Asia Pacific Journal of Educational Development 2:2 (December 2013): 39-55
DOI: 10.6228/APJED.02.02.04

Proximate, Isolate, and Raw Illiteracy Some Issues in Method and Measurement with Application to a State-Level Study from India

Atanu Sengupta[*] *and Tusher Mukherjee*
Economics Department, Burdwan University

Abstract

Literacy is an important concept in the development discourse. Unfortunately, this concept is still surrounded with mazes of ambiguity. For example, in addition to determining the number of adult literates, it may also be necessary to determine their influence on their families (Basu & Foster, 1998). Living in close proximity to literate persons will undoubtedly benefit an illiterate person. Most studies want to extend the simple literacy rate to include these externalities (Basu & Foster, 1998; Basu & Lee, 2008; Kell, 2008; Maddox, 2007; Mukherjee & Gupta, 2003; Subramanian, 2004, among others). The standard problem in these approaches is the specification of the externality parameters. Depending upon on the specified values, the resulting literacy indices vary considerably. In this paper, we have attempted to remove this drawback by introducing data-driven weights. Further, we consider some aspects of the dynamics of the literacy rate and its extended components. Our results depict some interesting dynamic features. Further, the data-driven weights bring less dramatic changes in the literacy rates.

Ever since independence, Indian policy makers have emphasized the attainment of a decent standard of living, its prime component being literacy. However most of the available studies on India's achievement on this front are one-dimensional. In recent years some other parameters of educational attainment (such as years of schooling) have been used to supplement this uni-dimensional measure. This new approach to literacy brings new depth to the assessment of literacy itself. Literacy is not merely a matter of how many (or what proportion of) people are literate, but where they are. It makes a lot of difference if the literate persons are well distributed across the families, rather than being concentrated in a few pockets. A person in a family with no literates is an "isolated literate," while those in a family with some literates are "proximate literates."

Keywords: literacy rate, externality, PCA, proximate and isolated illiteracy, two-dimensional dynamics

1 Contextualization of the Distributional Issues

In an interesting paper, Basu and Foster (1998) argue that the simple literacy rate cannot capture the effect of externality that it generates. Living in close proximity to literate persons will undoubtedly benefit an illiterate person. Of particular concern is the literacy status of the family in which an illiterate person dwells. A sharp distinction is made between proximate illiterates and isolated illiterates. The former refers to those illiterate persons who reside in a family with at least one literate person. The isolated illiterates are persons in a family with no literates. Effective literacy rates should take these externalities into considerations.

Thus Basu and Foster (1998) introduced a new component to literacy analysis. Three questions immediately emanate from such an analysis:

(1) An analysis into the dynamics of the new component?
(2) Implications of the new component in the dynamics of the traditional (raw) literacy rate?
(3) Incorporation of the new component into the traditional literacy framework?

Since then, a number of studies have been developed covering these issues. First, we considered the dynamics of this new component -- Isolated illiteracy. These dynamics have generally been neglected in the existing literature. Our data spanning over three points in time gives us ample scope for such a study.

Most such studies were, however, involved in the incorporation issue. They wanted to extend the simple literacy rate to include these externalities. The first attempt was by Basu and Foster (1998) themselves, whereby certain weights were granted for proximate illiterates as opposed to isolated illiterates. In an interesting development, Subramanian (2004) argues that such an approach pushes up the literacy rate without really changing anything. This may give misleading comfort to the policy makers by enhancing the literacy rate without any change in the reality. Subsequently, he posits a fine for isolated illiteracy and constrains the new measure below the simple rate. Kell (2008), on the other hand, describes literacy as a "distributed

* Corresponding Author: Dr. Atanu Sengupta (sengupta_atanu@yahoo.com)
Reader, Economics Department
Burdwan University

capacity" among persons who are not traditionally literate. Maddox (2007) in his anthropological study emphasizes other associations besides family that may be considered in defining proximate illiteracy.

Recently, Basu and Lee (2008) tried to make the concept flexible enough to capture non-family associations. However, a major lacuna of these studies is that they specify the proposed weights rather *a priori*. This neglect is strange given the huge policy implications that externally weighted, literacy generate.

There are a few studies stressing other family characteristics that may determine this ally weighted literacy (Mukherjee & Gupta, 2003). However, even this study uses *a priori* weights in its empirical exposition. Almost no attempt is made to utilize the data to generate the external weights, even though there is enough information regarding the factors that the authors themselves identified as important determinants.

The first study to develop the data-driven externality parameter (a) was that of Gibson (2001). From the logic of Basu and Foster (1998) it follows that the externality in literacy rate will be felt in terms of some measurable variables. For example the rate of adopting agricultural innovations will depend both on the number of literates and the influence they have on illiterate persons. In a society where there is no literacy, agricultural innovations may be difficult to adopt. However, if some members of the society attain literacy, then they can certainly help others adopt the new technology.

There are numerous studies incorporating the impact of family externality on literacy (Basu, Narayan, & Ravallion, 1999; Iversen & Palmer-Jones, 2008, etc.). Some have made effective literary a function of family literacy (Dutta, 2004; Sengupta, Sengupta, & Ghosh, 2004; Valenti, 2002). These measures are useful. But they need strong data, a requirement that may be beyond the capacity of most national-level macro surveys in underdeveloped economies.

For the empirical estimate, Gibson (2001) used the data of Papua New Guinea to generate such external effects. However, he argued that the value of a will depend on the choice of the dependent variables. For example, a is unlikely to be high if we consider individual returns, while it is higher if we take into account some aspects of family welfare (such as children's health).

Gibson's method, though noble, largely depends on the choice of variables. Moreover, it might not be easy to segregate between individual returns and collective welfare. It is also not advisable that the effective literacy measure depends on such case-specific parameters.

Sengupta, Sengupta, and Ghosh (2008) also used data-driven weights to generate the level of learning index depending on a set of family features. However, they did not use it to derive the extended literacy measures. As a practical user of these indices, the situation is highly unsatisfactory.

Another neglected issue in this entire discussion is that of dimensionality (the implication issue). The approach of Basu and Foster (1998) is novel not only because its gives us a better measure of literacy, but also extends the concept of literacy from one-dimension to a two-dimensional framework. Literacy does not merely mean the average proportion of persons who are literate, but also means how they are distributed among various families. Hence the dynamics of literacy becomes complex. It implies not merely a change in the magnitude of literacy proportion, but also a change in the inter-family distribution of literacy. This second aspect is completely neglected in the standard debate, where the emphasis is only on fathoming a "better" aggregative measure, rather than the simple literacy rate. In the present paper, we take up this neglected issue in analyzing the multidimensional dynamics of literacy.

In this paper we try to correct these shortcomings. This paper is divided into four sections. In the next section, we give a brief review of the existing literature on externally adjusted literacy rate and the related analytical questions. Section 3 gives the empirical results for the new types of "effective literacy rate" suggested by us. It also gives the multi-layered dynamics. The conclusion is provided in Section 4.

2 Technical Underpinnings

2.1 Externality Adjusted Literacy Rate – The Major Controversies

It was traditional to measure the adult literacy rate simply by determining the percentage of adult population who are literate. By "literacy" we mean the ability to read and write. It has recently been argued (Basu & Foster, 1998) that the impact of literacy can not be entirely captured by the simple measurement of literacy rate. The argument is that literacy has positive effects that overflow its measured dimension. In its stead, Basu and Foster (1998) introduced the concept of effective literacy (p). They distinguish between illiterate persons living in a family with at least one literate person and those where there is none. The former are referred to as "proximate illiterate," and the later as "isolated illiterate." The proximate illiterate enjoys some of the advantages of literacy. Hence, with a proper weighting, they should be included under the extended concept of literacy. Thus Basu and Foster use positive externality of literacy -- The benefit that an illiterate person gets by being in close proximity to a literate person. However, many dramatic changes in the ordinal ranking of units occur when we move from ordinary to effective literacy.

The raw measure of literacy is:

$$R = r/n \tag{1}$$

Where r is the total number of literate persons in a society, and n is the total number of adults in the society. This measure is one-dimensional. There is no degree of independence. R contains all information about literacy. Hence, the policy target should be exclusively on R itself.

In the new approach, we define proximate illiteracy rate as:

$$R^p = r^p/n \tag{2}$$

where
r^p = total proximate adult illiterates.

The externality attached literacy rate (R^E) is:

$$R^E = f(R, R^p) \tag{3}$$

Now the debate is on the exact form of the function f (…). The first of such a function form is given by Basu and Foster themselves (BF-1998). The BF measure of effective literacy is defined as:[1]

$$R_{BF} = R + aP \tag{4}$$

where P is the proportion of proximate illiterates.

However, there is a group of authors (Basu & Lee, 2008; Subramanian, 2008) who question the magnitude validity of the new measure with the argument that it is greater in dimension than ordinary literacy. This may give a false impression to policy makers, since it significantly increases the magnitude of literacy without anything happening to literacy. To quote Basu and Lee (2008), "as will be evident later, it is not clear why the normalization of our new measure should be such that effective literacy is higher than the standard literacy rate."

These considerations have given rise to "constrained externally induced literacy" (R^S) (Subramanian, 2004, 2008). The constrained effective literacy rate (R_S) according to Subramanian (2008) is:

$$R_S = R(1 - I) = R - RI \tag{5}$$

where I is the proportion of isolate literates.

These results penalize the region by deducting the proportion of isolated illiteracy. A region may have a high R, but a small R^* compared to another region if such comparison involves the weightings of isolated illiteracy.

[1] This measure satisfies a number of axioms proposed by the authors.

In a recent development, Basu and Lee (2008) generalized the externality in literacy from families to other social networks in which an individual may be webbed into. In this new concept the population size, number of literates, and number of isolated illiterates are the three constituent parts. Basu and Lee derived a mapping from the three-dimensional analogy to the two-dimensional fields. This permitted them to use a new measure of effective literacy (R_{BL}) that is significantly different from that used by Subramanian (2004).

Basu and Lee (2008)[2] constrained the effective literacy given by

$$\begin{aligned} R_{BL} &= \frac{(1-\alpha)R}{\{(1-\alpha)+\alpha I\}} \\ &= \frac{R}{\left\{1+\left(\frac{\alpha}{1-\alpha}\right)I\right\}} \qquad (\alpha \neq I) \end{aligned} \tag{6}$$

2.2 Data-Driven Weights and Externality in Literacy

In the standard exercises the parameter α is specified in an *a priori* manner. Various authors have used different values of literacy. These values produce indices which differ widely from the raw literacy rate. A lot of controversies have been generated as a result. It is our contention that much heat could have been avoided if the alternative values of α were based on some objective rational premises. In this regard the data-driven weights that are used in indexing in the social sciences could be helpful. Principal Component Analysis (PCA) has been used by many researchers to generate the proportional weights of various factors in constructing an overall index. The PCA technique may provide a means by which the debate on literacy can be amicably settled. In this paper, we turn to PCA to generate the standard weights that are used to measure the rate of externally induced literacy.

PCA is widely used to generate the weights that are necessary in the construction of several socio-economic indices (Johnson & Wichern, 2001). There are numerous instances where these weights give a better index than the alternative techniques. Algebraically, principal components are the linear combination of a set of random variables which are constructed according to a certain rule. Put simply, the rule is to choose the weight so that the variances are maximized, subject to certain constraints on the covariance and the weights themselves. It is possible to get a number of principal components from the available data set. Each PCA generates the weight by maximizing the corresponding variances, while constraining the relevant

[2] In fact, there are a large number of such measures, of which we here mention only a few.

covariances and the linear weights themselves. Empirically speaking, a set of eigen values are generated, from which weights are to be calculated. “The weights for each principal component are given by the eigenvectors of the correlation matrix, or if the original data were standardized, the co-variance matrix” (Vyas & Kumaranayake, 2006).[3] The calculation of weights becomes troublesome if some of them turn out to be negative (Vyas & Kumaranayake, 2006).[4]

Thus, in our case (4) can be written in a slightly different way

$$R^{*}_{BF} = R + w_1 P \qquad (4^{*})$$

Where w_1 is the weight generated by the PCA between R and P.

In our case we changed (5) slightly and weighted it in the form of

$$R^{*}_{S} = R - w_2 I \qquad (5^{*})$$

where W_2 is generated by the PCA of R and I.

In our case the slightly different form of effective literacy rate (6) is

$$R_{BL}^{\ *} = \frac{R}{(1 + W_3 I)} \qquad (6^{*})$$

where W_3 is a weight generated by the PCA of R and I.

In our paper we generate all the four measures of literacy (R, R_{BF}, R_S, R_{BL}) using the specification as suggested by the respective authors. However, in each case the externality weight is generated from the data rather than being fixed *a priori*. We used PCA to derive the weights in each case. Our measure is better than the traditional approaches, where it is specified abruptly. It is also better than that used by Gibson, in that it does not vary depending on the choice of a particular variable. We next turn to the issue of dimensionality and dynamics.

2.3 Dynamics of the Basic Indicators-Issue of Dimensionality

In the literature on education externality, the dimensionality and dynamics issues are neglected and pushed into background. However, they are of paramount importance. In order to understand the ground realities with respect to literacy, the movement of the literacy component over time is omnipotent. In the traditional approach, the focus is on the dynamics of simple literacy rate. However, under the new approach, this would be a wrong way to think about literacy.

Consider a country which is segmented into two population subgroups with respect to the literacy, one with a high level of literacy, and other with an abysmally low level. Now suppose over a period of time, the literacy rate of the endowed segment rises, while that of the deprived section fails to register much improvement. It is certain that the overall literacy rate of the country will rise. However, there is now a question of deprivation. Even though the second section has registered some meagre improvements in welfare, they might still feel that they have been “somewhat left behind” in the process of development. If we visualize social welfare as derived from individual welfare and not as an abstract concept (Sen, 1970), then this should have serious implications for development policies. In fact, many ethnic disturbances in India and elsewhere have been built up on such concepts of deprivation. In fact, the sense of deprivation may increase for those who are left behind. Hence the simple dynamics gives a distorted view of literacy gains.

Thus, in order to understand the dynamics properly, the movement of the literacy rate (R) should be corroborated with the dynamics of proportion of isolated illiteracy (I). Such a multi-level analysis will surely help us to understand the situation better. The externality debate raised by Basu and Foster has added an extra space to the conventional literature. We first turn to this multi-level analysis.

However, to understand the complexity of the dynamics of literacy, we should juxtapose the various dimensions of literacy as presented in the recent discussion. There are two dimensions that are very important -- The raw literacy rate (R) and the percentage of isolated literates (I). In order to bring these two dimensions within a single framework, we utilize the standard technique of comparison developed by Ranis and Stewart (2000, 2001).

A cross-sectional unit can have various combinations of R and I. Literacy has two dimensions -- Magnitude and spread. While the simple literacy rate measures the magnitude, the extent is measured by isolated illiteracy. Following Ranis and Stewart (2000, 2001), we can compare a unit’s performance in both R and I with respect to the average performance in these dimensions. If a unit has R above the average level, we denote it as “high R.” Similarly for I. In short, we can consider following four scenarios in respect to R and I.

(1) Virtuous -- High R, Low I (both attainment and distribution satisfactory).
(2) Lopsided R -- High R, High I (satisfactory attainment, unsatisfactory distribution).

[3] For a detailed mathematical derivation, one can consider Johnson and Wichern (2006). For socio-economic studies a good reference is Vyas and Kumaranayake (2006).

[4] In our exercise all the weights were positive.

(3) Lopsided I -- Low R Low I (attainment unsatisfactory, but distribution satisfactory).
(4) Vicious -- Low R High I (both attainment and distribution unsatisfactory).

The first situation is virtuous because the benefit of literacy has reached almost everybody and the unit is now on the low literacy path. The last case is vicious since here the attainment of literacy is high with a low excluded section. Thus there is very little awareness of the benefit of high literacy. The lopsided R and lopsided I are the medium cases. For lopsided R there is the possibility of social tension arising from the sense of deprivation -- The possibility of living in poverty around plenty. The lopsided I is very similar to the case of ancient egalitarian societies where whatever little there is, is divided among all, thus benefiting nobody disproportionately.

3 Data -- Empirical and Estimated

3.1 Measurement Issues

We have calculated the different measures of literacy with *a priori* weights as well as with PCA adjusted weights for three different sets of data. The first is the Census Report of 1981. This is an exhaustive report covering all the major states of India. The data was first used by Basu and Foster (1998).

For a more recent explanation, most authors have used NSSO data. This may be prompted by the availability of data in the published NSSO report. Mukherjee and Gupta (2003) used the NSSO 43rd round (1993-1994).[5] Further they used a different measure of adult literacy that is different from the standard.[6] We have also used NSSO 64th round unit level (2007-2008) data for our analysis. The data set allows us to estimate "raw" and externally adjusted literacy rates for different states of India. In this paper, we have taken the standard literacy rate (persons aged 7 years and above).

3.2 Data-Driven and the Traditional Measure of Literacy

For the reasons presented above, we have used the data-driven weights to generate the suggested literacy measures. We compare our measures with those obtained by the traditional indicators with the different weight structure. It is evident that our measures of data-driven R^* are well below the exogenous R for all the states for the two alternative specifications of α. Again, a main concern of the debate between the raw literacy rate (R) and the exogenous literacy rates was that the later unduly influenced the value of literacy, and that this might have negative policy implications. In fact, this was a main reason for the researchers adopting more sophisticated measures that are different from the exogenous R. For a data-driven measure this deficiency is much less. Hence the selection of weight structure plays a very important role in the entire debate.

> For example, in Arunachal Pradesh the raw R was 25.60% in 1981, and the traditional Basu and Foster effective literacy rate was 32.63% and 39.65% for two pre-specified values of α (0.25 & 0.5). But in our case the value is 27.80% in 1981. Again in 2007-2008 for the same state the raw literacy rate is 75.65%, the first traditional BF literacy rate is 76.86%, and the second traditional BF literacy rate is 82.14%, while ours is only 71.75%.

Further, the traditional BF measure of literacy penalizes advance states (states with high levels of raw literacy and low levels of proximate literacy) more than the so-called backward states (low level of R and high level of P). Thus BF measures of effective literacy rate with pre-specified weight structure over-compensates the backward states, which sometimes give a misleading picture. In our method the situation is somewhat different. Here the penalization is least in the two most advanced states (Kerala & Meghalaya) compared to other states, like Haryana or Bihar. According to our measure, in none of the states is the difference more than 0.5 percentile.

Similar results also follow from the traditional measure of Subramanian (R^S) and our measure of data-driven Subramanian (R^{*S}). The traditional R^S lies significantly below the raw R. In our case again, even though the deviation is there, the magnitude of the deviation is much less. In most of the cases the deviation is only 0.1 percentile. For example, for Haryana the raw R in 1981 is 43.9% but the R^S is 36.6%, whereas ours is 43.35%. The weights are again the main issue here. Thus Subramanian's measure over penalized the states where the dominance of isolate families is quite high compared to the other literacy-advanced states, which could be corrected in a more positive way if we take into account the data-driven weight structure in the Subramanian measurement of literacy.

The BL measure of literacy with two pre-specified values of α also gives the literacy percentage in various states that are much below than the traditional raw literacy rates of these states. But if we use our data-driven weight in the BL measure of literacy, then the data-driven BL literacy rates in various states come close to their raw literacy percentages. Thus, like the traditional Subramanian

[5] Some authors have wrongly referred to NSSO 43rd round as their source of data, instead of NSSO 50th round.

[6] Thus their measures are not strictly comparable with other measures in the literature.

measure of literacy, in the traditional BL measure of literacy the over penalization of backward states can be rectified to a great extent by adopting the data-driven weight structure in the model.

In order to test the above measures analytically, we have used a paired-t test to determine the difference between the raw literacy rate and the externally-adjusted rate along with our data-driven measures. This clearly vindicates the rationale of our measures. In almost all the cases, differences between the raw literacy rate and the data-driven literacy rate are insignificant. This is irrespective of the types of literacy rates used in our analysis. On the other hand, the differences between the raw measures and *a priori* specified, externally adjusted measures are significant. It is insignificant only in a few cases. Also, the data-driven measures significantly differ from the *a priori* measures in the majority of cases. Thus it may be safely argued, so far as the Indian data is concerned, that the heat generated in the externally adjusted literacy literature is largely a byproduct of *a priori* specification of the weight.

3.3 Relative Dynamics in Literacy

In the extended version of literacy there are three basic indicators -- The "raw" literacy rate (R), the proportion of proximate illiterates (P), and the proportion of isolated illiterates (I). In general, we observe that in all the states there has been a significant rise in R and a significant fall in I over the years 1981 to 2007-2008 (Table 1 and Table 2 in Appendix). The proportion of proximate illiterates also seems to have declined. This simply means that with the spread in literacy, the proportion of families with zero literacy have declined drastically. Again, the major portion of erstwhile proximate illiterates is changing into real literates. Thus there is expansion both in inter-family and intra-family literacy.

> However, the relative ranking shows a bit of interstate dynamics. Some states have improved their relative position, while a few have fallen behind. For example, Gujarat has drastically drifted down from 6th to 14th position in terms of literacy. On the contrary, Sikkim has improved from 15th to 6th position. Interestingly, there is a close contest between two states for the first position -- Kerala and Mizoram. These types of dynamics indicate the complex processes of development changes.

Next, we concentrate on the dynamics of the refined measure -- An issue that is often neglected in the foregoing debate. Though strictly speaking, due to variation of weight across points in time, our measure is not fully comparable, unlike the traditional measures, where the weight remains unchanged. There is, though, a general trend of the rising magnitude of the externally adjusted literacy rate. Again, there are differences in the interstate dynamics resulting in an improvement of position for one, and deterioration for others. However, this inter-state relative dynamics is the same for the refined measure as well as for the traditional measures. For example, if $\alpha = 0.5$, by the traditional measure of BF literacy, Meghalaya's rank improved from 17th in 1981 to 3rd in 2007-2008. Following the data-driven BF, the condition of Meghalaya improved from 16th to 3rd. Thus, though there are differences in absolute magnitude, the relative position remains the same.

We now turn to the multi-dimensional dynamics of R and I (Table 4 and Figure 1-Figure 3 in Appendix). From this table and figures, it is clear that these types of classifications yield a rich result. In 1981 there were 9 states in the vicious cycle category (including six major states). In 1993 this number was reduced to 8 and remains so in 2007-2008. It is to be noted that of the 9 states in 1981, Sikkim and Meghalaya became virtuous states in 1993 and 2007-2008, while Karnataka went from lopsided I to the vicious cycle category. The persistence of the major states, which are all in the so-called BOMARU region (Roy, 2010), within this category is a warning sign to policy makers in India. In a similar way, Gujarat and Punjab moved from virtuous to lopsided I, while Haryana was consistently in lopsided I. West Bengal moved from a lopsided R to virtuous in 1993, but retrograded to lopsided I in 2007-2008. The virtuous states of India include Kerala, Tamil Nadu, Himachal Pradesh, and many of the North Eastern hilly states.

4 Conclusion

There is a rigorous debate in the educational literature concerning the introduction of externality of education within the literacy index. However, the wide deviation of such an externally induced literacy rate has been a major concern for social scientists working in this field, particularly after the objection raised by Subramanian (2004). Our analysis seems to argue that this wide deviation rests on *a priori* specification of external parameters. However, introducing the data-adjusted literacy rate removes much of the dilemma. Thus this may be a realistic solution for introducing externality into the literacy rate without distorting the reality to a great extent. Our version of the story clearly argues in favor of reducing illiteracy without paying too much attention to the academic nuances of the externality debate in literacy. Since data-driven literacy rates are within the close neighborhood of the "raw" values, this debate looses much of its focus.

A major lacuna of this exercise is the neglect of the dimensionality issue raised in the externally adjusted literacy measurement. The paper by Basu and Foster (1998) opens up a new dimension in the literature on literacy measurement -- The proximity to a literate person. Unfortunately, all the effort up to now has been directed only in aggregating the various dimensions of literacy into a single homogeneous measure. This neglect is unfortunate, since it misses out on some important dynamics in these additional dimensions. The present paper tries to fill this gap by illuminating this neglected issue of dimensionality. Our results shows that there is a complexity in the dynamic structure that is not discernable in the unidimensional exercise. However, some further studies incorporating the micro-level features of literacy externality are required before any final conclusion can be reached (Tables 5-12).

References

Basu, K., & Foster, J. E. (1998). On measuring literacy. *Economic Journal*, *108*, 1733-1749.

Basu, K., & Lee, T. (2008). *A new and easy-to-use measure of literacy, its axiomatic properties and an application*. (CAE working Paper No. 08-04). Ithaca, NY: Center for Analytic Economics, Cornell University.

Basu, K., Narayan, A., & Ravallion, M. (1999). *Is knowledge shared within households?* (No. 2261). Washington, DC: World Bank, Office of the Senior Vice President and Chief Economist, Development Economics, and Development Research Group, Poverty and Human Resources.

Dutta, I. (2004). Generalized measures of literacy. *Mathematical Social Science*, *48*, 69-80.

Ghosh, M. (2006). Economic growth and human development in Indian states. *Economic and Political Weekly*, *41*, 3321-3327.

Gibson, J. (2001). Literacy and intrahousehold externalities. *World Development*, *29*, 155-166.

Iversen, V., & Palmer-Jones, R. (2008). Literacy sharing, assortative mating, or what? Labour market advantages and proximate illiteracy revisited. *Journal of Development Studies*, *44*, 797-838.

Johnson, R. A., & Wichern, D. W. (2001). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Kell, C. (2008). "Making things happen:" Literacy and agency in housing struggles in South Africa. *Journal of Development Studies*, *44*, 892-912.

Maddox, B. (2007). Worlds apart? Ethnographic reflections on "effective literacy" and intrahousehold externalities. *World Development*, *35*, 532-541.

Mukherjee, D., & Gupta, M. R. (2003). *Measures of effective Literacy: A theoretical note* (ISER Working Paper, No. 2003-10). Colchester, UK: Institute for Social and Economic Research, University of Essex.

Ranis, G., & Stewart, F. (2000). *Strategies for success in human development* (Center discussion paper, No. 808). New Haven, CT: Economic Growth Center, Yale University.

Ranis, G., & Stewart, F. (2001). *Growth and human development: Comparative Latin American experience* (Center discussion paper, No. 826). New Haven, CT: Economic Growth Center, Yale University.

Roy, S. (2010). *Backwaters of development: Six deprived states of India*. New Delhi, India: Oxford University Press.

Sen, A. (1970). *Collective choice and social welfare*. San Francisco, CA: Holden-Day.

Sengupta, A., Sengupta, M., & Ghosh, M. (2008). Economics of learning- extent versus level: A village level study based on two midnapore district, West Bengal. *Journal of Economics, Vidya Sagar University*, *13*, 111-123.

Subramanian, S. (2004). Measuring literacy: Some extensions of the Basu-Foster framework. *Journal of Development Economics*, *73*, 453-463.

Subramanian, S. (2008). Externality and literacy: A note. *Journal of Development Studies*, *44*, 839-848.

Valenti, P. M. (2002). *Should we be concerned about the distribution of literacy across households? An axiomatic characterization* (CAE Working Paper, No. 02-15). Ithaca, NY: Center for Analytic Economics, Cornell University.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*, *21*, 459-468.

Appendix

(All tables & figures are based on Census of India, 1981, NSSO 43rd round and 64th round respectively.)

Table 1 Comparison of Literacy Rate Over the Three Decades

Literacy Rate	1981	1993	2007-08
Less Than 30	AP*, RAJASTHAN	AP*	NIL
31-60	BIHAR, UP, MP, AP**, ORISSA, MEGHALOYA SIKKIM, HARYANA, KARNATAKA, PUNJAB, WB, MANIPUR, TRIPURA, NAGALAND, GUJRAT, TN, MAHARASTRA, HP.	RAJASTHAN,MP,AP*,BIAHAR,KARNATAKA,ORISSA,HARYANA,GUJRAT,PUNJAB,MAHARASTRA,TN,HP,WB.	RAJASTHAN
61 & Above	GOA, MEJORAM, KERALA.	MEGHALOYA,SIKKIM.MONIPUR,TRIPURA,GOA,NAGALAND,MEJORAM,KERALA	MP,AP*,BIAHAR,KARNATAKA,ORISSA,HARYANA,GUJRAT,PUNJAB,MAHARASTRA,TN,HP,WB. MEGHALOYA,SIKKIM. MONIPUR,TRIPURA,GOA,NAGALAND,MEJORAM,KERALA

AP*: ARUNACHAL PRADESH
AP**: ANDHRA PRADESH

Table 2 Comparison of Isolate Illiteracy (I) Rate Over the Three Decades

Isolate Illiteracy Rate	1981	1993	2007-08
0.00-5.00	KERALA,MEJORAM	KERALA, NAGALAND.	MEJORAM,MEGHALOYA,NAGALAND,KERALA,GOA,MAHARASTRA,TN,MANIPUR,SIKKIM,TRIPURA,PUNJAB,HARYANA,GUJRAT,WB.
6.00-20.00	GOA,HP,MAHARASTRA,MANIPUR,TN,GUJRAT,HARYANA,PUNJAB,NAGALAND,TRIPURA.	GOA,MEJORAM,MANIPUR,SIKKIM,TRIPURA,HP,MAHARASTRA,MEGHALOYA,HARYANA,PUNJAB,GUJRAT,TN,WB.	KARNATAKA,MP,AP**,AP*,ORISSA,UP,RAJASTHAN,BIHAR.
21 & Above	KARNATAKA,SIKKIM,WB,ORISSA,MEGHALOYA,UP,RAJASTHAN,BIHAR,AP*,AP**	KARNATAKA,UP,ORISSA,MP,RAJASTHAN,BIHAR,AP*,AP**	NIL

AP*: ARUNACHAL PRADESH
AP**: ANDHRA PRADESH

Table 3 Two-Dimensional Dynamics

	1981	1993	2007-08
Vicious	Sikkim, Meghaloy, Orissa, AP1, MP, UP, Bihar, Rajasthan, AP2	Orissa, Karnataka, UP, Bihar, AP1, MP, Rajasthan, AP2	Orissa, Karnataka, UP, Bihar, AP1, MP, Rajasthan, AP2
Low R, high (100-I)	Karnataka, Haryana	Maharashtra, Punjab, Gujarat, Haryana	Punjab, WB, Gujarat, Haryana
High R, Low (100-i)	West Bengal	Nil	Nil
Virtuous	Kerala, Mizoram, Goa, Maharashtra, TN, Gujarat, HP, Nagaland, Tripura, Manipur, Punjab	Kerala, Mizoram, Goa, TN, HP, Nagaland, Tripura, Manipur, Sikkim, Meghalaya, WB	Kerala, Mizoram, Goa, TN, HP, Nagaland, Tripura, Manipur, Sikkim, Meghalaya, Maharashtra

Source: Figure 1, 2, & 3.

Table 4 Comparison of Literacy (Proportion) -- 1981

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	0.357	0.43	0.50	0.23	0.32	0.26	0.38	0.35	0.35
Arunachal pradesh	0.256	0.33	0.40	0.14	0.22	0.17	0.28	0.24	0.25
Bihar	0.321	0.40	0.48	0.21	0.29	0.24	0.35	0.31	0.32
Goa	0.653	0.72	0.79	0.60	0.64	0.60	0.67	0.65	0.65
Gujarat	0.522	0.60	0.68	0.44	0.49	0.45	0.55	0.52	0.52
Haryana	0.439	0.54	0.64	0.37	0.42	0.38	0.47	0.43	0.44
Himachal Pradesh	0.512	0.60	0.70	0.45	0.49	0.46	0.54	0.51	0.51
Karnataka	0.462	0.54	0.62	0.36	0.43	0.38	0.49	0.45	0.46
Kerala	0.816	0.86	0.90	0.80	0.81	0.80	0.83	0.82	0.82
Madhya pradesh	0.342	0.42	0.51	0.23	0.31	0.26	0.37	0.33	0.34
Maharashtra	0.558	0.64	0.72	0.49	0.54	0.50	0.58	0.55	0.56
Manipur	0.497	0.59	0.68	0.43	0.48	0.44	0.53	0.49	0.49
Meghalaya	0.420	0.49	0.56	0.30	0.38	0.32	0.44	0.41	0.42
Mizoram	0.740	0.79	0.85	0.71	0.73	0.71	0.76	0.74	0.74
Nagaland	0.503	0.58	0.66	0.41	0.48	0.43	0.53	0.50	0.50
Orissa	0.410	0.49	0.57	0.30	0.38	0.32	0.44	0.40	0.41
Punjab	0.482	0.57	0.66	0.40	0.46	0.41	0.51	0.48	0.48
Rajasthan	0.301	0.39	0.48	0.20	0.27	0.22	0.33	0.29	0.30
Sikkim	0.420	0.51	0.60	0.33	0.39	0.34	0.45	0.41	0.42
Tamil Nadu	0.544	0.62	0.69	0.46	0.52	0.47	0.57	0.54	0.54
Tripura	0.501	0.58	0.65	0.40	0.47	0.42	0.52	0.49	0.50
Uttar Pradesh	0.334	0.42	0.51	0.23	0.30	0.26	0.36	0.32	0.33
West Bengal	0.486	0.56	0.63	0.38	0.45	0.40	0.51	0.48	0.48

Table 5 Comparison of Rank -- 1981

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	18	18	20	20	18	18	18	18	18
Arunachal pradesh	23	23	23	23	23	23	23	23	23
Bihar	21	21	21	21	21	21	21	21	21
Goa	3	3	3	3	3	3	3	3	3
Gujarat	6	7	8	7	6	7	6	6	6
Haryana	14	14	12	13	14	14	14	14	14
Himachal Pradesh	7	6	5	6	7	6	7	7	7
Karnataka	13	13	14	14	13	13	13	13	13
Kerala	1	1	1	1	1	1	1	1	1
Madhya pradesh	19	19	19	19	19	19	19	19	19
Maharashtra	4	4	4	4	4	4	4	4	4
Manipur	10	8	7	8	8	8	9	10	10
Meghalaya	15	17	17	17	16	17	16	16	16
Mizoram	2	2	2	2	2	2	2	2	2
Nagaland	8	9	9	9	9	9	8	8	8
Orissa	17	16	16	16	17	16	17	17	17
Punjab	12	11	10	11	11	11	11	12	12
Rajasthan	22	22	22	22	22	22	22	22	22
Sikkim	15	15	15	15	15	15	15	15	15
Tamil Nadu	5	5	6	5	5	5	5	5	5
Tripura	9	10	11	10	10	10	10	9	9
Uttar Pradesh	20	20	18	18	20	20	20	20	20
West Bengal	11	12	13	12	12	12	12	11	11

Table 6 Comparison of Literacy (Proportion) -- 1993

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	0.387	0.450	0.513	0.248	0.345	0.284	0.397	0.377	0.383
Arunachal pradesh	0.262	0.336	0.409	0.146	0.228	0.181	0.274	0.250	0.259
Bihar	0.390	0.455	0.521	0.254	0.349	0.289	0.401	0.381	0.386
Goa	0.725	0.781	0.837	0.689	0.713	0.690	0.735	0.724	0.725
Gujarat	0.509	0.585	0.661	0.413	0.479	0.428	0.521	0.504	0.506
Haryana	0.486	0.572	0.658	0.403	0.460	0.415	0.500	0.482	0.484
Himachal Pradesh	0.579	0.655	0.730	0.510	0.557	0.518	0.591	0.576	0.577
Karnataka	0.475	0.548	0.620	0.364	0.441	0.385	0.487	0.469	0.472
Kerala	0.904	0.924	0.944	0.890	0.899	0.890	0.907	0.903	0.903
Madhya pradesh	0.384	0.458	0.532	0.261	0.347	0.291	0.396	0.376	0.381
Maharashtra	0.560	0.628	0.696	0.466	0.530	0.479	0.571	0.555	0.557
Manipur	0.673	0.732	0.792	0.613	0.653	0.618	0.683	0.670	0.671
Meghalaya	0.611	0.666	0.721	0.508	0.578	0.523	0.620	0.606	0.608
Mizoram	0.888	0.903	0.917	0.841	0.873	0.843	0.891	0.887	0.887
Nagaland	0.801	0.843	0.886	0.777	0.793	0.778	0.808	0.800	0.800
Orissa	0.477	0.540	0.602	0.347	0.437	0.375	0.488	0.470	0.474
Punjab	0.522	0.597	0.672	0.429	0.493	0.443	0.535	0.517	0.520
Rajasthan	0.328	0.415	0.501	0.221	0.296	0.248	0.343	0.320	0.326
Sikkim	0.653	0.715	0.777	0.588	0.632	0.594	0.664	0.651	0.652
Tamil Nadu	0.567	0.628	0.688	0.459	0.533	0.476	0.577	0.562	0.564
Tripura	0.714	0.757	0.800	0.632	0.688	0.641	0.721	0.711	0.712
Uttar	0.424	0.502	0.581	0.313	0.390	0.336	0.437	0.417	0.421
West Bengal	0.591	0.645	0.700	0.478	0.555	0.496	0.600	0.586	0.588

Table 7 Comparison of Rank -- 1993

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	20	21	21	21	21	21	20	20	20
Arunachal pradesh	23	23	23	23	23	23	23	23	23
Bihar	19	20	20	20	19	20	19	19	19
Goa	4	4	4	4	4	4	4	4	4
Gujarat	14	14	14	14	14	14	14	14	14
Haryana	15	15	15	15	15	15	15	15	15
Himachal Pradesh	10	9	8	8	9	9	10	10	10
Karnataka	17	16	16	16	16	16	17	17	17
Kerala	1	1	1	1	1	1	1	1	1
Madhya pradesh	21	19	19	19	20	19	21	21	21
Maharashtra	12	11	11	11	12	11	12	12	12
Manipur	6	6	6	6	6	6	6	6	6
Meghalaya	8	8	9	9	8	8	8	8	8
Mizoram	2	2	2	2	2	2	2	2	2
Nagaland	3	3	3	3	3	3	3	3	3
Orissa	16	17	17	17	17	17	16	16	16
Punjab	13	13	13	13	13	13	13	13	13
Rajasthan	22	22	22	22	22	22	22	22	22
Sikkim	7	7	7	7	7	7	7	7	7
Tamil Nadu	11	12	12	12	11	12	11	11	11
Tripura	5	5	5	5	5	5	5	5	5
Uttar	18	18	18	18	18	18	18	18	18
West Bengal	9	10	10	10	10	10	9	9	9

Table 8 Comparison of Literacy (Proportion) -- 2007-2008

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	0.648	0.713	0.778	0.588	0.629	0.593	0.650	0.646	0.647
Aorunachal Pradesh	0.716	0.769	0.821	0.663	0.699	0.667	0.717	0.715	0.715
Bihar	0.576	0.643	0.709	0.485	0.547	0.498	0.579	0.574	0.575
Goa	0.847	0.881	0.915	0.833	0.842	0.833	0.848	0.846	0.846
Gujrat	0.752	0.803	0.854	0.719	0.741	0.720	0.753	0.751	0.751
Hariyana	0.732	0.788	0.844	0.700	0.721	0.701	0.733	0.731	0.731
Himachal Pradesh	0.798	0.842	0.886	0.777	0.791	0.778	0.799	0.798	0.798
Karnataka	0.715	0.773	0.831	0.676	0.702	0.678	0.716	0.714	0.714
Kerala	0.936	0.950	0.964	0.929	0.933	0.929	0.936	0.936	0.936
Madhaya Pradesh	0.705	0.760	0.816	0.654	0.688	0.658	0.706	0.704	0.704
Maharastra	0.815	0.854	0.893	0.791	0.807	0.792	0.816	0.815	0.815
Manipur	0.804	0.845	0.885	0.777	0.795	0.778	0.805	0.803	0.804
Meghalaya	0.927	0.943	0.960	0.920	0.924	0.920	0.927	0.927	0.927
Mizoram	0.962	0.970	0.978	0.957	0.960	0.957	0.962	0.962	0.962
Nagaland	0.915	0.934	0.952	0.906	0.912	0.906	0.915	0.915	0.915
Orissa	0.700	0.757	0.813	0.648	0.683	0.652	0.702	0.699	0.699
Punjab	0.771	0.817	0.864	0.738	0.760	0.740	0.772	0.770	0.770
Rajasthan	0.606	0.682	0.757	0.551	0.588	0.556	0.609	0.605	0.606
Sikkim	0.824	0.860	0.895	0.797	0.815	0.797	0.826	0.824	0.824
Tamil Nadu	0.808	0.848	0.888	0.782	0.799	0.782	0.809	0.807	0.807
Tripura	0.784	0.828	0.873	0.754	0.774	0.755	0.785	0.783	0.783
Uttar Pradesh	0.641	0.709	0.776	0.584	0.623	0.589	0.643	0.640	0.640
West Bengal	0.759	0.807	0.856	0.723	0.747	0.724	0.761	0.759	0.759

Table 9 Comparison of Rank -- 2007-2008

States	R	exo(.25) BF	exo(.5) BF	Subramaniam	BL(.25)	BL(.5)	Data Driven BF	Data Driven S	Data Driven BL
Andhra Pradesh	20	20	20	20	20	20	20	20	20
Aorunachal Pradesh	16	17	17	17	17	17	16	16	16
Bihar	23	23	23	23	23	23	23	23	23
Goa	5	5	5	5	5	5	5	5	5
Gujrat	14	14	14	14	14	14	14	14	14
Hariyana	15	15	15	15	15	15	15	15	15
Himachal Pradesh	10	10	9	9	10	10	10	10	10
Karnataka	17	16	16	16	16	16	17	17	17
Kerala	2	2	2	2	2	2	2	2	2
Madhaya Pradesh	18	18	18	18	18	18	18	18	18
Maharastra	7	7	7	7	7	7	7	7	7
Manipur	9	9	10	10	9	9	9	9	9
Meghalaya	3	3	3	3	3	3	3	3	3
Mizoram	1	1	1	1	1	1	1	1	1
Nagaland	4	4	4	4	4	4	4	4	4
Orissa	19	19	19	19	19	19	19	19	19
Punjab	12	12	12	12	12	12	12	12	12
Rajasthan	22	22	22	22	22	22	22	22	22
Sikkim	6	6	6	6	6	6	6	6	6
Tamil Nadu	8	8	8	8	8	8	8	8	8
Tripura	11	11	11	11	11	11	11	11	11
Uttar Pradesh	21	21	21	21	21	21	21	21	21
West Bengal	13	13	13	13	13	13	13	13	13

Table 10 Comparison of 't' Ratios

Year-1981			
Raw Vs. exogenous			
	T statistics	*P* value	T critical one tail
Raw VS BF(.25)	-2.01443	0.025051	1.68023
Raw VS BF(.5)	-4.12673	8.06E-05	1.68023
Raw VS Subramaniam	2.015152	0.025012	1.68023
Raw VS BL(.25)	0.665969	0.254454	1.68023
Raw VS BL(.5)	1.684879	0.049547	1.68023
Raw Vs. Data Driven			
Raw VS Data Driven BF(.25)	-0.61805	0.269866	1.68023
Raw VS Data Driven S	0.176309	0.43043	1.68023
Raw VS Data Driven BL	0.072487	0.471271	1.68023
Exogenous Vs. Data Driven			
BF(.25) VS Data Driven BF	1.397889498	0.084577089	1.68023
BF(.5)VS Data Driven BF	3.516240985	0.000514231	1.68023
Subramanian VS Data Driven Subramanian	-1.83359071	0.036742264	1.68023
BL(.25)VS Data Driven BL	-0.59346411	0.277954629	1.68023
BL(.5)VS Data Driven BL	-1.61211359	0.057044042	1.68023

Table 11 Comparison of 't' Ratios

Year-1993			
Raw Vs. exogenous			
	T statistics	*P* value	T critical one tail
Raw VS BF(.25)	-1.292823659	0.101409807	1.68023
Raw VS BF(.5)	-2.695595708	0.004957594	1.68023
Raw VS Subramaniam	1.633394945	0.054761229	1.68023
Raw VS BL(.25)	0.540306521	0.29585512	1.68023
Raw VS BL(.5)	1.380376758	0.08722183	1.68023
Raw Vs. Data Driven			
Raw VS Data Driven BF(.25)	-0.20721	0.418402	1.68023
Raw VS Data Driven S	0.100354	0.46026	1.68023
Raw VS Data Driven BL	0.04708	0.481331	1.68023
Exogenous Vs. Data Driven			
BF(.25) VS Data Driven BF	1.086732	0.141537	1.68023
BF(.5)VS Data Driven BF	2.493623	0.008239	1.68023
Subramanian VS Data Driven Subramanian	-1.53007	0.066579	1.68023
BL(.25)VS Data Driven BL	-0.49322	0.312155	1.68023
BL(.5)VS Data Driven BL	-1.33316	0.09467	1.68023

Table 12 Comparison of 't' Ratios

	Year-2007-08		
	Raw Vs. exogenous		
	T statistics	*P* value	T critical one tail
Raw VS BF(.25)	-1.598560552	0.058537827	1.68023
Raw VS BF(.5)	-3.462555926	0.000601627	1.68023
Raw VS Subramaniam	1.015053791	0.157814768	1.68023
Raw VS BL(.25)	0.353011486	0.362882734	1.68023
Raw VS BL(.5)	0.960665706	0.170984988	1.68023
	Raw Vs. Data Driven		
Raw VS Data Driven BF(.25)	-0.047653883	0.481103868	1.68023
Raw VS Data Driven S	0.02245551	0.491093065	1.68023
Raw VS Data Driven BL	0.015783118	0.493739384	1.68023
	Exogenous Vs. Data Driven		
BF(.25) VS Data Driven BF	1.551896	0.063926	1.68023
BF(.5)VS Data Driven BF	3.419552	0.000682	1.68023
Subramanian VS Data Driven Subramanian	-0.99273	0.163135	1.68023
BL(.25)VS Data Driven BL	-0.33726	0.368764	1.68023
BL(.5)VS Data Driven BL	-0.94507	0.174892	1.68023

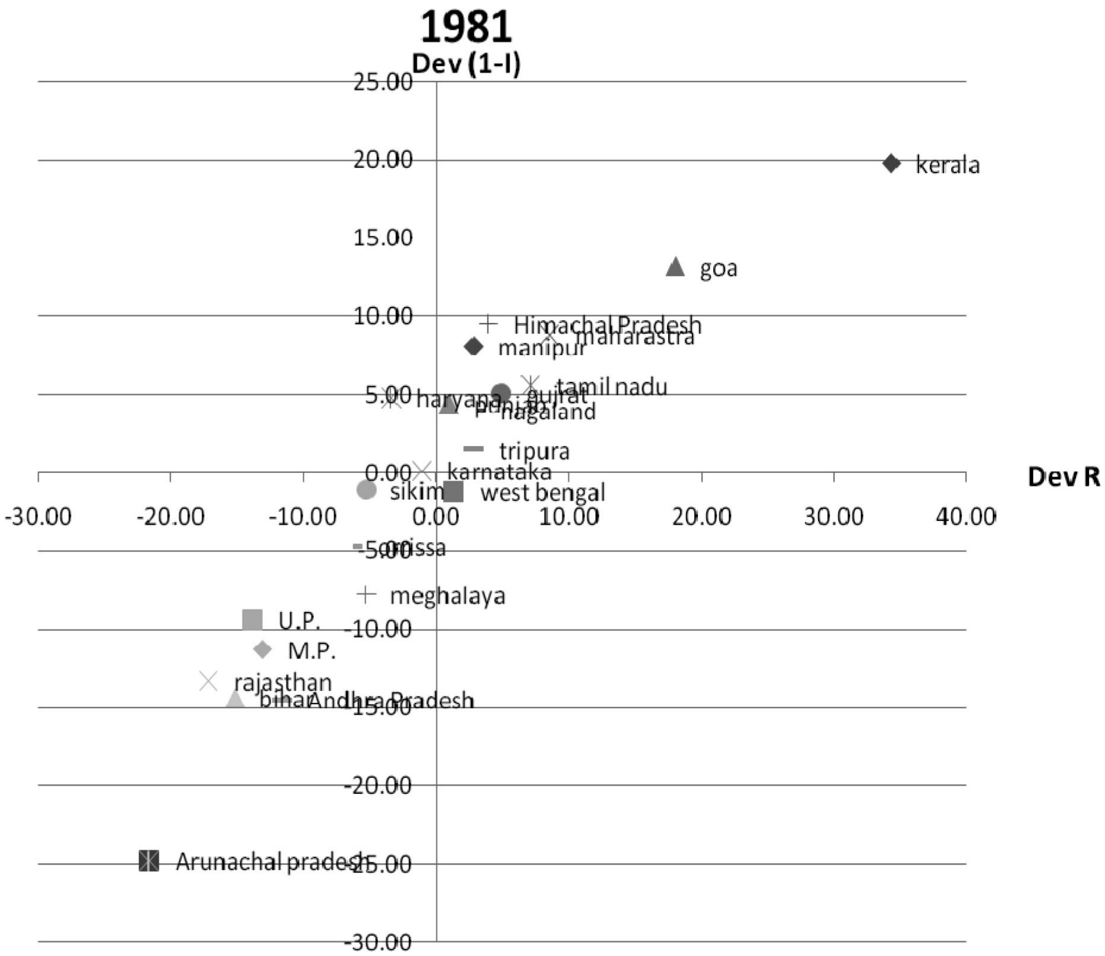

Figure 1 Deviation of R & I from Means, 1981

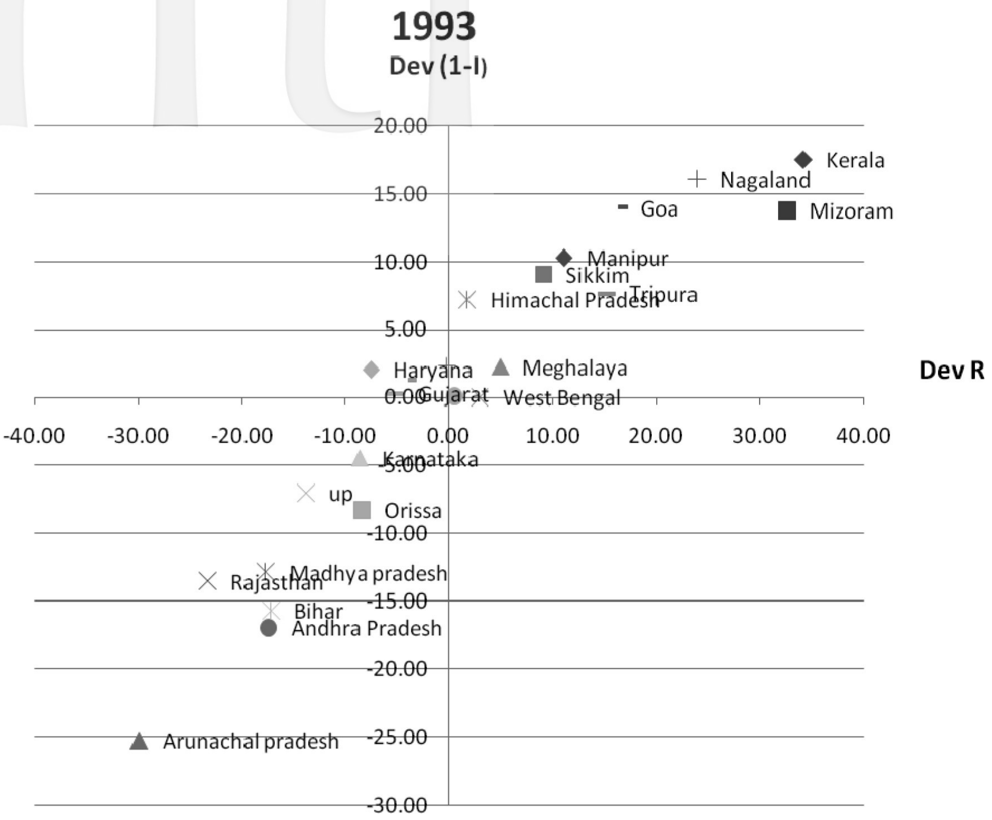

Figure 2 Deviation of R & I from Means, 1993

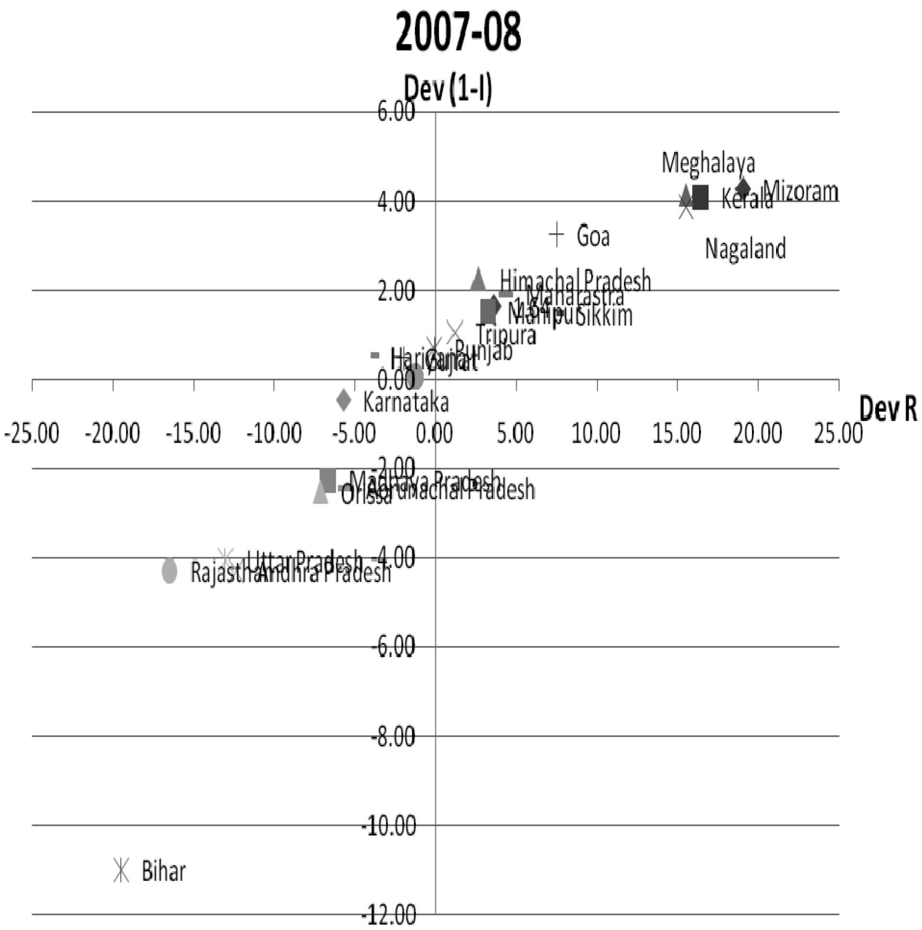

Figure 3 Deviation of R & I from Means, 2007-2008