

Chávez Juárez, Florian Wendelspiess

Article

Intergenerational transmission of education: The relative importance of transmission channels

Latin American Economic Review

Provided in Cooperation with:

Centro de Investigación y Docencia Económica (CIDE), Mexico City

Suggested Citation: Chávez Juárez, Florian Wendelspiess (2015) : Intergenerational transmission of education: The relative importance of transmission channels, Latin American Economic Review, ISSN 2196-436X, Springer, Heidelberg, Vol. 24, Iss. 1, pp. 1-44, <https://doi.org/10.1007/s40503-014-0015-1>

This Version is available at:

<https://hdl.handle.net/10419/108961>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Intergenerational transmission of education: the relative importance of transmission channels

Florian Wendelspiess Chávez Juárez

Received: 24 September 2013 / Revised: 4 September 2014 / Accepted: 12 November 2014
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract This paper aims at quantifying the relative importance of different transmission channels generating the high levels of intergenerational correlations in education, especially in Latin America. A simultaneous equations model is applied to rich survey data from Mexico. The results show that the economic situation of the family has the highest impact, even more than heritability of cognitive abilities. The long-run economic situation seems to matter more than the current consumption level. Parental education affects the schooling outcome directly but also indirectly through the economic situation, which is particularly true for the father.

Keywords Intergenerational transmission of education · Social mobility · IQ transmission · Inequality · Mexico

JEL Classification: D31 · I21 · I24 · I62

1 Introduction

Education is a main ingredient for a successful life in modern societies. However, in many countries the opportunities to get well educated strongly depend on the family background, particularly in Latin America. As a result, we observe very low intergenerational mobility. Understanding the mechanisms generating this intergenerational persistence in education is essential to target policy measures adequately.

F. Wendelspiess Chávez Juárez (✉)
Institute of Economics and Econometrics, University of Geneva, UNI-MAIL, 40 Bd. du Pont d'Arve, 1211 Geneva 4, Switzerland
e-mail: florian@chavezjuarez.com; florian.wendelspiess@unige.ch

In this paper, I try to shed some light on this mechanism by applying a simultaneous equations model to data from Mexico. The main goal of the study was to estimate the relative importance of different transmission channels and their interactions.

The three main mechanisms put forward in recent theoretical and empirical contributions can be broadly described as being the biological, the economic and the direct education-to-education channel.

The biological channel refers to the genetic transmission of ability, often measured by the IQ, which explains a part of the relationship (Anger and Heineck 2010; van Leeuwen et al. 2008; Björklund et al. 2010; Black et al. 2009).

Poor families facing credit constraints are an example for the economic channel, because they cannot borrow against the expected future earnings of their offspring, which generates a link between the socioeconomic situation of the parents and the schooling of their children (see for instance Black and Devereux 2010; Attanasio and Kaufmann 2009; Stinebrickner and Stinebrickner 2007; Carneiro and Heckman 2002; Alfonso 2009).

Finally, a higher return to education for children with highly educated parents is an argument for the direct education-to-education channel (Black and Devereux 2010). It might also include preferences for education, non-cognitive skills, aspirations and many other factors.

The empirical literature focuses on estimating the causal effect of parental education on children's education. Holmlund et al. (2011) revise this literature and propose a comparison of different methods applied to the same dataset from Sweden. They conclude that the estimates differ substantially across identification strategies and that no method is perfect. They find relatively modest causal effects of parental education and point to the importance of analyzing in more detail other mechanisms explaining the intergenerational transmission of education. They hypothesize about the possibility of an indirect effect of education through better socioeconomic environment that can be offered to children. This conclusion on the need of better understanding the mechanisms is shared by recent literature surveys such as Black and Devereux (2010), Björklund and Jäntti (2009) and Piketty (2000) who coherently argue that more empirical research must be undertaken to understand the mechanisms behind educational mobility and social mobility in general. Black and Devereux (2010, p. 69) conclude that "[...] *there is still much work to do to pin down which family background factors are most important*" and Björklund and Jäntti (2009, p. 516) argue that "*a major challenge for future research is to find out what in the family other than income is important for the future of children*".

This study tries to contribute in the proposed direction by moving from the estimation of one single causal effect to the estimation of a larger system of effects, incorporating simultaneously the three channels previously outlined. This focus on different mechanisms at the same time aims at getting a better understanding of the larger picture—the whole process of intergenerational transmission of education. In this respect, this approach should be seen as a complement to the single causal effect estimation and not as an alternative.

The choice of Mexican data is primarily motivated by the importance of high intergenerational correlations in education in Latin America. Hertz et al. (2007) compare educational mobility of 42 countries including 7 Latin American ones. These seven countries take the first seven places ranked according to their intergenerational education correlation. The country with the highest correlation is Peru (0.66), followed by Ecuador, Panama and Chile.

Within Latin American Countries, Mexico displays relatively high intergenerational persistence in education. Dahan and Gaviria (2001) compare 16 Latin American countries using data from the late 1990s and find that Mexico has the second lowest intergenerational mobility level behind El Salvador. de Hoyos et al. (2010) use recent data on social mobility in Mexico. They report correlations between the education in years of parents and children, finding the highest correlation of about 0.6 for the children cohorts born between 1942 and 1951, followed by a reduction of the correlation to about 0.5 for the cohort 1962–1971 and finally a new increase of the correlation to 0.55 for the youngest cohort, composed of children born between 1972 and 1981. According to the same authors, this recent increase is even higher when using different data sources. The same pattern of increasing educational mobility prior to the economic crisis in the 1980s and a subsequent decrease was found by Binder and Woodruff (2002), who use different cohorts to estimate the intergenerational link.

Hence, the Mexican case is not only interesting on its own but might be representative for other Latin American Countries. Moreover, another advantage of Mexico is the high-quality data available. The suitability of the survey for this study is underlined by the availability of cognitive ability scores. Nevertheless, the analysis faces a series of empirical challenges in the form of trade-offs. This study tries to find a balance between a more complete model with very high data requirements and a simpler model in which less data are lost due to unavailable information.

The main result is that even when controlling for parental education and ability, the economic situation of the family has the largest direct effect on the schooling outcome of children. This important source of inequality of opportunity could be reduced by policy interventions targeting the link between economic requirements and schooling. Moreover, the estimation shows that there are important interactions between the channels, suggesting that the exclusion of some channels might seriously bias estimates.

In Sect. 2 I will review the literature on the mechanisms of educational mobility motivating my empirical models I present in Sect. 3. In Sect. 4 I describe the data used and especially the needed transformations in detail and present some descriptive evidence. In Sect. 5 I present the main results, which are complemented by some figures in “Appendix B”. Finally, Sect. 6 concludes the paper.

2 Theory and empirical evidence on educational mobility

Intergenerational mobility in education is a complex phenomenon that does not rely on a single mechanism. The literature identified three main channels of transmission

(Chevalier 2004). The first channel is the biological transmission of ability, the second refers to the dependence of schooling outcome on the economic situation of the parents and the third deals with direct education-to-education effects. In this section, I present some theoretical and empirical contributions to the understanding of these channels.

2.1 Ability transmission through genes: the biological channel

The direct transmission of abilities, which is not limited to simple IQ transmission, represents a biological explanation of the phenomenon. For instance, Becker and Tomes (1979) use the term *endowments acquired from parents* to describe this direct transmission. They provide a theory of intergenerational transmission based on rational choices through a human capital theory approach, where the ability level of the offspring is a key determinant of the decision. Their model was consequently extended by Loury (1981) and Solon (2004) and serves as a benchmark in many analyses.

Empirically, much work has been done to determine the importance of this channel. In a meta-analysis of 212 IQ studies, Devlin et al. (1997) quantify the genetic transmission and find the broad-sense heritability of IQ to be 48 %.

Social scientists put more emphasis on quantifying the overall IQ correlations between parents and children, which might also include environmental effects in addition to the pure heritability measured by Devlin et al. (1997). For instance, Anger and Heineck (2010) use German panel-data with two ultra-short IQ-tests to estimate the parent-offspring relation. They find that a 1-point increase in parents' score results in a 0.45-point increase in the coding speed (inherent ability) and 0.50-point increase in word fluency scores. The estimated coefficients remain stable at the inclusion and exclusion of control variables. Björklund et al. (2010) use Swedish data from military IQ tests and official registers. They estimate intergenerational and sibling IQ correlations. The estimated values are all highly significant and attain values of 0.346 for father-son, 0.510 for siblings and 0.65 for twins. According to the authors, their estimations represent rather a lower bound of the true values. Black et al. (2009) find a similar father-son IQ-correlation (0.38) in a comparable study with Norwegian data.

van Leeuwen et al. (2008) go even further in the analysis of IQ-transmission by dividing it further up. No evidence for cultural transmission of the IQ was found and no indication that intelligent parents provide children with intelligence promoting circumstances. Individual differences in intelligence were found to be largely accounted for by genetic differences. Moreover, they find a spousal IQ-correlation of about 0.33 suggesting a relatively high degree of assortative mating.

2.2 Credit constraints and the economic situation: the economic channel

According to the logic of the economic channel the intergenerational correlation in education is the fruit of an underinvestment in education by poor families. This idea that poorer families might face credit constraints making the optimal investment in

the human capital of their offspring impossible can be found, for example, in Banerjee and Newman (1994) and Loury (1981).

Empirical research was not able to conclude on the exact importance of credit constraints and the economic environment in general. While it seems that the impact of credit constraints is relatively modest in richer countries, some evidence was found that in developing countries the effect is larger.

Stinebrickner and Stinebrickner (2007) analyze the situation at a college in the US using panel data of students. They find that a group of students is credit constrained in consumption during their stay at the college, but that many of them are not willing to borrow.

Carneiro and Heckman (2002) critically revise the literature on the question of credit constraints. They compute that using modern US data, only about 8 % of students really face short-term credit constraints. They argue that long-term effects, such as the family environment during the whole schooling period of children, play a much bigger role. Winter (2007) elaborates a computable general equilibrium model to evaluate the role of credit constraints in the decision whether to go to college. The model is calibrated for the US economy and predicts observed patterns quite well. The findings contrast the results of few credit constrained students found by other studies and argue that econometric estimates such as in Carneiro and Heckman (2002) are downward-biased. In line with the results of other studies is the observation that the share of people financially constrained has increased (dramatically) over the past decades.

Alfonso (2009) presents a study of 4 Latin American countries (Mexico, Chile, Colombia and Peru). She shows that the effect of credit constraints disappears in regression analysis when controlling for long run family variables (parental education, family assets, etc.). However, the relatively small effect of credit constraints increases from the oldest to the newest datasets used in the study.

Attanasio and Kaufmann (2009) use Mexican data to analyze the relationship between post-secondary school decisions and subjective expectations. Among other findings on the role of expectations, they show that credit constraints represent an important issue for poor Mexicans, in contrast to some literature coming from higher developed countries, where these effects do not seem to be as present.

To sum up, the literature finds evidence for the existence of the economic channel in producing high intergenerational correlations in education. It remains somewhat unclear if short-run credit constraints or the long-run economic situation are more important determinants. It seems, however, that in developing countries, both contribute to the low educational mobility.

2.3 Education to education transmission

Finally, the third channel considered is the direct effect of parental education on the schooling attainment of the children. This channel is generally known as the *nurture* effect, capturing the direct causal effect of parental education (Holmlund et al. 2011; Chevalier 2004). Dickson et al. (2013) show that this direct causal effect starts at very early ages and remains visible years later when comparing students' performances. Different explanations why parental education should have a direct

causal effect can be found in the literature. One possible explanation is that highly educated parents tend to encourage their children more to achieve high levels of education (Merton 1953; Boudon 1973, 1974; Sewel and Shah 1968). For instance, Steinberg et al. (1992) show that parental encouragement and parental school involvement have important effects on the school performance of children. Besides this active encouragement and involvement of parents, it can also be argued that the child’s aspirations increase when parents have more education (Sewel and Shah 1968; Ermisch et al. 2006). Ermisch et al. (2006) argue that parental education can alter the productivity of parents’ time investments in children. It could also be argued that expected returns to education depend on parental education. Jensen (2010) shows that students with higher educated parents tend to perceive higher returns to education. Hence, this third channel is motivated by a series of arguments and most likely composed of different sub-channels. The distinction of these different sub-channels is beyond the scope of this paper. In this paper I consider a compound channel linking high parental education to high offspring’s schooling.

3 Model

3.1 Conceptual model of educational mobility

Following the literature outlined in the previous section, we can easily illustrate the three transmission channels. Figure 1 displays the system of transmission in education suggested by the literature.

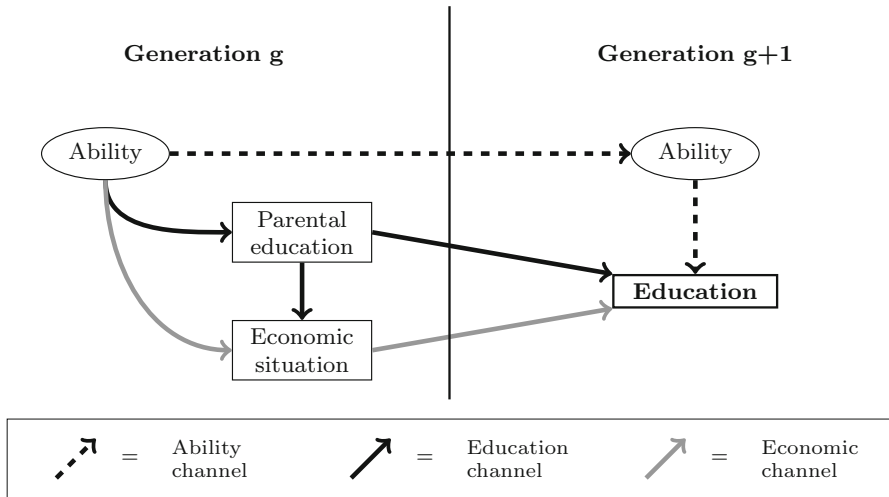


Fig. 1 Simplified conceptual system. The figure displays the conceptual framework of the analysis presented in this study. *Arrows* refer to direct causal effects. Only channels that are included in the empirical analysis are included in the scheme

First, there is a direct link between abilities of parents and children, presented with the dotted line. The ability of the parents influences the ability of children, which in turn increases their propensity for education. The economic channel is represented by gray arrows using the compound term *economic situation*, which includes short-term credit constraints and long-run effects of assets. Finally, the third channel illustrated with solid black arrows represents the direct education-to-education transmission, which is based on many different hypotheses as explained above.

A channel that I do not consider in this study is the health channel. The health status of the child is likely to be influenced by the family background on the one hand (Bradley and Corwyn 2002; Rosa Dias 2009; Delajara and Wendelspiess Chávez Juárez 2013), while it might also have important effects for education on the other hand (Case et al. 2005; Doyle et al. 2009). There are two reasons for not including this channel, both related to the data. First, there is a problem of timing in the health data. While the literature emphasizes on the importance of the prenatal period and early childhood in the health dimension (Heckman 2006; Doyle et al. 2009; Delajara and Wendelspiess Chávez Juárez 2013), in the data I can at most observe the current health status of the child. Unfortunately there is no retrospective information on the child's health conditions at birth available. Second, as I will highlight in Sect. 4 the inclusion of additional variables would seriously reduce the sample and increase the risk of sample selection problems. In the “Appendix E” I present some regressions where I include variables for health and personality traits to illustrate the problems just described.

Different strategies are possible to analyze such a framework empirically. One way is to focus on one particular link. For instance, Holmlund et al. (2011) present different methods to estimate the causal relationship between the education of parents and children. In this paper I use two different strategies. First, I focus on a single equation approach where I aim at estimating all the determinants of the child's education outcome in one regression. In a second step, I move to a simultaneous equations model where I estimate several equations to describe the whole system outlined in Fig. 1. I will now explain the two approaches along with their intuition, advantages and challenges.

3.2 The single regression approach

The main goal of the empirical application is to estimate the intergenerational links determining the educational outcome of the child. Therefore, a first empirical approach consists in regressing the educational outcome on the possible intergenerational determinants while controlling for some contemporaneous effects. The intergenerational determinants are parental education and the economic situation of the family. Moreover, I also control for parental age, whether the parents have an indigenous background and some variables capturing the family structure¹. The most important contemporaneous effect I control for is the cognitive ability of the

¹ These variables include the number of children (up to 12 years old) and teenagers (12–18 years old) in the household and a dummy for the first-born child.

child. Additionally, I also control for contemporaneous effects such as child labor, government program benefits, state fixed effects and indicators for girls and rural areas.

Besides its appealing simplicity, the single regression approach has the advantage that we do not have to impose a lot of structure in the model. Therefore, only the standard assumptions for ordinary least squared must be fulfilled. One concern that could arise is that some variables do not satisfy these conditions and are likely to be correlated with the error term. The most likely reason for this to happen in our context is an omitted variable bias. A first critical variable is the ability level of the child, which might also capture for instance motivation or the ability to perform well in a situation of examination. Both potentially omitted variables are also likely to influence the schooling outcome. To a large extent, these concerns are reduced by the type of cognitive ability measure I am using in this study. The ability measure is based on a short version of the Raven's progressive matrices test, which is one of the most culture-free and education independent IQ-tests (Désert et al. 2009). I discuss this in more detail when describing the data in Sect. 4. A second variable that might suffer an omitted variable bias is parental education, as parental education might also be influenced by preferences and taste for education. If these preferences and tastes are also transmitted to the next generation, we are likely to have an endogeneity problem as well.

In order to account for these potential endogeneity issues I also perform instrumental variable regressions for the single regression approach. Father's and mother's cognitive ability is used to instrument both the parental education and the child's ability level. The cognitive ability of the parents should be a strong predictor of both parental education² and the cognitive ability of the child. At the same time, the cognitive ability of the parents should not have a direct impact on the schooling outcome of the child. For the case of parental education, I additionally use information on the place of living of the parents when they were 12 years old. A dummy capturing whether they lived in a town or not is used.³ The idea is that parents living in cities had substantially more access to education than parents living in rural areas. At the same time, the place of living of the parents when they were 12 years old should not directly affect the education outcome of their children today. In the results section, I will present in detail the different tests for the instruments, which clearly indicate that the instruments are valid and strong.

Let me first introduce the second estimation approach, where I simultaneously estimate the whole system presented in Fig. 1.

² The use of parental cognitive ability to instrument parental education could be problematic if the ability measure was influenced by the education. However, as mentioned in the data section, the used RPM test is less education dependent than other measures of cognitive ability. Therefore, I argue that the assumption of no reversed causality might be reasonable. Additionally, I discuss the empirical tests related to the validity of the instruments.

³ The original variable included more categories to describe the situation outside towns. However, they were relatively unclear and did not provide additional explanatory power. For this reason, I regrouped all non-town answers.

3.3 The simultaneous regression approach

In addition to the single regression approach previously presented, I also use a simultaneous equations model approach to estimate not only the intergenerational links, but also the different transmission channels in more detail. There are two main advantages of focusing on the whole system. First, it allows us to estimate the relative importance of the three channels put forward by the literature. This is important because analyzing a specific channel and finding significant effects does not tell us much about the relative importance of the analyzed channel with respect to other possible channels. A channel might show very significant effects and at the same time be relatively irrelevant for the whole system. Second, estimating a system allows us to consider interactions between channels and as a consequence direct and indirect effects. For instance, parents' ability is likely to have both of them. The direct link effect of parental ability refers to the biological channel introduced earlier. The indirect effect goes through parental education and the economic situation. More able parents are likely to have more education and a better socioeconomic status. Behrman and Rosenzweig (2002) show that intergenerational correlations between mother's and children's education might be biased when such system aspects are not considered. For instance, they mention that such estimates are upward biased, when not controlling for the ability channel or for assortative mating.

On the other hand, the estimation of the model as a whole might also has some disadvantages. We might overestimate the effect of the three analyzed channels due to relevant but unobserved channels. Such unobserved channels might include the health channel discussed earlier or some soft skills like personality traits and non-cognitive abilities. I will discuss this issue in the description of the econometric model and the empirical analysis with more detail. First, I will formally introduce the econometric model, which can be written as follows:

$$\text{ability} = \psi_1 \text{ability}_f + \psi_2 \text{ability}_m + \mathbf{Z}\Lambda + \varepsilon_1 \tag{1}$$

$$\text{educ}_f = \delta_1 \text{ability}_f + \delta_2 \text{age}_f + \delta_3 \text{indi}_f + \delta_4 \text{city}_f + \varepsilon_2 \tag{2}$$

$$\text{educ}_m = \delta_5 \text{ability}_m + \delta_6 \text{age}_m + \delta_7 \text{indi}_m + \delta_8 \text{city}_m + \varepsilon_3 \tag{3}$$

$$\begin{aligned} \text{wealth} = & \gamma_1 \text{educ}_f + \gamma_2 \text{educ}_m + \gamma_3 \text{ability}_f + \gamma_4 \text{ability}_m \\ & + \gamma_5 \text{indi}_f + \gamma_6 \text{indi}_m + \gamma_7 \text{age}_f + \gamma_8 \text{age}_m + \varepsilon_4 \end{aligned} \tag{4}$$

$$\begin{aligned} \text{cons} = & \gamma_9 \text{educ}_f + \gamma_{10} \text{educ}_m + \gamma_{11} \text{ability}_f + \gamma_{12} \text{ability}_m \\ & + \gamma_{13} \text{indi}_f + \gamma_{14} \text{indi}_m + \gamma_{15} \text{age}_f + \gamma_{16} \text{age}_m + \varepsilon_5 \end{aligned} \tag{5}$$

$$\text{schooling} = \beta_1 \text{ability} + \beta_2 \text{educ}_f + \beta_3 \text{educ}_m + \beta_4 \text{wealth} + \beta_5 \text{cons} + \mathbf{Z}\Omega + \varepsilon_6 \tag{6}$$

This set of equations represent the simultaneous equations model of the above conceptual model of educational mobility. I take deviations from the mean to avoid constant terms and to simplify the notation. Subscript *f* refers to the father and

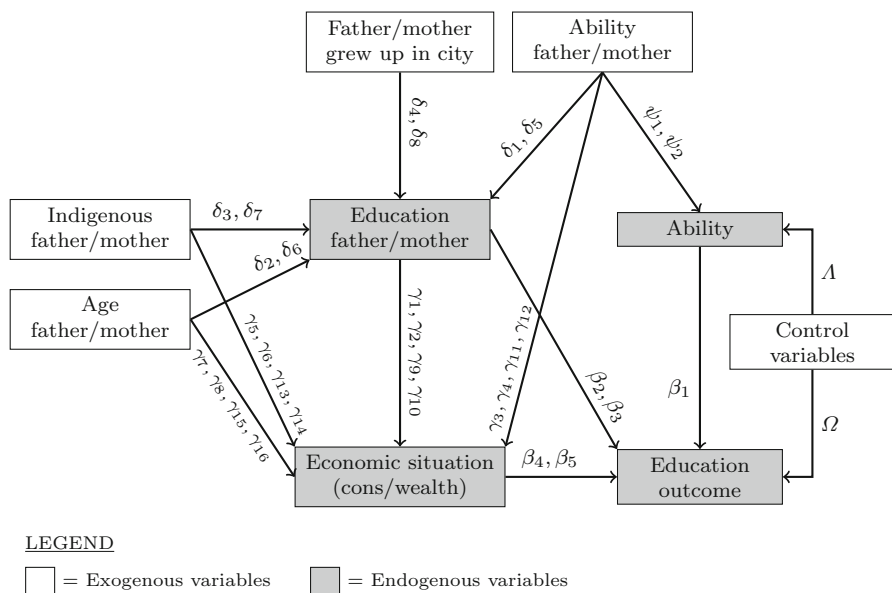


Fig. 2 Path diagram of the simultaneous equations model. The path diagram directly refers to the system of equations presented in Eqs. (1)–(6)

subscript *m* to the mother. Variables without subscript describe either the situation of the family or the child. An alternative way of presenting the model is the path diagram in Fig. 2.

The white filled boxes refer to exogenous variables, the gray boxes represent endogenous variables and the arrows describe direct effects. Note that for the sake of readability of the graph, I present both parents and both economic indicators together. Even though the graphical representation is more illustrative, I will now discuss the model mainly based on the equations presented above.

Equation (1) describes the genetic transmission of cognitive ability, where the main explanatory factors of the child’s ability are the parental cognitive ability scores.⁴ Additionally to the parental ability scores, I add some control variables such as the gender of the child, a dummy for first-born children and two dummies for children with a small (<20 years) and large (> 40 years) age difference to their parents. Parental education is excluded from this regression because we can assume that there is no direct effect on the cognitive ability score of the child. This assumption is based on the nature of the used cognitive ability test, which is education- and culture-independent (Désert et al. 2009). For the same reason, no feedback effect from education to ability is included. The economic situation is not included as an explanatory variable because the transmission through genes took

⁴ Note that ability refers to cognitive ability and does not include non-cognitive ability. For this reason, I do not include indicator variables for non-cognitive abilities and estimate a latent factor. This choice allows me to focus on the *nature* and not on the *nurture* effect.

place at birth, while the economic situation indicators are contemporaneous values and have, therefore, no direct effect.

Equations (2) and (3) are simplified education production functions for the father and the mother, respectively. The idea is to link parental ability to their schooling outcome and to control for cohort- and ethnicity-based differences in the educational level of parents. I also include the instrument of the single regression approach capturing the place of living of the parents when they were 12 years old. The idea is that parents that grew up in towns had substantially better access to schooling than those living in the countryside. Both indicators of the economic situation are excluded from this regression because of the timing of these variables. The economic situation today does not directly explain the educational achievement of the parents when they were in the schooling age.

Equations (4) and (5) estimate the effect of parental education and ability on the two indicators of the economic situation. The economic situation is split into consumption and a wealth index. This index was obtained by taking the first component of a principal component analysis on several indicators of durable good holdings and housing conditions.⁵ Taking this index instead of the full set of indicator variables allows me to reduce the dimensionality and to use the wealth index as an indicator for the long-run economic situation. The use of both, the wealth index and the current consumption level, is motivated by the findings in the literature, saying that the (long-run) economic environment is more important than current consumption. In addition to the exogenous variables, the economic situation is also influenced by parental education.

Finally, Eq. (6) is the main equation corresponding to the single equation approach outlined before. The explanatory variables of interest are parental education, the economic situation indicators and the ability score of the child. I also control for some contemporaneous effects by including control variables capturing the family structure, the place of living, the government program benefits, the working conditions of the child and the gender of the child.

To sum up, the coefficients of main interest are the β 's and to a lesser extent the ψ 's. The β 's estimate the direct impact of family background variables and the child's ability on educational attainment. The ψ 's permit us to estimate the relationship between parental and child's ability, i.e. estimating the biological transmission. Through the ψ 's and β_1 the total effect of the biological transmission on the educational outcome can be estimated. This setting allows us to estimate the relative importance of the different channels in the educational transmission. The model is estimated using the maximum likelihood method under normality assumptions (Muthén 2004). In contrast to the instrumental variables techniques used in the single equation approach, the identification of the simultaneous equation model is somewhat more complicated to show. The model presented in this paper is easily identified due to its quasi-recursive structure. I use the term quasi-recursive because I do not assume independence of the error terms for contemporaneous equations. As a consequence, I use more restrictive identifying conditions to show

⁵ In the "Appendix C" I describe the indicators and the estimation of the wealth index in detail.

that all parameters in all equations are identified. A detailed discussion of the identification conditions along with the proofs can be found in the “[Appendix D](#)”.

4 Data

4.1 Data description

The analysis of this paper requires very complete data at the micro level for both children and parents. The *Mexican Family Life Survey (MXFLS)*⁶ is a very rich and award-winning panel data project from Mexico and fits these requirements quite well. I use information from the first two waves (2002 and 2005), focusing on the latter wave. The panel structure was mainly used to reduce the amount of measurement errors, for instance, by identifying and correcting impossible values for time invariant variables.⁷ To the extent of my knowledge, this is the best data source from a Latin American Country for this kind of analysis, particularly because it includes short cognitive ability tests. Nevertheless, the data are not perfect and before starting with the analysis I discuss some trade-offs faced and the resulting decisions taken.

4.1.1 Choosing the age range of the sample and the schooling outcome variable

A first challenge is to choose correctly the age range of the primary units of analysis. In order to estimate properly the correlation of years in education one would have to limit the analysis to people having finished their education, i.e. mostly people over 25, implying two major problems.

First, older individuals are probably no longer living with their parents. However, as I do not have administrative data, I can only establish the link between children and parents when they are living in the same household. Those still living with their parents years after completing school are most likely not representative for the whole population.

The second problem is that the schooling period of older people having finished education lies potentially far in the past. Therefore, the economic situation for that time would be hard to proxy and the mechanisms I would analyze would be those prevailing some years ago, which is not necessarily very policy relevant.

At the lower bound of the age range we cannot include too young children, as they are only about to start their educational path. Therefore, the information on years of schooling is likely to be much less related to their final schooling outcome as compared to slightly older children.

For these reasons, I focus on children and young adults from 12 to 25 years and use a constructed education index instead of years of schooling. The idea behind this

⁶ The original Spanish name is *Encuesta Nacional sobre Niveles de Vida de los Hogares (ENNVIH)*.

⁷ For instance, if in one wave the father was younger than the child and in the other wave the difference was plausible, then only the plausible value was taken. However, if there was no plausible value, the observation was dropped.

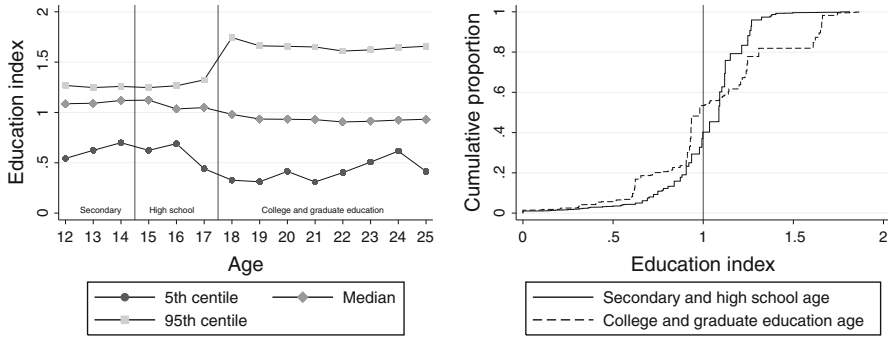


Fig. 3 Distribution of the education index. The figure displays the distribution of the main dependent variable (education index) as a function of age. Both graphics are based on the working sample used in the main regressions

index is very simple: instead of measuring the final outcome, I consider the delay in schooling that people have with respect to their peers. The index is computed by dividing an individual’s years of schooling by the average years of schooling of her age cohort. A value of 1 corresponds to a child that is just on time compared to its peers; a value below 1 suggests a delay.

Figure 3 displays some key statistics by age on the left side and the cumulative density function of the education index on the right side. The cumulative distribution function is depicted by age groups corresponding to students in the age of secondary and high school education (12–17 years) and tertiary education (18–25 years), respectively.

On the left graph we can see that the dispersion of the index is relatively stable for the ages corresponding to secondary and high school. For the ages corresponding to tertiary education the dispersion increases, especially the 95th percentile increases. This change is due to the fact that a substantial proportion of individuals do not continue education beyond the high school level. Therefore, the reference level remains at lower levels and those actually attending tertiary education achieve higher levels of the education index. On the right-hand side graph we can see that there is a considerable amount of variation in the index, starting at values close to zero for those with no or very little education and going up to almost 2. For the younger age group a stronger concentration around the value of 1 can be observed. This is due to the fact that less variation in the years of education is observed.⁸

The underlying assumption of this indicator is that a delay in schooling is translated later on in fewer years of schooling. In “Appendix A” I present some empirical evidence of this relationship and provide some additional information on the Mexican education system, along with some basic statistics such as enrollment rates.

⁸ Note that when plotting the cumulative density function by age, we even find for the youngest children a similar range of values.

4.1.2 Variable selection and construction

A second data challenge is to include as much relevant information as possible by minimizing the cost in terms of loss of observations due to missing values. My strategy in the variable selection process was to give absolute priority to the three main channels discussed earlier. At the same time, I tried to avoid unnecessary loss of observations due to less relevant variables. To face this trade-off I started by defining a set of absolutely needed variables which cannot be excluded from the analysis without seriously changing the model. For this type of variables, dropping observations due to missing values is unavoidable. A second series of interesting but not absolutely indispensable variables was selected trying to avoid variables that would cause a large loss of observations. In this respect, some variables potentially able to capture soft skills and personality traits were excluded, because too many observations would have been lost.⁹

One of the main reasons to use the MXFLS data is the availability of cognitive ability measures based on Raven's progressive matrices (RPM). According to Désert et al. (2009) RPM is a frequently used intelligence test with proven reliability and validity in measuring cognitive aptitudes and reasoning. Désert et al. (2009) further highlight that this IQ test is less education dependent than others, reducing the risk of feedback effects from education.¹⁰ Different versions of the test were applied to children (5–12 years) and adults (13–65 years). In order to have comparable scores across age groups, the values were normalized to a distribution with mean 100 and standard deviation 15 for each age. The choice to normalize to the mean and standard deviation of the IQ is essentially for illustrative purpose, but it does not imply that the cognitive ability scores can be seen as a complete measure of IQ. Moreover, the normalization is not relevant for the results, because I report only standardized coefficients, which are by definition independent of previous normalizations. Given the panel structure of the survey, two test scores per person are available, allowing us to compute the average score of the person to reduce measurement errors. Observations where the two scores had a difference of more than 2 standard deviations were dropped from the analysis. For people with only one valid test score this was taken to avoid losing too many observations.

Parental education in years was obtained by computing the average time spent in school to achieve the reported education. Repeated years are, therefore, not considered as schooling years, as one can argue that they do not provide additional human capital. Note that the question on the achieved education level is asked twice in the survey. Once it is asked in the roster questionnaire and once in the individual questionnaires. I primarily took the information from the individual data and completed it by the roster data when the individual data was missing.

The family log-consumption per capita was obtained out of a series of information on consumption and normalized to the consumption per equivalent

⁹ Experimental regressions were performed including such variables in order to see if their exclusion alters the results. I discuss these briefly in the “Appendix E”.

¹⁰ More details on Raven's progressive matrices and its implementation in the MXFLS can be found in Raven et al. (1986, 1983), Raven (2000), Rubalcava and Teruel (2006).

adult following the methodology proposed by Rojas (2007), who provides estimates for Mexico based on the subjective well-being approach.¹¹ The wealth index was obtained by taking the first component of a principal component analysis performed on several household assets and indicators of the housing conditions. A list of the included indicators and their relative importance for the wealth index and its possible relation with parental age are reported in the “Appendix C”. The remaining variables included in the study were constructed in a straightforward way according to the standards in the literature and are reported with a short description in Table 1.

4.1.3 Sample size and sample selection

Initially 11,273 children and young adults aged between 12 and 25 years were present in the database. From these only 8,155 individuals lived with both parents in the same household. This is a necessary condition for this study, since otherwise no cognitive ability scores of the parents are available. Proxies for other variables such as education of absent family members would be available, but the cognitive scores are not. Missing values in parents’ and children’s characteristics introduced another loss of observations, reducing the sample to 4,266 observations. The large loss of observations is not surprising considering the data requirements of the study and the fact that they are survey data from an emerging country. These data are obviously not as good as administrative data from European countries that were used in some other studies on the topic. It could be argued that the loss of observations introduces sample selection biases. Being fully aware of this fact, I try to show that the sample used in this study produces some very comparable results to findings in the literature and that the analysis is quite robust to changes in the sample. I also estimate the benchmark model with larger samples, where I relax some data requirements. For instance, excluding the channel of the father allows me to take into account the numerous single-mother households and merging the effects of the father and the mother, allows me to include every single-parent household. These larger sample regressions are reported in “Appendix B”.

4.2 Descriptive statistics

Let us now have a closer look at the data. Table 1 presents some univariate descriptive statistics of the sample I use in the econometric models. The different variables are divided into blocks corresponding to their role in the econometric model. The main dependent variable is the education index previously introduced. The index was constructed using the largest sample possible and not only the observations used in the econometric estimation. Therefore, the average value is slightly higher than 1. The same logic applies to the ability measures, which were estimated using all available information.

¹¹ Using the official Mexican equivalence scales based on CONEVAL (2008) gives essentially the same results where only the third digit after the comma changes by at most two units. I prefer to follow Rojas (2007) as his definition is concave in the number of people, while the official equivalence scales are not.

Table 1 Variables used in the study

Abbreviation	Description	Mean	SD
Dependent outcome variable			
Educ. index	Years of education divided by the average years of education of the age group	1.037*	0.310
Endogenous regressors			
Consumption	Average log consumption per equivalent adult	9.829	0.617
Educ. father	Father's years of education	6.348	4.362
Educ. mother	Mother's years of education	5.725	3.806
Ability	Child's ability measure	101.016*	12.910
Exogenous regressors and control variables			
Age father	Age of the father	43.856	7.876
Age mother	Age of the mother	40.300	7.139
Ability father	Ability measure for the father	100.614*	13.711
Ability mother	Ability measure for the mother	97.795*	13.469
Indig. father	Dummy variable for indigenous father	0.172	
Indig. mother	Dummy variable for indigenous mother	0.139	
Father city	Father grew up in a urban area	0.257	
Mother city	Mother grew up in a urban area	0.264	
Female	Dummy for girls (=1)	0.491	
Age	Age in years	17.321	3.696
Rural	Dummy variable for rural areas	0.483	
Work02	Dummy for working activities in 2002	0.166	
Work05	Dummy for working activities in 2005	0.250	
Number children	Number of children below 12 years	1.077	1.236
Number teenagers	Number of teenagers (12–18 years)	1.676	1.150
Program 1	Dummy for program <i>Alianza para el campo</i>	0.011	
Program 2	Dummy for program <i>Coinversión social</i>	0.003	
Program 3	Dummy for program <i>Crédito a la palabra</i>	0.012	
Program 4	Dummy for program <i>FONAES</i>	0.004	
Program 5	Dummy for program <i>Fondo para la Micro, Pequeña y Mediana Empresa</i>	0.006	
Program 6	Dummy for any other assistance program	0.029	
Program 7	Dummy for program <i>Programa de empleo temporal</i>	0.009	
Program 8	Dummy for program <i>PROCAMPO</i>	0.132	
Program 9	Dummy for program <i>VIVAH</i>	0.007	
Oportunidades	Dummy for program <i>Oportunidades</i>	0.255	

Descriptive statistics based on the sample of 4,266 observations

* Normalization was made with the full sample to use a maximum of information. The mean can deviate slightly from the normalized value due to missing values in other variables

The average and the standard deviation of the fathers' years of education are slightly higher than for the mothers. The proportion of indigenous parents is around 15 % which corresponds to the national average. The age of parents is measured in years and fathers are slightly older than mothers. About one-quarter of parents grew up in a city. The sample is strongly balanced between girls and boys and also between families living in rural and urban areas. The indicator of rural areas is based on the official definition of rural zones in Mexico and the place of living at the time of the survey. As people might have lived in a different place during their education, I use additional information on migration to correct the variable accordingly.¹² Two variables (*Work02* and *Work05*) capture whether the children were working in 2002 and 2005, respectively. The proportion grows from around 17 to 25 %, reflecting the aging of the cohort. The indicators on the number of children and teenagers allow us to control for the composition of the households. On average, there is one child below 12 years and about 1.6 teenagers present in a household. The set of dichotomous program variables captures the beneficiary status of families for different government programs. The proportions of beneficiaries are generally very low, with the exception of *Oportunidades* and *Procampo* where the proportion is above 10 %.

More interestingly than the averages of the variables are the relationships among them. I now present some simple linear correlations between important variables. They should provide a good impression of the data and outline some potentially interesting phenomena. On the other hand, they should give us an impression of comparability of the data with data used in other studies. I hope to reduce some concerns regarding the sample selection issues and the definition of some main variables by showing that the descriptive statistics are surprisingly comparable to other studies in the literature.

A first issue that one might discuss regarding the data is the use of Raven's progressive matrices test as a measure of cognitive ability or even IQ. In the sample, the correlation of the ability measure with the one of the father is 0.363. This value is very close to the 0.347 and 0.38 estimated by Björklund et al. (2010) and Black et al. (2009), respectively, both using more detailed IQ measures. The same correlation with respect to the mother was found to be 0.387, which is slightly higher than the father-son correlation. Considering only the two oldest siblings in a family gives a siblings IQ-correlation of 0.506, which is again relatively close to the values reported by Björklund et al. (2010) who find estimates between 0.473 and 0.510. Interestingly, and giving a first evidence for assortative mating, the spousal IQ-correlation is 0.400. The spousal education-correlation based on the years of schooling is 0.646, which is even higher and supports the idea of an important role of non-random spousal selection.

¹² Unfortunately, it is not possible to use exclusively the information on the place of living when people where at the age of education, because the variable is measured differently. I correct the variable rural only in cases where people reported that they lived in a city during education and living in rural areas at the time of the survey. The large majority of the individuals (around 90 %) never changed the place of living and, therefore, the information of the place of living at the time of the survey is accurate for the education period as well.

Table 2 Intergenerational correlation of education

Age group	Ages	Gender	N	Correlation with	
				Father	Mother
All	12–25	Girls	2,095	0.369	0.395
		Boys	2,171	0.389	0.377
Secondary and high school	12–17	Girls	1,153	0.312	0.307
		Boys	1,183	0.319	0.334
Tertiary	18–25	Girls	942	0.442	0.489
		Boys	988	0.465	0.447

The reported correlations refer to the correlation between the years of education of the father and mother, respectively, and the constructed education index. All correlations are computed for the working sample used in this study

Regarding the simple educational attainment correlation between parents and children, a very interesting pattern can be found when splitting the sample into age groups. Table 2 presents the correlation between the education index used in this study to proxy the educational attainment of children and their parents' years of education.

The correlations are substantially higher for the older age group as compared to the younger group.¹³ Several possible explanations for this can be found. First, the intergenerational transmission is likely to be a cumulative process, thus the older the children become, the larger is the relationship between parental education and the educational outcome of their children. Second, it could also be due to the precision of the education attainment indicator used in this study. The older the children are, the more values the indicator can take and, therefore, the correlations might be estimated with more precision.¹⁴

The correlations for the older age group are slightly below the correlation of 0.55 estimated by de Hoyos et al. (2010) for children born between 1972 and 1981 in Mexico. The likely reason for the difference is that de Hoyos et al. (2010) use older individual with finished education. Looking at the difference from the younger to the older age group, it is very likely to end up with similar values as de Hoyos et al. (2010) if we could include older individuals.

Finally, Table 3 gives a comparison between the used and the full sample for the main variables of interest. We can see that the differences are not statistically significant for father's education and both parental ability measures. For consumption the difference is only significant at the 10 % level. For other variables we observe statistically different means, which is not very surprising with that many observations. However, by looking at the column 'Diff/SD' we can see that the

¹³ Note that when computing the same correlation for younger children (say 7–11 years), the values are even lower.

¹⁴ This argument is particularly true when considering even younger children at the age of primary school. A previous version of this study included them. The decision to take them out of the study is mainly due to this argument saying that the precision of the education attainment indicator is not sufficient for the youngest individuals.

Table 3 Comparison of sample with excluded observations

Variable	Full sample		Used sample		Diff.	Diff/SD	P value
	Mean	SD	Mean	SD			
Education index	1.000	(0.331)	1.037	(0.310)	-0.037	-0.111	0.000
Log consumption	9.849	(0.641)	9.829	(0.617)	0.020	0.031	0.081
Father's education	6.416	(4.430)	6.348	(4.362)	0.068	0.015	0.424
Mother's education	5.941	(4.017)	5.725	(3.806)	0.216	0.054	0.003
Father's ability	100.620	(13.907)	100.614	(13.711)	0.006	0.000	0.982
Mother's ability	97.947	(13.832)	97.795	(13.469)	0.151	0.011	0.557
Ability of the child	100.292	(13.621)	101.016	(12.910)	-0.724	-0.053	0.003
Mother's age	40.713	(7.586)	40.300	(7.139)	0.413	0.054	0.003
Father's age	44.450	(8.645)	43.856	(7.876)	0.594	0.069	0.000
Age	18.104	(3.914)	17.321	(3.696)	0.783	0.200	0.000
Number of children	1.157	(1.278)	1.077	(1.236)	0.080	0.063	0.000
Number of teenagers	1.469	(1.186)	1.676	(1.150)	-0.207	-0.175	0.000
Indigenous father	0.160		0.172		-0.011		0.126
Indigenous mother	0.135		0.139		-0.004		0.559
Gender	0.510		0.491		0.019		0.033
Rural area	0.429		0.483		-0.054		0.000
First born	0.371		0.429		-0.059		0.000

The 'full sample' includes all individuals in the age range of the study with non-missing values. The column 'Diff/SD' refers to the difference between the two samples divided by the standard deviation of the full sample

difference in terms of standard deviation of the variable never exceeds 0.2; thus they are probably not as problematic as the statistical tests might lead us to think.

Overall, the data are certainly not perfect and do not attain the standards of high-quality administrative data from some European studies. Nevertheless, the working sample does seem to represent the full sample relatively well and permits us to carry out the analysis.

5 Results

In Sect. 3 I introduced the two approaches to estimate the intergenerational transmission in education. I now present the result following the same structure. First I present the single regression approach where I estimate simple OLS and IV models and then I move to the discussion of the simultaneous equations model.

5.1 Single regression approach

Table 4 presents the main regression results of the single-equation approach. The first column is a simple OLS estimation, followed by several IV estimates, all using

the robust estimator to account for heteroskedasticity. In the model IV-1 I instrument both parental education and the ability level of the child. For the models IV-2 and IV-3 I instrument child's ability and parental education separately. For presentational reasons, I do not report some control variables such as the age and the indigenous background of the parents and the indicator for rural areas. These variables are not significant. Additionally, I do not report the coefficients of the government program benefits and the state fixed effects to reduce the size of the table.

Let us first discuss the main coefficients of the OLS regression, which are all presented with standardized coefficients. The ability of the child has a strong and highly significant effect on the schooling outcome. The direct effect of parental education is also highly significant and positive. The effect of the mother is larger than the one of the father, which is in accordance with the literature. Both indicators for the economic situation of the family display positive and significant effects. Note that this estimation does not directly allow us to conclude about the biological channel, as we do not estimate the link between parental ability and child's ability. In order to see whether these OLS estimates are reliable, I move now to the discussion of the IV estimates. First, we can see in the models IV-1 and IV-2 that the coefficients of the child's ability does not change a lot as compared to the OLS estimates. By looking at the endogeneity test based on Baum et al. (2007), we can actually see that the variable is not endogenous and, therefore, instrumenting it is not required. However, this test is only valid under the assumption of valid instruments. To test the validity, I use the Hansen J statistic, which indicates that the instruments are valid.¹⁵ Hence, for the ability measure of the child we do not seem to have an endogeneity problem.¹⁶ As mentioned earlier, this might be due to a large extent to the nature of the cognitive ability test, which is much less related to education and cultural aspects than other ability measures.

Let us now turn to parental education. In the models IV-1 and IV-3 I instrument parental education. Contrary to the previous results, we find strong differences in the coefficients between the OLS estimation and the IV estimates. The coefficient for the father increases sharply while the coefficients for the mother becomes much smaller and insignificant.¹⁷ This is surprising and contrary to the findings in the literature where the maternal education seems to matter more. It is, therefore, important to understand where this result comes from. According to the Hansen J-statistic the instruments are valid and the weak instrument test does not point to a problem of weak instruments. In order to better understand the results, let us have a closer look to the first stage regressions presented in Table 5.

The Angrist-Pischke F-statistic is very large and suggests that the instruments are strong (Angrist and Pischke 2009). Thus, in terms of the standard test for IV-

¹⁵ The null hypothesis of the test is that the instruments are valid.

¹⁶ In the regression IV-2 we could reject the null hypothesis at the 10 % level. I, therefore, performed the endogeneity test on child's ability only in the model IV-1, where we have clearly valid instruments. The test shows also that child's ability is not endogenous.

¹⁷ However, the two coefficients are not significantly different from each other (p value of 0.344 and 0.320 for IV-1 and IV-3, respectively). Hence, they are not contradicting the results found in the OLS regression.

regression, these estimates seem to be valid. However, there is another problem stemming from the underlying nature of the analysis, which goes beyond the standard challenges of IV. Note that both father’s and mother’s education are correlated with the cognitive ability measure and the place of living of either parents. This is a direct result of assortative mating. It is clear that these correlations

Table 4 Single equation results (OLS and IV estimates)

	OLS	IV-1	IV-2	IV-3	IV-3b	IV-3c
Ability of the child ⁺	0.209*** (0.015)	0.232*** (0.084)	0.225*** (0.046)	0.204*** (0.018)	0.205*** (0.018)	0.220*** (0.017)
Father’s education ⁺	0.095*** (0.018)	0.198* (0.117)	0.093*** (0.019)	0.218** (0.098)	0.245*** (0.068)	
Mother’s education ⁺	0.143*** (0.018)	0.030 (0.094)	0.142*** (0.019)	0.042 (0.087)		0.156** (0.061)
Wealth index ⁺	0.134*** (0.020)	0.128*** (0.024)	0.132*** (0.021)	0.126*** (0.024)	0.128*** (0.024)	0.151*** (0.022)
Log consumption ⁺	0.069*** (0.018)	0.067** (0.028)	0.068*** (0.018)	0.062*** (0.024)	0.065*** (0.024)	0.090*** (0.022)
Worked in 2002	-0.386*** (0.044)	-0.388*** (0.048)	-0.384*** (0.044)	-0.385*** (0.046)	-0.389*** (0.046)	-0.402*** (0.046)
Worked in 2005	-0.076** (0.034)	-0.080** (0.037)	-0.075** (0.034)	-0.077** (0.036)	-0.081** (0.035)	-0.089** (0.036)
First born	0.076*** (0.029)	0.081*** (0.031)	0.076*** (0.029)	0.079*** (0.030)	0.082*** (0.029)	0.083*** (0.030)
Gender	0.072*** (0.026)	0.071*** (0.026)	0.073*** (0.026)	0.071*** (0.026)	0.070*** (0.026)	0.070*** (0.026)
Number of children	-0.062*** (0.012)	-0.063*** (0.012)	-0.061*** (0.012)	-0.064*** (0.012)	-0.065*** (0.012)	-0.061*** (0.012)
Number of teenagers	-0.073*** (0.013)	-0.071*** (0.013)	-0.073*** (0.013)	-0.071*** (0.013)	-0.070*** (0.013)	-0.077*** (0.013)
Constant	-0.075 (0.121)	-0.110 (0.196)	-0.078 (0.121)	-0.146 (0.164)	-0.135 (0.167)	0.064 (0.139)
Child’s ability instrumented	No	Yes	Yes	No	No	No
Parental education instrumented	No	Yes	No	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
N	4,266	4,266	4,266	4,266	4,266	4,266
Hansen J-statistic		1.658	2.728	1.842	0.864	0.150
Hansen J-statistic (p value)		0.198	0.099	0.398	0.353	0.699
Weak instr. test (statistic)		66.828	323.284	113.838	233.928	269.153
Weak instr. test (p value)		0.000	0.000	0.000	0.000	0.000

Table 4 continued

	OLS	IV-1	IV-2	IV-3	IV-3b	IV-3c
Endogeneity test (χ^2 stat.)		2.053	0.175	1.860	1.981	0.186
Endogeneity test (p value)		0.562	0.676	0.394	0.159	0.666

All regressions include also additional control variables such as mother's and father's age and indigenous background, an indicator for rural areas (none of these is significant). Additionally, state fixed effects and control variables of government aid programs are included but not reported. Robust standard errors are reported and robust endogeneity test following Baum et al. (2007) is used, ⁺denotes standardized coefficients

Significance level * 10 %, ** 5 %, *** 1 %

are not causal. The consequence of this is that instrumented variables of the two parents very strongly correlate. I computed the predicted education of the father and the mother using the first stage regression and find a correlation of nearly 0.95, while the actual parental education correlation is close to 0.65. Hence, in the main regression, we have a strong problem of multicollinearity, which can explain why the increase in the coefficient related to the father is compensated by the coefficient related to the mother. This problem can also be highlighted with the additional regressions IV-3b and IV-3c reported in Table 4. In these two regressions I excluded one of the two parents and used only the instruments related to the parent included in the regression. We can see that in both regressions the parameter of the parents is highly significant.

Overall, the single equation approach provided very coherent and expected results. The education of the mother seems to matter slightly more than the education of the father. The wealth index seems to matter more than the short-run consumption and the cognitive ability of the child is also an important predictor of the schooling outcome. Finally, the endogeneity tests performed on the IV estimates did not allow us to conclude that we have a serious problem of endogeneity. I now move to the simultaneous equations model which will allow us to learn more about the different channels and their relative importance.

5.2 Simultaneous regression approach

Let me now turn to the results of the simultaneous equation model introduced in Sect. 3.3. The possibility of estimating several channels simultaneously permits us not only to avoid some biases due to omitted variables, but also to quantify these biases by running regressions with some excluded variables on the same data. This idea influenced the estimation strategy and made it straightforward to estimate some simplified models alongside the complete model. This first set of estimation results is reported in Table 6. All models are estimated on exactly the same sample to avoid confounding potential differences in the coefficients due to changes in the model on the one hand and due to changes in the sample on the other. Standardized coefficients are reported and should be interpreted as changes in standard deviations

Table 5 First stage regressions

Dependent variable	IV-1 Ability	IV-1 Educ fath.	IV-1 Educ moth.	IV-2 Ability Ability	IV-3 Educ fath.	IV-3 Educ moth.
Instruments						
Father's ability measure ⁺	0.198*** (0.015)	0.182*** (0.013)	0.063*** (0.013)	0.179*** (0.015)	0.169*** (0.013)	0.055*** (0.013)
Mother's ability measure ⁺	0.215*** (0.015)	0.062*** (0.013)	0.179*** (0.013)	0.204*** (0.016)	0.048*** (0.014)	0.171*** (0.013)
Father grew up in a town	-0.032 (0.036)	0.205*** (0.037)	0.133*** (0.035)		0.207*** (0.037)	0.134*** (0.035)
Mother grew up in a town	-0.035 (0.036)	0.041 (0.036)	0.250*** (0.034)		0.043 (0.036)	0.252*** (0.034)
Control variables						
Wealth index ⁺	0.129*** (0.021)	0.179*** (0.016)	0.133*** (0.016)	0.110*** (0.021)	0.171*** (0.016)	0.128*** (0.016)
Log consumption ⁺	0.052*** (0.018)	0.208*** (0.015)	0.178*** (0.016)	0.028 (0.019)	0.204*** (0.015)	0.176*** (0.016)
Worked in 2002	-0.099** (0.043)	-0.192*** (0.033)	-0.217*** (0.033)	-0.078* (0.044)	-0.185*** (0.033)	-0.213*** (0.033)
Worked in 2005	-0.073** (0.035)	-0.139*** (0.028)	-0.182*** (0.029)	-0.058* (0.035)	-0.134*** (0.028)	-0.179*** (0.029)
First born	0.029 (0.029)	0.075*** (0.025)	0.118*** (0.025)	0.020 (0.029)	0.073*** (0.025)	0.117*** (0.025)
Gender	-0.076*** (0.027)	-0.041* (0.023)	-0.059** (0.024)	-0.072*** (0.027)	-0.036 (0.023)	-0.056** (0.024)
Number of children	-0.043*** (0.012)	-0.002 (0.010)	-0.019* (0.010)	-0.043*** (0.012)	0.001 (0.010)	-0.018* (0.010)
Number of teenagers	-0.024* (0.014)	-0.034*** (0.011)	-0.013 (0.011)	-0.021 (0.014)	-0.032*** (0.011)	-0.012 (0.010)
Father's education ⁺				0.078*** (0.019)		
Mother's education ⁺				0.018 (0.019)		
Ability of the child ⁺					0.065*** (0.013)	0.041*** (0.013)
Constant	0.641*** (0.129)	1.586*** (0.108)	1.179*** (0.110)	0.444*** (0.130)	1.545*** (0.107)	1.153*** (0.110)

Table 5 continued

Dependent variable	IV-1 Ability	IV-1 Educ fath.	IV-1 Educ moth.	IV-2 Ability Ability	IV-3 Educ fath.	IV-3 Educ moth.
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Angrist-Pischke F-statistic	69.165	42.813	64.246	206.105	40.996	52.094
Angrist-Pischke <i>p</i> value	0.000	0.000	0.000	0.000	0.000	0.000
<i>N</i>	4,266	4,266	4,266	4,266	4,266	4,266
Adj. R2	0.254	0.460	0.446	0.258	0.463	0.447

State fixed effects, government program benefits, the rural dummy and parental ethnicity and age are not reported. + denotes standardized coefficients. Significance level: * 10 %, ** 5 %, *** 1 %

Table 6 Estimation results of Eq. (6)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Consumption	0.073*** (0.018)	0.156*** (0.018)				0.086*** (0.018)
Wealth index	0.146*** (0.019)	0.227*** (0.019)				0.175*** (0.019)
Cognitive ability	0.218*** (0.015)		0.287*** (0.014)			
Father's education	0.086*** (0.018)			0.272*** (0.016)		0.114*** (0.018)
Mother's education	0.136*** (0.018)				0.282*** (0.016)	0.151*** (0.018)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
<i>R</i> ²	0.249	0.242	0.230	0.204	0.208	0.231
<i>N</i>	4,266	4,266	4,266	4,266	4,266	4,266

Standardized coefficients. Standard deviations in parentheses. Dependent variable: education index of the child. The full system of equations was estimated simultaneously, but only the coefficients of Eq. 6 are reported in this Table. Significance level *** 1 %, ** 5 % and * 10 %

of the dependent variable upon a one standard deviation change of the continuous regressors or upon a unit change in dichotomous regressors.

Model 1 is the complete model including ability, father's and mother's education and the economic situation proxied by two variables. These main regressors were accompanied by control variables such as gender, a rural area dummy, state fixed effects, social program dummies and child labor indicators which are not reported in Table 6. The full estimation results of model 1, including the remaining equations of the model, can be found in Table 11 in "Appendix B".

Considering model 1, the estimation is quite precise and all coefficients are significant at the 1 %-level. The coefficient related to the child's ability measure attains with 0.218 the highest value. Both father's and mother's education have a highly significant and positive effect. The size of the coefficient for the mother is substantially higher than the one for the father. With respect to the economic variables we can also observe a significant difference between the two. The effect of the wealth index is substantially higher than the one of consumption. This finding is coherent with the findings by Carneiro and Heckman (2002) who argue that the long run economic environment matters more than short-term credit constraints. When considering both economic effects, we see that the economic situation has the largest direct intergenerational effect on the schooling outcome of the child. In general, these results are relatively close to what was found in the OLS regression in Table 4.

Model 1 is estimated on a sample of 4,266 individuals and could potentially suffer from a sample selection bias. As discussed in the data section, I also estimate the full model relaxing some requirements on the data. In a first step, I include single mother households by dropping the channel of the father increasing the sample size to 6,547 individuals. In a second step, I include all households where data are available on either of the parents and taking the maximum value when both are available. This allows us to include 143 additional individuals, because in model 1 some were dropped just because of one missing parental characteristic. These larger sample estimates are reported in Table 11 in the "Appendix B". The general pattern is very encouraging, as almost no changes in the main regression are observed. Most coefficients increase slightly, but remain at very similar levels. The relative importance of the effects remains unchanged. Overall, these additional regressions give some support on the validity of model 1 since the results hold even when changing the sample a lot. In what follows, I take model 1 as the benchmark model, as it is the only one allowing us to control for all different channels.

5.2.1 Direct versus indirect effects

Let us now return to the discussion of model 1 from Table 6. An interesting feature of simultaneous equation models is that they enable us to compute direct and indirect effects. For example, it is clear that parental education does not only affect the schooling outcome through the direct effect discussed before, but also through the economic situation of the family. Figure 4 shows the direct and indirect effects based on the results of model 1, fully reported in Table 11. As in the discussion before, one can easily see that the ability measure of the child has the largest direct effect (black bars). The wealth index has the second largest direct effect, followed by the mother's education. However, the total effect of mother's education is larger than the total effect of the wealth index. This is due to the fact that besides the direct effect we also have an indirect effect of maternal education through both economic indicators. The same is true for the father, where the relative importance of the indirect effect is even bigger. Nevertheless, the total effect of father's education remains smaller than the one of the mother.

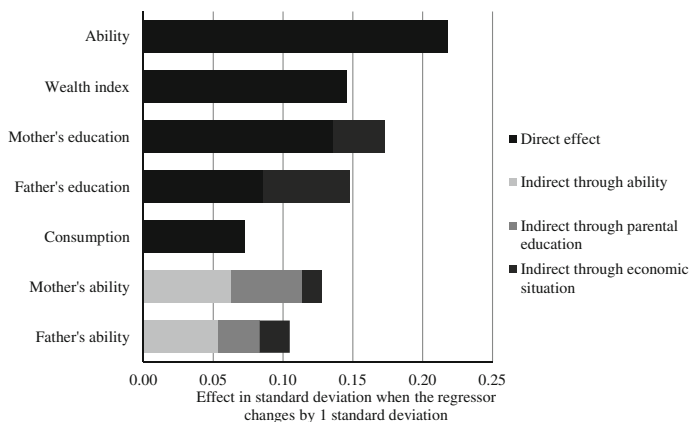


Fig. 4 Direct and indirect effects on child's schooling attainment. The figure displays the direct and indirect effects of the main variables of interest. The values are based on Model 1 reported in Table 6

Finally, parental ability has no direct effect but only indirect effects through the genetic transmission and the other two channels. The total effects attain values of about 0.13 for the mother and 0.10 for the father.

5.2.2 Biases when neglecting channels

Models 2–5 in Table 6 only include one of the four possible channels assuming the others to have no impact. The last model includes the often available data on the education of parents and the economic situation but not the ability measures. We can notice that the one-covariate models always give strongly upward biased estimators of the coefficients, when comparing them to the benchmark model in the first column. Not surprisingly, the bias in relative terms is lower for the important channels, namely ability and the wealth index, where the new coefficient is roughly 1.5 times higher than in model 1. The upward bias of parental education is much more important, since the coefficient attains 2–3 times higher values for mother's and father's education, respectively. However, the biases become much smaller when all but the ability measure are included. Due to missing information of ability measures in most of the surveys, this setting corresponds to the best we can normally do. The coefficients are about 20 % higher than that in the benchmark model, which is considerably less than in models 2–5. More importantly, the relative importance of the coefficients is very similar in model 6 as compared to model 1.

5.2.3 Regression by age groups

Based on the descriptive findings presented in Table 2 of increasing intergenerational education correlations with age, a second set of estimation results is presented in Table 7. Model 1 is estimated for different age groups and additionally for girls and boys separately. The age groups are chosen in a way that they

correspond to the age when people are normally in secondary (including high school) and tertiary education.

As for the simple correlation, I find differences between the two age groups. In general, the coefficients are slightly higher for the older group. A sharp increase is observed for the effect of mother’s education. The model fit also increases substantially from the younger the older age group. As for the simple correlations presented before, there are several possible explanations. First, it could be argued that this is due to a more precise measure of the dependent variable for the older age group. The second explanation is that the inequalities in education are a cumulative process and that the relative importance of the channels can evolve with the age of the child. Most likely both phenomena are present in these results. The fact that all indicators become more important supports the idea that the measurement is more precise for the older group. However, the fact that not all explanatory variables increase their effect in the same way points to something beyond this argument. In particular, the coefficient of the mother’s education increases substantially more than that of the others. Hence, we might have reasons to believe that the impact of the mother becomes more important with age. This could be due to the role of the mother in pushing the child to continue at school. Of course, additional research is required to confirm this conclusion, because the results could also be driven by the larger precision of the dependent variable.

5.2.4 Regression by gender

Given that mother’s and father’s education have different effects, it might be interesting to see whether the effects are also different for boys and girls. The last two columns in table 7 present model 1 for girls and boys, respectively. We can see that the two economic indicators have slightly higher coefficients for boys while the child’s ability seems to matter a bit more for girls. The education of the father is

Table 7 Estimation results of Eq. (6) for different samples

Sample	Age groups		By gender	
	12–17	18–25	Girls	Boys
Consumption	0.065*** (0.025)	0.078*** (0.025)	0.071*** (0.026)	0.079*** (0.025)
Wealth index	0.144*** (0.026)	0.151*** (0.027)	0.137*** (0.027)	0.150*** (0.027)
Cognitive ability	0.208*** (0.021)	0.232*** (0.021)	0.232*** (0.021)	0.196*** (0.021)
Father’s education	0.077*** (0.025)	0.096*** (0.026)	0.075*** (0.026)	0.094*** (0.025)
Mother’s education	0.099*** (0.025)	0.194*** (0.025)	0.172*** (0.026)	0.095*** (0.025)
Control variables	Yes	Yes	Yes	Yes
R ²	0.203	0.331	0.256	0.252
N	2,336	1,930	2,095	2,171

Standardized coefficients. Standard deviations in parentheses. Dependent variable: education index of the child. The full system of equations was estimated simultaneously, but only the coefficients of Eq. 6 are reported in this Table. Significance level *** 1 %, ** 5 %, * 10 %

somewhat more important for boys. A large difference can be observed for the role of mother's education, which has an almost twice as large effect for girls as compared to boys. The exact reasons for this difference are beyond the scope of this analysis, but it might be a very interesting question for future research.

6 Conclusion

The present study tries to contribute to the understanding of the mechanisms generating the high intergenerational education correlations observed all over the world and especially in Latin American countries. A particularly important issue is to distinguish the different channels of transmission outlined by the literature over the past years. Using very rich data from Mexico, a simultaneous equations model of the educational transmission can be estimated, allowing me to distinguish between the different channels: the biological transmission of ability, transmission through the economic situation and the education-to-education channel. Additional channels such as health or non-cognitive abilities are not considered in this study. Unfortunately, the data and especially the unavailability of retrospective information on health did not allow me to include such channels. However, these channels might be important as they might upward bias the importance of the included channels, particularly the education-to-education channel. This caveat must be kept in mind when discussing the results.

The results suggest that the economic situation of the family is the most important direct intergenerational channel, which has an even larger effect than the ability of the child when considering the effects of both economic indicators together. I distinguish between consumption as a proxy of the current economic situation and a wealth index to capture the long-term economic situation in the analysis. I find a larger effect of the wealth index, which is in accordance with findings in the literature. Parental education matters to explain children's schooling but not in a very strong way as the intergenerational correlation might lead us to expect. The mother's education directly and significantly influences the schooling outcome of children. The education of the fathers also affects the schooling outcome directly and has additionally a strong indirect effect through the economic situation of the family.

The finding that the economic situation plays an important role suggests that the current situation is likely not to be efficient. This is due to the non-optimal investment in education of the poorer children and, therefore, they cannot exploit their potential. On the other hand, the finding is encouraging in the sense that the low educational mobility does not seem to be a fatality. The strong influence of the economic situation of the parent can be targeted by public policies. In this respect, cash transfer programs (conditional or not) might help us to increase social mobility as they allow poorer families to invest more in education. Over the recent years many programs were implemented and it is, therefore, possible that they already generate beneficial effects in terms of educational mobility.

In addition to the main results I also performed the same analysis on sub-samples. These additional estimates provided interesting insights.

First, the intergenerational links are higher for older children. All coefficients increase with the age group and particularly the education of the mother becomes much more important with age. This result might suggest that the intergenerational links are following a cumulative process, suggesting that even at higher ages policy interventions can be useful. However, the differences found for the different age groups could also be due to the more precise measure of the educational attainment for the older age group. Additional research is required to distinguish the two possible explanations found in this study.

Second, I find differences in the relative importance of transmission channels between girls and boys. The economic situation of the family matters slightly more for boys while the ability of the child is somewhat more important for girls. The biggest gender difference is found for maternal education. The effect of maternal education is almost twice as high for girls as compared to boys.

The analysis demonstrates that estimates ignoring important alternative channels of transmission tend to overestimate the effects of the analyzed variables. Remaining unobserved channels such as personality traits could upward bias my results to a certain extent. Nevertheless, the used data does not allow me to consider additional channels as they would imply a large drop in the sample sizes and increase the problem of a non-random sample. Finally, the analysis should be seen as a piece among others in the recent literature aiming at understanding the mechanisms of educational mobility. For future research I see mainly three interesting directions. First, it would be useful to conduct similar analyses for other countries with low educational mobility to see whether the findings hold also outside the Mexican context. Second, the results suggest that cash transfer programs could potentially help us to increase educational mobility. Future research could look at this effect and try to find out more about the most effective specificities of such programs. Third, while most effects were relatively stable across sub-samples, the effect of maternal education changes substantially with age and gender of the child. It would, therefore, be interesting to further investigate the role of the mother in the intergenerational transmission of education.

Acknowledgments I am grateful for very helpful comments on earlier versions of this paper by Tobias Müller, Jaya Krishnakumar, Marcelo Olarreaga, Dirk Van de gaer, Duncan Thomas and the participants at conferences and seminars in Geneva, Buenos Aires, Bordeaux, Washington DC, and Mexico City and two anonymous referees. A special thanks to Isidro Soloaga for enlightening discussions on the topic and to Ian MacKenzie for correcting the writing of an earlier version.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix A: The education index: distribution and relevance

The goal of this appendix is to show that the education index used in the study is a good proxy for the final number of years of education and to provide some

additional information on the Mexican education system. I start by discussing how an educational delay is related to the final level of education.

Relevance of the education index

As outlined in Sect. 4.1.1, it is assumed that the education index is related to the years of education once the individual leaves school. That is, a delay in school at an early age should be translated in getting less education. School delay at early ages can arise from late entry to the educational system or repeating grades. In Table 8 I present a simple OLS regression of the years of education on the number of grade repetitions in primary school and the years of delay to start school. The data come from the same survey as the main analysis of the paper, but here I use only people no longer attending school.

We can see that already a late entry of 1 year is related to a decrease of total schooling of about 1 year, while students entering the system 2–4 years later have on average between 3 and 4 years less years of education in the end. Grade repetition has also a negative effect on total schooling, where 1 repetition is broadly related with one year less of schooling. It is important to notice that I do not claim that this regression identifies causal effects, which is actually not needed to show the utility of the chosen educational index.

The Mexican education system

The Mexican education system is characterized by 6 years of primary education, followed by 6 years of secondary education. Secondary education is divided in 3 years of lower secondary education (*secundaria*) and 3 years of upper secondary

Table 8 Effect of late entry and grade repetition on education outcome

	Coeff.	Std. Err
Delay of school entry		
Entry on time (6 years old)	Reference	
1 year later	-1.077***	(0.082)
2 years later	-3.005***	(0.152)
3 years later	-3.472***	(0.259)
4 or more years later	-4.027***	(0.237)
Number of grade repetition in primary		
No repetition	Reference	
1 repeated grade	-0.743***	(0.113)
2 repeated grades	-2.525***	(0.229)
3 repeated grades	-2.928***	(0.354)
4 or more repeated grades	-4.379***	(0.627)
Constant	8.598***	(0.040)
<i>N</i>	11,360	
Adj. <i>R</i> ²	0.0954	

Source Author's calculation using data from MxFLS.

Standard errors in parenthesis.

Dependent variable: years of education. Significance levels at

* 10 %, ** 5 %, *** 1 %

Table 9 Key statistics of the Mexican education system

Indicator	2002	2005
Primary		
Primary completion rate, total (% of relevant age group)	95.98	96.64
School enrollment, primary (% net)	96.51	93.75
Private schools (% of total primary enrollment)	7.95	8.05
Repeaters, primary (% of total enrollment)	5.68	4.64
Lower secondary completion rate (% of relevant age group)	65.26	72.93
Secondary		
School enrollment, secondary (% net)	58.66	63.84
Progression to secondary school (%)	92.86	93.79
Repeaters (% of total enrollment)	2.10	2.05
Tertiary		
School enrollment, tertiary (% gross)	21.07	23.31

Source World Development Indicators (<http://data.worldbank.org/country/mexico>)

Table 10 Distribution of entry age to the education system

	Entry age	Proportion (%)	Cumulative proportion (%)
	5	9.0	9.0
	6	72.7	81.7
	7	14.5	96.2
	8	2.1	98.3
	9+	1.71	100

Source Author’s calculations based on the MxFLS using 5,791 individuals between 15 and 25 years old

education (*preparatoria*). Table 9 provides additional background information of the Mexican education system for the years 2002 and 2005.

Finally, Table 10 presents statistics on late entry to the education system based on the Mexican Life Family Survey (MxFLS).

Appendix B: Full estimation results and larger sample regressions

Table 11 displays in the first column the full estimation results of model 1 already reported in Table 6 including all control variables.¹⁸

Let me highlight some interesting results I did not discuss in the main body of the paper. The education production function estimates are very similar for the mother and the father. The cognitive ability and the place of living at the age of 12 years have the large positive and the age and the indigenous background negative effects. For both the long- and the short-run economic situation the fact of being indigenous has a negative impact, even when controlling for education.

¹⁸ Except the state fixed effects and dummies for beneficiaries of government programs other than Oportunidades.

Table 11 Full estimation results of model 1-different samples

	Benchmark results		Including mothers		Highest parental values	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Equation 1 (Ability measure)						
Father's ability	0.245***	0.015				
Mother's ability	0.288***	0.014	0.385***	0.011		
Parents' ability (highest)					0.407***	0.010
Female	-0.032**	0.014	-0.015	0.011	-0.012	0.011
Old parents	-0.047***	0.014	-0.052***	0.011	0.018	0.011
Young parents	0.015	0.015	0.021*	0.012	0.020*	0.012
First born	0.021	0.015	0.020*	0.012	0.022*	0.012
Equation 2 (Father's education)						
Father's ability	0.347***	0.013				
Father's age	-0.195***	0.013				
Indigenous father	-0.053***	0.013				
Father grew up in city	0.185***	0.013				
Equation 3 (Mother's education)						
Mother's ability	0.374***	0.013	0.357***	0.010		
Mother's age	-0.211***	0.012	-0.239***	0.010		
Indigenous mother	-0.071***	0.012	-0.076***	0.010		
Mother grew up in city	0.186***	0.013	0.254***	0.010		
Equation 3' (For highest parental education)						
Parents' ability (highest)					0.365***	0.010
Parents' age (highest)					-0.246***	0.010
Indigenous parents (at least 1)					-0.078***	0.010
Parents grew up in city					0.234***	0.010
Equation 4 (Wealth index)						
Father's education	0.275***	0.016				
Mother's education	0.155***	0.017	0.329***	0.012		
Parents' education (highest)					0.362***	0.011
Father's ability	0.034**	0.014				
Mother's ability	0.118***	0.015	0.154***	0.012		
Parents' ability (highest)					0.148***	0.011
Indigenous father	-0.121***	0.016				
Indigenous mother	-0.208***	0.016	-0.276***	0.010		
Indigenous parents (at least 1)					-0.256***	0.010
Father's age	0.049**	0.021				
Mother's age	0.155***	0.021	0.133***	0.011		
Parents' age (highest)					0.117***	0.011
Equation 5 (Consumption)						
Father's education	0.300***	0.016				

Table 11 continued

	Benchmark results		Including mothers		Highest parental values	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Mother's education	0.198***	0.017	0.407***	0.011		
Parents' education (highest)					0.444***	0.011
Father's ability	0.042***	0.015				
Mother's ability	0.130***	0.015	0.154***	0.012		
Parents' ability (highest)					0.138***	0.011
Indigenous father	-0.053***	0.017				
Indigenous mother	-0.116***	0.017	-0.156***	0.011		
Indigenous parents (at least 1)					-0.140***	0.010
Father's age	0.087***	0.022				
Mother's age	0.054**	0.021	0.107***	0.011		
Parents' age (highest)					0.102***	0.011
Equation 6 (Schooling outcome)						
Cognitive ability	0.218***	0.015	0.252***	0.012	0.246***	0.012
Wealth index	0.146***	0.019	0.152***	0.015	0.155***	0.015
Consumption	0.073***	0.018	0.094***	0.015	0.086***	0.014
Father's education	0.086***	0.018				
Mother's education	0.136***	0.018	0.158***	0.013		
Parents' education (highest)					0.172***	0.013
Female	0.040***	0.014	0.043***	0.011	0.046***	0.011
Rural area	-0.007	0.017	-0.007	0.014	0.002	0.014
Oportunidades	0.057***	0.017	0.041***	0.013	0.041***	0.013
Worked in 2002	-0.143***	0.015	-0.128***	0.012	-0.122***	0.012
Worked in 2005	-0.029*	0.015	-0.036***	0.012	-0.037***	0.012
First born	0.029**	0.014	0.028**	0.011	0.027**	0.011
Number children	-0.090***	0.014	-0.080***	0.011	-0.083***	0.011
Number teenagers	-0.100***	0.015	-0.090***	0.012	-0.090***	0.012
Government program dummies	Included		Included		Included	
State fixed effects	Included		Included		Included	
Sample size	4,266		6,547		6,690	
R^2	0.249		0.256		0.264	

The dependent variable for the structural equation is reported in parenthesis in the title of each panel. The coefficients are reported in the first column and the standard errors in the second for each model. All coefficients are standardized, meaning that for continuous regressors the coefficients measure how many standard deviations the dependent variable changes when the regressor changes one standard deviation and for dichotomous variables when the variables turn from zero to one. Significance levels *** 1 %, ** 5 % and * 10 %

The second and third regressions are based on enlarged samples. The first enlarged sample includes also the single-mother households and excludes as a consequence the channel of the father. The second enlarged sample considers always the highest values of either the mother or the father. As already mentioned in the main text, the results do not change a lot despite the substantial change in the sample size.

The role of mother's ability increases in the ability equation when not controlling for the ability level of the father. This is due to the correlation among the IQ of both parents. The effect of consumption is slightly higher in the large sample regressions, but remains always considerably smaller than the effect of the wealth index. When merging the education of the two parents, the combined effect is somewhat larger than the effect of the mother in the main regression, but does not attain the sum of the two parental coefficients. The relatively stable results give additional credibility to the results on the main sample used in the study.

Appendix C: Discussion of the wealth index

In the analysis, I use a wealth index to approximate the long run economic situation of the household. In this appendix, I first present the way it was constructed and then I discuss the concern that such a wealth index could actually capture an age effect of the parents.

Construction of the wealth index

Let me first discuss how the wealth index is constructed. I use the first component of a principal component analysis performed on various indicators. This allows me to reduce the dimensions and to have a single indicator. In this appendix, I present the descriptive statistics of the indicator variables and the composition of the used wealth indicator. Table 12 displays the mean of each variable and its relative contribution to the wealth index.

Table 12 Construction of the wealth index

Indicator	Mean	Contribution (%)
Dummy for electricity	0.987	0.94
Dummy for a clean floor	0.847	15.83
Dummy for good quality roof	0.890	8.52
Number of bedrooms	2.321	10.10
Household has a phone	0.346	16.69
Household has a kitchen	0.932	2.46
Clean drainage of feces	0.845	11.57
Clean garbage evacuation	0.657	16.88
Clean cooking energy (gas or electricity)	0.826	17.01

The principal component analysis was performed using the polychoric correlation matrix to account for the non-continuity of the indicator variables

From these figures, we see that the contribution substantially varies across indicators and therefore a simple average of the indicator variable would probably not well describe wealth. Among the most important contributors, we find the clean cooking energy, the clean garbage evacuation, the availability of a phone and the indicator for a clean floor. Rather of minor importance are the indicators for having a kitchen in the household and whether the household has access to electricity. The reason for these low contribution levels is the almost full coverage among the Mexican population and the resulting small variance in these variables.

The relationship with the age of the parents

The main results of the study show that the long run economic situation proxied by the wealth index is a main channel of transmission from one generation to the next. A potential concern with this variable stems from the fact that household assets could be directly linked to the age of the parents. If older parents have systematically more of these goods and therefore a higher wealth index, then we might actually capture an age effect rather than an effect of the economic channel. In the regressions, I control for the parental age to deal with this concern. Nevertheless, a closer look at the relationship between the wealth index and parental age can help us to reduce the concerns even more.

Figure 5 displays the non-parametric regression of degree 1 of the standardized wealth index (left axis) as a function of average parental age (left graph) and the child–parents age differential (right graph). The dashed line in each graph is a density estimation of the average parental age and the child–parents age differential, respectively.

We can observe that the average of the wealth index is slightly below 0 for the youngest parents. However, the density of such young parents is rather small. For the remaining part of the parental age distribution the average is very close to zero. This is also true in the right graph where the variable on the x-axis is the age differential between parents and the child. The estimator of the mean is very close to zero for all values of the age differential.

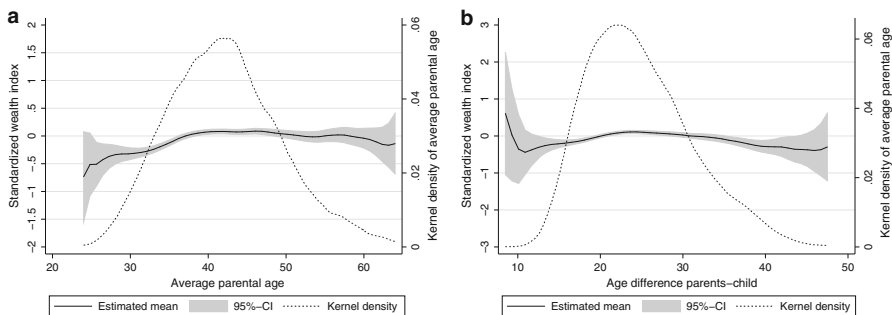


Fig. 5 Relationship between average parental age and the wealth index. The figure displays non-parametric estimates of the relationship between the wealth index and age of parents and the distribution of parental age. The estimates are based on the working sample used in the main regressions

In general, we cannot observe a strong relationship between the wealth index and the age of the parents. Thus, it is unlikely that the wealth index actually captures an age effect of the parents.

Appendix D: Identification of the simultaneous equations model

In this appendix, I discuss the identification of the model presented in Eqs. (1)–(6) in the main body of the article. To simplify the notation, the endogenous left-hand side variables of the Eqs. (1)–(6) can be combined in the matrix \mathbf{Y} and all exogenous variables in matrix \mathbf{X} , which includes all exogenous variables in white boxes in Fig. 2. This allows us to rewrite the model in the standard simultaneous equation model (SEM) notation:

$$\mathbf{Y} = \mathbf{B}\mathbf{Y} + \mathbf{A}\mathbf{X} + \boldsymbol{\zeta} \quad (7)$$

where $\boldsymbol{\zeta}$ is a vector containing the error terms of the equations, \mathbf{B} is a zero-diagonal coefficient matrix for the endogenous variables and \mathbf{A} is the coefficient matrix for the exogenous variables. Let me further define Φ to be covariance matrix of \mathbf{X} and Ψ to be the covariance matrix of the disturbance terms $\boldsymbol{\zeta}$. We impose a condition that the exogenous variables \mathbf{X} are uncorrelated with the error terms in $\boldsymbol{\zeta}$. The model described before, has the following matrix \mathbf{B} and I assume the following covariance matrix¹⁹ Ψ :

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ 0 & \gamma_9 & \gamma_{10} & 0 & 0 & 0 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & 0 \end{bmatrix} \quad \Psi = \begin{bmatrix} \Psi_{11} & 0 & 0 & 0 & 0 & 0 \\ & \Psi_{22} & \Psi_{23} & \Psi_{24} & \Psi_{25} & 0 \\ & & \Psi_{33} & \Psi_{34} & \Psi_{35} & 0 \\ & & & \Psi_{44} & \Psi_{45} & 0 \\ & & & & \Psi_{55} & 0 \\ & & & & & \Psi_{66} \end{bmatrix} \quad (8)$$

This model has a lower triangular matrix \mathbf{B} which greatly simplifies its identification. However, it is not a recursive model because I do not assume Ψ to be diagonal.²⁰ While a recursive model would be automatically identified, the conditions for this model are somewhat more complicated. To discuss the identification of the model I follow Heuchenne (1997) and Paxton et al. (2011). Heuchenne (1997) proposes a sufficient rule for identification based on \mathbf{B} and Ψ exclusively. He

¹⁹ The covariance matrix is symmetric. For presentational purpose, I only present the upper triangular version.

²⁰ I only assume uncorrelated error terms for the ability transmission Eq. (1) and the child's education Eq. (6). This last assumption is confirmed by the IV estimates I will present in Sect. 5.1. Note, however, as I will show below, these restrictions are not required for the identification.

proposes to combine the lower triangle of \mathbf{B} with the upper triangular of Ψ , which gives us:

$$\mathbf{B} \setminus \Psi = \begin{bmatrix} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{2} & \Psi_{23} & \Psi_{24} & \Psi_{25} & 0 \\ 0 & 0 & \mathbf{3} & \Psi_{34} & \Psi_{35} & 0 \\ 0 & \gamma_1 & \gamma_2 & \mathbf{2} & \Psi_{45} & 0 \\ 0 & \gamma_9 & \gamma_{10} & 0 & \mathbf{3} & 0 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \mathbf{5} \end{bmatrix} \tag{9}$$

The number (in bold) on the diagonal indicate how many excluded parameters are above in the same column and to the left of the same row. For instance, the value of 3 in the third row is obtained by counting the two zero values in the third row left to the diagonal and the zero value in the third column above the diagonal. Heuchenne (1997) shows that equation k is identified whenever the corresponding value on the diagonal is bigger or equal to $(k - 1)$. We can see that all but Eqs. (4) and (5) satisfy this sufficient condition. Hence, we can conclude that all but the economic situation equations are identified based on the lower triangular form of \mathbf{B} and the structure of Ψ .

To verify if also Eqs. (4) and (5) are identified, we have to use conditions that are not only based on \mathbf{B} but also on \mathbf{A} , the coefficient matrix of the exogenous variables. All equations in the model pass the order condition saying that an equation is identified if the number of excluded exogenous variables is equal or greater than the number of endogenous variables in that equation minus one (Paxton et al. 2011). The order condition is, however, only a sufficient condition. A stronger necessary condition is the equivalent structures approach, which is an algebraic identification technique (Paxton et al. 2011). For this, let us rewrite Eq. (7) by regrouping all parameters related to the vector Y on the left-hand side.

$$BY = AX + \xi \tag{10}$$

By doing so, matrix \mathbf{B} has a unit diagonal and all off-diagonal elements change the sign. This change of notation does not change the system but is more convenient for the computation of the equivalent structures approach. Let us now define a general matrix \mathbf{M} and define the following set of equations:

$$MA = A \quad MB = B \quad M\Sigma M' = \Sigma \tag{11}$$

The model is fully identified if the only solution to this system obtained by using the restrictions on \mathbf{A} , \mathbf{B} and Ψ is the identity matrix ($\mathbf{M} = \mathbf{I}$). Let us start with the expression of $\mathbf{MB} = \mathbf{B}$

$$\begin{aligned}
 MB &= \begin{bmatrix} m_{11} - \beta_1 m_{16} & -\beta_2 m_{16} - \gamma_9 m_{15} - \gamma_1 m_{14} + m_{12} & -\beta_3 m_{16} - \gamma_{10} m_{15} - \gamma_2 m_{14} + m_{13} & m_{14} - \beta_4 m_{16} & m_{15} - \beta_5 m_{16} & m_{16} \\ m_{21} - \beta_1 m_{26} & -\beta_2 m_{26} - \gamma_9 m_{25} - \gamma_1 m_{24} + m_{22} & -\beta_3 m_{26} - \gamma_{10} m_{25} - \gamma_2 m_{24} + m_{23} & m_{24} - \beta_4 m_{26} & m_{25} - \beta_5 m_{26} & m_{26} \\ m_{31} - \beta_1 m_{36} & -\beta_2 m_{36} - \gamma_9 m_{35} - \gamma_1 m_{34} + m_{32} & -\beta_3 m_{36} - \gamma_{10} m_{35} - \gamma_2 m_{34} + m_{33} & m_{34} - \beta_4 m_{36} & m_{35} - \beta_5 m_{36} & m_{36} \\ m_{41} - \beta_1 m_{46} & -\beta_2 m_{46} - \gamma_9 m_{45} - \gamma_1 m_{44} + m_{42} & -\beta_3 m_{46} - \gamma_{10} m_{45} - \gamma_2 m_{44} + m_{43} & m_{44} - \beta_4 m_{46} & m_{45} - \beta_5 m_{46} & m_{46} \\ m_{51} - \beta_1 m_{56} & -\beta_2 m_{56} - \gamma_9 m_{55} - \gamma_1 m_{54} + m_{52} & -\beta_3 m_{56} - \gamma_{10} m_{55} - \gamma_2 m_{54} + m_{53} & m_{54} - \beta_4 m_{56} & m_{55} - \beta_5 m_{56} & m_{56} \\ m_{61} - \beta_1 m_{66} & -\beta_2 m_{66} - \gamma_9 m_{65} - \gamma_1 m_{64} + m_{62} & -\beta_3 m_{66} - \gamma_{10} m_{65} - \gamma_2 m_{64} + m_{63} & m_{64} - \beta_4 m_{66} & m_{65} - \beta_5 m_{66} & m_{66} \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -\gamma_1 & -\gamma_2 & 1 & 0 & 0 \\ 0 & -\gamma_9 & -\gamma_{10} & 0 & 1 & 0 \\ -\beta_1 & -\beta_2 & -\beta_3 & -\beta_4 & -\beta_5 & 1 \end{bmatrix} = B
 \end{aligned}
 \tag{12}$$

From the restrictions in the last column of **B** we directly determine m_{16} to m_{66} , which greatly simplifies MB to:

$$MB = \begin{bmatrix} m_{11} & -\gamma_9 m_{15} - \gamma_1 m_{14} + m_{12} & -\gamma_{10} m_{15} - \gamma_2 m_{14} + m_{13} & m_{14} & m_{15} & 0 \\ m_{21} & -\gamma_9 m_{25} - \gamma_1 m_{24} + m_{22} & -\gamma_{10} m_{25} - \gamma_2 m_{24} + m_{23} & m_{24} & m_{25} & 0 \\ m_{31} & -\gamma_9 m_{35} - \gamma_1 m_{34} + m_{32} & -\gamma_{10} m_{35} - \gamma_2 m_{34} + m_{33} & m_{34} & m_{35} & 0 \\ m_{41} & -\gamma_9 m_{45} - \gamma_1 m_{44} + m_{42} & -\gamma_{10} m_{45} - \gamma_2 m_{44} + m_{43} & m_{44} & m_{45} & 0 \\ m_{51} & -\gamma_9 m_{55} - \gamma_1 m_{54} + m_{52} & -\gamma_{10} m_{55} - \gamma_2 m_{54} + m_{53} & m_{54} & m_{55} & 0 \\ m_{61} - \beta_1 & -\beta_2 - \gamma_9 m_{65} - \gamma_1 m_{64} + m_{62} & -\beta_3 - \gamma_{10} m_{65} - \gamma_2 m_{64} + m_{63} & m_{64} - \beta_4 & m_{65} - \beta_5 & 1 \end{bmatrix}$$

Using the first 5 rows of columns 1, 4 and 5 we can further simplify to:

$$MB = \begin{bmatrix} 1 & & m_{12} & & m_{13} & & 0 & 0 & 0 \\ 0 & & m_{22} & & m_{23} & & 0 & 0 & 0 \\ 0 & & m_{32} & & m_{33} & & 0 & 0 & 0 \\ 0 & & -\gamma_1 + m_{42} & & -\gamma_2 + m_{43} & & 1 & 0 & 0 \\ 0 & & -\gamma_9 + m_{52} & & -\gamma_{10} + m_{53} & & 0 & 1 & 0 \\ m_{61} - \beta_1 & -\beta_2 - \gamma_9 m_{65} - \gamma_1 m_{64} + m_{62} & -\beta_3 - \gamma_{10} m_{65} - \gamma_2 m_{64} + m_{63} & m_{64} - \beta_4 & m_{65} - \beta_5 & 1 \end{bmatrix}$$

Finally, using the first three rows of column 2 and 3 in **B** allows us to simplify to:

$$MB = \begin{bmatrix} 1 & & 0 & & 0 & & 0 & 0 & 0 \\ 0 & & 1 & & 0 & & 0 & 0 & 0 \\ 0 & & 0 & & 1 & & 0 & 0 & 0 \\ 0 & & -\gamma_1 + m_{42} & & -\gamma_2 + m_{43} & & 1 & 0 & 0 \\ 0 & & -\gamma_9 + m_{52} & & -\gamma_{10} + m_{53} & & 0 & 1 & 0 \\ m_{61} - \beta_1 & -\beta_2 - \gamma_9 m_{65} - \gamma_1 m_{64} + m_{62} & -\beta_3 - \gamma_{10} m_{65} - \gamma_2 m_{64} + m_{63} & m_{64} - \beta_4 & m_{65} - \beta_5 & 1 \end{bmatrix}$$

Hence, using the restrictions of **B** we are able to uniquely identify most of the elements in matrix **M**:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & m_{42} & m_{43} & 1 & 0 & 0 \\ 0 & m_{52} & m_{53} & 0 & 1 & 0 \\ m_{61} & m_{62} & m_{63} & m_{64} & m_{65} & 1 \end{bmatrix}$$

To identify the remaining elements of M , I use the condition $MA = A$. Unfortunately, it is impossible to display the full matrix due to its dimension. I only display the last three rows and transpose them for presentational purpose²¹:

$$\begin{bmatrix} \gamma_3 + m_{42} \delta_1 & \gamma_{11} + m_{52} \delta_1 & m_{65} \gamma_{11} + m_{64} \gamma_3 + m_{62} \delta_1 + m_{61} \varphi_1 \\ \gamma_7 + m_{42} \delta_2 & \gamma_{15} + m_{52} \delta_2 & m_{65} \gamma_{15} + m_{64} \gamma_7 + m_{62} \delta_2 \\ \gamma_5 + m_{42} \delta_3 & \gamma_{13} + m_{52} \delta_3 & m_{65} \gamma_{13} + m_{64} \gamma_5 + m_{62} \delta_3 \\ m_{42} \delta_4 & m_{52} \delta_4 & m_{62} \delta_4 \\ \gamma_4 + m_{43} \delta_5 & \gamma_{12} + m_{53} \delta_5 & m_{65} \gamma_{12} + m_{64} \gamma_4 + m_{63} \delta_5 + m_{61} \varphi_2 \\ \gamma_8 + m_{43} \delta_6 & \gamma_{16} + m_{53} \delta_6 & m_{65} \gamma_{16} + m_{64} \gamma_8 + m_{63} \delta_6 \\ \gamma_6 + m_{43} \delta_7 & \gamma_{14} + m_{53} \delta_7 & m_{65} \gamma_{14} + m_{64} \gamma_6 + m_{63} \delta_7 \\ m_{43} \delta_8 & m_{53} \delta_8 & m_{63} \delta_8 \\ 0 & 0 & \beta_6 + m_{61} \varphi_3 \\ 0 & 0 & m_{61} \varphi_4 \\ 0 & 0 & m_{61} \varphi_5 \\ \vdots & \vdots & \vdots \end{bmatrix}'$$

$$= \begin{bmatrix} \gamma_3 & \gamma_{11} & 0 \\ \gamma_7 & \gamma_{15} & 0 \\ \gamma_5 & \gamma_{13} & 0 \\ 0 & 0 & 0 \\ \gamma_4 & \gamma_{12} & 0 \\ \gamma_8 & \gamma_{16} & 0 \\ \gamma_6 & \gamma_{14} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}'$$

where m_{42} to m_{62} , m_{43} to m_{63} and m_{61} can be directly identified and are all equal to zero. This simplifies the remaining elements considerably:

²¹ I excluded the parameters of some control variables to save space. They are not required for the identification.

$$\begin{bmatrix} \gamma_3 & \gamma_{11} & m_{65} \gamma_{11} + m_{64} \gamma_3 \\ \gamma_7 & \gamma_{15} & m_{65} \gamma_{15} + m_{64} \gamma_7 \\ \gamma_5 & \gamma_{13} & m_{65} \gamma_{13} + m_{64} \gamma_5 \\ 0 & 0 & 0 \\ \gamma_4 & \gamma_{12} & m_{65} \gamma_{12} + m_{64} \gamma_4 \\ \gamma_8 & \gamma_{16} & m_{65} \gamma_{16} + m_{64} \gamma_8 \\ \gamma_6 & \gamma_{11} & m_{65} \gamma_{11} + m_{64} \gamma_6 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}' = \begin{bmatrix} \gamma_3 & \gamma_{11} & 0 \\ \gamma_7 & \gamma_{15} & 0 \\ \gamma_5 & \gamma_{13} & 0 \\ 0 & 0 & 0 \\ \gamma_4 & \gamma_{12} & 0 \\ \gamma_8 & \gamma_{16} & 0 \\ \gamma_6 & \gamma_{14} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}'$$

Finally, we have six equations with two unknown, which makes it very easy to determine the two remaining elements of \mathbf{M} . For instance, using the first equation we can define $m_{65} = -\frac{\gamma_3}{\gamma_{11}} m_{64}$ and plug this into the second row to find that $m_{64} = m_{65} = 0$.

Note that using the restrictions on \mathbf{A} and \mathbf{B} allows us to solve the whole matrix \mathbf{M} and we find $\mathbf{M} = \mathbf{I}$. The full model is therefore identified.

Appendix E: Regressions with additional dimensions

The results in the main body of the article include only three channels. However, there might be other important channels affecting the intergenerational transmission of education. For instance, health, non-cognitive abilities or personality traits could be transmitted from one generation to the next and affect education.

Unfortunately, the data used in this study do not allow to include such channels. There are two main reasons why I cannot include these channels in the analysis. First, including indicators on personality traits would substantially reduce the sample size and, therefore, increase the risk of sample selection biases. Second, for the health dimensions, no retrospective information about the health status of parents is provided. Hence, we could observe the health status of parents at the time of the survey, but not at the relevant time when children were attending school.

In this appendix, I present experimental regressions to show what would happen, if despite the problems we would try to include these channels. Table 13 displays the OLS regression reported in Table 4 in the main body of the text and an augmented version, where I include some parental health and behavioral variables.

The mental health indicator is based on 18 questions about the emotional situation of individuals and combined through a factor analysis. A higher value indicates more emotional problems. Parental height can have an influence as it directly affects birth weight of children (Delajara and Wendelspiess Chávez Juárez

Table 13 Single equation results including some health and non-cognitive abilities

	Model A	Model B	Model B'	Model B''	Model C
Ability measure of the child ^a	0.209*** (0.015)	0.211*** (0.019)	0.210*** (0.019)	0.213*** (0.019)	0.212*** (0.019)
Father's education ^a	0.095*** (0.018)	0.073*** (0.022)	0.075*** (0.022)	0.079*** (0.022)	0.081*** (0.022)
Mother's education ^a	0.143*** (0.018)	0.134*** (0.021)	0.133*** (0.022)	0.136*** (0.021)	0.134*** (0.021)
Wealth index ^a	0.134*** (0.020)	0.105*** (0.024)	0.106*** (0.024)	0.111*** (0.023)	0.112*** (0.024)
Log consumption ^a	0.069*** (0.018)	0.065*** (0.023)	0.062*** (0.023)	0.070*** (0.023)	0.067*** (0.023)
Gender	0.072*** (0.026)	0.068** (0.032)	0.068** (0.032)	0.067** (0.032)	0.067** (0.032)
Mental health issues (mother)		-0.002 (0.018)	-0.003 (0.018)		
Mental health issues (father)		-0.019 (0.020)	-0.020 (0.020)		
Mother's height (in m)		0.158 (0.296)	0.158 (0.296)		
Father's height (in m)		0.537** (0.266)	0.556** (0.267)		
Self confidence (father)		-0.094 (0.068)		-0.097 (0.067)	
Self confidence (mother)		-0.068 (0.078)		-0.073 (0.078)	
Respecting rules important (father)		0.122** (0.055)		0.124** (0.055)	
Some financial planning		0.019 (0.033)		0.021 (0.033)	
Constant	-0.075 (0.121)	-1.092* (0.641)	-1.174* (0.636)	0.086 (0.181)	0.033 (0.150)
Additional control variables	Yes	Yes	Yes	Yes	Yes
N	4266	2701	2701	2701	2701
Adjusted R ²	0.306	0.303	0.302	0.302	0.301

Model A is the same model as reported in Table 4 as OLS. Model B includes additional explanatory variables and Model C uses the variables of Model A and the sample of Model B. All regressions include also additional control variables such as mother's and father's age and indigenous background, an indicator for rural areas (none of these is significant). Additionally state fixed effects and control variables of government aid programs are included but not reported. Robust standard errors are reported

^a Standardized coefficients. Significance level: * 10 %, ** 5 %, *** 1 %

2013), which in turn affect the schooling outcome (Behrman and Rosenzweig 2004; Black et al. 2007). In respect to personality traits and behaviors, I include a dichotomous variables on the self-reported self confidence, on the importance of respecting rules and whether at least one of the parents aims at planning their financial situations more than just a couple of days ahead.

The results of the table underline the discussed difficulties. First, we can observe a sharp drop in the sample size, which substantially increases the problem of a non-random sample. Second, the coefficients of the newly added variables are mostly not significant. This can be due to the quality of the indicators themselves, but also to the fact that we do not observe the values for the relevant period. For instance, the mental health today is much less relevant than the mental health when the child was at school. Finally, the parameters of main interest for the study are only very little affected by the inclusion of these variables. When comparing to the baseline model estimated on the same sample as the augmented model, we can see only very little variation. Of course, this is also due to the poor explanatory power of the included variables.

References

- Alfonso M (2009) Credit constraints and the demand for higher education in Latin America. In: Proceedings of Inter-American Development Bank, Education Division, SCL, working paper #3
- Anger S, Heineck G (2010) Do smart parents raise smart children? The intergenerational transmission of cognitive abilities. *J Popul Econ* 23:1255–1282
- Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton
- Attanasio O, Kaufmann K (2009) Educational choices, subjective expectations, and credit constraints. In: Proceedings of NBER, working paper 15087
- Banerjee AV, Newman AF (1994) Poverty, incentives, and development. *Am Econ Rev* 84(2):211–215
- Baum CF, Schaffer ME, Stillman S (2007) Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata J* 7(4):465–506
- Becker GS, Tomes N (1979) An equilibrium theory of the distribution of income and intergenerational mobility. *J Polit Econ* 87(6)
- Behrman J, Rosenzweig M (2004) Returns to birthweight. *Rev Econ Stat* 86(2):586–601
- Behrman JR, Rosenzweig MR (2002) Does increasing women's schooling raise the schooling of the next generation? *Am Econ Rev* 92(1):323–334
- Binder M, Woodruff C (2002) Inequality and intergenerational mobility in schooling: the case of Mexico. *Econ Dev Cult Change* 50(2):249–267
- Björklund A, Jäntti M (2009) Intergenerational income mobility and the role of family background. In: Salverda W, Nolan B, Smeeding TM (eds) *The Oxford handbook of economic inequality*, chap 20, Oxford University Press, Oxford
- Björklund A, Hederos Eriksson K, Jäntti M, (2010) IQ and family background: are associations strong or weak? *BE J Econ Anal Policy* 10(1)
- Black SE, Devereux PJ (2010) Recent developments in intergenerational mobility. In: Proceedings of prepared for the handbook in labor economics, IZA discussion paper no 4866
- Black SE, Devereux PJ, Salvanes KG (2007) From the cradle to the labor market? The effect of birth weight on adult outcomes. *Q J Econ* 122(1):409–439
- Black SE, Devereux PJ, Salvanes KG (2009) Like father, like son? A note on the intergenerational transmission of IQ scores. *Econ Lett* 105(1):138–140
- Boudon R (1973) *L'inégalité des chances*. Armand Colin, Paris
- Boudon R (1974) *Education, opportunity and social inequality*. Wiley, New York
- Bradley RH, Corwyn RF (2002) Socioeconomic status and child development. *Annu Rev Psychol* 53:371–399

- Carneiro P, Heckman JJ (2002) The evidence on credit constraints in post-secondary schooling. *Econ J* 112(482):705–734
- Case A, Fertig A, Paxson C (2005) The lasting impact of childhood health and circumstance. *J Health Econ* 24:365–389
- Chevalier A (2004) Parental education and child's education: a natural experiment. In: Proceedings of ISA discussion paper no 1153
- CONEVAL (2008) Metodología para la medición multidimensional de la pobreza en México. Consejo Nacional de Evaluación de la Política de Desarrollo Social. <http://www.coneval.gob.mx/cmsconeval/rw/pages/medicion/multidimensional/index.es.do>
- Dahan M, Gaviria A (2001) Sibling correlations and intergenerational mobility in latin america. *Econ Dev Cult Change* 49(3):537–554
- Delajara M, Wendelspiess Chávez Juárez F (2013) Birthweight outcomes in bolivia: the role of maternal height, ethnicity, and behavior. *Econ Hum Biol* 11(1):56–68
- Désert M, Préaux M, Jund R (2009) So young and already victims of stereotype threat: socioeconomic status and performance of 6 to 9 years old children on raven's progressive matrices. *Eur J Psychol Educ* 24(2):207–218
- Devlin B, Daniels M, Roeder K (1997) The heritability of iq. *Nature* 338:468–471
- Dickson M, Gregg P, Robinson H (2013) Early, late or never? When does parental education impact child outcomes? In: Proceedings of ISA discussion paper no 7123
- Doyle O, Harmon CP, Heckman JJ, Tremblay RE (2009) Investing in early human development: timing and economic efficiency. *Econ Hum Biol* 7(1):1–6
- Ermisch J, Francesconi M, Siedler T (2006) Intergenerational mobility and marital sorting. *Econ J* 116(513):659–679
- Heckman JJ (2006) Skill formation and the economics of investing in disadvantaged children. *Science* 312(1900):1900–1902
- Hertz T, Jayasundera T, Piraino P, Selcuk S, Smith N, Verashchagina A (2007) The inheritance of educational inequality: international comparisons and fifty-year trends. *BE J Econ Anal Policy* 7(2):1–46
- Heuchenne C (1997) A sufficient rule for identification in structural equation modeling including the null b and recursive rules as extreme cases. *Struct Equ Model* 4(3)
- Holmlund H, Lindahl M, Plug E (2011) The causal effect of parents' schooling on children's schooling: a comparison of estimation methods. *J Econ Lit* 49(3):615–651
- de Hoyos R, Martínez de la Calle JM, Székely M (2010) Educación y movilidad social en México. In: Serrano J, Torche F (eds) *Movilidad social en México, Población, desarrollo y crecimiento*. Centro de Estudios Espinosa Yglesias, Mexico City
- Jensen R (2010) The (perceived) returns to education and the demand for schooling. *Q J Econ* 125(2):515–548
- van Leeuwen M, van den Berg S, Boomsma D (2008) A twin-family study of general iq. *Learn Individ Differ* 18:76–88
- Loury GC (1981) Intergenerational transfers and the distribution of earnings. *Econometrica* 49(4):843–867
- Merton R (1953) Reference group theory and social mobility. In: Bendix R, Lipset S (eds) *Class: status and power*. The Free Press, New York
- Muthén BO (2004) *Mplus technical appendices*. Muthén & Muthén, Los Angeles. <http://www.statmodel.com>
- Paxton P, Hipp JR, Marquart-Pyatt S (2011) Nonrecursive models: endogeneity, reciprocal relationships, and feedback loops. In: Proceedings of quantitative applications in the social sciences, series/number 08–168, Sage Publications, Los Angeles
- Piketty T (2000) Theories of persistent inequality and intergenerational mobility. In: Atkinson A, Bourguignon F (eds) *Handbook of income, vol 1. Distribution*, North Holland
- Raven J (2000) The raven's progressive matrices: change and stability over culture and time. *Cognit Psychol* 41:1–48
- Raven J, Court J, Raven J (1983) *Manual for Raven's progressive matrices and vocabulary scales (section 3), coloured progressive matrices, 1983rd edn*. Lewis, London
- Raven J, Court J, Raven J (1986) *Manual for Raven's progressive matrices and vocabulary scales (section 2), coloured progressive matrices (1986 edition with US norms)*. Lewis, London
- Rojas M (2007) A subjective well-being equivalence scale for mexico: estimation and poverty and income-distribution implications. *Oxf Dev Stud* 35(3):273–293

- Rosa Dias P (2009) Inequality of opportunity in health: evidence from a UK cohort study. *Health Econ* 18(9):1057–1074
- Rubalcava L, Teruel G (2006) Gufa del usuario para la primera encuesta nacional sobre niveles de vida de los hogares. <http://www.enmvih-mxfls.org>
- Sewel WH, Shah VP (1968) Social class, parental encouragement, and educational aspirations. *Am J Sociol* 73(5):559–572
- Solon G (2004) A model of intergenerational mobility variation over time and place. In: Corak M (ed) *Generational income mobility in North America and Europe*. Cambridge University Press, Cambridge
- Steinberg L, Lamborn SD, Dornbusch SM, Darling N (1992) Impact of parenting practices on adolescent achievement: authoritative parenting, school involvement, and encouragement to succeed. *Child Dev* 63(5):1266–1281
- Stinebrickner TR, Stinebrickner R (2007) The effect of credit constraints on the college drop-out decision: a direct approach using a new panel study. In: *Proceedings of NBER working paper 13340*
- Winter C (2007) Accounting for the changing role of family income in determining college entry. European University Institute, working paper ECO 2007/49