Policy Research Working Paper 8816

Education Spillovers in Farm Productivity

Revisiting the Evidence

Véronique Gille

WORLD BANK GROUP
Development Economics Vice Presidency
Strategy and Operations Team
April 2019

Policy Research Working Paper 8816

Abstract

This paper exploits the social organization of India to revisit the question of education spillovers in farm productivity. The fact that social interactions mainly occur within castes in rural India provides tools to show that the observed correlation between farm productivity and neighbors' education is likely to be a spillover effect. In particular, there are no cross-caste and no cross-occupation effects, which underlines that, under specific assumptions, which are stated and explored in the paper, the education of neighbors does not capture the effect of group unobservables. This evidence is complemented by separate estimations by crops, which show results that are consistent with education spillovers. The strategy used in this paper helps understand and interpret previous findings from the literature.

This paper is a product of the Strategy and Operations Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/research. The author may be contacted at gille@dial.prd.fr.

Education Spillovers in Farm Productivity: Revisiting the Evidence

Véronique Gille

JEL Classification: D13, I25, O12, Q12

Keywords: education externalities, rural India, farm productivity, caste

1. Introduction

This paper investigates the question of social returns to education by studying the existence and the magnitude of education externalities in agriculture in rural India. Quantifying education externalities is key to the design and the evaluation of schooling systems and education policies: if social returns of education are higher than private ones, schools should be publicly financed or students subsidized such that individual investments in education become socially optimal. It is of particular relevance in India, where the number of private schools has soared in the last 10 years. The percentage of 6-to-14-year-old children enrolled in private schools almost doubled between 2005 and 2014, going from 16.3 percent to 30.8 percent.[1] Understanding the importance of education externalities helps the evaluation of the long-term consequences of these structural changes in the schooling system. The focus of this paper on agriculture in rural areas is motivated by the fact that 70 percent of the population of India lives in rural

Véronique Gille is a research fellow at Développement, Institutions et Mondialisation (DIAL), Laboratoire d'économie de Dauphine (LEDa), Institut de Recherche pour le Développement (IRD), Université Paris-Dauphine, Université PSL, 75016 Paris, France; her e-mail address is gille@dial.prd.fr. The author thanks the editor and three anonymous referees for their comments and suggestions, which have considerably improved the paper and Flore Gubert and Kaivan Munshi for insightful discussions. The author also thanks Andrew Foster for providing the data.

1 Annual Status of Education Report (2005, 2014).

areas, of which 72 percent depends on agriculture (Census of India, 2011). Agriculture accounts for 17 percent of GDP and employs 51 percent of the total labor force (World Bank, 2010).[2]

This paper studies the productivity effects of the mean education of neighbors in the agricultural sector, which can take place through learning spillovers. First, farmers can learn from their neighbors about agricultural technology. The fact that there are learning spillovers when there is a technological change has been widely asserted in the literature (see for example in the case of agriculture in developing countries Foster and Rosenzweig [1995]; Munshi [2004]; Conley and Udry [2010]). Second, neighbors' education can influence the *adoption* of technologies, through their own adoption of technology. It can also have a positive impact on the *use* of technologies, or more generally, neighbors' education can increase efficiency in the use of inputs (Weir and Knight 2007). It can be the case if all farmers are not on the production frontier, which would be a signal that certain inputs, such as fertilizers, manure or grains are over- or underused. Finally, neighbors can have an *allocative effect*: their education level would have an impact on farmers through learning about which input or output to choose, given their relative prices (Kumbhakar and Lovell 2003).

The existing empirical literature mainly focuses on education externalities in cities or in firms in developed countries (see for example Acemoglu and Angrist 1999; Moretti 2004a, b). Very few papers study education externalities in agriculture in developing countries, with a majority subject to limitations inherent to the study of peer effects, namely the problem of definition of peers, and the identification challenge due to group-level unobservables. Appleton and Balihuta (1996) introduce neighbors' average level of education in the production function of farmers in rural Uganda. They find that the proportion of farmers with primary schooling is positively and significantly related to farmers' productivity. However, they cannot control for unobserved community effects by adding community dummies as neighbors' education is calculated at this level. Weir and Knight (2007) estimate the average and stochastic frontier production function with neighbors' education as a control variable in rural Ethiopia. Here, neighbors are defined as groups of households within «sites», so they are able to control for site dummies. They find a positive relation between neighbors' education and average production, but they do not find any impact of neighbors' education on farmers' efficiency. Again, although some unobservables are controlled for by the site dummies, it cannot be ruled out that the observed relation between neighbors' education and average production is driven by unobservables at a smaller level than the site. Asadullah and Rahman (2009) use the same methodology to study external returns of education in agriculture in Bangladesh. They control for village fixed effects while defining neighbors at a smaller level called «Bari». In contrast to Appleton and Balihuta (1996) and Weir and Knight (2007), they find no evidence of external returns of education on farm productivity. However, the (absence of) results may be driven by the sample design of the data: their database only reports information on two households per Bari. This definition of neighbors allows for village dummies but may be too restrictive to capture any external effect. Gille (2012), using spatial econometrics tools to study education externalities in farm productivity in India, also finds evidence of a positive correlation between neighbors' education and productivity. But the strategy used does not rule out unobservables at the group level. These four papers illustrate the current trade-off in the empirical literature: It is necessary to control for geographical fixed effects to prove that the captured effect does not reflect a spurious correlation, but controlling for geographical effects may lead to a too restrictive definition of neighbors.

This paper overcomes this trade-off by following the strategy of Munshi and Myaux (2006), who study the impact of other women's behavior towards contraceptives on own adoption of contraceptives. The authors take advantage of the fact that women in rural Bangladesh interact solely within their religious group to define women's reference group and to rule out that the results are driven by group unobservables

2 Agriculture represented 1.8 percent of total GDP in France in 2009 and 1.2 percent in the United States in 2011.

that have some correlation across groups. Rural India provides a similar context, with an organization of social interactions within castes. Although social mobility and social mixing have increased in urban India, life in villages is still organized along caste lines (Damodaran 2008; Munshi 2011; Munshi, Myaux, and Rosenzweig 2016). Education spillovers are therefore expected to occur within caste, while cross-caste effects should be absent.

This prediction is tested using data from the 2006 round of the Additional Rural Incomes Survey and Rural & Demographic Survey (ARIS-REDS), a household survey conducted in rural India. This dataset has several advantages over the datasets used in previous literature. First, the round of 2006 has a *listing* of every household in the surveyed villages with basic demographic characteristics information on households and heads of household, such as their caste, gender, age, education level, size of the household and main occupation. This is very valuable information as it makes it possible to compute the average level of education of members from the same caste using all the heads of households from the same caste in the village and not only the heads of household that have also been sampled as is usually done in the literature. The education of neighbors is therefore very precisely measured. The other advantage of the village listing is that each sampled household is surveyed twice, once in the survey, and once in the listing, which is useful when dealing with potential measurement error as explained in section 4. Second, households report their subcaste name, which makes it possible to define the caste group at a very thin level. Finally, this round is part of a panel with earlier rounds that are exploited for robustness checks.

The findings show that the education level of households from the same caste is positively correlated with farm productivity, but there are no cross-caste effects, and no effects from members from the same caste that do not have agriculture as their main activity. These results are robust to changes in specifications, and in particular to a specification with household fixed effects, using the round of 1982. The absence of cross-caste and cross-occupation effects rules out the possibility that the estimated education spillover effect is driven by village or caste-level unobservables that are correlated within castes. Indeed, as shown by Munshi and Myaux (2006), and further elaborated on in section 2, these results can only be obtained without any education spillover if the omitted determinants of farm productivity are totally uncorrelated across castes, as well as within castes across occupations. The panel specifications also rule out that the positive correlation between education of the caste and productivity is driven by unobservables that are fixed over time. Additionally, there is an effect of the education of members from the same caste only for wheat growers, while there is no effect for rice growers. This pattern is consistent with a spillover effect as argued in section 4 and highlights that the omitted factors correlated with neighbors' education should only determine wheat production but not rice production to explain the positive effect of neighbors' education on productivity. Finally, the effect of the education of members from the same caste solely comes from members from the same subcaste, which shows that interactions with respect to learning happen at the level of the subcaste in rural India.

Section 5 discusses the properties that the unobservables need to have to explain away the results. In particular, the methodology suggested by Oster (2017) is used to compute the relative degree of selection on unobservables to selection on observables for which the true effect of the mean education level of the group would be zero. In most cases the selection on unobservables needs to be larger than the selection on observables, which can be interpreted as an indication that the true effect of the education of members from the same group is nonzero, as suggested by Oster (2017).

This paper therefore provides evidence showing that there is an impact of neighbors' education when neighbors are adequately defined and measured, and that this result is unlikely to be driven by omitted variables. It is, however, important to keep in mind that the identification relies on the assumption that time-varying omitted variables correlated with the educational level of groups and to agricultural productivity are also correlated within villages across castes or within castes across occupations. If this is not the case, it is possible that the education level of members from the same caste proxies for other caste-level unobservable characteristics, and causality cannot be inferred.

The rest of the paper is organized as follows. Section 2 describes the theoretical framework, the empirical strategy and discusses the identification challenges. Section 3 describes the data and provides some descriptive statistics. Section 4 shows the results. Finally, section 5 discusses the identification assumptions and section 6 concludes.

2. Theory and Empirics

Identifying peer effects is a well-known empirical challenge (see for example Manski 1993; Brock and Durlauf 2001). The three main issues are the endogeneity of network formation, the problem of correlated effects, and the reflection problem (Manski 1993). The endogeneity of network formation refers to the problem arising from the fact that individuals may choose the individuals with whom they interact according to specific characteristics related to the output of interest (Jackson 2008). For example, if households with a high productivity choose to interact with households with a high education level, the estimate on peers' education will be biased upward. The problem of correlated effects arises when the peers' variable that one is interested in captures the effect of other group-level variables that cannot be controlled for. The reflection problem refers to the fact that if your neighbor influences you, then it is most likely that you also influence your neighbor.

This section, after presenting the theoretical framework from which the empirical specification is derived, explains how the identification challenges arising from the endogeneity of network formation and the problem of correlated effects are addressed. The reflection problem, which arises when the group variable under study is the dependent variable aggregated at the group level (e.g., when it is the impact of the agricultural productivity of the reference group on the agricultural productivity of households) is here avoided because the group variable is neighbors' education level. This variable is predetermined and different from the dependent variable, namely the agricultural output of households.

Theoretical Framework

Agricultural output is produced according to a Cobb-Douglas production function, where productivity is allowed to depend on neighbors' education. Namely, output y of household i is produced with three factors of production, land l, labor n, and physical capital k, as follows:

$$y_i = A_i k_i^{\alpha} n_i^{\beta} l_i^{\gamma} \tag{1}$$

where each household has a specific productivity A_i. The external effect of human capital can be captured by allowing A_i to depend on the surrounding human capital, as in Moretti (2004a). In other words, A_i can be written as

$$A_i = f(\bar{E}) \tag{2}$$

where $\bar{E}$ is the average human capital in the neighborhood of the household farm. It is important to note that this specification assumes that neighbors' education augments the productivity of the three factors of production. Other assumptions are possible: neighbors' education could only increase labor productivity, for example, but as argued by Moretti (2004a), it is empirically hard to distinguish between all alternative explanations. Moreover, given that in this particular context education externalities go through learning spillovers, it is not unlikely that neighbors' education increases the productivity of the three factors of production. Upon taking logs, equation (1) reads:

$$\ln y_i = \ln A_i + \alpha \ln k_i + \beta \ln n_i + \gamma \ln l_i \tag{3}$$

In turn, agricultural productivity is allowed to depend on neighbor's education E_c and on household characteristics X_i in a linear way, such that

$$\ln A_i = \delta_c E_c + X_i \rho \tag{4}$$

Definition of Neighbors

How to define neighbors? From whom will individuals most likely learn about agricultural productivity? The social structure of India, the caste system, is exploited to define neighbors as in Gille (2012). The problem of the endogeneity of network formation is therefore avoided by using a definition of peers that is exogenously determined. Using an exogenous definition of the peer group is common practice in the literature. However, because of the lack of precise data, the chosen definition is often geographical, such as the village (e.g. Case 1992; Foster and Rosenzweig 1995; Munshi 2004) or the neighborhood (Weir and Knight 2007; Asadullah and Rahman 2009).

Caste is defined as an "hereditary, endogamous, usually localized group, having a traditional association with an occupation, and a particular position in the local hierarchy of castes. Relations between castes are governed, among other things by the concepts of pollution and purity, and generally maximum commensality[3] occurs within the caste" (Srinivas 1962).[4] Although customs and traditions evolve quickly in India, these changes are mostly happening in urban India. In rural India, caste is still the group of reference within which interactions occur (Damodaran 2008; Munshi 2011; Munshi, Myaux, and Rosenzweig 2016).

Here, two different definitions of caste groups are used. In most specifications, households are grouped into four caste groups: the Scheduled Castes (hereafter SC), the Scheduled Tribes (ST), the Other Backward Classes (OBC) and the High Castes. SC and ST are at the bottom of the hierarchy. OBC are slightly higher in the hierarchy, but are still considered as low castes. High Castes as their name indicate are at the top of the hierarchy. Under this definition neighbors are members from the same broad caste group in the same village.

However, recent literature shows that the relevant reference group is the subcaste or jati when commitment or trust is an issue or when social norms are being established (Munshi, Myaux, and Rosenzweig 2006, 2016). With respect to learning spillovers, whether it is the broad caste group or the subcaste or jati level that matters is not so clear. In the survey, individuals were asked to identify the five most reliable persons in the village that they contacted to get information on various topics. Overall, approximately 70 percent of the persons mentioned are from the same broad caste group as the respondent, and 50 percent are from the same subcaste group. When focusing only on information flows related to agriculture, the proportions are similar. This paper tentatively answers whether it is the jati level that matters or the broad caste group by defining neighbors as members from the same jati in some specifications. These two definitions of «neighbors» are of course an approximation of real interactions. It may be that there are households from the same caste that do not influence each other or households from another caste group that have an impact on productivity. Falsely excluding households that have an impact from the group of neighbors or falsely including households that have no impact is similar to a classical measurement error. Under the assumption that the measurement error is uncorrelated with the regressors, the estimates should be downward biased (Wooldridge 2003).

The Problem of Group-Specific Omitted Variables

As mentioned in section 2, another difficulty in the estimation of peer effects is the problem of group-specific omitted variables, also called correlated effects in the literature (Manski 1993). The education

3 That is to say the act of eating together.

4 For more information on castes, see Deliege (2004).

of members from the same caste may be correlated with unobserved caste characteristics that also have an impact on productivity. If this is the case, the coefficient on the education of members from the same caste will be biased. The identification strategy that is used to partially deal with this issue is taken from Munshi and Myaux (2006). In their paper, they study the impact of the use of contraception of women from the same community on own use of contraception. They find that the coefficient on the use of contraceptives of women from the same community is positive and significant, whereas the use of contraceptives of women from the other community in the village has no impact. They argue that this result could only reflect a spurious correlation between contraceptive behavior at the community level and contraceptive behavior that is in fact driven by unobservables if unobservables are fully uncorrelated across communities in the same village. This section reproduces their argumentation and applies it to the specification of this paper. It therefore closely follows the wording of their paper.

Equation (5), which is derived from equations (3) and (4), illustrates the problem of omitted variables:

$$\ln y_i = \theta + \delta_c E_c + \delta_o E_o + X_i \rho + F_i \zeta + U_c + v_i \tag{5}$$

$\ln y_i$ is the agricultural production of household i, θ is a constant, E_c is the average education of caste members in the same village (which excludes the education of household i), X_i are individual characteristics, F_i represents the factors of production (land, labor and physical capital) and U_c the caste-level characteristics that cannot be controlled for in the estimation because U_c is unobserved. v_i is the error term. Cross-caste effects are introduced through E_o, which is the mean education level of village members from other castes. Because interactions are happening within castes, the education level of households from the same caste is expected to have a positive and significant impact, whereas the education level of households from other castes in the village should have no impact. In other words, we expect $\delta_c > 0$ and $\delta_o = 0$.

It is easy to see that if there is no impact of neighbors' education on agricultural productivity, the true specification is:

$$\ln y_i = \theta + X_i \rho + F_i \zeta + U_c + v_i \tag{6}$$

In this case, E_c in equation (5) proxies for the unobserved U_c if E_c and U_c are correlated. Put differently, a positive estimate of δ_c when the true value is zero could be obtained if the education level of individuals from the same caste in the village is correlated with unobserved caste characteristics that have an impact on agricultural productivity. As E_c cannot by itself perfectly proxy for U_c, in a model where there is no education spillovers E_o is an additional proxy for U_c.

Can we get $\hat{\delta}_c > 0$ and $\hat{\delta}_o \approx 0$, the expected result from a model with education spillovers, in a model without education spillovers? Yes, but only if E_o does not provide any information on U_c. This could only be the case if the unobserved characteristics of members from the same caste and members from other castes in the village, respectively U_c and U_o, are orthogonal. Therefore, to explain the pattern where $\hat{\delta}_c > 0$ and $\hat{\delta}_o \approx 0$ *without education spillovers*, U_c and U_o must be totally uncorrelated within villages. When estimating equation (5), getting $\hat{\delta}_c > 0$ and $\hat{\delta}_o \approx 0$ consequently rules out that E_c, the education level of households from the same caste, captures caste-level omitted variables *correlated across castes* in the village.

However, it does not rule out the case where E_c captures omitted variables which are uncorrelated across caste groups. To take that into account, the impact of households from the same caste that have agriculture as their main activity, whose education level is E_{AGRc}, is now estimated. The other group is members from the same caste that do not have agriculture as their main activity, and their education level is noted $E_{nonAGRc}$. The equation to estimate is:

$$\ln y_i = \theta + \delta_{AGRc} E_{AGRc} + \delta_{nonAGRc} E_{nonAGRc} + X_i \rho + F_i \zeta + U_{AGRc} + v_{ic} \tag{7}$$

The same reasoning applies here. If $\hat{\delta}_{AGRc} > 0$ and $\hat{\delta}_{nonAGRc} \approx 0$, it can be ruled out that E_{AGRc} captures the unobserved characteristics U_{AGRc}, as long as U_{AGRc} is not totally uncorrelated with the unobserved characteristics of households from the same caste but not having agriculture as their main activity, $U_{nonAGRc}$. Given the social structure of India, assuming that unobserved characteristics are correlated within castes across occupations seems reasonable. This assumption is discussed in more detail in section 5.[5]

3. Data and Descriptive Statistics

Data and Variables Definition

The data used to conduct this study are the ARIS-REDS data from the National Council of Applied Economic Research (NCAER). Since 1971, the NCAER has been conducting surveys on a *sample* of households in 232 villages in the 17 major states of India. Most of the results come from the round of 2006, which has the advantage of providing a *listing* of every household in the surveyed villages with basic information on households and heads of household such as their demographic characteristics (gender, age, education level, size of the household) and their main occupation. The average level of education of members from the same caste can therefore be computed using all the heads of the households from the same caste in the village and not only the heads of the households that have been sampled, which makes the measure very precise. The information on education from the listing is also useful to deal with measurement error in education of the head of household, as is shown in section 4. The round of 1982 is used for robustness checks.

In the round of 2006, detailed information about agricultural inputs, such as cultivated area or labor use, is not provided in the listing, so the analysis is restricted to households that cultivate land in the *sample*. The identification strategy also requires at least two castes per village and that at least one caste has households that do not have agriculture as their main activity and households that have agriculture as their main activity. The final sample is composed of 4405 households in 223 villages.

The first set of variables includes the agriculture-related variables, used to estimate the production function. The dependent variable is total agricultural output of households.[6] As most households cultivate several crops, the different crops are aggregated, using their value in rupees declared by the household. When the production function is estimated separately for wheat and rice, total agricultural output is a quantity, measured in quintals (100 kilograms). The factors of production are the area of land cultivated (in acres), the number of days worked on the land (this measure includes family workers as well as hired labor), and variables proxying for the capital used, namely mechanized assets and nonmechanized assets. Mechanized assets and nonmechanized assets are dummy variables equal to one if the household respectively owns at least one mechanized or one nonmechanized asset.[7] Additional controls include the

5 It is also worth noting that $\hat{\delta}_o$ and $\hat{\delta}_{nonAGRc}$ provide insights into the bias in the estimates of the other parameters of the production function. Production functions are typically hard to estimate because inputs are endogenous. Finding that $\hat{\delta}_o \approx 0$ and $\hat{\delta}_{nonAGRc} \approx 0$ mean that the estimates of the other coefficients of the production function can be interpreted as the true coefficients if the omitted determinants of productivity are correlated across castes or within castes across occupations.

6 The focus is only on field crops. Plantation crops, such as coffee or tea are excluded, to keep some homogeneity in the sample. It is however not a strong restriction because only six households in the data cultivate plantation crops.

7 Mechanized assets include tractors, trailers, threshers, electrified motors, nonelectrified motors (oil engine), gauge wheels, plows, tractors with trailer, tractors with thresher, tractors with oil engine, tractors with gauge wheel, tractors with plough, disc harrows, tillers or cultivators, plow discs or mould boards seed drills, power tillers, power sprayers, chaff cutters, cane crushers, combine harvesters. Nonmechanized assets include iron ploughs, cultivators, harrows, levelers, hoes or manual earth removers, seed drills, weeders, sprayers or dusters, winnowers or pitch forks, bullock carts, chaff cutters, sickles, scissors.

characteristics of the soil,[8] the proportion of cultivated land that is irrigated and the proportion of land owned (as opposed to rented) by the household.

The second set of variables includes the demographic characteristics, namely the age of the head of household, her/his gender, her/his education level (in years of schooling) and the number of members in the household. A set of dummies indicates the caste group of the household. Village characteristics, that is, the population of the village (in log) and the amount of land in the village, are also controlled for. Neighbors' education level is the mean education level of the head of household from the reference group, which depends on the specification. In all cases, the education level of the head of household is excluded from the mean calculation. Neighbors' demographic characteristics are computed the same way. Finally, some specifications control for the wealth of neighbors, proxied by the average amount of land owned in the group, and for group transfers, which is the amount of money received as a loan from children, parents, or relatives and from friends within the village, aggregated at the group level. This variable proxies for the intensity of insurance networks.

There is less information in the round of 1982. In particular, there is no information about hired labor time, quality of land, and group transfers. There is also no village listing with detailed demographic characteristics available, so the caste education level can only be computed using sampled households. This reduces the precision of the measure of education and of the other group level controls. All the other variables are built as in the 2006 round. To build the panel, the sample is restricted to households that are observed in both rounds and that declare the same caste in the two rounds. To preserve the maximum number of observations, households that have split between the two rounds are considered as one household.

Descriptive Statistics

Table 1 reports descriptive statistics for the sample of households. On average the head of household is around 51 years old and has primary education (5 years of education). Only 5 percent of the heads of household are women, and on average a household is composed of 6 individuals. In terms of agricultural equipment, only a third of the households have mechanized assets, but almost every household has at least one nonmechanized asset (96 percent). Interestingly, a large proportion of the cultivated land is owned by the household (93 percent) and irrigated (64 percent). Also, new technologies are widely used: 63 percent of the households have at least one crop for which they are using high-yielding varieties (thereafter HYV). However, caste groups are very heterogenous. Although pure demographic characteristics such as age, household size, and percentage of households with a female head of household do not differ much from one group to the other, the other variables such as education or cultivated areas are very different and follow the traditional hierarchy. SC and ST have a lower education level and cultivate less land than OBC, who are themselves worse off than high castes. Similarly, the proportion of low-caste land that is irrigated or the probability that a low caste has mechanized equipment is much lower than for a high caste. The fact that human capital and wealth are distributed along caste lines is not very surprising. It is a well reported phenomenon in the literature (see e.g. Kijima 2006). These statistics confirm that in rural India, the economic situation is still highly related to the caste one is born in.

Neighbors' characteristics are calculated with the *listing* that provides information for every household in the village. The first set of neighbors' descriptive statistics is when neighbors are defined as members from the same caste, whatever the occupation. Neighbors are slightly different from the sample of farmers. They have smaller households, and a lower education level on average. Reassuringly, when looking only

8 The characteristics for which there is information are the depth (up to 1 foot, 1 to 3 feet, more than 3 feet), the color (red, black, grey, yellow, brownish black, off-white), the characteristics (sand, loam, light clay, heavy clay, gravel, latrite), the salinity (nil, moderate, high), the rate of percolation after one round of irrigation (fast, medium, slow) and the ease of drainage in case of heavy rainfall (easy, moderate, difficult).

Table 1. Descriptive Statistics

	Full sample		SC		ST		OBC		High castes	
Age (Years)	50.84	(13.15)	50.41	(13.31)	47.50	(12.65)	50.47	(12.73)	52.42	(13.63)
Household size	6.09	(3.28)	5.78	(2.93)	5.83	(3.45)	6.26	(3.22)	6.02	(3.42)
Female head	0.05	(0.23)	0.06	(0.24)	0.06	(0.23)	0.05	(0.22)	0.06	(0.23)
Head education level	5.15	(4.94)	4.07	(4.49)	2.89	(3.86)	5.02	(4.93)	6.32	(5.03)
Farm characteristics										
Cultivated area (in acres)	6.92	(10.20)	3.75	(4.62)	6.03	(6.54)	6.69	(9.22)	8.64	(13.09)
Days worked on field	129	(139)	97	(102)	92	(74)	132	(147)	147	(148)
Mechanized assets	0.32	(0.47)	0.12	(0.32)	0.15	(0.35)	0.31	(0.46)	0.46	(0.50)
Nonmechanized assets	0.96	(0.21)	0.95	(0.22)	0.97	(0.16)	0.96	(0.20)	0.95	(0.22)
Proportion of land owned	0.93	(0.24)	0.88	(0.30)	0.98	(0.13)	0.92	(0.25)	0.94	(0.22)
Proportion of irrigated land	0.64	(0.46)	0.58	(0.49)	0.44	(0.47)	0.68	(0.45)	0.66	(0.46)
High-yielding variety	0.63	(0.48)	0.55	(0.50)	0.64	(0.48)	0.57	(0.49)	0.73	(0.45)
Characteristics of neighbors										
All neighbors										
Mean age	48.12	(3.49)	46.98	(2.88)	45.44	(2.72)	47.57	(2.70)	50.05	(3.93)
Mean household size	5.50	(1.01)	5.25	(0.84)	5.23	(0.85)	5.59	(1.03)	5.54	(1.05)
Female head	0.09	(0.06)	0.10	(0.06)	0.09	(0.06)	0.08	(0.05)	0.09	(0.06)
Mean education level	4.59	(2.04)	3.50	(1.55)	2.73	(1.99)	4.32	(1.71)	5.87	(1.89)
Neighbors in agriculture										
Mean age	49.59	(3.88)	50.41	(4.38)	46.97	(4.07)	49.18	(3.15)	50.61	(4.16)
Mean household size	5.79	(1.13)	5.61	(1.14)	5.55	(0.95)	5.95	(1.14)	5.69	(1.13)
Female head	0.06	(0.06)	0.07	(0.08)	0.07	(0.08)	0.06	(0.05)	0.06	(0.07)
Mean education level	4.77	(2.23)	3.70	(2.05)	2.77	(2.10)	4.58	(2.02)	5.97	(1.94)
Observations	4405		511		381		2085		1428	

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.
Note: The table reports the mean for each variable. Standard deviations are in parentheses. The first column reports the mean for the full sample. Columns 2 to 4 divide the sample into castes groups. SC stands for "Scheduled Castes," ST for "Scheduled Tribes," OBC for "Other Backward Classes."

at neighbors with farming as their main activity, neighbors are really similar to the sample of farmers. The only difference is on the education level, which is lower for the sample of farmers. This may be due to the fact that the definition of farmers is different in the *sample* and in the *listing*. In the sample a farmer is anybody that has some agricultural output, whereas in the listing farmers additionally include those that declare agriculture as their primary occupation.

4. Results

Empirical Specification

The baseline equation to be estimated for agricultural productivity is derived from equations (3) and (4):

$$\ln y_i = \theta + \delta_c E_c + \delta_o E_o + \alpha k_i + \beta \ln n_i + \gamma \ln l_i + X_i \rho + X_v \zeta + C + D_s + v_i \tag{8}$$

θ is a constant. y_i is the agricultural production of household i. Neighbors' education E is calculated over individuals from the same caste c in the same village (excluding household i) and E_o, the mean education level of members from other castes in the same village is also included. The agricultural production depends on factors of production: k_i is a dummy variable, which equals one if the household owns any agricultural equipment, n_i is the number of days worked on the land, l_i is the amount of land cultivated.

X_i is a vector of household and land characteristics,[9] X_v of village characteristics.[10] The heterogeneity in productivity across castes and States is taken into account by controlling for Caste dummies C and State dummies D_S. The characteristics of members from the same caste and from other castes are also added in some specifications.

Standard errors are corrected for clustering at the village level, since individuals are likely to be correlated within villages.

Caste and Cross-Caste Impact

This section first looks at the impact of the education of members from the same caste and of members from other castes in the village on agricultural output. Table 2 presents four different specifications, and each column has a different set of controls. Column 1 only controls for farm and demographic characteristics as well as village characteristics (population of the village in log and land area). Column 2 additionally controls for variables at the caste in the village level, namely the number of caste members in the village and the caste mean value of demographic characteristics (mean age of the head of household in the caste, percentage of female heads of household in the caste, and mean household size in the caste). Column 3 controls for the same variables calculated over members of the other castes in the village. Column 4 additionally controls for the amount of land owned by caste members and by members of the other castes as a proxy for the wealth of the caste and of the other caste, as well as for the amount of transfers within the caste as a proxy for the strength of the insurance network in the caste. To keep the table easy to read, the coefficients on demographic and farm characteristics are reported in table S1.1 in the appendix.

The first specification in table 2 is the baseline specification: It controls for State dummies in addition to the controls cited above. The mean level of education of members from the same caste in the same village is positive and significant: an additional year in the mean level of education of members from the same caste increases agricultural productivity by 4.4 to 6.1 percent. On the contrary, the education level of members from other castes in the village has no impact. Whatever the set of controls, the coefficient is close to zero and not significant. The education of the head of household does not seem to have any impact on agricultural production.

However, there may be measurement error in the education of the head of household. It could be that the education of members from the same caste is a better measure of the education of the head of household than her/his own response. Specification 2 instruments the education level of the head of household, taken from the *survey*, by the measure of the education level of the head of household in the *listing*. Instrumenting by a measure of one variable by another measure of the same variable corrects for measurement error if the two measures have uncorrelated measurement errors (Wooldridge 2010). The household survey and the listing are two sets of data on the same households that are independently collected. The two measures of the education of the head of household are strongly but not perfectly correlated: the correlation between the two measures is 0.78. The first stages are reported in table S1.2 in the supplementary appendix S1. In the second stage the coefficients change very little compared to specification 1. There is still a strong effect of the education level of members from the same caste but no cross-caste effects. Moreover, instrumenting the education level of the head of household only marginally changes its coefficient estimate.

Specification 3 further checks if the correlation between the education level of members from the same caste and the agricultural productivity of the household is driven by unobservables by controlling for district dummies. The sample only includes districts for which there is information about at least two

9 Age, gender, education level of the head of household, household size, quality of the soil, the share of land that is owned, and the share of irrigated land.

10 Population and area of land.

Table 2. Within Villages: Caste and Cross-Caste Effect

Dependant variable:	**Agricultural production**			
	(1)	**(2)**	**(3)**	**(4)**
Specification 1: Baseline—State dummies				
Head education level	0.0025	0.0030	0.0036	0.0037*
	(0.0023)	(0.0023)	(0.0022)	(0.0022)
Caste mean education level	0.055***	0.061***	0.048***	0.044***
	(0.015)	(0.018)	(0.017)	(0.016)
Other castes mean education level	−0.0048	−0.0042	0.00098	0.0045
	(0.014)	(0.014)	(0.014)	(0.015)
Observations	4405	4405	4405	4405
R^2	0.82	0.82	0.82	0.82
Specification 2: Instrument for head's educ				
Head education level	0.0021	0.0026	0.0036	0.0037
	(0.0030)	(0.0031)	(0.0030)	(0.0030)
Caste mean education level	0.055***	0.061***	0.048***	0.044***
	(0.015)	(0.018)	(0.017)	(0.016)
Other castes mean education level	−0.0047	−0.0041	0.00098	0.0045
	(0.014)	(0.014)	(0.013)	(0.015)
Observations	4405	4405	4405	4405
R^2	0.82	0.82	0.82	0.82
F statistic (KP)	1949	1912	1906	1902
Specification 3: District Dummies				
Head education level	0.0035*	0.0033*	0.0034*	0.0032*
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Caste mean education level	0.029***	0.028**	0.024**	0.020*
	(0.010)	(0.012)	(0.012)	(0.011)
Other castes mean education level	−0.0015	−0.0014	−0.0026	0.0011
	(0.011)	(0.011)	(0.011)	(0.011)
Observations	4309	4309	4309	4309
R^2	0.87	0.87	0.87	0.87
Specification 4: Panel 1982–2006—Household FE				
Head education level	0.00040	0.00095	0.00097	0.00065
	(0.0062)	(0.0062)	(0.0061)	(0.0061)
Caste mean education level	0.045***	0.043***	0.047***	0.028**
	(0.012)	(0.013)	(0.013)	(0.012)
Other castes mean education level	−0.012	−0.014	−0.015	−0.0079
	(0.010)	(0.010)	(0.010)	(0.0098)
Observations	1318	1318	1318	1318
R^2	0.91	0.91	0.92	0.92
Household level controls	Yes	Yes	Yes	Yes
Agricultural controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Control for other group	No	No	Yes	Yes
Additional group controls	No	No	No	Yes

Source: Author's analysis based on the 1982 and 2006 rounds of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. The coefficients on the other controls are reported in table S1.1 in the supplementary appendix S1. The reference group is defined as individuals from the same caste. The other group refers to individuals from other castes. Group controls and controls for other groups include the number of individuals in each group (in log) and the mean value for the demographic characteristics in each group. Additional group controls include the average amount of land in each group and the amount of transfers within groups.

caste groups. The estimated coefficient is smaller, between 2 and 2.9 percent, but the results confirm the previous findings: the higher the education level of the caste, the higher the productivity. Interestingly, the coefficient on the education of the head of household is now significantly different from 0, although not so different in size compared to specification 1. The results also confirm that there is no cross-caste effect.

Specification 4 additionally uses the round of 1982 to estimate the impact with household fixed-effects. The effect is identified from changes in the education level of members from the same caste between 1982 and 2006. The sample only contains households that are observed twice, and households that have split after 1982 are pooled as explained in section 3. In a large number of households, the head of household has changed between the two waves, so the characteristics of the head of household are also controlled for. This makes it possible to estimate the effect of the education of the head of household as in the other specifications. The number of observations that can be used is much smaller, but the results are similar: there is a large effect of the education level of members from the same caste, but no cross-caste effect.

For the other coefficients, reported in table S1.1 in the supplementary appendix S1, the results are as expected. Land is the most important factor of production, followed by labor. Having irrigation also matters a lot, but owning the land that is cultivated instead of renting it does not have an impact on agricultural production. The only demographic characteristic that has an impact on productivity in all the specifications is household size. For a given quantity of labor, having more family members increases productivity. This effect may be due to hired labor being less productive than family members, which is in line with what Rosenzweig and Foster (2010) find in the context of agriculture in India.

Finally, the robustness of the results to the choice of the functional form is tested in the supplementary appendix S1. Table S1.3 estimates the effect using a production frontier approach, where the production and efficiency parameters are estimated simultaneously. The agricultural efficiency of the household and the error term are allowed to depend on the education of members from the same caste. The results are similar. The education of members from the same caste increases productivity by diminishing inefficiencies in the production process.

Within Castes: Same Occupation and Cross-Occupation Effect

The results in table 2 have shown that there is a positive impact of neighbors' education on agricultural productivity. The fact that the education of other castes has no impact on agricultural productivity rules out the possibility that the estimated caste effect is driven by unobserved caste variables *correlated* across castes as explained in section 2. The results from the panel specification with household fixed effects also rules out caste-level unobservables that do not change over time. However, one can still think that the impact is due to time-varying unobserved caste variables *uncorrelated* across castes. Potential caste-level confounders are already controlled for in column 4, but other varying caste-level unobservables that are correlated with caste-level education and productivity may exist and cannot be excluded.

As explained in section 2, a way to check if the measured spillovers come from unobservables correlated with the education of the caste and productivity is to only focus on households from the same caste. Caste members are now divided between those that have agriculture as their main activity and those that do not have agriculture as their main activity. As underlined in the introduction, the channels leading to education spillovers in agriculture are agriculture-specific, so education spillovers are not expected to go from households not having agriculture as their main occupation to households having agriculture as their main occupation. The interpretation of the results is as previously. If there is no cross-occupation impact, then it can be ruled out that the coefficient on neighbors' education reflects a spurious correlation due to unobserved characteristics correlated across occupations within caste.

Table 3 shows the results. The specifications are the same as in table 2. The coefficients on demographic and farm characteristics are reported in column 2 of table S1.1 in the supplementary appendix S1. Again, the results are consistent with an education spillover effect: households from the same caste that have agriculture as their main activity have a strong and significant impact on agricultural productivity

Table 3. Within Caste: Same Occupation and Cross-Occupation Effect

Dependent variable:	**Agricultural production**			
	(1)	**(2)**	**(3)**	**(4)**
Specification 1: Baseline—State dummies				
Head education level	0.0026	0.0032	0.0034	0.0037
	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Same occup caste mean education level	0.047***	0.049***	0.054***	0.051***
	(0.016)	(0.016)	(0.016)	(0.015)
Other occup caste mean education level	0.0018	0.0027	−0.0030	−0.0032
	(0.012)	(0.013)	(0.015)	(0.015)
Observations	4405	4405	4405	4405
R^2	0.82	0.82	0.82	0.82
Specification 2: Instrument for head's educ				
Head education level	0.0022	0.0030	0.0032	0.0033
	(0.0031)	(0.0031)	(0.0032)	(0.0031)
Same occup caste mean education level	0.048***	0.049***	0.054***	0.051***
	(0.016)	(0.016)	(0.016)	(0.015)
Other occup caste mean education level	0.0019	0.0028	−0.0029	−0.0031
	(0.012)	(0.013)	(0.015)	(0.015)
Observations	4405	4405	4405	4405
R^2	0.82	0.82	0.82	0.82
F statistic (KP)	1920	1866	1848	1831
Specification 3: District Dummies				
Head education level	0.0037**	0.0035*	0.0035*	0.0035*
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Same occup caste mean education level	0.020*	0.019*	0.020*	0.019*
	(0.012)	(0.012)	(0.012)	(0.011)
Other occup caste mean education level	0.0055	0.0045	0.0062	0.0043
	(0.0090)	(0.0092)	(0.0099)	(0.0098)
Observations	4309	4309	4309	4309
R^2	0.87	0.87	0.87	0.87
Specification 4: Panel 1982–2006—Household FE				
Head education level	−0.00071	0.00033	0.0027	0.0029
	(0.0073)	(0.0075)	(0.0073)	(0.0072)
Same occup caste mean education level	0.035*	0.029	0.030*	0.033*
	(0.018)	(0.018)	(0.018)	(0.019)
Other occup caste mean education level	−0.0090	−0.0043	−0.0070	−0.015
	(0.0095)	(0.0099)	(0.010)	(0.010)
Observations	862	862	862	830
R^2	0.90	0.90	0.90	0.90
Household level controls	Yes	Yes	Yes	Yes
Agricultural controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Control for other group	No	No	Yes	Yes
Additional group controls	No	No	No	Yes

Source: Author's analysis based on the 1982 and 2006 rounds of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.

Note: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. The coefficients on the other controls are reported in table S1.1 in the supplementary appendix S1. The reference group is defined as individuals from the same caste and the same occupation. The other group refers to individuals from the same caste that do not have agriculture as their main occupation. Group controls and controls for other groups include the number of individuals in each group (in log) and the mean value for the demographic characteristics in each group. Additional group controls include the average amount of land in each group and the amount of transfers within groups.

in the four specifications. On the contrary, the variable that measures the education level of neighbors from the same caste but who do not have agriculture as their main activity has no impact: the coefficient estimate is very close to zero and not significant. The coefficient on caste mean education level in column 2 in the panel specification is not significantly different from zero, which may be due to a lack of power given that its size is in line with the coefficients in the other columns. The results are also robust to the use of a frontier approach as shown in table S1.3 in the supplementary appendix S1.

The absence of cross-occupation impact shows that the previously measured impact of neighbors' education is not driven by unobservables that are correlated within castes across occupations.

Within Caste: Same Jati and Cross-Jati Effect

This section now looks at the impact of members from the same caste and the same subcaste (or jati) and of members from the same caste but not from the same subcaste. In section 2, the definition of households' group of reference was driven by the idea that social proximity is required for learning to happen and the results in the previous sections have confirmed that other caste groups in the village do not matter. This section is interested in understanding if, in the context of learning spillovers, interactions cross jati boundaries or only happen within the jati.

Table 4 shows the results. Only households that have agriculture as their main activity are considered. As expected, the education of households from the same jati has a positive impact on agricultural productivity. One additional year in the educational level of households from the same jati increases households' productivity by 1.6 to 3.4 percent. More interestingly, households from the same caste but not from the same jati have no impact on household productivity. This result shows that the positive effect of the education of members from the same caste and occupation measured in tables 2 and 3 is driven by members from the same jati. It also underlines that learning interactions, similarly to other types of interactions, do not cut across jatis boundaries.

Who Benefits from Education Spillovers? Heterogeneity of the Impact

On average, there is an education spillover effect. The mean education level of households from the same caste and jati and from the same occupation has a positive impact on agricultural productivity. However, previous sections of this article do not differentiate across crops, even if the extent to which learning can happen for a specific crop may depend on the difficulty of cultivating this crop, as well as the level of technology used for cultivation. This section looks at the impact of neighbors' education separately for two crops, rice (paddy) and wheat.

Rice and wheat are the two main crops cultivated in India. In the sample, 27 percent of the households only cultivate wheat, 31 percent only cultivate rice, and 21 percent cultivate both. Only 21 percent of the farmers cultivate neither wheat nor rice. Additionally, except for the states on the Eastern Ocean side (Andhra Pradesh, Orissa, Tamil Nadu and West Bengal) and Kerala where no wheat is cultivated, most of the states cultivate both crops.[11] The diffusion of HYV during the Green Revolution, which began in the 1970s, drastically increased the productivity of these two crops. However, the adoption of HYV has been slower for rice, because the productivity of the first generation of rice HYV was very dependent on local conditions. Munshi (2004) shows that learning from neighbors was lower for rice HYV than for wheat HYV during the Green Revolution. This difference across crops may explain the differences in findings in the previously quoted literature on education spillovers. Appleton and Balihuta (1996) and Weir and Knight (2007) find an impact of neighbors' education on respectively agricultural productivity in general[12] and cereals productivity. On the contrary, Asadullah and Rahman (2009) look at the impact of

11 Rajasthan is an exception: in the sample only two households cultivate rice. Therefore it is excluded from the rice estimation.

12 The production of different crops is aggregated.

Table 4. Within Caste: Same Jati and Cross-Jati Effect

Dependent variable:	**Agricultural production**			
	(1)	**(2)**	**(3)**	**(4)**
Specification 1: Baseline—State dummies				
Head education level	0.0042*	0.0041*	0.0036	0.0040*
	(0.0026)	(0.0025)	(0.0024)	(0.0024)
Same jati mean education level	0.030***	0.030***	0.033***	0.029***
	(0.011)	(0.011)	(0.011)	(0.011)
Other jatis mean education level	0.0011	0.00068	−0.0041	−0.0046
	(0.0087)	(0.0090)	(0.0090)	(0.0087)
Observations	3333	3333	3333	3333
R^2	0.83	0.83	0.83	0.83
Specification 2: Instrument for head's educ				
Head education level	0.0029	0.0027	0.0025	0.0025
	(0.0040)	(0.0040)	(0.0039)	(0.0039)
Same jati mean education level	0.031**	0.031**	0.034***	0.031***
	(0.012)	(0.012)	(0.012)	(0.012)
Other jatis mean education level	0.0013	0.00082	−0.0040	−0.0044
	(0.0087)	(0.0089)	(0.0089)	(0.0086)
Observations	3333	3333	3333	3333
R^2	0.83	0.83	0.83	0.83
F statistic (KP)	1264	1183	1174	1175
Specification 3: District dummies				
Head education level	0.0040**	0.0036*	0.0034*	0.0034*
	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Same jati mean education level	0.016**	0.016**	0.016**	0.017**
	(0.0069)	(0.0069)	(0.0067)	(0.0067)
Other jatis mean education level	0.0058	0.0054	0.0029	0.0030
	(0.0063)	(0.0062)	(0.0067)	(0.0068)
Observations	3238	3238	3238	3238
R^2	0.88	0.88	0.88	0.88
Household level controls	Yes	Yes	Yes	Yes
Agricultural controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Controls for other group	No	No	Yes	Yes
Additional group controls	No	No	No	Yes

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. The coefficients on the other controls are reported in table S1.1 in the supplementary appendix S1. The reference group is defined as individuals from the same jati and the same occupation. The other group refers to individuals from the same caste that have agriculture as their main occupation, but not from the same jati. Group controls and controls for other groups include the number of individuals in each group (in log) and the mean value for the demographic characteristics in each group. Additional group controls include the average amount of land in each group and the amount of transfers within groups.

education spillovers on rice productivity and do not find any evidence that neighbors' levels of education matter. This absence of education spillovers may be due to a lower learning potential in rice production. Table 5 provides separate estimations for rice and wheat, focusing on households that only cultivate wheat or rice. As before, the dependent variable is the agricultural output of the household, and the coefficients on demographic characteristics and farm characteristics are reported in table S1.1 in the

Table 5. Impact for Wheat and Rice

Dependent variable:	**Agricultural production**					
Groups:	**Caste**		**Caste & Occup**		**Jati & Occup**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Crop: WHEAT						
Head education level	0.0013	0.00076	0.00061	0.000036	−0.00098	−0.0014
	(0.0042)	(0.0040)	(0.0043)	(0.0040)	(0.0051)	(0.0051)
Group mean education level	0.040**	0.048**	0.041*	0.025	0.030**	0.032**
	(0.019)	(0.021)	(0.021)	(0.019)	(0.013)	(0.014)
Other group mean education level	−0.0010	−0.012	0.0025	0.026**	0.0013	−0.0010
	(0.014)	(0.014)	(0.010)	(0.012)	(0.011)	(0.011)
Observations	1165	1165	1165	1165	872	872
R^2	0.845	0.852	0.846	0.854	0.845	0.846
Crop: RICE						
Head education level	0.0079***	0.0085***	0.0085***	0.0090***	0.0094***	0.010***
	(0.0029)	(0.0028)	(0.0030)	(0.0029)	(0.0032)	(0.0032)
Group mean education level	0.0037	0.0070	−0.00054	0.000025	−0.0068	−0.0052
	(0.011)	(0.011)	(0.013)	(0.015)	(0.0093)	(0.0088)
Other group mean education level	0.013	0.013	0.0028	0.0018	−0.0044	−0.00096
	(0.0095)	(0.011)	(0.013)	(0.015)	(0.0098)	(0.0080)
Observations	1362	1362	1362	1362	1110	1110
R^2	0.849	0.851	0.848	0.850	0.841	0.845
HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Agr. controls	Yes	Yes	Yes	Yes	Yes	Yes
Group controls	No	Yes	No	Yes	No	Yes
Controls for other group	No	Yes	No	Yes	No	Yes

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.
Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. The coefficients on the household controls (demographic and farm characteristics) are reported in table S1.1 in the supplementary appendix S1. Group controls and controls for other groups include the number of individuals in each group (in log) and the mean value for the demographic characteristics in each group.

supplementary appendix S1. The specification used is the baseline specification with state dummies, and the table reports the results with and without controls at the group and other-group levels. In columns 1 and 2 the level of aggregation is the caste level, in columns 3 and 4 it is the caste × occupation level and in columns 5 and 6 it is the jati × occupation level.

The impact of neighbors' education is very different for wheat and rice. For wheat, the impact is large and similar to what was previously measured, and there is no impact of the education of the head of household. On the contrary, for rice there is no impact of the education level of members from the same group, but there is a strong and significant impact of the education of the head of household. These results are in line with Munshi (2004), who finds that there is learning from neighbors for wheat, but that rice growers experiment more on their own land to compensate for the lack of learning. This heterogeneity of impact across crops therefore provides additional evidence that the positive impact of neighbors' education is likely to be a spillover effect.[13]

13 For the sample of households that cultivate both rice and wheat, there is a positive effect of members from the same group on both crops. This may be due to the fact that households that cultivate both crops are different from households that cultivate only rice or wheat or that households that cultivate rice in addition to wheat benefit from spillovers from wheat cultivation.

5. Discussion

Section 4 provides consistent evidence that the mean education level of households from the same jati and occupation is positively correlated with agricultural productivity at the household level, while there is no cross-caste, no cross-jati, and no cross-occupation effect within castes. These results are robust to different specifications and in particular to the addition of an earlier round of data that makes it possible to estimate the effect using household fixed effects.

Could group-level unobserved factors be driving the observed positive correlation? The set of specifications and results from section 4 makes it possible to rule out a large number of potential unobserved factors. In particular, as shown in section 2, the absence of cross-caste and cross-occupation effects within castes rules out the possibility that the effect is driven by unobserved factors that are correlated across castes or within castes across occupations. Unobserved factors such as the supply of education (for example, the presence of a school in the village, the distance to the closest city, or schooling quality) that are arguably correlated across castes within villages or within castes across occupations, can therefore not explain the observed correlation. The results from the household fixed-effects estimation also rule out group-level unobservables that are fixed over time, including inherited characteristics, such as the quantity and quality of land, inherited ability, or inherited health.

The correlation is also robust to the inclusion of group-level variables that are potential uncorrelated within caste across occupations, such as group wealth proxied by the average amount of land owned or the intensity of insurance networks (column 4 of tables 2, 3, and 4). Moreover, the results on rice and wheat show that these unobservable factors should only be correlated with group-level education and agricultural productivity for wheat growers but not for rice growers to be driving the effect. However, the set of results cannot fully rule out caste- or jati-level unobservables that are uncorrelated across caste and, jatis and across occupations within caste, and that are not fixed over time.

While it is hard to come up with variables that would have these characteristics, the possibility that they exist cannot be excluded. However, it is possible to have an idea of how likely they are to fully explain the observed correlation by calculating how correlated with the mean education level of the group these unobservables should be to explain away the results. To do so, the methodology suggested in Oster (2017) is used. The intuition of the method is that one can derive information about unobserved confounders by exploiting information about the relation of observed covariates to the "treatment" variable, in this case the education of members from the same group, along with information about how much of the variance of the outcome variable is explained by the observed covariates. The underlying assumption is that the relationship of observed covariates to the treatment variable does contain information about the relationship of unobservables to the treatment. More precisely, the assumption is that the covariance of observed covariates with the treatment (scaled by the variance of the observed covariates) is proportional to the covariance of unobserved covariates with the treatment (scaled by the variance of the unobserved covariates). Consequently, the set of variables included in the list of covariates that are assumed to provide information about the covariance between unobserved covariates and the treatment is a key ingredient of the calculation.

Table 6 reports the "coefficient of proportionality," that is, the relative degree of selection on unobservables to selection on observables, for which the true effect of the mean education level of the group would be 0, calculated with two alternative sets of variables.In the first set of results (first four rows) the assumption is that all the covariates provide information about the unobserved factors. In other words, the covariance between all the observed covariates and the education of members from the same group is assumed to be related to the covariance between the unobserved covariates and the education of members from the same group. In the second set of results (last four rows) the education variables (of the head of

Table 6. Coefficient of Proportionality

Group	(1)	(2)	(3)	(4)
No excluded variable				
Caste	1.34	1.27	0.92	0.79
Caste & occup	0.84	0.83	0.81	0.71
Jati & occup	0.82	0.79	0.89	0.74
Caste & occup, WHEAT	7.76	7.47	2.27	2.88
Excluded var: Head educ and Other group mean educ				
Caste	2.05	1.93	1.35	1.14
Caste & occup	3.07	3.00	2.90	2.39
Jati & occup	1.87	1.80	2.04	1.65
Caste & occup, WHEAT	17.93	16.64	3.34	4.36
Household controls	Yes	Yes	Yes	Yes
Agr. controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Controls for other group	No	No	Yes	Yes
Add. controls	No	No	No	Yes

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.

Note: This table reports coefficients of proportionality calculated following the methodology in Oster (2017). Coefficients of proportionality represent the relative degree of selection on unobservables to selection on observables that would imply a true effect of the education of the group of 0. The first set of coefficients has been estimated using all the control variables in the set of observables; the second set of coefficients exludes the education of the head of household and the mean education level of the other group.

household and of the other group) are excluded, as it is unlikely that unobserved factors would be as correlated with the education of members from the same group as these variables, so they do not contain useful information about the relationship of unobservables to the treatment. In both sets of results the maximum possible R^2 is assumed to be 1, which renders the results conservative. What the table shows is that the relationship between unobserved factors and education of the group needs to be quite high—between 0.71 and 22.71 times the relation between observed factors and education of the group—to fully drive the correlation between education of the group and productivity. Most of the coefficients are higher than 1, which, as suggested by Oster (2017), can be interpreted as an indication that the true effect is nonzero. All the coefficients that are lower than 1 are estimated using the education variables in the set of controls, which, as argued earlier is very conservative as the unobserved characteristics are unlikely to be as correlated with the education of the group as these two variables. Moreover, the coefficients have been computed under the assumption that the variance in the productivity can be fully explained if the unobservables could be controlled for ($R^2 = 1$), which is an upper bound, as productivity may be measured with error or some of the variance in productivity may be idiosyncratic. If this assumption is relaxed, all the coefficients are higher than 1 for a R^2 higher or equal to 0.94. Finally, a large set of unobservables has been ruled out as potential confounders such that the pool of potential unobservables is quite small. Together, these findings corroborate that the positive effect of neigbors education is indeed likely to be a spillover effect.

6. Conclusion

This paper takes advantage of the social structure in India to study spillovers of education in farm productivity. The results show that the mean education level of members from the same caste is positively related to agricultural productivity. On the contrary, there is no cross-caste effect. The results also show that within castes, only households that have agriculture as their main activity have a positive impact on the agricultural productivity of their neighbors. This pattern of results rules out the possibility that the measured impact is a spurious correlation due to unobservable characteristics that have some correlations across groups. Additionally, this paper shows that spillovers are only present for wheat but not for rice, which is consistent with the fact that learning from neighbors is limited when growing conditions strongly depend on land characteristics.

These findings confirm that education externalities do not only exist in urban contexts and education spillovers do not only occur between workers of the manufacturing and service sectors. But there are also spillovers in more traditional sectors such as agriculture. Therefore, improving education in developing countries should continue to be a priority, because education has a multiplicative effect, even in rural areas. Moreover, this paper underlines that education has to be publicly financed, because social returns to education are higher than private returns. The boom of private schools in rural India cannot be a long-term solution.

Maybe even more important, the fact that the external effect of education is happening solely within social groups calls for an equality of access to education across social groups. In India, the law imposing that 25 percent of classroom seats should be reserved for children from poorer or disadvantaged families in the neighborhood is a step in that direction. More generally, affirmative action programs in favor of low castes should continue to be developed.

These findings also have implications for other developing countries, where school enrollment is low and inequality across ethnic groups is also very salient.

References

Acemoglu, D., and J. Angrist 1999. "How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws." NBER Working Paper No. 7444, National Bureau of Economic Research, Cambridge, MA.

Appleton, S., and A. Balihuta. 1996. "Education and Agricultural Productivity: Evidence from Uganda." *Journal of International Development* 8 (3): 415–44.

Asadullah, M. N., and S. Rahman. 2009. "Farm Productivity and Efficiency in Rural Bangladesh: The Role of Education Revisited." *Applied Economics* 41 (1): 17–33.

Annual Status of Education Report. 2005. Published by ASER Center, New-Delhi.

Annual Status of Education Report. 2014. Published by ASER Center, New-Delhi.

Brock, W. A., and S. N. Durlauf. 2001. "Interactions-Based Models." In *Handbook of Econometrics*, vol. 5, edited by J. J. Heckman and E. E. Leamer, 3297–380. Amsterdam: Elsevier.

Case, A. 1992. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics* 22 (3): 491–508.

Census of India. 2011, http://censusindia.gov.in/2011-Common/CensusData2011.html.

Conley, T. G., and C. R. Udry. 2010. "Learning About a New Technology: Pineapple in Ghana." *American Economic Review* 100 (1): 35–69.

Damodaran, H. 2008. *India's New Capitalists: Caste, Business, and Industry in a Modern Mation*. London: Palgrave Macmillan.

Deliège, R. 2004. *Les Castes en Inde Aujourd'hui*. Paris: Presses Universitaires de France.

Foster, A. D., and M. R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103 (6): 1176–209.

Gille, V. 2012. "Education Spillovers: Empirical Evidence in Rural India." *Indian Growth and Development Review* 5 (1): 4–24.

Jackson, M. O. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Kijima, Y. 2006. "Caste and Tribe Inequality: Evidence from India, 1983–1999." *Economic Development and Cultural Change* 54 (2): 369–404.

Kumbhakar, S. C., and C. A. K. Lovell. 2003. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.

Moretti, E. 2004a. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*, vol. 4, edited by V. Henderson and J. F. Thisse, 224–91. Amsterdam: Elsevier.

——. 2004b. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94 (3): 656–90.

Munshi, K. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics* 73 (1): 185–213.

Munshi, K. 2011. "Strength in Numbers: Networks as a Solution to Occupational Traps." *Review of Economic Studies* 78 (3): 1069–101.

Munshi, K., and J. Myaux. 2006. "Social Norms and the Fertility Transition." *Journal of Development Economics* 80 (1): 1–38.

Munshi, K., J. Myaux, and M. R. Rosenzweig. 2006. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review* 96 (4): 1225–52.

——. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap." *American Economic Review* 106 (1): 46–98.

Oster, E. 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 1–18.

Rosenzweig, M. R., and A. D. Foster. 2010. "Is There Surplus Labor in Rural India?" Discussion Paper No. 991, Economic Growth Center, Yale University, New Haven, CT.

Srinivas, M. N. 1962. *Caste in Modern India and Other Essays*. Bombay, Calcutta, New Delhi: Asia Publishing House.

Weir, S., and J. Knight. 2007. "Production Externalities of Education: Evidence from Rural Ethiopia." *Journal of African Economies* 16 (1): 134–65.

Wooldridge, J. M. 2003. *Introductory Econometrics: A Modern Approach*, 2nd ed. Sydney, Australia: South-Western College Publication.

——. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

World Bank. 2010, https://data.worldbank.org/.

Supplementary Appendix

Education Spillovers in Farm Productivity: Revisiting the Evidence

Véronique Gille

Supplementary Appendix S1: Additional Tables

Table S1.1. Impact of Other Controls, Tables 2–5

Dependant variable:	Agricultural production				
	Caste	Caste & occup	Jati & occup	WHEAT	RICE
Demographic characteristics					
Household size	0.011***	0.012***	0.0079*	0.0056	0.014***
	(0.0039)	(0.0039)	(0.0045)	(0.0047)	(0.0045)
Age	0.00058	0.00066	0.0015*	0.00044	0.0019**
	(0.00072)	(0.00073)	(0.00082)	(0.0015)	(0.00096)
Scheduled Caste	0.12*	0.095	0.034	−0.11*	−0.065
	(0.067)	(0.062)	(0.071)	(0.064)	(0.060)
Scheduled Tribe	0.24**	0.21**	0.16	0.12	−0.0092
	(0.092)	(0.095)	(0.11)	(0.21)	(0.050)
Other Backward Classes	0.13**	0.10**	0.12**	0.025	0.041
	(0.050)	(0.050)	(0.048)	(0.052)	(0.043)
Farm characteristics					
Cultivated area (in log)	0.75***	0.75***	0.75***	0.78***	0.79***
	(0.041)	(0.042)	(0.047)	(0.038)	(0.032)
Days worked (in log)	0.25***	0.24***	0.22***	0.17***	0.15***
	(0.036)	(0.036)	(0.036)	(0.034)	(0.032)
Mechanized assets	0.17***	0.16***	0.15***	0.018	0.022
	(0.031)	(0.032)	(0.034)	(0.034)	(0.037)
Nonmechanized assets	−0.071	−0.063	−0.076	0.11	−0.13**
	(0.064)	(0.064)	(0.074)	(0.16)	(0.054)
Irrigated land	0.24***	0.24***	0.26***	0.15**	0.15***
	(0.066)	(0.065)	(0.076)	(0.068)	(0.042)
Soil characteristics					
Soil depth: up to 1 feet	−0.10	−0.093	−0.12*	−0.33***	0.013
	(0.069)	(0.068)	(0.070)	(0.084)	(0.067)
Yellow soil	−0.15**	−0.15*	−0.10	−0.11	0.053
	(0.075)	(0.074)	(0.067)	(0.082)	(0.082)
Brownish black soil	−0.31***	−0.30***	−0.25***	−0.20**	0.12
	(0.081)	(0.076)	(0.066)	(0.079)	(0.080)
Off-white soil	−0.18**	−0.17**	−0.17**	−0.13	0.058
	(0.072)	(0.070)	(0.067)	(0.087)	(0.082)
Loam soil	−0.13	−0.14	−0.0082	0.17**	0.00086
	(0.15)	(0.15)	(0.15)	(0.071)	(0.098)
Light clay soil	−0.026	−0.038	0.10	0.23***	−0.074
	(0.15)	(0.15)	(0.15)	(0.073)	(0.10)
Easy drainage soil	0.16	0.17	0.15	0.23***	0.050
	(0.12)	(0.12)	(0.13)	(0.085)	(0.087)
Moderate drainage soil	0.13	0.13	0.13	0.21**	0.080
	(0.13)	(0.13)	(0.13)	(0.081)	(0.082)

Table S1.1. Continued.

Dependant variable:	Agricultural production				
	Caste	Caste & occup	Jati & occup	WHEAT	RICE
Village characteristics					
Village population (in log)	0.084^{*}	0.090^{**}	0.093^{**}	0.11^{**}	0.11^{***}
	(0.044)	(0.044)	(0.044)	(0.051)	(0.036)
Village land	-0.000040^{*}	-0.000039^{*}	-0.000044^{*}	-0.000081^{*}	-0.000070
	(0.000022)	(0.000022)	(0.000026)	(0.000042)	(0.000056)
Observations	4405	4405	3333	1165	1362
R^2	0.819	0.819	0.829	0.846	0.848

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.
Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. This table provides the coefficients on demographic and farm characteristics estimated in tables 2 to 5, baseline specification. I do not report the gender of the head of household and the share of land owned, as they are not significant. For soil characteristics, I only report the variables that have a significant impact on productivity at the 5 percent level.

Table S1.2. First Stages—Specification 2, Tables 2, 3, and 4

Dependant variable:	Educ (survey)			
Group: Caste				
Educ (listing)	0.72^{***}	0.72^{***}	0.72^{***}	0.72^{***}
	(0.016)	(0.016)	(0.016)	(0.017)
Group: Caste & Occup				
Educ (listing)	0.72^{***}	0.72^{***}	0.72^{***}	0.72^{***}
	(0.017)	(0.017)	(0.017)	(0.017)
Group: Jati & Occup				
Educ (listing)	0.71^{***}	0.71^{***}	0.71^{***}	0.71^{***}
	(0.020)	(0.021)	(0.021)	(0.021)
Household level controls	Yes	Yes	Yes	Yes
Agricultural controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Controls for other group	No	No	Yes	Yes
Additional group controls	No	No	No	Yes

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.
Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. This table provides the first stages of specification 2 in tables 2, 3, and 4.

Table S1.3. Production Frontier

Dependant variable:	Agricultural production			
	(1)	(2)	(3)	(4)
Group: Caste				
Usigma				
Head education level	−0.019	−0.027	−0.026	−0.031
	(0.015)	(0.020)	(0.020)	(0.022)
Caste mean education level	-0.47^{***}	-0.77^{***}	-0.68^{***}	-0.67^{**}
	(0.14)	(0.21)	(0.19)	(0.30)
Other castes mean education level	0.020	0.047	0.044	0.065
	(0.083)	(0.096)	(0.092)	(0.10)
Vsigma				
Head education level	−0.0072	−0.010	−0.011	−0.010
	(0.011)	(0.0098)	(0.0099)	(0.0099)
Caste mean education level	0.15^{***}	0.13^{**}	0.14^{**}	0.17^{**}
	(0.058)	(0.061)	(0.067)	(0.073)
Other castes mean education level	0.036	0.00079	0.0085	−0.036
	(0.057)	(0.053)	(0.055)	(0.064)
Observations	4405	4405	4405	4405
Log pseudolikelihood	−3211	−3102	−3067	−3038
Group: Caste & Occup				
Usigma				
Head education level	−0.0154	−0.0192	−0.0246	0.0147
	(−1.00)	(−1.35)	(−1.19)	(0.55)
Same occup caste mean education level	-0.343^{***}	-0.393^{***}	-0.486^{**}	-2.124^{**}
	(−3.39)	(−3.63)	(−2.91)	(−3.17)
Other occup caste mean education level	−0.102	−0.0985	−0.144	−0.402
	(−0.98)	(−0.84)	(−1.31)	(−1.71)
Vsigma				
Head education level	−0.00690	−0.0108	−0.00745	−0.0126
	(−0.64)	(−1.09)	(−0.81)	(−1.57)
Same occup caste mean education level	0.148^{**}	0.159^{**}	0.132^{**}	0.108^{*}
	(2.93)	(3.00)	(2.76)	(2.51)
Other occup caste mean education level	0.00887	−0.00778	−0.0271	−0.0343
	(0.20)	(−0.16)	(−0.50)	(−0.70)
Observations	4405	4405	4405	4405
Log pseudolikelihood	−3210	−3164	−3106	−3046
Household level controls	Yes	Yes	Yes	Yes
Agricultural controls	Yes	Yes	Yes	Yes
Group controls	No	Yes	Yes	Yes
Controls for other group	No	No	Yes	Yes
Additional group controls	No	No	No	Yes

Source: Author's analysis based on the 2006 round of the Additional Rural Incomes Survey and Rural Economic & Demographic Survey (ARIS-REDS) from the National Council of Applied Economic Research.

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the village level are reported in parentheses. This table reports the effect of the education of neighbors on agricultural productivity using a frontier approach. Usigma is the inefficiency term, and Vsigma is the error term, which is also expressed as a function of neighbors' education. The controls are the same as in the baseline specification.