubeuf, S., F. Jusot, and D. Bricard. 2012, June. "Mediating Role of Education and Lifestyles in the Relationship Between Early-life Conditions and Health: Evidence From the 1958 British Cohort." *Health Economics* 21 (S1): 129–150. https://doi.org/10.1002/hec.v21.S1.

UCLA. 2023. *A Practical Introduction To Factor Analysis: Exploratory Factor Analysis*. https://stats.oarc.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/.

Wang, L.-Y., E. Kick, J. Fraser, and T. J. Burns. 1999. "Status Attainment in America: The Roles of Locus of Control and Self-esteem in Educational and Occupational Outcomes." *Sociological Spectrum*19 (3): 281–298. https://doi.org/10.1080/027321799280163.

Wolinsky, F. D., M. W. Vander Weg, R. Martin, F. W. Unverzagt, S. L. Willis, M. Marsiske, G. W. Rebok, et al. 2010. "Does Cognitive Training Improve Internal Locus of Control Among Older Adults?" *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 65 (5): 591–598. https://doi.org/10.1093/geronb/gbp117.