Affimation Action, Education and Gender: Evidence from India

G. Cassan

CRED – January 2014

Department of Economics
Working Papers
2014/01

UNIVERSITÉ DE NAMUR

Affirmation Action, Education and Gender: Evidence from India*

Guilhem Cassan†

January 6, 2014

Abstract: *We use a unique natural experiment in order to assess the impact of positive discrimination in India on targeted groups' educational attainment. We take advantage of the harmonization of the Scheduled Castes lists within the Indian states taking place in 1976 to measure the increase of the educational attainment of the new beneficiaries. We show that this policy had heterogenous effects across genders, with males benefiting from the SC status and females remaining essentially unaffected. We show that this translated into a differential increase in literacy and numeracy, and propose a novel method to measure the latter.*

JEL Classification: I24; O15; H41
Keywords: scheduled caste; quota; positive discrimination; gender.

*I am grateful to Denis Cogneau, Christelle Dumas, Hemanshu Kumar, Eliana La Ferrara, Sylvie Lambert, Andreas Madestam, Annamaria Milazzo and Rohini Somanathan as well as seminar participants at AMID Summer School, Bocconi University, Louvain La Neuve, Bristol, Graduate Institute Geneva, Paris I, CERDI, EUDN Conference and Paris School of Economics for useful comments. Marion Leturcq and Lore Vandewalle comments on an earlier draft vastly improved this paper. I am indebted to Julien Grenet for his help in the data collection, and to Ashwini Natraj, with whom I started this project. This paper is produced as part of the project "Actors, Markets, and Institutions in Developing Countries: A micro-empirical approach" (AMID), a Marie Curie Initial Training Network (ITN) funded by the European Commission. Funding from the CEPREMAP is gratefully acknowledged. All remaining errors are mine. All remaining errors *in this paper*, obviously.

†University of Namur, CRED. Email: guilhem.cassan@unamur.be.

Introduction

This paper proposes the first evidence allowing to infer the causal role of affirmative action policies on educational attainment in India, using a nation wide natural experiment. It shows that if those policies did indeed lead to an average increase in education, the women were excluded from those benefits.

Several countries have put in place affirmative action programs targeting disadvantaged minorities. Many universities in the United States provide quotas for Black students, and so do certain universities in Brazil. Among those programs, the largest one has been established in India. It consists in a set of various policies targeting low castes ("Scheduled Castes" or SC), in three main domains: politics, public employment, and education. As is the case in other countries with affirmative action policies such as the US, with an increasing competition in the access to education, the affirmative action policies in education are heavily contested in India. However, to the best of our knowledge, the impact of the affirmative action programs on the educational attainment of the SC has not been rigorously studied (Chalam (1990), Chitnis (1972), Galanter (1984), Kumar (1992)).

Only the first two domains of the Indian affirmative action policies have been rigorously evaluated: Pande (2003) studied the impact of reservations of seats in legislative assemblies for low castes. Howard and Prakash (2008) and Prakash (2009) chose to focus on the effect of reservation of public employment. On education however, most of the research has focused on the equity aspect of positive discrimination, with the case study of Bertrand et al. (2010) providing a thorough analysis of the question. As a consequence, we have little knowledge of the extent to which the SC as a whole benefited from the affirmative action programs in terms of educational attainment, as well as of the differential effect of those policies by gender.

This paper takes advantage of a nationwide natural experiment on the beneficiaries of caste-based positive discrimination in order to measure its impact on educational attainment. Indeed, the list of the castes considered as SC were drawn by each Indian state at the Independence, with variations in the list across and even within states. In 1956, the borders of the Indian states have been redrawn, while the lists of SC remained unchanged. This created a situation where, within a State, the "Scheduled Caste" status of the members of a same caste could vary across areas. For administrative reason,

this situation lasted until 1976, when the list of castes considered as SC were finally harmonized within each state, allowing 2.4 million individuals (Government of India, ed, 1978) to have access to the benefits of the SC status[1]. This unique historical event allows us to compare the fate of the individuals who had access to the SC status in 1976 to those that were considered SC since the Independence within a same state and a same caste. A difference in difference approach comparing the cohorts too old to benefit from the policy to those young enough to be exposed to this effect, allows us to identify the effect of the SC status on educational attainment from various other policies targeting poor households at the time.

It will be seen that overall, access to the SC status leads to an increase of 0.3 years of education, very unevenly distributed across genders, with males capturing essentially all the increase in the educational attainment and females remaining unaffected. This paper thus provides a new angle to the affirmative action debate, by assessing a differential impact by gender. This angle is rarely taken when discussing the effect of affirmative action policies[2]. Moreover, the dismal quality of the Indian education system is well documented since at least the PROBE report of 1999 (Dreze and Kingdon, 2001) and many times since Duflo et al. (2010). Hence, we turn to measures of skills to evaluate if the increase in schooling years led to an increase in human capital. We show that the access to the SC status led to an increase of both literacy and numeracy. To evaluate the latter, we propose a measure inspired by the approach of Mokyr (1983)[3] and widely used by economic historians, but which, to our knowledge, has not yet been applied at the household level. We build a household level measure of age heaping in the declared ages of household members as a measure of the respondent's lack of numeracy. While this index is generally used to assess the quality of age data in surveys, economic historians have often used it as a measure of numeracy. We propose to use it as a proxy for the numeracy of the respondent. We find evidence that this gender differentiated increase in educational attainment also translated in an increase in hard skills as measured by literacy and numeracy.

This paper is organized in the most traditional manner. You are approaching the end of the introduction, and we will now turn to a first section in which we will examine the context and the natural experiment exploited in this paper. We will then describe

[1] "The removal of area restrictions in respect of SC and ST will no doubt enable the members of these communities, who were deprived so far of the benefits and concessions given by the Central and State Governments to get their due share of educational, economic and political safeguards."(Government of India, ed, 1978)

[2] Even if gender bias in educational attainment has of course been widely studied (Kingdon (2005), Kingdon (2007).

[3] I thank Gani Aldashev for having mentioned this work to me.

the data as well as our empirical strategy, which will open the way to the presentation of the results, in the following section. i-In order to enhance our confidence in the results, various robustness checks are proposed (we will test the parallel trends assumption, run placebo regressions and provide a check for selective migration) in the subsequent section. And we will finally conclude. The interested reader is also invited to have a look at the appendices, in which additional robustness checks are provided, and a more detailed presentation of the use of the Whipple index in this context is also given, as we believe that this use of the Whipple index is also an important contribution of this paper.

1 Context

1.1 Affirmative action in India

Even if the first affirmative action policies were implemented under the British rule, it is not before the Independence that a systematic positive discrimination policy was implemented across India. Reservations concerned 3 main items: legislative seats, education and public employment. Affirmative action in education consists in various policies. The most famous one is quotas in higher education institutions, but secondary schooling was also made free, while each state also has various policies (specific scholarships, schools and hostels, free mid day meals, etc). All in all, the various schemes targeting the education of the SC are the main expenditures in the budget for the Welfare Scheme for SC of the Indian Plan: they represented 33% of its amount in 1951-1956 and 55% in 1974-79. Positive discrimination might thus affect schooling through various channels. First of all, by reducing the cost of education, it favors longer studies in the cost-benefit arbitrage of the household. Also, the quotas in higher education will allow, among those that had made the choice to pursue their studies up until this level, to effectively have access to it. Finally, the quotas in public employment also are an incitation to pursue longer studies, as they increase the probability of employment in the formal sector, and thus, increase the returns to education. Hence, this paper does not evaluate the effect of affirmative action *in* the educational sector but the effect of the affirmative action policies *on* educational attainment.

1.2 The definition of the Scheduled Castes

There is no precise definition of the criteria making a caste eligible to the SC status, leaving the door open to some arbitrariness in the lists. Indeed, the constitution of the

list of castes considered as SC has been and still is the subject of debates. As a result, it has been modified on several occasion. This section, drawing from the work of Galanter (1984) provides a short history of the list of Scheduled Castes.

One of the main problems with the making of such classification is that the definition of "untouchability", the criteria to be considered a SC, is not straightforward: as "untouchability" varies in its meaning across the sub continent, it is hard to create a definition that would apply to the whole country. Indeed, whiole untouchable castes are relatively well identified in the South and West of India, it is not the case in the other parts of the country. Hence, the Constitution of 1950 avoids to define a clear concept and only provides a procedure of designation[4] that each State is to follow. This allowed for the possibility of inconsistencies across States[5] and even within States, as certain States decided to give the SC status to certain castes only in certain areas. Despite those inconsistencies, the lists were revised only three times since the Independence, but with only one revision being of real importance[6]. With an increase of 2.4 million SC over an original population of 80 million SC, the Scheduled Castes and Scheduled Tribes (Amendment) Act of 1976 was the most dramatic change in the list of SC in India.

1.3 The Scheduled Castes and Scheduled Tribes (Amendment) Act of 1976

In 1956, India reorganized the borders of its States along linguistic lines[7]. But, as the borders of the States were redefined (see Figure 1), the State-wise SC lists remained unchanged. This led to a situation of large discrepancies between State borders and the lists of SC. The latter ones not being defined at the State level anymore, but by area within each State, areas corresponding to the pre-1956 borders. Hence, from 1956, the number of castes considered as SC in one part of a State and non SC in an other part of the same State vastly increased.

[Figure 1 about here.]

The reason for the list not to be adjusted to the new borders is the slowness of the administration: *"It has been mentioned in the last report that the President has issued*

[4] "*castes, races or tribes or parts of or groups within castes, races and tribes which shall for purposes of this Constitution be deemed to be Scheduled Castes in relation to that State.*"

[5] Bayly (1999) gives the example of the Khatik caste, considered as SC in Punjab, but classed as a "forward" caste in Uttar Pradesh, a neighboring State at the time of the establishment of the lists..

[6] The change of 1956 mainly affected Rajasthan and Uttar Pradesh, and also allowed all Sikh untouchable castes to claim SC status, while the change of 1990 allowed the Buddhists to have access to the SC status in all the States, while only Maharashtra had recognized them before.

[7] And in 1960, the states of Maharashtra and Gujarat were created from the former state of Bombay.

the SC and ST Lists (Modification) Order, 1956, specifying the SC and ST in the re-organized States. As these lists had to be issued urgently for the re-organized States, it was not possible to prepare comprehensive and consolidated lists and therefore, the SC and ST had to be specified in these list territory-wise within each re-organized State" (Government of India, ed, 1958). But not only did the administration fail to change the lists on time, it failed to do so for a period of twenty years. The yearly reports of the Commissioner on SC and ST are particularly telling in this aspect, as many of its yearly occurrences refer to the fact that *"[...]the question of preparation of comprehensive lists of SC and ST for the reorganized States [...] remained pending [...]"* (Government of India, ed, 1960). It is only under the emergency rule of Indhira Gandhi that the SC lists were harmonized within states with the SC and ST (Amendment) Act of 1976. Hence, due to administrative reasons, a situation was created in which individuals from the same caste could be considered as SC in one part of a State but not in an other. Only 20 years after were they finally being granted the SC status.

As the pre-1956 borders were not drawn according to the linguistic and cultural areas of India, certain castes were split in two across two States, facing different policies in terms of their SC status. With the reorganization of the State borders, they were facing different SC status within a single State. This situation thus creates a natural experiment setting in which an identical caste faced a different treatment with respect to positive discrimination due to an historical accident. In 1976, the Area Restriction Removal Act removed almost all intra-State restrictions[8]. This removal of restrictions led to an increase of 2.4 million of the SC population (3% of the 1971 SC total population), mainly in the States affected by the reorganization of borders of 1956. Table 1 lists the States in which the SC lists were modified in 1976, along with the corresponding increase in their SC population. Among those States, only Bihar, Himachal Pradesh and Uttar Pradesh were not affected by the reorganization of the borders in 1956, but removed restrictions that had been present since the Independence.

[Table 1 about here.]

[8]According the Galanter (1984) their number dropped from 1,126 to 64.

Hence, this paper focuses on 98.4% of the SC who obtained the SC status in 1976[9]. This unique historical event allows us to compare the fate of individuals who had access to the SC status since the Independence to those whose SC status was delayed until 1976. We thus use plausible exogenous variation in the SC status across individuals, allowing us to assess the causal impact of the SC status on educational attainment.

2 Data and Empirical Strategy

2.1 Data and descriptive statistics

The Indian Demographic and Health Survey of 1998-99 is to our knowledge the only dataset offering both the precise caste name (the "jati") of respondents, their district of residence and a sufficient sample size to perform the analysis. Using the 1971 and 1981 Census lists of Scheduled Castes and the district of residence of households, we are able to identify the households that were granted the SC status in 1976. As each caste can have various synonyms, which can vary locally, the coding of the treatment status has used the synonyms provided in the SC lists as well as in the project People of India (Singh, ed, 1996)[10]. Table 2 provide the summary statistics for the variables used throughout the paper.

[Table 2 about here.]

2.2 Identification Strategy

The SC being a very different population from the general population, much poorer in particular. Thus comparing the new SC to the general population does not allow to identify the specific effect of the access to the SC status on educational attainment from

[9]The analysis excludes Himachal Pradesh, as when this Union Territory was granted the State status in 1966, large portions of the state of Punjab were also transferred to it: hence, between 1956 and 1966, the population living in the contemporary borders of Himachal Pradesh were exposed to different State policies, preventing us to identify the sole effect of the access to the SC status. Results are unaffected by the inclusion of Himachal Pradesh.

[10]Thus, we use the jati name to identify if an individual is a SC or not, and not the answer to the question "are you a SC" that is present in the questionnaire. All regressions control for the declaration of the SC status, but do not remove individuals that do not declare to be a SC, as this is likely endogenous (Gille (2012), Moffit (1983)). Only 18% of the sample does not declare to be a SC. Online Appendix shows that removing those does not change the results, and increases the precision of the estimates. An additional concern is the tendency not to declare the precise "jati" name. Indeed several SC prefer to answer a generic name such as "SC", "Dalit" or "Harijan" instead of their exact "jati" name. This can be a concern if this tendency is different between treatment and control. Online Appendix 7 shows that this is unlikely to lead to biased estimates.

other social policies, as various pro-poor policies were launched at that period.
Hence, the only credible counterfactual are the SC already in the lists in 1976. The identification strategy of this paper will thus rely on comparing cohorts too old to benefit from the access to the SC status in 1977 (year of implementation of the 1976 change) from the point of view of their educational attainment to those that were young enough, for individuals that had access to the SC status since the Independence versus individuals that were granted the access in 1977 only.

The difference in the timing of access to the SC status suggests that if the policy had an impact, the evolution of the educational status of the two groups should diverge for the cohorts at school age between 1950 and 1976, when their treatment status differed and converge for the cohorts at school age after 1976, when both groups are treated. Figure 2 shows that indeed, starting from the cohorts born in the mid 40's there has been a divergence between the two SC groups, and this divergence begins to fade out for the cohorts born in the mid 60's. However, Figure 3, also shows that this evolution is in fact driven by males, while no clear pattern seems to be emerging for females. While the divergence taking place for the cohorts born in the mid 40's is illustrative, using the cohorts born before 1950 is problematic on two accounts. First, the cohorts born before 1950 are aged 48 and above at the time of the survey, meaning that selective mortality might be significant in those cohorts. Second, the cohorts born before 1950 were older than 6 in 1956, when the States borders have been reorganized, and thus did not face the same institutional environment than the future cohorts while at school age. For this reason, the reminder of the paper will focus on cohorts born from 1950 onwards[11].

[Figure 2 about here.]

[Figure 3 about here.]

To see if the general patterns seen in Figures 2 and 3 survive the inclusion of controls, one would ideally want to follow a non parametric identification strategy such as the one used in Duflo (2001), allowing the effect of the access to the SC status to vary for each cohort. Due to sample size constraints, this is not possible here, as each cohort-treated group cell contains on average only 50 individuals, making it impossible to precisely estimate each coefficient, even less so if the the treatment effect is to be differentiated by gender. Figure 4 nonetheless presents the coefficients of such a specification:

[11] And old enough at the time of survey to have completed the level of education considered: 14 and above for primary school and 18 and above for secondary schooling.

$$Edu_{idt} = constant + \sum_{k=1950}^{1980} \delta_k NewSC_{id} * cohort_{ik} + \sum_{k=1950}^{1980} \gamma_k cohort_{ik} + \lambda X_{idt} + \epsilon_{idt} \quad (1)$$

Where Edu_{idt} is the number of years of education up to secondary schooling completion of individual i residing in district d born in the two year cohort t, $NewSC_{id}$ is a dummie taking a value 1 if individual's i caste was added on the SC list for district d in 1976, $cohort_{ik}$ is a dummie that indicates whether individual i is born in the two years cohort k and X_{idt} is a set of controls variables including district FE and trend, jati FE, religion FE, female head of household FE, urban FE, size of the household at time of the survey and a dummie for declaration to be SC.
As can be seen in Figure 4, there is a clear jump in the coefficients for the males born from the cohorts 1967-8 (aged 10 or below in 1977): while New SC males born before 1967 had on average 0.5 years of schooling less than other SC, this difference disappears starting with the 1967 cohort, i.e the last cohort still at primary school age in 1977. For females however, no clear pattern seems to be emerging, with new SC females having on average 0.5 years of schooling above their counterpart throughout the period[12].

[Figure 4 about here.]

Hence, throughout the paper, we will consider as treated the cohorts aged 10 years old or below in 1977, still at primary school age in the year of implementation of the New SC lists. We will thus run regressions of the type:

$$Edu_{idt} = constant + \beta NewSC_{id} + \delta NewSC_{id} * post1967_i + \gamma post1967_i + \lambda X_{idt} + \epsilon_{idt} \quad (2)$$

Where Edu_{id} is a measure of educational attainment of individual i born in year t and residing in area of restriction d, $NewSC_{id}$ a dummy indicating whether individual i residing in district d is member of a caste added to the SC list in 1976, $post1967_i$ a dummie taking value 1 if individual i is born in year 1967 or after and X_{id} a set of control variables.

[12]The confidence intervals are not represented as none of the coefficients are significant at standard levels but the coefficient for the cohorts 1955-6 and 1975-6 for males

3 Results

3.1 Years of Education

Table 3 presents the results of the estimation of Equation 1. We can see that the access to the SC status led to an increase in the number of years of schooling of 0.3 to 0.4 years. The coefficient of interest on New SC*post 1967 remains stable throughout specifications: district trend in column 2, within Jati estimation in column 3, within household in column 4 and within household and state-cohort FE, in column 5. The latter, the most demanding one, is our preferred specification, as it allows to control for both all unobserved fixed characteristics of household and flexibly controls for state level changes, in addition to district specific linear changes.

[Table 3 about here.]

However, the picture changes dramatically once the coefficient is allowed to vary across gender, as can be seen in Table 4. Indeed, it becomes clear that only males seem to benefit from their inclusion in the lists of SC. With an increase of their educational level of 0.7 to 0.8 years of schooling, they capture all the increase in schooling that had been measured in Table 3, while women remain essentially unaffected.

[Table 4 about here.]

3.2 Educational level

Focusing on the number of years of education might hide various effects of the access to the SC status on the different levels of education that could also be potentially different across genders. Table 5 presents the results of our preferred specification on educational levels: schooling, primary completion, some secondary schooling and secondary completion[13]. It can be seen that the educational attainment of males has increased throughout all the steps of education but secondary school completion, and this increase is particularly strong - and precisely estimated - for primary completion and some secondary schooling. Moreover, the absence of impact of the SC status on the level of education of women remains true throughout the spectrum of education levels.

[Table 5 about here.]

[13]The very similar - but slightly more significant - results for all the specifications can be found in Online Appendix 2.

Hence, it is now clear that while the access to the SC status does indeed lead to an increase in the educational attainment of the SC population, this increase is in fact completely captured by males, females being completely excluded from it.

3.3 Skill acquisition

However, as has been discussed in the introduction, increasing the number of years spent at school might not lead to an increase in human capital if the time spent at school is completely wasted. Hence, we will now turn to the evaluation of the impact of the access to the SC status on two skills: literacy and numeracy.

3.3.1 Literacy

Table 6 presents the result of various specifications on literacy outcomes. It can be seen that there is indeed a male specific increase in literacy for new SC born after 1967, but the coefficient loses its significance when turning to more demanding specifications.

[Table 6 about here.]

3.3.2 Numeracy

However, there is no measure of numeracy in the DHS data. Hence, to measure to numeracy skills of the respondent of the household (i.e., not of all household members), we will turn to a proxy based on the tendency to declare numbers as multiple of 5. This intuition has been formalized in the Whipple index, which measures the tendency to declare ages which are multiple of 5. It is the share of declared aged between 23 and 62 that are a multiple of 5, multiplied by 500[14]. This index, along with similar measures, has been widely used by economic historians to measure the evolution of numeracy. Bachi (1951),Myers (1954) and Mokyr (1983) were pioneers in the use of this method, which is still regularly used (A'Hearn et al., 2009).

We propose here to exploit this measure at the household level to proxy for the numeracy of the respondent, the persons that declares the age of all the other members of the household. To our knowledge, this use of the Whipple index has never been done before. This measure has the major advantage to be based on revealed numeracy instead of declared numeracy. However, it comes at the expense of being available only at the household level and thus to be relevant only for the respondent of the household.

[14]Hence, it varies between 0 and 500, where 0 means that no one declares an age multiple of 5 and 500, the opposite. If there is no tendency of age heaping, the Whipple index is of 100 (20% of the population should have an age that is a multiple of 5 is there is no misdeclaration).

Appendix 2 describes the age heaping in the data, and how our measure correlates with measures of schooling.

To compute our measure of numeracy, we calculate the Whipple index of the ages declared by the respondent of each household. Those declared ages can be multiple of 5 for three reasons. First, one fifth of the household members obviously should have an age which is a multiple of 5. Second, the respondent lacks numeracy skills and rounds the age to a multiple of 5. Third, the respondent is numerate, but does not know the exact age of the household members (because they themselves are not numerate, or do not know their date of birth), and tends to round them. The Whipple index of the household is thus a function of the size of the household and of the numeracy of the respondent, of the numeracy of other household members and of their parents (as the knowledge of the year of birth depends on them).

As we are interested in the numeracy skill of the respondent, we want to control for the first and third possibilities. The first point in particular could be an issue for small households, as the Whipple index will tend to take more extreme values. To attenuate this measurement bias, we will control for the number of individuals aged 23 to 62 in the household. For the third point, we will control for the number of household members aged 23 to 62 who were born when the respondent was 15 and above (i.e when she was old enough to remember the year of birth of the respondent). In addition, as we are willing to assess the effect of the access to the SC status on the respondent's numeracy, we will control for the number of other household members aged 23 to 62 born after 1966, as they also could have seen their numeracy affected by their access to the SC status, and thus indirectly affect the ages declared by the respondent. Additionally, we control for the average years of education of the household members aged 23 to 62 (except of the respondent).

Finally, the identity of the respondent is heavily selected , separating the coefficient between males and females would not be interpretable, and we will only present average effects on respondents. We will consider as numerate respondent whose Whipple index is below 175[15] as this is the threshold used by the United Nations (2000) to declare that the data is "very badly" biased towards multiples of 5.

Table 7 presents the result of various specifications on numeracy outcomes. As there is only one observation per household, the within household estimation can obviously not be implemented here. It can be seen that there is indeed a positive effect on access

[15]A Whipple Index of 175 means that 35% of the population is declared to have an age which is a multiple of 5, i.e 15 percentage points above what the distribution of ages is expected to be absent age heaping.

to the SC status on the numeracy skills of the respondent.

[Table 7 about here.]

4 Robustness Checks

4.1 Parallel trends

As our estimates are difference in difference, they rely on a common trends assumption. While we cannot test this hypothesis, we can see if the trends were indeed parallels for the cohorts born between 1950 and 1966, i.e before the "New SC" were treated. Table 8 regresses the number of years of education on differential trends by status. It can be seen that prior treatment, the trends were very similar across SC and "New SC", as well as across gender, comforting our estimation strategy.

[Table 8 about here.]

4.2 Placebo regressions

To show that the effect estimated can indeed be attributed to the access to the SC status, we will now turn to placebo regressions, to show that it is only from the cohorts born in 1967 that the increase in educational attaintment of males "New SC" can be found. While this could already be seen in Figure 4, we propose here another method. We run 15 regressions varying the treatment year from 1958 to 1972 and restrict the sample to individuals born 8 years before or after this treatment year. Due to the large reduction in sample that this leads to, we will use the within Jati - district trend specification. Those regressions are run separately for males and females, and the coefficients on the interaction between "New SC" and the cohort of treatment are pictured in Figure 5. For females, as expected the coefficient remains relatively stable and close to zero for all regressions. For males, the coefficient is roughly 0 when the treatment cohort is between 1958 and 1964 and starts increasing with the cohort 1965 to 1967, before it decreases again. The largest coefficient is when 1967 is assumed to be the cohort of treatment. The picture is in line with the effect of the SC status having an effect only on the cohorts born from 1967 on. Indeed, we expect the coefficient to be at its maximum when the treatment year is the "real" treatment year, as it is only in that specification that both control and treatment groups are "clean".

Hence, this clearly shows that it is only from the cohort 1967 that a differential effect in terms of educational attainment takes place, showing that it is only the access to the SC status that explains the evolution previously estimated.

[Figure 5 about here.]

4.3 Selective migration

One last concern with our estimation is selective migration. Indeed, we attribute the "New SC" status based on the district of residence of respondents. However, it is possible that before 1976, households that would be considered SC in a part of a state would have migrated in order to benefit from the SC status. However, migration is relatively low in India (Munshi and Rosenzweig, 2009) and particularly so in the 1970's and before. In addition, if such migration was to be important, then the household choosing to migrate would probably be the ones that would have benefited the most from the access to the SC status had they not migrated before 1976. That means that, if anything, this selective migration is likely to bias our estimates downwards.

To confirm that selective migration is not an issue, one would need migration information, and in particular the year of arrival and the place of origin of the household. This information is unfortunately not available in the DHS. This survey only provides information on whether ever-married women aged 15-49 in the household "have been living continuously in their current place of residence", when they arrived in that place, and if they were coming from a rural or from an urban area. Given the Indian marriage pattern[16], under that definition of migration, more than 80% of the women have migrated, and the question can not be exploited as such. We propose to exploit the rural/urban dimension of the data as a check, relying on the intuition that households that would migrate in response to the geographic restrictions on the SC status would probably have a migration pattern different from the classic rural to rural or urban to urban migration. Hence, Table 9 presents the results of our preferred specification on the different levels of education, reducing the sample to household in which the wife of the household head declares not having migrated, having migrated from a rural to a rural area or from an urban to an urban area. We can see that the results remain qualitatively identical to our main results, suggesting that selective migration does not affect our estimates.

[Table 9 about here.]

[16] In which women join their husband's household after marriage.

5 Conclusion

This paper studies the impact of the positive discrimination policy in education conducted by the Indian Government since the Independence using a natural experiment. It shows that the impact of reservations in educational attainment is mixed. Indeed, while the males young enough to benefit from the SC status indeed see their schooling level increase, along with their actual literacy and numeracy skills, no such effect can be detected among females. Hence, while the affirmation action policies do indeed seem to have been successful in increasing the average education level of the SC population, they have not been able to reach its most backwards sub-population: the women. While the "creamy layer" debate has captured most of the attention in the affirmative action debate in India, this paper suggests that the gender dimension also deserves a much larger attention.

References

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen, "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital," *Journal of Economic History*, 2009, *69* (4).

Bachi, R, "The Tendency to Round Off Age Returns: Measurement and Correction.," *Bulletin of the International Statistical Institute*, 1951, *33*.

Bayly, Susan, *Caste, Society and Politics in India*, Cambridge University Press, 1999.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan, "Affirmative Action in Education: Evidence from College Admissions in India," *Journal of Public Economics*, 2010, *94* (1-2), 16–29.

Chalam, K.S., "Caste Reservations and Equality of Opportunity in Education," *Economic and Political Weekly*, 1990, *25* (41).

Chitnis, Suma, "Education for Equality: Case of Scheduled Castes in Higher Education," *Economic and Political Weekly*, 1972, *7* (31-33).

Dreze, Jean and Geeta Kingdon, "School Participation in Rural India," *Review of Development Economics*, 2001, *5* (1).

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *4* (91), 795–813.

_ , Rema Hanna, and Stephen Ryan, "Incentives Work: Getting Teachers to Come to School," *Working Paper*, 2010.

Galanter, Marc, *Competing Equalities: Law and the Backward Classes in India*, University of California Press, 1984.

Gille, Véronique, "Stigma in positive discrimination application? Evidence from Quotas in Education in South India," *Working Paper*, 2012.

Government of India, ed., *Report of the Commissioner for Scheduled Castes and Scheduled Tribes, 1957-1958.*, Manager Govt. Of India Press, 1958.

_ , ed., *Report of the Commissioner for Scheduled Castes and Scheduled Tribes, 1958-1959.*, Manager Govt. Of India Press, 1960.

_ **, ed.**, *Report of the Commissioner for Scheduled Castes and Scheduled Tribes, 1975-1977.*, Manager Govt. Of India Press, 1978.

Howard, Larry and Nishit Prakash, "Does Employment Quota Explain Occupational Choice Among Disadvantaged Groups? A Natural Experiment from India," *Working Paper*, 2008.

Kingdon, Geetah, "Where Has All the Bias Gone? Detecting Gender Bias in the Household Allocation of Educational Expenditure in India," *Economic Development and Cultural Change*, 2005, *53* (2).

_ , "The Progress of School Education in India," *Oxford Review of Economic Policy*, 2007, *23* (2).

Kumar, Dharma, "The Affirmative Action Debate in India," *Asian Survey*, 1992, *32* (3), 290–302.

Moffit, Robert, "An Economic Model of Welfare Stigma," *American Economic Review*, 1983, *73* (5).

Mokyr, Joel, *Why Ireland Starved*, George Allen & Unwin, 1983.

Munshi, Kaivan and Mark Rosenzweig, "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth," *Working Paper*, 2009.

Myers, R, "Accuracy of Age Reporting in the 1950 United States Census," *Journal of the American Statistical Association*, 1954.

Pande, Rohini, "Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence ? Theory and Evidence from India.," *American Economic Review*, 2003, *93*, 1132–1151.

Prakash, Nishit, "The Impact of Employment Quotas on the Economic Lives of Disadvantaged Minorities in India," *Working Paper*, 2009.

Singh, K.S., ed., *People of India: Communities, Segments, Synonyms, Surnames and Titles*, Anthropological Survey of India, 1996.

United Nations, "Demographic Yearbook - Special Issue," http://unstats.un.org/UNSD/Demographic/products/dyb/dybcens.htm 2000.

Figure 1: Variation in States' borders between 1951 and 1961.

Indian States in 1951.

Indian States in 1961.

Figure 2: Evolution of educational attainment by treatment status. Local Mean Smoothing

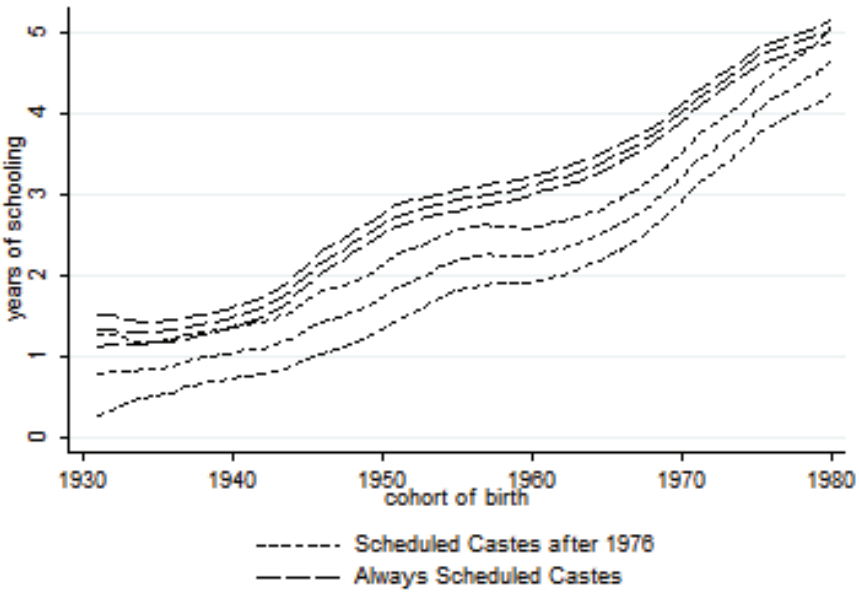
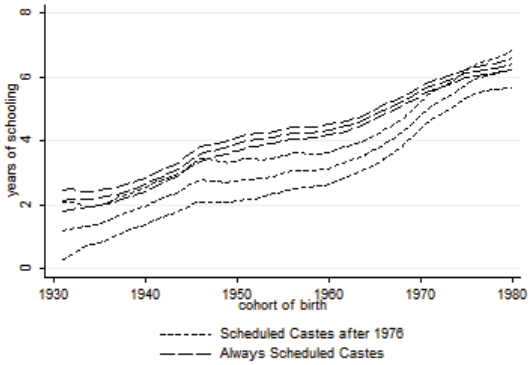


Figure 3: Evolution of educational attainment by treatment status and gender. Local Mean Smoothing

Males



Females

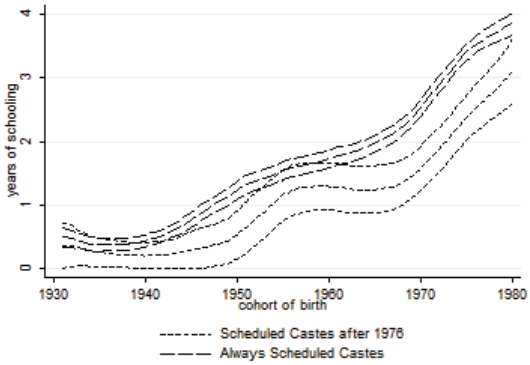
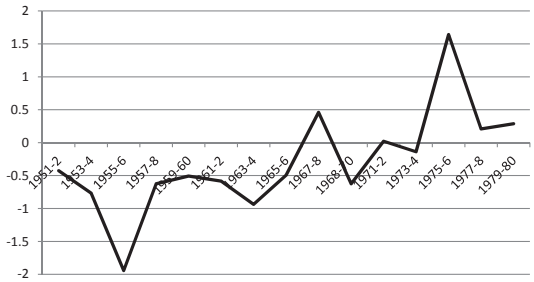


Figure 4: Coefficients of the interaction of two-years cohort dummies and New SC status,

Males



Females

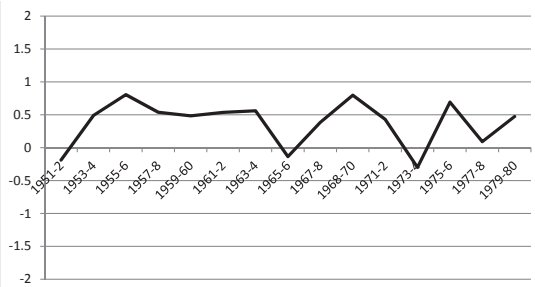


Figure 5: Placebo regressions, by gender

Males

Females

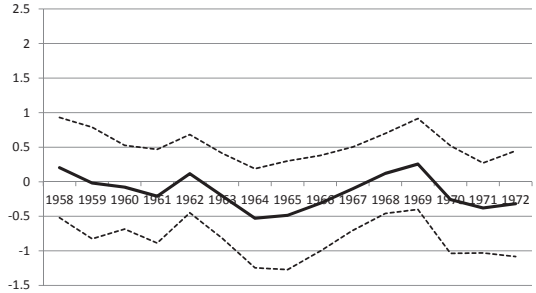


Table 1: 1976 Increase in SC population, by State.

	Original SC population	Revised SC population	Difference	Share in total increase
Andhra Pradesh	57.75	58.16	0.41	*1.7%*
Bihar	*79.51*	*83.86*	*4.35*	*17.6%*
Gujarat	18.26	18.9	0.64	*2.6%*
Karnataka	*38.5*	*42.77*	*4.27*	*17.2%*
Kerala	*17.72*	*20.02*	*2.3*	*9.3%*
Madhya Pradesh	*54.54*	*57.52*	*2.98*	*12.0%*
Maharashtra	*30.26*	*31.77*	*1.51*	*6.1%*
Rajasthan	*40.76*	*42.16*	*1.4*	*5.7%*
Uttar Pradesh	*185.49*	*190.95*	*5.46*	*22.1%*
Tamil Nadu	73.16	73.38	0.22	*0.9%*
West Bengal	88.16	89	0.84	*3.4%*
Himachal Pradesh	7.7	8.08	0.38	*1.5%*
Total	**691.81**	**716.57**	**24.76**	**100%**

Source: Report of the Commissionner on SC and ST, 1975-77

Table 2: Descriptive Statistics.

	All		Males				Female			
			Always SC		New SC		Always SC		New SC	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Education variables										
Years of Primary Education	2.66	2.39	3.43	2.19	3.09	2.30	1.94	2.34	1.46	2.19
Schooling	0.58	0.49	0.74	0.44	0.68	0.47	0.43	0.50	0.34	0.47
Primary Completion	0.47	0.50	0.62	0.49	0.54	0.50	0.34	0.47	0.25	0.43
Years of Schooling up to Secondary	3.94	4.14	5.35	4.06	4.60	4.13	2.66	3.76	1.93	3.42
Incomplete Secondary	0.38	0.48	0.52	0.50	0.45	0.50	0.24	0.43	0.18	0.39
Secondary Completion	0.18	0.38	0.26	0.44	0.20	0.40	0.11	0.31	0.09	0.28
Born before 1967										
Years of Primary Education	2.01	2.34	2.84	2.34	2.04	2.31	1.19	2.03	0.87	1.80
Schooling	0.46	0.50	0.63	0.48	0.49	0.50	0.28	0.45	0.21	0.41
Primary Completion	0.34	0.47	0.49	0.50	0.33	0.47	0.19	0.39	0.13	0.34
Years of Schooling up to Secondary	3.02	3.84	4.37	4.04	3.08	3.85	1.68	3.09	1.18	2.63
Incomplete Secondary	0.27	0.44	0.40	0.49	0.28	0.45	0.14	0.35	0.09	0.29
Secondary Completion	0.12	0.33	0.19	0.39	0.13	0.34	0.06	0.23	0.03	0.17
Born after 1966										
Years of Primary Education	2.95	2.36	3.72	2.06	3.56	2.14	2.25	2.39	1.72	2.30
Schooling	0.64	0.48	0.80	0.40	0.76	0.42	0.50	0.50	0.39	0.49
Primary Completion	0.53	0.50	0.68	0.47	0.64	0.48	0.40	0.49	0.30	0.46
Years of Schooling up to Secondary	4.49	4.21	5.99	3.95	5.54	4.01	3.19	3.98	2.35	3.72
Incomplete Secondary	0.44	0.50	0.60	0.49	0.55	0.50	0.30	0.46	0.23	0.42
Secondary Completion	0.21	0.41	0.30	0.46	0.24	0.43	0.14	0.34	0.12	0.32
Individual and household variables										
New SC	0.09	0.29								
Female	0.50	0.50								
Household size	6.69	3.43	6.67	3.47	6.86	3.39	6.69	3.39	6.77	3.45
Urban	0.29	0.45	0.29	0.46	0.27	0.44	0.28	0.45	0.25	0.43
Woman head of household	0.08	0.27	0.07	0.26	0.05	0.21	0.10	0.30	0.07	0.26
Hindu	0.90	0.30	0.89	0.31	0.98	0.15	0.89	0.31	0.97	0.17
Muslim	0.01	0.10	0.01	0.11	0.01	0.07	0.01	0.10	0.01	0.08
Christian	0.04	0.19	0.04	0.19	0.02	0.13	0.04	0.20	0.02	0.15
Sikh	0.00	0.05	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.00
Buddhist/Neo Buddhist	0.05	0.22	0.05	0.22	0.00	0.00	0.06	0.23	0.00	0.00
Zorastrian/Parsi	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00
No Religion	0.00	0.01	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.03
N (aged 14+ at time of survey)	20,576		9,415		973		9,307		881	
N (aged 18+ at time of survey)	17,207		7,808		797		7,848		754	

Table 3: Effect of access to the SC status on years of schooling. OLS regressions.

	(1)	(2)	(3)	(4)	(5)
New SC*post 1967	0.230	0.357**	0.428**	0.311	0.389*
	[0.150]	[0.163]	[0.172]	[0.201]	[0.210]
New SC	-0.108	-0.211	-0.406		
	[0.192]	[0.200]	[0.281]		
District FE and Trend	N	Y	Y	Y	Y
Within Jati	N	N	Y	N	N
Within Household	N	N	N	Y	Y
State-Cohort FE	N	N	N	N	Y
r2	0.38	0.40	0.32	0.35	0.37
N	17201	17201	17195	16328	16328
Number of groups			220	6076	6076

Standard errors two way clustered at the jati and district level in brackets. All regressions include districts FE, birth cohort FE, state trends, gender FE, religion of household head FE, urban FE, household size, declaration to be SC FE and woman head of household FE. Population aged 18 and above at time of the survey and born after 1950.

Table 4: Effect of access to the SC status on years of schooling, by gender. OLS regressions.

	(1)	(2)	(3)	(4)	(5)
New SC*post 1967	0.735***	0.834***	0.900***	0.713***	0.782***
	[0.211]	[0.217]	[0.227]	[0.238]	[0.246]
New SC * Female * post 1967	-1.047***	-1.009***	-0.999***	-0.875**	-0.845**
	[0.292]	[0.286]	[0.297]	[0.353]	[0.338]
New SC * Female	0.634**	0.610**	0.577**	0.515*	0.453
	[0.274]	[0.274]	[0.277]	[0.284]	[0.290]
Female * post 1967	-0.038	-0.044	-0.006	0.088	0.129
	[0.166]	[0.167]	[0.170]	[0.228]	[0.219]
New SC	-0.403	-0.490*	-0.675**		
	[0.271]	[0.282]	[0.330]		
Female	-2.780***	-2.783***	-2.801***	-2.932***	-2.963***
	[0.135]	[0.140]	[0.136]	[0.140]	[0.140]
District FE and Trend	N	Y	Y	Y	Y
Within Jati	N	N	Y	N	N
Within Household	N	N	N	Y	Y
State-Cohort FE	N	N	N	N	Y
r2	0.38	0.40	0.32	0.35	0.37
N	17201	17201	17195	16328	16328
Number of groups			220	6076	6076

Standard errors two way clustered at the jati and district level in brackets. All regressions include districts FE, birth cohort FE, State-Cohort FE, religion of household head FE, urban FE, household size, declaration to be SC FE and woman head of household FE. Population aged 18 and above at time of the survey and born after 1950.

Table 5: Educational attainment, by gender. OLS regressions.

	Schooling	Primary	Some Secondary	Secondary
	(1)	(2)	(3)	(4)
New SC*post 1967	0.050	0.091**	0.104***	0.013
	[0.038]	[0.037]	[0.035]	[0.037]
New SC * Female * post 1967	-0.093*	-0.108***	-0.074*	0.006
	[0.052]	[0.040]	[0.040]	[0.041]
New SC * Female	0.043	0.073**	0.043	0.013
	[0.047]	[0.034]	[0.033]	[0.032]
Female * post 1967	0.061***	0.030	-0.014	0.009
	[0.021]	[0.023]	[0.033]	[0.019]
Female	-0.370***	-0.324***	-0.290***	-0.166***
	[0.022]	[0.018]	[0.017]	[0.014]
r2	0.33	0.30	0.29	0.17
N	19986	19986	16336	16336
Number of groups	6433	6433	6079	6079

Standard errors two way clustered at the jati and district level in brackets. All regressions are within household, and include district trends and state-cohort FE. Population aged 14+ (columns 1 to 3) or 18+ (columns 4 and 5) at time of the survey and born after 1950.

Table 6: Literacy skills, by gender. OLS regressions.

	(1)	(2)	(3)	(4)	(5)
New SC*post 1967	0.078**	0.072	0.083**	0.058	0.056
	[0.035]	[0.437]	[0.038]	[0.040]	[0.042]
New SC * Female * post 1967	-0.134***	-0.127*	-0.119**	-0.099*	-0.094*
	[0.048]	[0.074]	[0.047]	[0.058]	[0.057]
New SC * Female	0.065	0.057	0.046	0.049	0.044
	[0.046]	[0.047]	[0.044]	[0.050]	[0.048]
Female * post 1967	0.056***	0.055*	0.059***	0.055**	0.060***
	[0.018]	[0.029]	[0.019]	[0.022]	[0.021]
New SC	-0.031	-0.027	-0.051		
	[0.037]	[0.359]	[0.044]		
Female	-0.358***	-0.359***	-0.360***	-0.367***	-0.372***
	[0.018]	[0.039]	[0.018]	[0.019]	[0.020]
District FE and Trend	N	Y	Y	Y	Y
Within Jati	N	N	Y	N	N
Within Household	N	N	N	Y	Y
State-Cohort FE	N	N	N	N	Y
r2	0.33	0.35	0.28	0.31	0.33
N	20573	20573	20567	19981	19981
Number of groups			220	6431	6431

Standard errors two way clustered at the jati and district level in brackets.All regressions include districts FE, birth cohort FE, state trends, religion of household head FE, urban FE, household size, declaration to be SC FE and woman head of household FE. Population aged 14 and above at time of the survey and born after 1950.

Table 7: Numeracy skills. OLS regressions.

	(1)	(2)	(3)	(4)
New SC*post 1967	0.066**	0.078**	0.091**	0.113***
	[0.031]	[0.039]	[0.042]	[0.039]
New SC	-0.012	-0.015	0.044	0.032
	[0.033]	[0.035]	[0.050]	[0.048]
District FE and Trend	N	Y	Y	Y
Within Jati	N	N	Y	Y
State-Cohort FE	N	N	N	Y
r2	0.38	0.42	0.32	0.40
N	4989	4989	4933	4933
Number of groups			151	151

Standard errors two way clustered at the jati and district level in brackets.All regressions include districts FE, birth cohort FE, state trends, religion of household head FE, urban FE, household size dummies, number of household members aged 23 to 62 (and among those, the number who were born when the respondent was 15 or above, and the number of those born from 1967), average education years of non respondent household members aged 23 to 62, declaration to be SC FE and woman head of household FE. Population aged 18 and above at time of the survey and born after 1950.

Table 8: Test of parallel trends. OLS regressions.

	(1)	(2)
New SC * trend	0.005	-0.008
	[0.039]	[0.056]
New Sc * Female * Trend		-0.000
		[0.056]
Female * trend		-0.011
		[0.013]
District FE and Trend	Y	Y
Within Jati	Y	Y
r2	0.33	0.33
N	6371	6371
Number of groups	172	172

Standard errors two way clustered at the jati and district level in brackets. All regressions include birth cohort FE, gender FE, religion of household head FE, urban FE, household size, declaration to be SC FE and woman head of household FE. Population born between 1950 and 1966.

Table 9: Migration robustness check. OLS regressions.

	Numeracy	Literacy	Schooling	Primary	Some Secondary	Secondary	Years of Education
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New SC*post 1967	0.107	0.091*	0.108**	0.100*	0.171***	-0.033	1.147**
	[0.069]	[0.054]	[0.049]	[0.060]	[0.058]	[0.056]	[0.490]
New SC * Female * post 1967		-0.007	-0.029	-0.078	-0.033	0.044	-0.475
		[0.093]	[0.078]	[0.076]	[0.070]	[0.048]	[0.527]
New SC * Female		-0.007	0.002	0.066	0.026	0.022	0.313
		[0.071]	[0.068]	[0.058]	[0.060]	[0.036]	[0.460]
Female * post 1967		0.014	0.013	0.027	-0.021	-0.004	-0.168
		[0.033]	[0.033]	[0.037]	[0.038]	[0.030]	[0.276]
Female	-0.065**	-0.334***	-0.334***	-0.311***	-0.279***	-0.161***	-2.739***
	[0.031]	[0.025]	[0.025]	[0.021]	[0.023]	[0.021]	[0.153]
New SC	-0.008						
	[0.058]						
r2	0.51	0.40	0.40	0.37	0.38	0.29	0.45
N	2779	8599	8601	8601	6997	6997	6993
Number of groups	119	3204	3205	3205	3051	3051	3049

Standard errors two way clustered at the jati and district level in brackets. All regressions are within household (except column 1, within jati) and include district trends, and state-cohort FE. Population aged 14+ (columns 2 to 4) or 18+ (columns 1 and 4 to 6) at time of the survey and born after 1950. Column 1 additionnaly controls for number of household members aged 23 to 62 dummies (and among those, the number who were born when the respondent was 15 or above, and the number of those born from 1967), average education years of non respondent household members aged 23 to 62.