Measures of Effective Literacy: A Theoretical Note

Diganta Mukherjee[a,b] and Manash Ranjan Gupta[a]

a: Economic Research Unit, Indian Statistical Institute, Kolkata - 700 108, INDIA
b: Institute for Social and Economic Research, University of Essex, UK

Abstract

In this paper, some axiomatic extensions of **Basu - Foster (1998)** measure of effective literacy are made in the light of endogenizing the external effect on proximate illiterates relating to several characteristics of the literates in the household, e.g. age, sex, education etc.

Key Words: Proximate Illiteracy, Measurement, Age, Gender, Education.

JEL classification numbers: I21

Address for Correspondence:
Diganta Mukherjee
Institute for Social and Economic Research
University of Essex
Colchester, Essex CO4 3SQ, United Kingdom
Tel: +44 (0)1206 872631
Fax: +44 (0)1206 873151
e-mail address: diganta@isical.ac.in, digantam@hotmail.com

1 Introduction

Traditionally the literacy rate is defined as the ratio of number of adult literates to the total number of adults. **Basu and Foster (1998)** (henceforward **BF**) have suggested an alternative measure of effective literacy in which the distinction is made between proximate illiterates and isolated illiterates.[1] A proximate illiterate is assumed to be α-equivalent to a literate member of the household with $0 < \alpha < 1$. The significance of this alternative measure in designing literacy education programme is also documented by **Basu et al (1999)**. Some theoretical modifications and comments on this measure are also available in **Chakravarty and Mazumder (2001)**, **Subramanian (2001)** and **Mishra (2001)**.

However, in all these works, the external effect of literacy on the proximate literates is exogenous and is independent of the set of characteristics of the literate members of the household.[2] In reality, the magnitude of the external effect of literacy on the proximate illiterate depends on the various characteristics of the literate members in the household. Age of the literate member is one such determinant because an elder literate member usually exerts a greater external effect on the illiterate than an younger literate member. A female literate member should have a larger effect than a male literate member because the females play a more active role in the domestic activities of the household. The literate member who stays in the home should have a greater effect than a migrant literate member. Also the level of education of the literate member is an important determinant of α as an illiterate member has a greater respect for the literate member with higher educational qualification.

In this note we extend the **BF** measure of effective literacy in the light of endogenizing α in terms of the above mentioned characteristics of the literate members of the household. We propose a set of axioms that the ideal measure of literacy should satisfy; and which postulates the effect of the characteristics like age, sex and level of education of literate members on a proximate illiterate. Our suggested extensions of **BF** measure satisfy these axioms; and are reduced to **BF** measure itself in the special case when insensitivity of the external effect to the relevant characteristics is assumed.

The paper is organized as follows. In section 2, the axioms are described and preliminary observations are made. The formulae and the relevant lemmas and characterisation theorems are presented in section 3. Concluding remarks are made in section 4.

[1] Proximate illiterates are the illiterate members of the household with at least one literate member. Isolated illiterates come from a household with no literate member.

[2] **BF** partially takes care of this point in footnote 5, section 5 and 6 (last paragraph) of their paper.

2 Preliminaries

We will now describe the relevant **set of variables** that we deem are important characteristics of a literate person with respect to the externality effect that she exerts on an illiterate member of the same household. For any person i, $i = 1, ..., N$, in an N-person household, we define the variables in the following way.

(i) Level of education, $e_i \in E = \{0, 1, 2, ..., k\}$ or $[0, k]$ for some $k \geq 1$, integer. That is, we consider either several discrete levels of education or may allow education to be measured on a continuous scale.

(ii) Sex, $s_i \in \{m, f\}$.

(iii) Age, $y_i \in R_+$.

Thus, for our purpose, person i is now completely described by the characteristic vector (e_i, y_i, s_i).[3] Define the set $C^N = (E \times \{m, f\} \times R_+)^N$ for positive integer N (household size), as the set of all possible characteristic vectors for a given household size. Also define $C = \bigcup_{N \geq 1} C^N$ as the union of C^N over all positive integers N. Therefore we can define the measure of household level literacy as a function P defined over all possible household sizes and the set of characteristic vectors for each individial in the household to a real number between 0 and 1. More precisely, $P : Z_+ \times C \rightarrow [0, 1]$.

We now introduce the set of basic assumptions on our literacy measure P.

A1 (Anonymous additivity): The aggregate literacy status of the household is the average of each person's literacy status in the household.

$$P(.) = \frac{1}{N} \sum_{i=1}^{N} p(N; (e_i, s_i, y_i), (e_j, s_j, y_j), j = 1, ..., N, j \neq i) = \frac{1}{N} \sum_{i=1}^{N} p(N; \Omega_N)$$

where $p : Z_+ \times C \rightarrow [0, 1]$ is the identical literacy indicator function for each of the members of the household and $\Omega_N = \{(e_i, s_i, y_i), i = 1, ..., N\}$, the set of characteristics of all members in the household.

Thus, assumption (A1) provides us with a convenient breakdown of the general effective literacy measure of the household, $P(.)$, in terms of the effetive literacy status of each individual in the household, $p(.)$, and postulates that $P(.)$ is a simple average of the $p(.)$ values of all the members of the household.

A2 (Maximality): The maximum value (= 1) of literacy status for the individual i is attained when $e_i > 0$. That is,

$p(N; (e_i, s_i, y_i), .) = 1$ if and only if $e_i > 0$.

[3]Whatever follows is actually much more general, as any set of individual characteristics including the level of education may be considered as the relevant characteristic set. Our use of age and sex is just one of many possibilities that may be considered relevant. Analysis with just one additional characteristic is also feasible but we have considered two to allow for interactions between characteristics which are often considered relevant.

The minimum value (= 0) is attained when there are no literate members in the household. That is,

$p(N;(e_i,s_i,y_i),(e_j,s_j,y_j),j=1,...,N,j\neq i)=0$ if and only if $e_j=0$ for all $j=1,...,N$.

For all other cases, $0<p(.)<1$.

A3 (Monotone externality of literates): The effective literacy status of any person is determined by his own characteristics and that of the literate members (if any) only and is non-decreasing in the cardinality of the set of literates.

That is, we can write $P(.)$ as

$$P(.)=\frac{1}{N}\sum_{i=1}^{N}p(N;(e_i,s_i,y_i),L_N), \tag{1}$$

where $L_N=\{(e_j,s_j,y_j)\in\Omega_N|e_j>0\}=$ the set of characteristics of the literate members in the household. Denote the size of this set by $l=|L_N|$.

(A3) postulates that $p(N;(e_i,s_i,y_i),L_N)$ is nondecreasing in l. Each proximate illiterate's effective literacy status (weakly) improves if the number of literate members in the household increases.

(A1) - (A3) is our basic set of axioms. Note that (1) is the most general form of effective literacy measure that we consider.

We will now introduce the additional set of axioms which are dependent on alternative judgments about the externality effect we are trying to model.

A4 (Education level sensitivity) : For any $(e_j,s_j,y_j)\in L_N$

$$\frac{\partial p}{\partial e_j}(N;(0,s_i,y_i),L_N)>0.$$

That is, the externality effect of any literate member j on the illiterate member i in the household, given other things, is increasing in the level of education of person j.

We define the contrary position to **(A4)** by the following

A4' (Education level insensitivity) : The externality effect of any literate member j on the illiterate member i in the household is independent of the level of education of j.

A5 (Gender sensitivity) : There can be several variants of such axiom. We present two of those.

(i) For any $(e_j,m,y_j),(e_k,f,y_k)\in L_N$

$$\frac{\partial p}{\partial e_j}\leq\frac{\partial p}{\partial e_k},$$

for illiterate person i. That is, female literates has higher externality on the literacy status of illiterates in the household than the male literates. Alternatively, one can think of f as

(m, δ) where δ is a hypothetical entity adding the extra effect. As females spend more time in household activities, this structure is justified.

(ii) : Given other factors, female literates has a higher externality on female illiterates than on male illiterates. Similarly for a male literate.

Again we define **A5' (Gender insensitivity)** in the same manner as in **(A4')**.

A6 (Age ordering sensitivity) : Again, for an illiterate person,

$$\frac{\partial p}{\partial e_j} \geq \frac{\partial p}{\partial e_k},$$

where $(e_j, s_j, y_i + h), (e_k, s_k, y_i - k) \in L_N$ and $h, k > 0$. That is, externality effect is more effective on younger persons than on elders. (Eg. Justifying adult education programmes.)

Again, we can define **(A6')**.

A7 (Insensitivity to multiple identical literates) : For $(e_1, s_1, y_1), (e_2, s_2, y_2) \in L_N$ with $e_1 = e_2$, $s_1 = s_2$ and $y_1 = y_2$

$$p(N; (0, s_i, y_i), (e_1, s_1, y_1), (e_2, s_2, y_2)) = p(N; (0, s_i, y_i), (e_1, s_1, y_1)).$$

This axiom says that multiplicity of identical literates is ineffceitve with respect to externality effect on illiterates. The additional literates do not exert any externality on the illiterate members.

If (A4'), (A5') and (A6') are always satisfied, then all literates are treated as identical - similar to that in **BF** and **Subramanian (2001)** (See proposition 2 and 5 below).

It is routine to check the following relations between Basu-Foster (BF) axioms and ours. We omit the proof as it is straightforward.

Proposition 1 :

(i) **BF** Normalisation $\Leftrightarrow$ (A2)

(ii) **BF** Additivity and Decomposability $\Leftrightarrow$ (A1)

(iii) **BF** Externality and Monotonicity $\Leftrightarrow$ (A3) and (A7)

Before we proceed to discuss our results in the next section,let us recall the salient measures of proximate literacy proposed in the literature. The first two are due to **BF** and are given by (using our notation)

$$P_\alpha = \frac{l + (N - l)\alpha}{N}$$

and their gender sensitive prescription

$$\begin{aligned} P_{\alpha_m, \alpha_f} &= \frac{l + \alpha_m(N - l)}{N} \quad \text{if there are no female literates in the household,} \\ &= \frac{l + \alpha_f(N - l)}{N} \quad \text{otherwise.} \end{aligned}$$

Here $0 < \alpha < 1$, and $0 < \alpha_m < \alpha_f < 1$.

The other important measure of effective literacy that we want to relate our findings to is by **Subramanian (2001)**, defined as

$$P_S = \frac{l + (N-l)\frac{l}{N}}{N}.$$

For a discussion of the properties of this measure, op. cit.

3 The Results

In the following propositions, we identify precise sets of conditions that characterise alternative forms of the effective literacy measure.

Proposition 2 : Under (A1) - (A3), (A4'), (A5'), (A6') and (A7), the form of the effective literacy measure becomes identical to the **BF** P_α formulation.

Proof : For any proof, we assume that $l > 0$ as otherwise the proof becomes trivial as $p(.) = 0$. In that case the household has no literates and the question of proximate literacy does not arise.

Under (A1-3), the form of the individual effective literacy measure becomes as given by (1). Now, if we impose (A5') and (A6') then the measure becomes independent of the sex and age information of the individuals concerned. So, we can redefine the effective literacy measure $p(N; (e_i, s_i, y_i), L_N)$ as

$$p'(N; 0, e_1, e_2, ..., e_l) \tag{2}$$

where l is the number of literates in the house hold. Now, if we invoke (A4') then the function will become independent of the level of education of each literate and only the information that they are literate will be important. So, effectively, the function will now become

$$p'(N; 0, 1, 1, ..., 1).$$

Now we finally use (A7) and, as a consequence, the function becomes independent of N and the number of literates, so we finally get the following form for the effective literacy measure,

$$p"(0, 1) = \alpha, \tag{3}$$

say, where $0 < \alpha < 1$ by (A2). Hence, using (1), the effective literacy status of the household finally becomes,

$$P = \frac{l + (N-l)\alpha}{N},$$

which is just the measure P_α. ■

Before we go on to characterise alternative measures of effective literacy, let us first demonstrate an interesting consequence of (A5) and (A7) which will be useful for proving our subsequent results.

Lemma 1 : Under (A4'), (A5(i)), (A6') and (A7) we have the following restriction on the individual externality function.

$$p"(f,m) = p"(m,\delta,m) = p"(m,\delta) = p"(f)$$

Under other combinations and (A7), the form of the externality function would be analogous.

Proposition 3 : Under (A1) - (A3), (A4'), (A5), (A6') and (A7), the form of the effective literacy measure becomes identical to the **BF** gender sensitive P_{α_m,α_f} formulation.

Proof : Again, we start with (1) under (A1-3) and invoke (A4') and (A6') to get, after a similar redefining, the effective literacy measure as

$$p'(N;(0,s_i),(1,s_1),(1,s_2),...,(1,s_l)). \tag{4}$$

Now, there are two possibilities.

(a) There are no female literates in the household. That is, $s_1 = s_2 = ... = s_l = m$. In that case, invoking (A7) on (4) reduces it to $p"((0,s_i),(1,m)$ which, under (A5), can be written as $\alpha_m(s_i)$ say.

Now, if we consider the axiom (A5(i)), then the externality effect becomes independent of s_i and hence can be taken as α_m.

(b) There is at least one female literate in the household. In that case, again invoking (A7) we can consider the case of one female and (possibly) one male literate only as the source of externality. Then using Lemma 1 to demonstrate the consequence of (A5(i)), (4) reduces to α_f which is, due to (A5(i)), $> \alpha_m$.

So, finally (1) becomes

$$\begin{aligned} P_{\alpha_m,\alpha_f} &= \frac{l+\alpha_m(N-l)}{N} \quad \text{if there are no female literates in the household,} \\ &= \frac{l+\alpha_f(N-l)}{N} \quad \text{otherwise.} \end{aligned}$$

This is evidently same as the Basu-Foster gender sensitive prescription. ■

We will now look at form of the effective literacy measure that arises when we consider effect of the education level of the literates.

Proposition 4 : Under (A1) - (A4), (A5'), (A6') and (A7), the form of the effective literacy measure becomes

$$P = \frac{l + (N-l)g_1(e_1, e_2, ..., e_l)}{N}, \tag{5}$$

for a suitably defined function g_1.

Proof : Under (A1-3), (A5') and (A6'), the individual externality effect for each illiterate takes the form (2), as already shown in Proposition 2. Now, this effect is identical for each illiterate, so aggregating, the form of (1) becomes

$$P(.) = \frac{1}{N}(l + (N-l)p'(N; 0, e_1, e_2, ..., e_l))$$

which can be rewritten as (5) for $g_1 : \bigcup_{l \geq 1} R_+^l \rightarrow [0, 1]$. ■

One can consider several interesting special cases of (5). Below, we illustrate with two of these.

Example 1 : Suppose we postulate $\frac{\partial g_1}{\partial e_j} = \beta$, say. This can be thought of as a special case of (A4) when the marginal contribution of education is constant across educational levels and across individuals. Then, by repeated integration[4], one obtains

$$g_1(.) = \beta \sum_{j=1}^{l} e_j.$$

Due to (A2), $0 < \beta \sum_{j=1}^{l} e_j < 1 \ \forall e_1, e_2, ..., e_l$. So, we get $0 < \beta lk < 1$. So, one may take $\beta = \frac{1}{kN}$ to get

$$P = \frac{l}{N} + \frac{(N-l)}{N} \frac{\sum_{j=1}^{l} e_j}{Nk}. \tag{5.1}$$

Example 2 : $g_1(.) = \frac{1}{k} e_{Max}$, where $e_{Max} = Max_{1 \leq j \leq l} \ e_j$. This is similar to (A7) when education is considered to be decomposable in some abstract sense. Only the maximum level of education matters. Then (1) reduces to

$$P = \frac{l}{N} + \frac{(N-l)}{N} \frac{e_{Max}}{k}. \tag{5.2}$$

The above two examples are very important in the sense that they are simply parametrized, easily computable and hence very useful for policy purposes. These two would be amenable to empirical exercises using real life data, for measuring the impact of literacy programmes and evaluating related policy.

We now discuss education level dependence more explicitly in the following proposition.

Proposition 5 : Under (A1) - (A3), (A4'), (A5') and (A6'), the form of the effective literacy measure becomes a generalisation of the Subramanian measure.

[4]This demonstration is similar to the proof of Lemma 3 in **Mukherjee (2001)**. We do not discuss it in detail here.

Proof : Proof is immediate, using (2) and invoking (A4') we get the desired form to be,

$$p'(N; 0, 1, 1, ..., 1).$$

Now, the only variable parts in the arguments' set are N and l, the number of "1"s. So, one can rewrite this as

$$g_2(N; l), \tag{6}$$

for $g_2 : Z_+^2 \to [0, 1]$; with $\frac{\partial g_2}{\partial l} \geq 0$, due to (A3). If we consider the particular case $g_2(N; l) = \frac{l}{N}$, the traditional measure of literacy, (1) simplifies to the **Subramanian (2001)** measure,

$$P_S = \frac{l + (N - l)\frac{l}{N}}{N}.$$

■

In the following proposition, we capture the effect of both sex and age ordering simultaneously and demonstrate the resulting form of the effective literacy measure. To do this in a very general manner, we consider the axiom (A5(ii)). That is, we allow the effects to depend on the sex of both the literate and the illiterate member.

Proposition 6 : Under (A1) - (A3), (A4'), (A5(ii)), (A6) and (A7), the form of the effective literacy measure becomes completely characterised by a 8 parameter family.

Proof : Under (A1-3) and (A4'), one can simplify and redefine the individual effective literacy measure as

$$p'(N; (s_i, y_i), (s_1, y_1), ..., (s_l, y_l)) \tag{7}$$

We now invoke (A6) on (7). Now, given (A6), we need only consider whether, for each literate j, $y_j > y_i$ or not. We denote this event by a variable I_J in the set of arguments of p' that takes a value of "1" when the condition hold and in case of the converse event it equals "-1". As before, we redefine p' and write it as

$$p"(N; (s_i, y_i), (s_1, I_1), ..., (s_l, I_l)) \tag{8}$$

Now we will have to consider the male and female illiterates separately as effects differ. Let's consider the case when $s_i = m$ first.

Invoking (A7) in a suitable manner (presence of literates who are identical with respect to the relevant characteristics), one can potentially have 15 (= 2^4 - 1) possible combinations of effects given by the presence of arbitrary non-empty subsets of $\{(m, 1), (f, 1), (m, -1), (f, -1)\}$ in the arguments of $p"$. But, using (A5(ii)), for a male illiterate, one can reduce the possibilities to the following set of 5.

$$\{(m,1), \quad (f,1), \quad (m,-1), \quad ((f,1),(m,-1)), \quad (f,-1)\}$$

Hence, the alternative effective literacy terms can be written and ranked as, after suitable redefinition, one of the following.

$$p^*_m(m,1) > p^*_m((f,1),(m,-1)) = max\{p^*_m(f,1), p^*_m(m,-1)\} \geq min\{p^*_m(f,1), p^*_m(m,-1)\} > p^*_m(f,-1).$$

Here $p^*_m(f,1)$ is the external effect of an older female literate on a male illiterate family member. The other terms are similarly defined.

One can argue analogously regarding the female illiterates. The corresponding terms for a female illiterate will be given by

$$p^*_f(f,1) > p^*_f((f,-1),(m,1)) = max\{p^*_f(m,1), p^*_f(f,-1)\} \geq min\{p^*_f(m,1), p^*_f(f,-1)\} > p^*_f(m,-1).$$

The measure will be completely characteried by these parameters. The actual form will depend on the presence of certain combinations in the set of literates and the sex of the illiterate member. The parameters can be ordered unambiguously if we impose more structure on comparability, for example being able to rank the externality effect of $(f,-1)$ and $(m,1)$ from a female illiterate's point of view.

Below, we illustrate the above proposition by an example.

Example 3: Consider the situation where
$p^*_f(f,1) = p^*_m(f,1) = \alpha_f$, $p^*_f(f,-1) = p^*_m(f,-1) = \beta\alpha_f$,
$p^*_f(m,1) = p^*_m(m,1) = \alpha_m$ and $p^*_f(m,-1) = p^*_m(m,-1) = \beta\alpha_m$.
Here, $0 < \beta < 1$ and $0 < \alpha_m < \alpha_f < 1$.

4 Conclusion

This note wants to pass on the message that, while measuring effective literacy, one should not only look at the literacy status of the members of the household but also look at the gender, age ranking and educational qualification of the literate members. Some extensions of **BF** measure have been developed in this paper; and all of them can be used for the measurement of effective literacy rate in a country like India using either Census or NSS data. **BF** measure with $\alpha = 1$, the traditional literacy rate, always appears as a special case of the class of measures suggested and characterized here.

In **BF**, α is exogenous and its value is arbitrarily chosen. So the literacy ranking of different regions (districts, states, countries etc.) may be different for different arbitrarily chosen values

of α. So choice of α is an important task in measuring the extent of effective literacy. We have suggested some relevant individual characteristics that may impact on the externality effect of a literate person on an illiterate member of the family. In particular, we have characterised certain classes of effective literacy measures which are generalisations of those proposed by **Basu and Foster (1998)** and **Subramanian (2001)** and other general class of measures.

These measures are useful in testing for the effect of the characteristics under discussion on effective literacy and hence for devising appropriate polcy in this direction. For example, if education level sensitivity is high, then higher education programme plays a significant role even in basic literacy improvement, in the proximate sense. Similarly, female literacy campaign should be more important for the purpose of generating higher externality if gender sensitivity is significant.

Although our proposed measures are very general in nature and allows many alternative possibilities, the final choice of a measure by a practisioner will depend on her subjective value judgement about the acceptability of alternative axioms set out in this paper.

References

1. Basu, K. and J.E. Foster, 1998, On Measuring Literacy; Economic Journal, 108 (November), 1733 - 49.

2. Basu, K., J.E. Foster and S. Subramanian, 1999, Isolated and Proximate Illiteracy and Why These Concepts Matter in Measuring Literacy and Designing Education Programme; Economic and Political Weekly, 35, 1 and 2,35 - 9.

3. Chaktavarty, S.R. and A. Majumdar, 2001, Issues in the Measurement of Literacy, mimeo, Indian Statistical Institute, Kolkata.

4. Mishra, S., 2001, Isolated and Proximate Illiteracy; Economic and Political Weekly, 36, 22, 2003 - 8.

5. Mukherjee, D., 2001, Measuring Multidimensional Deprivation; Mathematical Social Sciences, 2001, 42, 3, 233 - 51.

6. Subramanian, S., 2001, Equity, Efficiency and the Measurement of Literacy; Some Extensions of the Basu-Foster Framework; mimeo, Madras Institute of Development Studies, India.