Towards Conceptualization of a Household Educational Indicator: Incorporating Externalities

Upasak Das¹ · Udaya S. Mishra²

Accepted: 25 February 2020 / Published online: 25 March 2020
© The Author(s) 2020

Abstract

Though education is essentially an individual attribute, its positive externality goes beyond the individual to household conditioned by its gendered and generational features. Further, distributional aspects of education levels among the household members are also important in terms of its effects on its welfare outcomes. Hence, education of the head or highest educational level within the household may not adequately represent the household educational levels. In this context, we construct a household educational index that combines three aspects: distribution of education among household members, highest education level among the male and female members and difference in highest education level attained between generations. The generalised final index is found to satisfy all the axioms or the intuitive properties: monotonicity, anonymity, normalisation and uniformity. Additionally, we examine the values of the proposed indicators against the commonly used indicators including education of the household head and the highest education within the household using a representative survey from India. The empirical results indicate considerable difference between the proposed indicator and these common indicators. We also observe a strong association between our proposed indicator and welfare indicators like learning outcomes among children and having a toilet in the household. This implies usage of common indicators may underestimate the significance of education in terms of its bearing on welfare outcomes. These findings lay emphasis on the significance of conceptualizing a household educational indicator that accommodates the complexities of welfare externalities of individual educational levels within the household.

Keywords Education · Household index · Externality · Proximity · Gender · India

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11205-020-02301-2>) contains supplementary material, which is available to authorized users.

✉ Upasak Das
upasak.das@gmail.com; upasak.das@manchester.ac.uk
Udaya S. Mishra
mishra@cds.ac.in

¹ Global Development Institute, University of Manchester, Arthur Lewis Building, Oxford Road, Manchester, UK

² Centre for Development Studies, Medical College, Prasanth Nagar, P.O. Ulloor, Thiruvananthapuram, Kerala, India

1 Introduction

In this era of sustainable development goals, monitoring of development indicators have become more frequent in comparison with the past.¹ Hence conceptualization and formulation of these indicators is necessary. Among the wide range of development indicators conceived within the scheme of sustainable development goals, a majority of them are population level indicators derived as an aggregate of individual outcomes. For example, considering those set on quality education, one of the targets is ensuring “all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes” by 2030. The indicator it monitors is “Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex”.² Notably this is an individual level indicator aggregated at the population level as discussed. Though such indicators may be efficient in monitoring changes towards a set target, yet its implications on overall development remain complex.

Education is essentially an individual outcome and its immediate well-being implication is for the individual. But the broader literature on returns to education refers to market and non-market returns to education (Schultz 1988). The contribution of education to market measured output is the human capital that forms a part of production function in the endogenous growth model (Lucas 1988). More specifically, it indicates that investment in human capital, skills and knowledge are the main contributors of economic growth. The labour market returns to education have been well documented (Becker 1962; Krueger and Ashenfelter 1992; Card 1999). The non-market contribution of education is viewed in the household production function which may yield private non-market benefits. These non-market benefits can vary from health to choices and participation in criminal activities (Wolfe and Haveman 1984). These might as well be the associated private and social benefits that others in the society derive because of one’s own education, which are known as educational externalities (McMahon 2004). Notably, these benefits are enjoyed by co-resident household members and translate into overall household welfare outcomes. Given such complexities of returns to education and its linkage with the development outcomes, there arises a need to verify educational indicators conceived at varying levels ranging from individuals to that of all household members. The commonly used indicators for household educational level namely education of the household head and the highest educational level among members in the household may not address such complexities.

This paper intends to address these externalities and complexities emanating from age and gender and attempts to develop a household education indicator, going beyond the above mentioned indicators. More specifically, it takes into account three aspects: education of all household members, education level of the female and male members and education level of the younger and older generation. It combines these aspects and then proposes a household educational index that can be generalized across different societies. We test some of the common axioms and make empirical illustration using data obtained from a nationally representative survey from India to show the differences between the proposed indicator and the commonly used indicators. The empirical illustration involves examining the bearing of these indicators on two welfare outcomes: education and sanitation. The

¹ The list of sustainable goals can be obtained from <https://www.undp.org/content/undp/en/home/sustainable-development-goals.html> (last accessed on March 20, 2019).

² This can be accessed from <https://sustainabledevelopment.un.org/sdg4> (accessed on December 30, 2019).

findings suggest expectedly stronger association of the proposed indicator with these welfare outcomes in comparison with the two commonly used indicators.

The structure of the paper is as follows. Section 2 discusses the need of a household educational indicator. Section 3 presents the process of formulation of the index and then discusses its validation in terms of the axioms. Section 4 is an empirical exploration based on a national representative data (discussed elaborately) from India which compares the proposed index with the commonly used indicators. Section 5 presents the regression estimates of the association of children learning outcomes and household access to toilets separately with the household educational indicators after discussing the regression strategy and the variables. Section 6 concludes with a discussion.

2 The Need of a Household Educational Indicator

Role of education in enhancing development outcomes at the macro as well as micro level are most common in literature. Educational level of individuals with more decision-making power has been shown to influence parameters such as risk of poverty, nutrition and health. Cutler and Lleras-Muney (2006), for example found “An additional 4 years of education lowers 5-year mortality by 1.8 percentage points; it also reduces the risk of heart disease by 2.16 percentage points, and the risk of diabetes by 1.3 percentage points”. Further, it has been observed that mother’s schooling plays an important role in improving household health and nutrition status. (Behrman and Wolfe 1987a, b, 1989). Also a positive impact of parental education on infant and child mortality has also been established (Caldwell and McDonald 1982). Hence recognition of education in defining human and social development makes it a crucial dimension in the comprehension of human development. Higher level of maternal education is also associated with higher bargaining power, which in turn leads to better educational outcomes for the children (Behrman et al. 1999; Schultz 2002; Chudgar 2009, 2011). Hence, educational profile of a household is vital to display its positive externality on household well-being.

When determining the household level indicators of education, one of the most preferred attributes becomes the educational level of the head of the household. Such a preference is based on the understanding that head of the household influence or shape most of the household decisions. Literature has shown that household headship matters for better welfare outcomes across different countries (Johnson and Rogers 1993; Handa 1996; Sebens 2009; Singh et al. 2013). This is because more often than not the head of the household is the oldest member in the household and he/she would be the prime decision making authority in the household. However because the oldest surviving individual of the household is the head, he/she may have an educational level much lower compared with the younger members of the household. Hence it may be argued that the overall household’s education potential may not be necessarily reflected when education of the household head is chosen as the household educational indicator. Indeed, considering the head’s educational level to influence welfare outcomes may perhaps be inadequate and conditioned by the educational level of all other individuals in the household as well. Here one needs to imagine the process in which education influences outcomes. Education reduces information asymmetry, conformity with norms as well as has the potential to recognize implication of welfare interventions.

To address this concern, the other preferred indicator often used in literature is the highest available level of education in the household since household welfare could very well

be dictated by such levels of education. Considering a household unit with varying levels of educational attainment by its members, individual with highest available level of education is undoubtedly important as this person is likely to be the one influencing household level decisions. The highest level of education available within a household could very well represent the household educational level but as far as its implication regarding household welfare is concerned, education level of the other members along with their attributes like age and sex could be equally important. In other words, while it can be argued that education has always a systematic gradient of association with household ill-fare/welfare outcomes, strength of such gradient largely depends on the household in question, wherein the age and sex of the individual remains pertinent. In fact, household ill-fare/welfare outcomes like health care utilization, access to safe drinking water or sanitation are not governed by any one individual's educational level, rather than a composition of the educational level of all the members of the household. It is in this context that the proximate advantage of educational level or educational externality gains becomes relevant as it is not only conditioned by the highest level of education but also the gender gap in education as well as generational gaps in educational attainment. Further, it is therefore important to read every individual's educational level along with the available highest educational level that shapes welfare dividend collectively influenced by the composition of educational level of all the members of the household.

Literature offers sufficient evidence to the fact that welfare indicators are often enhanced with higher level of education for the female household members, keeping other factors constant (Tansel 2002; Glick and Sahn 2000; Kambhampati and Pal 2001). Further co-residence with individuals with better level of education enhances the educational dividend of welfare for individuals with lower level of education. This is otherwise understood as proximate positive externality as propounded by Basu and Foster (1998). They argue that if there is a literate member in the household, there can be substantial difference for all the other illiterate members in accessing information and performing task that require literacy skills. They list a number of studies which support this argument (Green et al. 1985; Foster and Rosenzweig 1996).

Given the importance of education among household members as well as female education, co-residence might generate sub-optimal welfare dividend of female educational level when they co-reside with males having lower level of education. Similarly, the generational gaps in educational levels too have causal effects on the optimal welfare indicators. In some societies, educational outcomes of younger generation is deemed as more important in the household decision making process. However in certain societies, elderly members of the households may have relatively higher bargaining power in the family. Therefore it is important that designing a household educational indicator needs accounting for all these complexities.

3 The Index

In formulation of a household educational indicator, we consider three aspects. Firstly, there is a need to account for educational levels of all the members in the household (Basu et al. 2000; Mishra 2001). As discussed, household and to some extent individual decisions are often not necessarily taken by the household head or by the individual with highest level of education within the household. Consider a household with four members with education levels given by the vector (E, x_2, x_3, H) . Consider another similar household with

same characteristics in all respect but with education levels (E, y_2, y_3, H) such that $y_2 > x_2$ and $y_3 > x_3$ with H as the highest education level of both the households and E as the educational level of the household head. Despite similar levels of the head's education and highest education in both the households, welfare outcomes are likely to be better in the second household as compared to the first one. It is in this context that it becomes pertinent to include educational levels of the all the individuals residing in the household. Accordingly our proposed index is an increasing function of average education among the household members.

The second aspect we account for is the gender construct of educational attainment. This is because of the fact that the welfare dividend of female education is greater as compared to a household with a similar level of education among males. Formally, this implies other things remaining same, a household of four male members with education level as (x_1, x_2, x_3, x_4) is likely to have lower welfare outcomes as compared to another household with three male members and one female member with same levels of education. However this incremental positive externality of female educational level may not be optimal unless coexisting with parallel levels of male education. Therefore a gender composition of highest available educational level within a household is considered in making an educational indicator that would not only account for suppression of optimality of female education externality but also recognize the incremental positive externality of female educational level in contrast with the male educational level. Thus, in the proposed educational index, we account for these complexities.

The third and equally important consideration is that of the generational aspect in educational attainment. In some context, it might be the middle aged individual, who would have greater say in the decision making process within the household as compared to the younger generation. In other societies, the elderly would be influential in the household. Accordingly we consider highest education among household members belonging to different age cohorts.

Hence as discussed, the Household Education index (hereupon referred to as HEX), will include three aspects, in consideration of the following:

1. Education of all members in the household
2. Highest education among the female and male members of the household
3. Highest education of the household members from different age cohorts.

Including these aspects in the proposed HEX would take into account the heterogeneous educational attainment of different cohorts of household members that may have differential bearing on externalities. The often used indicators of household education namely highest education and education of the household head fail to cover these aspects and hence the possible externalities of education.

The process of formulation of the index is as follows:

3.1 Aspect 1: Taking into Account the Education of all the Members in the Household

Let a household has members among which the highest education is given by H . The average education of all members is given by e . Then the educational index taking into account this aspect of including education of all members in the household is given by:

$$E_1 = (H)^{(e/H)} \quad \text{when } H \geq 1 \text{ and } e \geq 1$$

and $E_1 = e$ if $e < 1$ and $H = 1$ and $E_1 = 0$ if $H = 0$

E_1 is increasing in $e \forall e \in (0, 1)$ & $H = 1$ as well as for $H > 1$ implying as the average education of the household members increases, E_1 would also show an increase. The range of e would lie from 0 to H .

3.2 ASPECT 2: Taking into Account Gender Implications of Education in a Household

For this aspect, consider the household where highest education among the female members is given by H_f and the highest education among the male members is given by H_m . The unadjusted educational index encompassing this aspect is given by:

$$E_2(\text{unadjusted}) = [H_f^{\alpha_1} + H_m^{\alpha_2}]/2$$

where $\alpha_1, \alpha_2 \in (0, 2)$; $\alpha_1 + \alpha_2 = 2$; $\alpha_1 \geq \alpha_2$

The parameters, α_1 and α_2 assign the relevant weights of importance in terms of externality. For similar educational level, literature indicates that female education has greater welfare externality in comparison to that of males. Hence we ensure this condition with the weak inequality condition, $\alpha_1 \geq \alpha_2$. If H_h is the highest education level that can be attained, then final index capturing aspect 2 is given by:

$$E_2 = E_2(\text{unadjusted}) * H_h / [(H_h^{\alpha_1} + H_h^{\alpha_2})/2]$$

E_2 is increasing in H_f and H_m though the rate of change might be different depending on the parameters, α_1 and α_2 .

It should be noted that education of the household head's partner is taken into account through aspect 1 and possibly aspect 2 if the partner's education is highest among his/her gender category. Arguably, the partner's education is an indication of the bargaining power within the household as well as that on partnership. However our final index captures her education as well.

3.3 Aspect 3: Taking into Account Implications of Education Across Different Age Groups

For this aspect, consider a household where the household members are divided into n cohorts based on age groups. H_i is the highest education among the members in the i^{th} cohort. Then the unadjusted educational index encompassing this aspect is given by:

$$E_3(\text{unadjusted}) = \sum_{i=1}^n H_i^{\alpha_i} / n$$

where the parameter, α_i give the relative importance in the household of the i^{th} cohort. The following condition should as well hold: $\sum_{i=1}^n \alpha_i = n$

If H_h is the highest education level that can be attained, then final index capturing aspect 3 is given by:

$$E_3 = E_3(\text{unadjusted}) * H_h / \left[\sum_{i=1}^n H_h^{\alpha_i} / n \right]$$

As in the earlier case, please note E_3 is increasing in H_i . The rate of change in E_3 with respect to H_i for various i s is different depending on the parameters, α_i s.

3.4 The Final Index

After the three aspects, E_1, E_2 and E_3 have been normalized, they are aggregated to obtain a single HEX through three methods:

Linear averaging: This is derived from simple arithmetic mean of the three aspects.

$$E = (E_1 + E_2 + E_3)/3$$

Geometric mean: This is derived from geometric mean of the three aspects.

$$E = (E_1.E_2.E_3)^{1/3}$$

In the aggregation through linear averaging, there remains the problem of perfect substitutability, which implies a loss in one aspect is made up by a similar gain in the other with the index value remaining unchanged. Fundamentally, since this may not make sense, we aggregate by taking geometric mean. Notably the assumption of perfect substitutability has been criticized in literature (Desai 1991; Hopkins 1991; Palazzi and Lauri 1998; Sagar and Najam 1998; Nathan et al. 2008; Herrero et al. 2010). In contrast geometric mean gives higher weightage to the dimension having lower performance and penalizes imbalanced achievement across the aspects.

Displaced Ideal: Following Mishra and Nathan (2018), we aggregate E_1, E_2 and E_3 in the following way:

$$E = H_h - [\{(H_h - E_1)^2 + (H_h - E_2)^2 + (H_h - E_3)^2\}^{1/2} / (3)^{1/2}]$$

where $\{(H_h - E_1)^2 + (H_h - E_2)^2 + (H_h - E_3)^2\}^{1/2}$ is the euclidean distance from the ideal, which is H_h . In the three dimensional space, this is normalized by dividing with $(3)^{1/2}$. Hence a household which has a value of one of the aspect closer to H_h , the higher would be the measure of E .

As with the case with all indices, the HEX should as well satisfy a number of axioms or intuitive properties:

Monotonicity: As one can observe, HEX, obtained based on all the three aggregation methods satisfy the monotonicity axiom as the index gives higher (lower) value if the achievement in one aspect increases (decreases) with the value in the other two aspects remaining same. In fact, if educational level of a household member increases (decreases), E_1 would increase and accordingly, the final HEX would also increase since HEX is an increasing function of the three aspects.

Anonymity: We find that the HEX from all the three aggregation methods satisfy the anonymity condition as the proposed index should be indifferent to swapping of values across the aspects. For two households, i and j , $E_i = E_j$ if values are interchanged across the two aspects and in the other aspect, the value remains same. Hence if $E_{2i} = E_{3j}$ and $E_{2j} = E_{3i}$ and $E_{1i} = E_{1j}$, then $E_i = E_j$.

Normalization: The proposed educational index, HEX, from all the three aggregation method has a minimum of 0 and a maximum of H_h , which is the maximum possible education one can achieve, so $HEX \in (0, H_h)$. If no one in the household is educated, HEX obtains a value of 0 as values of all the three aspect separately would be 0. If all are educated at the highest level, all the aspect separately would return a value of 16 and hence HEX would also return a value of H_h .

Uniformity: The proposed HEX from the geometric mean and displaced ideal method is found to satisfy the uniformity axiom, which says for a given mean value of HEX across dimension, μ , higher (lower) dispersion across dimension would return a lower (higher) value of HEX. For example, for a given E_1 , if for a household, the value of $E_2 = 8$ and $E_3 = 14$ and for another household with same E_1 , $E_2 = 9$ and $E_3 = 13$, The HEX value for the second household would be higher. This axiom penalizes unbalanced development across dimensions. In other words, a household with higher educational achievement across all members, taking care of the gender as well age cohorts would have higher value. As mentioned, the HEX derived through linear averaging does not satisfy this axiom; however that from the other two methods satisfy.

4 Empirical Exploration

We use the developed HEX in the Indian context to look at the difference that may crop up while using the proposed index as against the normally used index of highest education among the household members or education of the household head.

4.1 Data and Variables

Data from the Indian Human Development Survey conducted in 2011-2012, generated jointly by National Council of Applied Economic Research (NCAER) and University of Maryland has been used for the empirical exploration.³ The survey covered over 40,000 households and gathered data on education, health, economic wellbeing, social status, and various other domains. In particular, the survey collected information on the number of standard years completed by each family member of the surveyed households in terms of education. Hence this variable is coded as “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation. The analysis in this paper is based on this variable. The survey data collected information from all the states of India and is considered to be representative nationally as well as at the state level.

For our analysis, we consider only those individuals who are above 6 years of age and have not reported “studying” as their current principal activity status. This indicates that the individual have completed his/her education and hence have reported a current principal activity other than studying, which may be cultivation, agricultural wage labour or household work among others. Accordingly, we include all individuals who have attained the minimum age of attending schools (6 years of age and above in general) and have

³ Full description of the survey can be obtained from <https://ihds.umd.edu/> (last accessed on March 22, 2019).

Fig. 1 Kernel density plot for E_1 (aspect1), highest education and household head's education. *Note.* The figure is generated using all the eligible sampled households covered in the IHDS 2011–12. The command used in STATA is “*kdensity*”. The unit of educational indicator (x-axis) is number of standard years completed. Here “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation

completed education (hence reported a principal activity other than “studying”) for the purpose of our constructing the indicator as well as analysis.

4.2 Calculating the Indices Capturing Three Aspects

We develop the indices capturing the three aspects separately. For the first aspect, education of all the members in the household is taken into account. For calculating e as well as H , we consider all the household members who are above 6 years of age and have not reported “studying” as the primary activity status. The pair-wise correlation of E_1 with the highest education among the household members and education of the household head is found to be 0.73 and 0.78 respectively. Figure 1 presents the kernel density plots for the three indicators. The plots indicate considerable difference between E_1 and the two commonly used indicators. Notably, we find distribution of the highest education in the household and household head's education to be pretty close, which is not largely the case with E_1 .

The index capturing the second aspect (E_2) is then calculated with the value of $\alpha_1 = 1.1$ and $\alpha_2 = 0.9$. The pair-wise correlation of E_2 with the highest education among the household members is found to be as high as 0.89 but with education of the household head it is found to be 0.59. Figure 2 presents the kernel density plots for the three indicators. As evident from the figure, we find considerable difference between E_2 and these commonly used household educational indicators.

Please note that the values are sensitive to the selection of α_1 and α_2 . If we increase the value of α_1 , it serves towards more weightage being assigned to female education in comparison to male education in the household. In that case, the difference between the E_2 s and the educational indicators would increase. Figure 3 gives the kernel density plots of E_2

Fig. 2 Kernel density plot for E_2 (aspect2), highest education and household head's education. *Note* The figure is generated using all the eligible sampled households covered in the IHDS 2011–12. The command used in STATA is “kdensity”. The unit of educational indicator (x-axis) is number of standard years completed. Here “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation

Fig. 3 Kernel density plots of E_2 with various sets of α_1 and α_2 . *Note* The figure is generated using all the eligible sampled households covered in the IHDS 2011–12. The command used in STATA is “kdensity”. The unit in the x-axis is number of standard years completed. Here “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation

s with various sets of α_1 and α_2 . As one can observe, the E_2 with $\alpha_1 = 1.5$ deviates considerable with the other set of E_2 s with lower values of α_1 .

Fig. 4 Kernel density plot for E_3 (aspect3), highest education and household head's education. *Note* The figure is generated using all the eligible sampled households covered in the IHDS 2011–12. The command used in STATA is “*kdensity*”. The unit in the x-axis is number of standard years completed. Here “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation

The third aspect (E_3) is calculated with four groups: household members who have completed education and are below 25 years of age; those between 25 and 45 years of age; those between 46 and 65 years of age and those above 65 years old. The parameters are as follows: $\alpha_1 = 0.8$; $\alpha_2 = 0.9$; $\alpha_3 = 1.1$ and $\alpha_4 = 1.2$. Because in India, household decision are mainly taken by the elder members of the family, we take their weight age in the index to be the highest (1.2). The weight age for the youngest group is the lowest (0.8). We observe after calculation of E_3 that the pair-wise correlation with highest education and education of the household is close to 0.6 and 0.74 respectively. Figure 4 presents the kernel density plots for the three indicators. As evident from the figure, we find considerable difference between E_3 and these commonly used household educational indicators.

Having computed the three aspects separately, adopting the three aggregation methods as discussed, we calculate the overall HEX, E . For the households in which members of both the genders do not reside or even if they reside, one group has not completed education (E_2 is missing), we calculate E as:

$$\text{Linear average: } E = (E_1 + E_2)/2$$

$$\text{Geometric mean: } E = (E_1 \cdot E_3)^{1/2}$$

$$\text{Displaced ideal: } E = 16 - \{[(16 - E_1)^2 + (16 - E_3)^2]^{1/2} / (2)^{1/2}\}$$

Table 1 shows the pair-wise correlation of E derived based on adopting all the three aggregate methods and the separate indices representing the three aspects along with the commonly used indicators, highest education in the household and education of the household head. The results indicate expected strong correlation with aspect 1. It reduces for aspects 2 and 3 and further reduces for the commonly used indicators. Figure 5 further shows through the kernel density plot that considerable difference exists

Table 1 Pairwise correlation of the index covering the aspects with the overall hex calculated from the three methods

Indicator	Linear average	Geometric mean	Displaced ideal
Aspect 1 (E_1)	0.9571	0.969	0.9662
Aspect 2 (E_2)	0.8982	0.862	0.8722
Aspect 3 (E_3)	0.8870	0.903	0.8979
Highest education	0.8392	0.804	0.8167
Education of the household head	0.8079	0.813	0.8085

The correlation is calculated using all the sampled households covered in the IHDS 2011–2012

Fig. 5 Kernel density plot of the final index through the three methods (geometric mean, linear average, displaced ideal), highest education and household head's education. *Note* The figure is generated using all the eligible sampled households covered in the IHDS 2011–2012. The command used in STATA is “kdensity”. The unit in the x-axis is number of standard years completed. Here “1” for standard 1; “2” for standard 2 and so on. It codes “0” for those who could not complete the first standard and codes “16” for those went on to complete education above graduation

between the proposed HEX measures and the commonly used indicators. Notably one can find that the density reduces substantially for higher values of the proposed HEX measures, which is not the case with the indicators like highest education within the household and education of the household head. Hence these commonly used household educational indicators, which generally represent an overall measure educational attainment of the household in true sense are often overestimated. This is further validated through the descriptive statistics of these indicators among the sampled household. This is presented in Table 2. As one can observe, the proposed HEX measures have lower mean values as compared to the other two commonly used indices.

Table 2 Descriptive statistics of the overall hex calculated from the three methods

Indicators	Total		Urban		Rural	
	Mean	SD	Mean	SD	Mean	SD
HEX (linear average)	5.201	3.585	6.971	3.740	4.262	3.116
HEX (geometric mean)	4.921	3.540	6.678	3.758	3.990	3.031
HEX (displaced ideal)	4.994	3.449	6.689	3.615	4.096	2.991
Highest education	8.520	4.743	10.418	4.350	7.514	4.634
Education of the head	5.665	4.927	7.682	4.991	4.597	4.543
Observations	42,128		14,591		27,537	

The mean and standard deviation is calculated using all the sampled households covered in the IHDS 2011–2012

5 Regression Analysis

We now demonstrate the relevance of the proposed household educational indices in contrast with the two commonly used indicators: highest educational level within in the household and education of the household head. This is revealed with two sets of regressions. The first set examines the association of children educational learning outcomes with these indices separately, controlling for a set of confounding factors. Given wide literature on household member's role in raising its children as discussed, we hypothesize a positive relationship between the two. The second one inspects the possible association of the household education indices with the probability of a household possessing a latrine. This is important as we find prevalence of open defecation to be high in India. As of 2015, almost 40% of the Indian population reported defecating in the open.⁴ One of the many reasons of this persistent open defecation is due to the fact that many households do not have access to a latrine. The IHDS survey, for example shows about 45 percent of Indian Households do not have a latrine in their house. Given widespread literature on the ill effects of open defecation, we hypothesize that awareness of this aspect would be higher in households which are well educated in comparison to those who are less educated. Hence we hypothesize a positive relationship between household education indicator and the probability of having access to a latrine.

5.1 Outcome Variable

The IHDS survey administers short tests for children aged 8–11 years capturing learning outcomes on reading, math and writing. These simple tests have been conducted in 14 languages and the concerned is free to choose a language which he/she prefers to write the test in. These tests have been successfully administered to over 11,500 children at their homes.

Outcomes on reading skills have been coded into five categories from 0 to 4, which are grouped as:

⁴ This can be seen from the website <https://data.worldbank.org/indicator/SH.STA.ODFC.ZS?locations=IN> (last accessed on March 22, 2019).

- (i) 0: those who cannot read at all
- (ii) 1: those who can recognize letters but not words.
- (iii) 2: those can read words but not a paragraph
- (iv) 3: those who can read a paragraph but not a story
- (v) 4: those who can read a story.

Math scores are coded into four categories:

- (i) 0: those who are unable to recognize numbers
- (ii) 1: those who recognize numbers but are unable to do arithmetic
- (iii) 2: those who can do a subtraction problem but not division
- (iv) 3: those who can solve a division problem

Writing has been coded into three categories:

- (i) 0: those who cannot write
- (ii) 1: those who can write a sentence but make one or two mistakes
- (iii) 2: those who write without mistakes.

We use average outcomes from these three tests for every child as our dependent variable for the first set of regressions.

For the second set of regression, the dependent variable as discussed is whether there is a toilet in the household. This is a dichotomous variable which takes the value of “0” if the variable is coded as “No facility belonging to household (or open fields)” and “1” if the variable is coded as either “Traditional pit latrine”, or “Semi-flush (Septic tank) latrine” or “Flush toilet”. We drop the households which has given no response (about of 0.4% of the total sample of households).

5.2 Independent Variables

The main explanatory variables of interest in both the set of regressions are highest education level within the household, education of the household head and our proposed set of educational indices. These variables are introduced separately in three different regression models for each set.

In terms of independent control variables, we include those which are relevant to the Indian context that may affect learning outcomes and these control variables remain same across the three regressions for the first set. Gender and age of the child are direct confounders and hence have been included. Along with that the grade of the child, caste and religion, household size, the state where household resides in, age and of the household head have been included. Household economic factors like yearly per capita consumption expenditure, television ownership and type of walls (whether concrete or not) have been added as controls. In the regression model, we also control for confounding variables like time spent in school, usage of computers, private coaching and for doing homework. Further, school management (private or government run schools) is also controlled as a number of studies have found improved performance of children studying in private schools (Rob and Kingdon 2010; Muralidharan and Sundararaman 2015). Other independent variables include whether the child has suffered from short-term illness or fever in the last 30 days prior to the survey, medium of instruction in the school, distance of the school and gender of the teacher since all these variables can directly

influence learning outcomes. Children having major morbidity problems such as mental illnesses, cancer, paralysis and heart diseases, and those who are not attending any schools have been dropped from the analysis due to limited number of observations.

For the second set of regressions, we consider a different set of control variables, some of which are common in both the sets. For example, yearly per capita consumption expenditure, television ownership and type of walls (whether concrete or not) have been included in the model since these variables largely are indicators of economic condition of the household, which in turn would likely to be correlated with possession of a toilet. Studies have shown religion and caste to be important determinants of toilet usage and hence presence of one in the household. For example toilet usage in India is found to be higher among Muslims (Geruso and Spears 2018). Similarly, it is also found that barriers to toilet usage is largely associated with the notions of purity, which is also entrenched in caste (Coffey et al. 2017). Accordingly we include socio-religious group dummies in the model. Main occupation of the household along with gender of the household head is also used as controls since both are determinants of presence of toilets in the households. Given utility of toilets being higher among females, it is likely that a female headed household would have higher probability of having a toilet in the household. Hence we also include number of adolescent females in the household along with number of elderly members. The regressions are run with state fixed effects and the standard errors are clustered at the PSU level.

5.3 Regression Model

As mentioned earlier, for the first set of regressions, the dependent variables are average learning outcomes from reading, mathematics and writing skills for children from 8 to 11 years. Since this is continuous variable, we apply Ordinary Least Squares (OLS) regression to get an estimate of the association of learning outcomes and household educational variables.

For the second set of regressions, since the dependent variable (presence of a toilet in the household) is a dichotomous variable, we run a logistic regression. The estimates are obtained from maximum likelihood.

For both, the model can be specified as:

$$y_{ij} = \lambda \cdot EI_j + \beta X_i + \varepsilon_i \quad (1)$$

Here y_{ij} is the learning score for child, i coming from household, j for the first set of regression and whether there is a toilet in the household, j for the second set. In each set of regressions, EI_j is the education of the household head for the first regression, highest education in the household for the second regression and in the third to fifth regressions it is the three set of educational index., aggregated by linear average method, geometric mean method and displaced ideal method. We compare λ of the three regressions. X_i is the matrix of corresponding household and child level control variables pertaining to child, i for the first set and matrix of confounding household characteristics for household, j (replace X_i by X_j) for the second set. The random error term is given by ε_i .

5.4 Regression Results

Figure 6 presents the point estimates of the coefficients of the three educational indices along with the confidence level at 95% level of significance from the first set of regressions to examine the association of the educational indicators with educational or learning

Fig. 6 Regression estimates of regression of learning outcomes on the educational outcomes. *Note* The coefficients along with the 95% confidence intervals are presented in the figure. The dependent variable is the average score obtained by the child in reading, writing and mathematics standardized test. The sample includes all children from 8 to 11 years surveyed in the IHDS 2011–2012. We ran separate OLS regressions keeping the above five variables in the regression separately as independent variables. The controls in all these regressions are same. The coefficients for the control variables are given in the appendix (Table A1). STATA software has been used to generate the regression coefficients

outcomes of the children. As one can observe, the association of the set of proposed HEX measures are significantly higher (at 95% level) than that of both the commonly used indicators: education of the household head and highest education in the household. When we compare the other two, we did not find a significant difference. In terms of the point estimates, an unit increase in our proposed HEX measures is associated with an increase of educational scores by about 4.4 percentage and this is consistent across all the indices derived from the three aggregation method. However, a unit increase in education of the household head and highest education is found to be associated with a rise in learning scores by 2.7 and just below 3 percentage points respectively.

We run the same regressions for rural areas alone. The results are shown in Fig. 7. Like in the case above, we get a significant difference between the commonly used indicators and the proposed HEX measures derived from the three aggregation methods. A unit increase in education of the household head or highest education in the household is associated with an increase in average test scores for the rural children by less than 3 percentage points. However for the HEX measures, an unit increase is associated with about 4.2 percentage point increase in test scores.

Having verified the differential bearing of household educational indicator with educational performance of children, we examine the relationship between the educational indicators and the probability of presence of toilet facility in the household. Figure 8 presents the point estimates of the odds ratio from the three logistic regressions along with the confidence intervals at 95% level of significance. An odds ratio of more than 1 indicates a positive relationship whereas a negative relationship is indicated by an odds ratio of less

Fig. 7 Regression estimates of regression of learning outcomes on the educational outcomes (for rural areas). *Note* The coefficients along with the 95% confidence intervals are presented in the figure. The dependent variable is the average score obtained by the child in reading, writing and mathematics standardized test. The sample includes all rural children from 8 to 11 years surveyed in the IHDS 2011–2012. We ran separate OLS regressions keeping the above five variables in the regression separately as independent variables. The controls in all these regressions are same. The coefficients for the control variables are given in the appendix (Table A2). STATA software has been used to generate the regression coefficients

Fig. 8 Odds ratio from logistic regression of having a toilet in the household on the educational outcomes. *Note* The odds-ratios along with the 95% confidence intervals are presented in the figure. We ran separate logistic regressions keeping the above five variables in the regression separately as independent variables. The controls in all these regressions are same. The coefficients for the control variables are given in the appendix (Table A3). The dependent variable is whether the household has access to a private latrine (dichotomous). The sample includes all households surveyed in the IHDS 2011–2012. STATA software has been used to generate the regression coefficients

Fig. 9 Odds ratio from logistic regression of having a toilet in the household on the educational outcomes (in rural areas). *Note* The odds-ratios along with the 95% confidence intervals are presented in the figure. We ran separate logistic regressions keeping the above five variables in the regression separately as independent variables. The controls in all these regressions are same. The coefficients for the control variables are given in the appendix (Table A3). The dependent variable is whether the household has access to a private latrine (dichotomous). The sample includes all rural households surveyed in the IHDS 2011–2012. STATA software has been used to generate the regression coefficients

than 1. The findings indicate more than 16% increase of the odds of having toilet facility in the household with an unit increase in the proposed household educational index. With the unit increase in highest education and education of the household head, the odds of having toilet facility increases by 9 and 7 percentage points respectively. This indicates our proposed index fetching higher intrinsic value of education which has its impact over different household welfare indicators. The fact that our index encompasses the intrinsic value of education among all the members assigning greater weight age to female members owing to higher positive externality and to elderly members get reflected in better welfare achievements.

We obtain similar findings when the regressions are run for rural areas, the results of which are shown in Fig. 9. Here an unit increase in the proposed HEX measures is found to be associated with an increase in odds of a household having toilet facilities by close to 17%. When the commonly used indicators are taken instead, we find the increase in odds by less than 10 percentage points.

6 Conclusion

The proposition of a household educational indicator is motivated by the complexities of educational attribute in terms of its welfare dividends. While education is essentially an individual attribute, its welfare dividend goes beyond the individual to household and to the community at large. Given the recognition of educational externality from aspects of proximity as well as its gender facet, the welfare outcomes at the household may not be linked solely with education of the household head or the highest education level in

the household, which are mostly used as an indicator of household educational level. In this context, this paper constructs a household educational index, which combines three aspects. Following Basu and Foster (1998), firstly it accounts for distribution of education among household members. Secondly to account for female education externality, highest education level among the male and female members are considered and thirdly highest education level between different generations within the household is included to capture the aspect of inter-generational educational externality.

The empirical verification based on a nationally representative survey conducted in 2011–12 in India indicate considerable difference between the proposed household educational indicator and the commonly used indicators including education of the household head and highest education in the household. We also find stronger association of the proposed indicator with the welfare outcomes like learning outcomes among children and having access to toilet facility in the household.

The significance of the study is immense. Considering the complexities of educational dividend as discussed, this exercise offers a comprehensive household educational indicator in accommodation of its numerous aspect of externality. The fact that it is found to exhibit a better strength of association with welfare outcomes is indicative of the improvement made in obtaining the new indicator involving educational profile of all household members. Beyond its comparative strength, the conceptual content of the indicator has the virtue to be considered holistic and sensitive. The empirical illustration establishes the advantage of such an indicator on one hand and confirms the limitation of existing indicators in terms their bearing on welfare outcomes on the other. More importantly the comprehensive feature of the newly developed measure can always be preferred over others in describing micro as well as macro level outcomes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Basu, K., & Foster, J. E. (1998). On measuring literacy. *Economic Journal*, 108(6), 1733–1749.
- Basu, K., Foster, J. E., & Subramanian, S. (2000). Isolated and proximate illiteracy and why these concepts matter in measuring literacy and designing education programmes. *Economic and Political Weekly*, 35, 8–14.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2), 9–49.
- Behrman, J. R., Foster, A. D., Rosenweig, M. R., & Vashishtha, P. (1999). Women's schooling, home teaching, and economic growth. *Journal of Political Economy*, 107(4), 682–714.
- Behrman, J. R., & Wolfe, B. L. (1987a). How does mother's schooling affect family health, nutrition, medical care usage, and household sanitation? *Journal of Econometrics*, 36(1–2), 185–204.
- Behrman, J. R., & Wolfe, B. L. (1987b). Investments in schooling in two generations in pre-revolutionary Nicaragua: The roles of family background and school supply. *Journal of Development Economics*, 27(1–2), 395–419.

- Behrman, J. R., & Wolfe, B. L. (1989). Does more schooling make women better nourished and healthier? Adult sibling random and fixed effects estimates for Nicaragua. *Journal of Human Resources*, 24(4), 644–663.
- Caldwell, J., & McDonald, P. (1982). Influence of maternal education on infant and child mortality: Levels and causes. *Health Policy and Education*, 2(3–4), 251–267.
- Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics* (Vol. 3, pp. 1801–1863). Elsevier, Amsterdam.
- Chudgar, A. (2009). Does adult literacy have a role to play in addressing the Universal Elementary Education challenge in India. *Comparative Education Review*, 53(3), 403–433.
- Chudgar, A. (2011). Female headship and schooling outcomes in rural India. *World Development*, 39(4), 550–560.
- Coffey, D., Gupta, A., Hathi, P., Spears, D., Srivastav, N., & Vyas, S. (2017). Understanding open defecation in rural India: Untouchability, pollution, and latrine pits. *Economic and Political Weekly*, 52(1), 59–66.
- Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: evaluating theories and evidence*. National Bureau of Economic Research (No. w12352).
- Desai, M. (1991). Human development: Concepts and measurement. *European Economic Review*, 35(2–3), 350–357.
- Foster, A. D., & Rosenzweig, M. R. (1996). Technical change and human-capital returns and investments: Evidence from the green revolution. *The American Economic Review*, 86, 931–953.
- Geruso, M., & Spears, D. (2018). Neighborhood sanitation and infant mortality. *American Economic Journal: Applied Economics*, 10(2), 125–162.
- Glick, P., & Sahn, D. E. (2000). Schooling of girls and boys in a West African country: The effects of parental education, income, and household structure. *Economics of Education Review*, 19(1), 63–87.
- Green, S. E., Rich, T. A., & Nesman, E. G. (1985). Beyond individual literacy: The role of shared literacy for innovation in Guatemala. *Human Organization*, 44(4), 313.
- Handa, S. (1996). Expenditure behavior and children's welfare: An analysis of female headed households in Jamaica. *Journal of Development Economics*, 50(1), 165–187.
- Haveman, R. H., & Wolfe, B. L. (1984). Schooling and economic well-being: The role of nonmarket effects. *Journal of Human Resources*, 19, 377–407.
- Herrero, C., Martínez, R., & Villar, A. (2010). *Improving the measurement of human development*. UNDP-HDRO Occasional Papers, (2010/12).
- Hopkins, M. (1991). Human development revisited: A new UNDP report. *World Development*, 19(10), 1469–1473.
- Johnson, F. C., & Rogers, B. L. (1993). Children's nutritional status in female-headed households in the Dominican Republic. *Social Science and Medicine*, 37(11), 1293–1301.
- Kambhampati, U. S., & Pal, S. (2001). Role of parental literacy in explaining gender difference: Evidence from child schooling in India. *The European Journal of Development Research*, 13(2), 97–119.
- Krueger, A., & Ashenfelter, O. (1992). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, 84(5), 1157–1173.
- Lucas, R. E., Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.
- McMahon, W. W. (2004). The social and external benefits of education. In *International handbook on the economics of education* (p. 211).
- Mishra, S. (2001). Isolated and proximate illiteracy. *Economic and Political Weekly*, 2003–2008.
- Mishra, S., & Nathan, H. S. K. (2018). A manush or humans characterisation of the human development index. *Journal of Human Development and Capabilities*, 19(3), 398–415.
- Muralidharan, K., & Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in India. *The Quarterly Journal of Economics*, 130(3), 1011–1066.
- Nathan, H. S. K., Mishra, S., & Reddy, B. S. (2008). *An alternative approach to measure HDI* (pp. 1–23). Mumbai: Indira Gandhi Institute of Development Research (IGIDR).
- Palazzi, P., & Lauri, A. (1998). The human development index: Suggested corrections. *PSL Quarterly Review*, 51, 205.
- Rob, F., & Kingdon, G. (2010). *The Relative Effectiveness of Private and Government Schools in Rural India: Evidence from ASER Data*. 10-03. DoQSS Working Papers. Department of Quantitative Social Science - UCL Institute of Education, University College London. <https://ideas.repec.org/p/qss/dqssw/p/1003.html>.
- Sagar, A. D., & Najam, A. (1998). The human development index: A critical review. *Ecological Economics*, 25(3), 249–264.
- Schultz, T. P. (1988). Education investments and returns. *Handbook of Development Economics*, 1, 543–630.

- Schultz, T. P. (2002). Why governments should invest more to educate girls. *World Development*, 30(2), 207–225.
- Seebens, H. (2009). *Child welfare and old age security in female headed households in Tanzania*. IZA discussion paper no. 3929. Bonn: Institute for the Study of Labor.
- Singh, A., Gaurav, S., & Das, U. (2013). Household headship and academic skills of Indian children: A special focus on gender disparities. *European Journal of Population/Revue européenne de Démographie*, 29(4), 445–466.
- Tansel, A. (2002). Determinants of school attainment of boys and girls in Turkey: individual, household and community factors. *Economics of Education Review*, 21(5), 455–470.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.