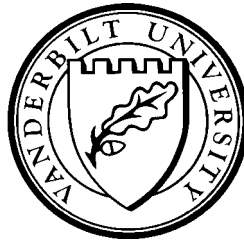# ISOLATED AND PROXIMATE ILLITERACY

AND WHY THESE CONCEPTS MATTER IN MEASURING LITERACY AND DESIGNING

EDUCATION PROGRAMMES

by

Kaushik Basu, James E. Foster and S. Subramanian

Working Paper No. 00-W02

January 2000

ISOLATED AND PROXIMATE ILLITERACY

And why these concepts matter in measuring literacy and designing education programmes

by

Kaushik Basu, James E. Foster and S. Subramanian

1. INTRODUCTION

Literacy is fundamental to many state-sponsored interventions in less developed countries because of its pervasive influence on economically relevant variables, such as productivity, health and earnings, quite apart from its intrinsic value as a vitally important goal of development. Traditionally, a society's literacy has been measured by the `literacy rate', that is, the percent (or, equivalently, fraction) of the adult population that is literate. The present paper maintains that the *distribution* of literates across households also matters, due to the external effects of literacy–the benefits that illiterate members of a household derive from having a literate person in the family. This is the argument used by Basu and Foster (1998), who suggest a new measure of literacy--the `effective literacy' rate–to which further amendments have been recently proposed by Subramanian (1999). The present paper reviews this argument, draws out its policy implications, and presents some suggestive data from Bangladesh to lend substance to the hypothesis that an illiterate belonging to a household with no literates is more deprived than an illiterate belonging to a household with at least one literate member.

2. LITERACY, EXTERNALITY AND MEASUREMENT

Economists use the term `externality' to convey the idea that the actions of an economic agent could sometimes have beneficial or harmful implications for other agents that are not reflected in market prices. A polluting industrial plant confers negative externalities on other agents (that is, agents who had no role in the decision to set up the plant), just as a public good like a highway confers positive externalities on other agents. The simple idea underlying this paper is that literacy, too, is something like a public good in that a literate agent confers a positive externality on the illiterate agents in the household by sharing the benefits of his or her literacy. This could happen for reasons of conscious altruism, unwitting munificence, osmosis or socio-cultural dispositions arising from group affiliation. Literate members of the same region, community, caste or family could be expected to positively affect the literacy status of their respective cohorts. The unit of aggregation within which such external effects of literacy might be expected to be most salient is that of the household, which is where we shall concentrate our attention in this paper (although generalizing to other groups is straightforward).

There are various contexts in which the intra-household externality can arise. The government circulates an order intimating the availability of social assistance to physically handicapped people, widows and accident victims. Agricultural extension workers disseminate printed information on new technology relating to irrigation and high-yielding crop varieties. Leaflets are distributed by a non-governmental voluntary agency advising rural people of their specific rights to information. The village moneylender doctors the statements of his borrowers' liabilities to his own advantage. The public health office puts out a simple printed bulletin on the advantages of oral rehydration. In every one of these cases, an illiterate person is poorly placed in the matter of availing himself of useful information or resisting misinformation. The problem is, in all likelihood, more acute if the illiterate person in question happens to belong to a household with no literate members. In a general way, an illiterate person's ability to transform various kinds of informational inputs into what Amartya Sen (1985) has called functionings is tied to the literacy status of the *household* to which he or she belongs. Indeed, apart from the sorts of `direct' effects we have illustrated, it could also be true that (a) the respect and consideration with which an illiterate is treated by society at large is an increasing function of the presence of literate members in his household; and (b) the person's own sense of advantage and self-respect is often mediated by the literacy status of the household to which the person belongs (Basu, 1989). In the broadest possible sense, the intra-household externality from literacy contributes to an expansion of each illiterate member's capability and welfare.

What implications could the intra-household externality arising from literacy have for the measurement of literacy? Consider a society consisting of M households, N persons and L literates. The number of illiterates in the society is clearly N-L. Now any given illiterate belongs either to (a) a household which has one or more literate members or (b) a household without any literate member. Following Basu and Foster (1998), call the first type of illiterate a *proximate illiterate* (to suggest his or her proximity, from an intra-household perspective, to a literate person), and the second type of illiterate an *isolated illiterate*. Of the N-L illiterates in the society under review, let *P* be the number of proximate illiterates and *I* the number of isolated illiterates.

The measure of literacy most widely employed is the familiar literacy rate *r*, which is the proportion of the adult population that is literate:

$$(1) \qquad r \ = \ L/N.$$

Underlying the measure *r* is the notion of a 0-1 classification: each literate person counts for one, and each illiterate person – irrespective of whether the person is a proximate or an isolated illiterate – counts for zero. By contrast, our approach to measuring literacy allows for a differentiation in the literacy status of an illiterate according to whether the person is a proximate or an isolated illiterate. In particular, we propose that each proximate illiterate person count for 'e' literate persons, where e is a number greater than zero and less than one. The idea is to suggest that there is a positive externality conferred by the presence of a literate person on the illiterate members of the household such that, in `literacy-equivalent' terms, each proximate illiterate enjoys a status which lies somewhere between that of complete literacy and that of complete illiteracy. The resulting `effective literacy rate', taking due account of the number of literate persons in `literacy-equivalent' terms, is given by

$$(2) \qquad r^* \ = \ (L+eP)/N.$$

Letting p stand for the rate *P/N* of proximate illiterates, from (1) and (2) we obtain:

$$(3) \qquad r^* \ = \ r + ep.$$

In other words, the effective literacy rate is the usual literacy rate plus e times the rate of proximate illiterates.

Basu and Foster (1998) propose r* as a way of capturing the externality from literacy and provide a set of axioms which exactly characterizes *r**. As far as e is concerned, they assert that it lies somewhere between 0 and 1, arguing that the exact value of e can be determined only through empirical estimation. Of course, the potential usefulness of the effective literacy approach depends on the hypothesis that there are advantages to being a proximate illiterate, and hence that e is strictly larger than 0. A recent study by John Gibson (1998), using literacy data from Papua New Guinea, offers a preliminary confirmation of this conjecture for certain anthropomorphic indicators of well-being. In section 4, we obtain further corroboration using household survey data from Bangladesh.

The advantage of r* over r is easily illustrated by considering an hypothetical example. Suppose there are M = 50 households, each having two members, so that overall population size is N = 100. Further, assume the number of literate persons to be L = 60, and let the externality from literacy within the household be e = 0.3. We consider two alternative regimes. Under regime A, all illiterates are isolated, so that P = 0 and I = 40; under B, all illiterates have a literate partner so that P = 40 and I = 0. In regime A the households are polarized into two subsets, while in regime B each household has at least one literate

member. Let r(A) and r(B) be the `standard' measures of literacy for regimes A and B respectively, and let r*(A) and r*(B) be the `effective' literacy measures for the two regimes. Then, it is easy to see, given (1) and (3), that r(A) = r(B) = 0.6, while r*(A) = 0.6 and r*(B) = 0.72. The measure r* will certify regime B to be literacy-wise superior to A, while the measure r will declare the two regimes to be literacy-wise equivalent. It is the ability of r* to discriminate between alternative distributions of literates across households which renders it more informative than r.

Notice that from (1) and (3) (and also from the example just discussed), the effective literacy rate r* is always greater than or equal to r. Is this to be construed as implying that the level of literacy is in general higher than what r would have us believe, or that we ought to take a greater satisfaction from the higher rate r*? Indeed, one implication of this approach is that certain people have greater access to literacy than is generally acknowledged, and this should be incorporated into the assessment of a population's literacy level. As for satisfaction with this observation, we would advise against jumping to any rash conclusions. An increase in 5.9 percentage points in going from r to r* (as is the case with Arunachal Pradesh in our second empirical illustration below) is a good deal less satisfying when we realize that it is among the smallest of such increases, and results in the lowest r* value of all. If we return to the above example where regimes A and B share r levels, the key comparison is between r*(A) and r*(B) and not between, say, r(B) and r*(B). The measure r* concludes that regime A has less effective literacy than B, and this is because A, in contradistinction to B, has a less even distribution of literates across households and consequently fewer proximate illiterates.

3. EMPIRICAL ILLUSTRATIONS FROM BANGLADESH AND INDIA

First, we provide some suggestive evidence drawn from Bangladesh data, to support the hypothesis that an intra-household externality from literacy not only exists, but is substantial. Our data are derived from the **Household Expenditure Survey for Bangladesh, 1995-96**, which covers a sample of 7,420 households. While these data are currently in the process of formal econometric analysis (Kaushik Basu, Ambar Narayan and Martin Ravallion, 1999), our more-restricted concern here is to look for certain leads which have a bearing on the externality hypothesis underlying this paper. *With this in mind, we focus on the sub-sample of 3002 individuals who live in households with one or no literate member (to rule out

cases where there is more than one literate in the household). The individuals are then divided into three categories: literates (298 observations), proximate illiterates (645 observations) and isolated illiterates (2059 observations). For each category of individuals we regress their earnings against their age in quadratic form. That is, letting Y stand for earnings and A for age, we estimate the regression equation $Y = a + bA + cA^2$. The quadratic curve is fitted for the following sub-groups: (I) all earners, (II) earners living in urban areas and (III) earners living in rural areas. Within each of these sub-groups, separate curves are fitted for literate, proximate illiterate and isolated illiterate earners. Thus the curves represent the predicted earnings corresponding to the age of the individual for each sub-group. As is expected, we obtain inverted U-shaped curves for each of these categories.

What is of interest in the present context is the heights of the curves for the three categories of individuals. As Figure 1 indicates, the fitted curve for literates lies everywhere above that for proximate illiterates which, in turn, lies everywhere above that for isolated illiterates. While recognizing that income is just one of a number of different possible indicators of advantage, the pattern revealed by Figure 1 does suggest that literates fare better than proximate illiterates who, in turn, fare better than isolated illiterates– which is consistent with the notion of an intra-household externality from literacy. Figures 2 and 3 do the same exercise but with the sample broken up into urban and rural earners. The qualitative results remain unchanged, though it is evident that the advantages of literacy (both direct and through externality) are smaller for those living in the rural sector.

Further corroboration is available from an elementary exercise entailing the calculation of group averages. Table 1(a) indicates that the average earnings of individuals are pronouncedly higher for the sample of literates than for the sample of proximate illiterates, while the average for proximate illiterates is (in relative terms, more modestly) higher than the average for isolated illiterates. Pretty much the same picture emerges when the exercise is repeated separately for the rural and the urban populations (see Tables 1b and 1c respectively). While this is surely an elementary exercise, it does suggest the presence of an intra-household externality from literacy. Moreover, income is not the only measure of well-being likely to be enhanced by literacy. It should be possible to track this externality by studying other indicators, such as diminished morbidity or improved nutritional status, as Gibson (1999) has done.

Next, we look at some illustrative data from secondary sources on literacy in India, in order to shed some light on the measurement concerns reviewed earlier. Using 1981 Census data, Basu and Foster have

provided state-wise data on literacy in the Indian Union. For each state, they have furnished estimates of the crucial literacy-related quantities r, p, i and (for differing assumed values of e) r*. Using the same data set (which is regrettably not amenable to updating since comparable data from the 1991 Census are not yet available), we present a similar profile of literacy for India and its states (see Table 2). For the purposes of this exercise, we quantify the value of e in a particularly simple-minded way. From the Bangladesh sample data we have reviewed earlier (see Table 1a), it is clear that the average earnings for the sample of literates (call it Y_L) is Rs.4569, while the corresponding averages for proximate illiterates and isolated illiterates (call these Y_P and Y_I respectively) are Rs.3512 and Rs.3240 respectively. To a first order of approximation, e could plausibly be given by the quantity $(Y_P-Y_I)/(Y_L-Y_I)$, which, for the particular set of values these variables assume for the Bangladesh sample, works out to roughly 0.21. It is as well to re-emphasize that no particular sanctity is attached to the valuation of e we have resorted to: the object of the exercise is simply to provide an illustrative example of our measurement concerns, while aiming for a degree of specificity in so doing. Another possibility is to consider several values of e, as Basu and Foster have done, and to do a sensitivity analysis.

Table 2 of this paper furnishes state-wise information on the values of the measures r, p, i and r* (for e = 0.21). The following observations can be made from the data presented in Table 2:

(a) While the ranking of states according to r is in general very similar to the ranking according to r*, there are a few instances of rank-reversal one can observe, such as with the pairs (West Bengal, Punjab), (Gujarat, Himachal Pradesh), (Nagaland, Manipur) and (Meghalaya, Sikkim). The possibility of rank-reversal in specific pairwise comparisons thus suggests that there is an operational sense in which the information conveyed by the measure r* can be different to the information conveyed by r.

(b) As far as the inter-regional picture is concerned, the data presented in Table 1 confirm the stereotyped impressions one has about the regional distribution of literacy achievement in India: whether literacy is measured by r or r*, Kerala leads all the other states, while Bihar, Madhya Pradesh, Rajasthan and U.P. together with Arunachal Pradesh bring up the rear.

(c) The inter-state variability in literacy attainment turns out to be lower when literacy is measured by r* than when it is measured by r: the squared coefficient of variation for the r* series, at 0.055, is about 70 per cent of the squared coefficient of variation, at 0.078, for the r series.

(d) The state-wide variability in the indicator p is a matter of some interest. While the three states with the highest values of r–Kerala, Mizoram and Goa–have, somewhat naturally, the lowest values of p, there is no obvious relationship between r and p for the remaining states. Thus, while both Gujarat and Himachal Pradesh have r-values in excess of the all-India average, Gujarat's p-value falls short of, and H.P.'s p-value exceeds, the corresponding all-India p-value. While Rajasthan and Arunachal Pradesh both have r-values lower than the all-India figure, Rajasthan's p-value exceeds, and Arunachal Pradesh's p-value falls short of, the corresponding p-value for India. Identification of the source of the variability in p across states would require deeper investigation. A speculative explanation may point to factors such as state-wise differences in the extent to which literacy is related to caste or gender (see Basu and Foster, 1998 for some discussion), in the structure of families (joint/nuclear), and in the age-structure of literacy.

4. INTRA-HOUSEHOLD EXTERNALITY AND THE DISTRIBUTION OF LITERACY ACROSS HOUSEHOLDS

As we have seen in Section 3, the measure r* penalizes literacy regimes in which the inter-household distribution of literates is concentrated *vis a vis* those regimes in which the inter-household distribution of literates is more evenly spread. The most `efficient' inter-household distribution of literates, i.e. the distribution best calculated to reap the advantage of the intra-household externality from literacy, is also the one in which literacy is spread equitably across the households. In this particular context, therefore, the goals of efficiency and equity are perfectly congruent.

This has important policy implications. Consider the problem of devising and administering literacy campaigns and programmes. Specifically, suppose that a village-level literacy programme aims at conferring literacy on a certain number of persons, say K, in the reference population. Evaluation of the programme's outcome, when the latter is measured in terms of the index r, would be compatible with rendering literate any set of K individuals. In contrast, evaluation in terms of the measure r* would afford an incentive for targeting, whereby one person from each of K households is rendered literate. Briefly, r* points the way to a stratagem of policy planning which is mediated by an instrumental justification for inter-household equality, in a way that the measure r does not. In a more general context, one could also extend the sorts of considerations dealt with earlier to other possible groupings, wherein it is not the

household but caste or ethnicity or sector of origin which serves as the classificatory criterion for partitioning the population. The precise choice of grouping one resorts to would have to be informed by a reasonably clear understanding of the social reality of cohesion, i.e. by a reasonable presumption that the reference groups are sufficiently close-knit for the intra-group externality from literacy to play a definitive role. Failing this, the role of externality – which is eventually the driving force of the present analysis—would become very attenuated, and the underlying motivation for the exercise would be poorly reflected in its application.

5. INTRA-HOUSEHOLD EXTERNALITY AND GENDER

In the preceding sections we have worked with a simple model of the intra-household externality from literacy – one in which the magnitude of the externality, e, is taken to be independent of and invariant with respect to all household characteristics. The real picture is likely to be a good deal more complex. The value of e, for example, is likely to be influenced by whether the household in question is one whose history of literacy is of recent or considerable vintage. Similarly, the `literacy coefficient' for a household – defined as the proportion of literates in the household – might be expected to affect the value of e: typically, the higher the literacy coefficient, the higher is e likely to be (this complication is dealt with in Subramanian, 1999). A third and crucially important household characteristic has to do with the gender of the literate member(s) of a household (this issue has been analyzed in Basu and Foster, 1998): at the margin, the external effects from literacy are likely to be larger if the source of the externality is a female rather than a male.

The crucial role of female literacy in expanding the capability status of a household, and in particular, the superiority of female literacy over male literacy, is revealed, with reference to one specific dimension, in certain statistics relating to infant mortality among households classified by the educational standing of the mother and the father. In general, the stronger positive externality of the mother's literacy over the father's literacy has been well-established (see Caldwell, 1979).

The status of female literacy in India, as it actually obtains on the ground, is a poor advertisement for its potential importance in the scheme of things. Table 3 is based on a part of Table 2 in Majumdar and Subramanian (1998), wherein a group-wise disaggregated analysis of adult literacy is undertaken. The table

presents information on the literacy rates for eight subgroups of the population, labeled (L,R,F), (L,R,M), (L,U,F), (L,U,M), (H,R,F), (H,R,M), (H,U,F) and (H,U,M) respectively, where `L' stands for `"Low" Caste' (Scheduled Castes and Tribes), `H' stands for '"High" Caste' (Non-Scheduled Castes and Scheduled Tribes), `R' stands for `Rural', `U' for `Urban', `F' for `Female' and `M, for `Male'. From these eight subgroups we can generate four pairs of `gender-variants', namely four pairs of groups such that the groups in each pair differ only with respect to gender, holding caste and sector of origin constant. Table 3 brings out clearly the systematically inferior status experienced by women: in every pair of gender-variants, the standard literacy rate of the group containing females is lower than that of the group containing males. Basu and Foster (1998) also provide state-wise evidence of gender-disparity in the attainment of literacy.

A recognition of the importance of female literacy, and its incorporation into the measurement of overall literacy, would lead to the following sorts of considerations. First, define P_f and P_m to be the numbers, respectively, of `female-proximate' and `male-proximate' illiterates: following Basu and Foster, a female- (respectively, male-) proximate illiterate is an illiterate who belongs to a household with at least one female literate (respectively, a household with at least one male literate and no female literate). Let $p_f := P_f/N$ and $p_m := P_m/N$ be, respectively, the female-proximate and male-proximate headcount ratios of illiteracy. Let e_f (respectively, e_m) be the magnitude of externality conferred on a female-proximate (respectively, male-proximate) illiterate. In line with earlier considerations, it would be reasonable to postulate that $0 < e_m < e_f < 1$. Precisely analogously to the way in which r* has been derived in (3), one can now present a measure of literacy, r**, which emphasizes the gender-differentiation in the intra-household source of externality:

(4) $r** = r + e_f p_f + e_m p_m$.

Quite apart from the distributional considerations spelt out in this paper, the measure r** emphasizes the need to lay special emphasis on female literacy..

7. CONCLUDING OBSERVATIONS

This paper has been concerned to advance one simple idea – namely that in assessing the literacy status of a society it is important to reckon the intra-household externality arising from literacy. Allowing for such externality has implications for the measurement, for the inter-household distribution, and for the

gender dimension of literacy. While these implications have been spelt out in the main text of the paper, an important message which emerges is that in order best to take advantage of the intra-household externality from literacy, a special effort would have to be made in ensuring an equitable distribution of literacy across households, and in especially promoting female literacy.

References

Basu, Kaushik (1989), "A Theory of Association', <u>Oxford Economic Papers</u>, vol. 38.

Basu, Kaushik and Foster, James E. (1998), 'On Measuring Literacy', <u>Economic Journal</u>, vol. 108, November.

Basu, Kaushik, Narayan, Ambar and Ravallion, Martin (1999), 'Is Knowledge Shared within Households?', mimeo: The World Bank.

Caldwell, John (1979), 'Education as a Factor in Mortality Decline: An Examination of Nigerian Data', <u>Population Studies</u>, Vol. 33.

Gibson, John (1999), 'Literacy and Intrahousehold Externalities', mimeo: University of Waikato (forthcoming <u>World Development</u>).

Majumdar, Manabi and Subramanian, S. (1998), 'Capability Failure and Group Disparity: Measurement and Application', Harvard Center for Population and Development Studies Working Paper 98.2.

Sen, Amartya (1985), <u>Commodities and Capabilities</u>, Amsterdam: North-Holland.

Subramanian, S. (1999), 'Basu and Foster on 'Measuring Literacy': Some Very Simple Further Considerations', mimeo: Madras Institute of Development Studies.