

Literacy and Intrahousehold Externalities

John Gibson[†]
University of Waikato

Abstract

The size of the intrahousehold externality created when a literate household member provides literacy services to illiterate members is estimated in this paper, using a model of children's height-for-age in Papua New Guinea. This externality appears to be large. These results support the use of a new measure of literacy, developed by Basu and Foster (1998) to take account of the gains when illiterates live in households where at least one person is literate. Regional rankings change when the new measure of literacy is used, suggesting that policies guided by the usual measure of adult literacy may overlook an important pathway for human development.

JEL: I20, I32

Keywords: Development indicators, Externalities, Human development index, Literacy, Papua New Guinea, Pacific Islands

[†] Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand.
Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz.

1. INTRODUCTION

Economists are paying increased attention to the impact of literacy on economic development. In addition to the role that literacy plays in output growth (Mitch, 1984; Gould, 1995) and structural transformation (Lin, 1997), improvements in literacy may also help in the adoption of innovations (Green *et. al.*, 1985) and with the health of children (Sandiford, 1995). Partly for these reasons, and also because improved literacy is an important end in itself, the Human Development Index places a high weight on a country's adult literacy rate.¹

One problem with usual measures of literacy – such as the adult literacy rate – is that the external benefits created by literate members of households for non-literate members are ignored. Thus, the literacy rate cannot distinguish between a society where every household contains at least one literate member and another society where the same total number of illiterates are concentrated in households where every member is illiterate. A new measure of literacy that incorporates the external benefits provided by literate members of households has recently been proposed by Basu and Foster (1998). This measure distinguishes *proximate illiterates*, persons who cannot read but who live with someone who can, from the *isolated illiterates* who live in households where nobody can read.² Each proximate illiterate is assigned an effective literacy level, a which depends on the strength of the external benefits provided by literate members. However, Basu and Foster do not show how a can be empirically estimated, simply using assumed values for a in their illustration of the new literacy measure.

In common with other types of positive externalities, the presence of these external benefits of literacy may cause the equilibrium level of literacy training to be too low, in other words there is a potential market failure. Adult educators who train illiterates may have in mind a model

where only the individual who is taught to read benefits from their teaching – as is implied by the usual adult literacy rate which assumes that $\alpha=0$ – so they fail to take into account the benefits to proximate illiterates. Because adult literacy training often takes place in a non-market setting, with government or non-profit groups such as churches and aid organisations as the providers, the views by suppliers about the likely demand for their services may influence the level of supply. But if these suppliers, who may be the determining side of the market if there is excess demand, compare marginal costs with the private benefits of further literacy training rather than with the (presumed higher) social benefits, they may provide a sub-optimal level of literacy training. On the other hand, once literacy advances beyond the point where there is at least one literate person per household, failure to account for the benefits already provided to proximate illiterates may cause the social value of further literacy training to be overstated.

This paper discusses a simple method of estimating the size of the intrahousehold externality created when literate household members provide literacy services to illiterate members. This method is based on an analogy with the literature on measuring economies of household size, where a ratio of food Engel curve coefficients can be used to establish the rate at which person numbers translate into effective household size (Lanjouw and Ravallion, 1995). In the same manner, the effective literacy level of proximate illiterates, α can be established from a ratio of coefficients in equations predicting outcomes that depend upon literacy skills.

If household members do not fully pool their resources, for which there is growing evidence (Haddad, 1999), the estimated value of α may understate the effective literacy level of proximate illiterates. Lack of pooling may indicate that household members bargain over resource allocations, and anything that makes an individual better off outside the household – such as literacy – also

increases their bargaining power within the household (McElroy, 1997). Therefore, two changes may occur when a person in a previously illiterate household learns to read: (i) the capabilities of the remaining illiterates are augmented, with the strength of this effect determining the magnitude of α (Basu and Foster, 1998), and (ii) household resources shift towards the person who learns to read. It is difficult to observe these effects separately. Instead, the observed relationship between the literacy of certain household members and the welfare of other members is the *net* effect of intrahousehold external benefits and any offsetting resource flows caused by changes in the balance of bargaining power. Hence, the net effect will tend to understate the capability-augmenting effect and therefore understate the magnitude of α .

In the empirical section of the paper, the size of the intrahousehold externality arising from the presence of literate household members is measured using a model of children's height-for-age in Papua New Guinea (PNG). The issues of child height (as an indicator of long-run health and nutritional status) and adult literacy are both topical in PNG. There is widespread concern that Papua New Guinea's abundant resources have not produced a higher or more even level of human development (Yala and Levantis, 1998). Partly as a consequence, the first national *Human Development Report* for PNG is currently in preparation. This report will show a low overall level of literacy, with high regional (and gender) inequalities. It is therefore interesting to see how the effective literacy level and the rankings of regions change when a literacy measure that takes account of intrahousehold externalities is used.

In addition to policy interest in measures of literacy there is also considerable interest in the relationship between education, literacy and the health of young children, both in PNG (Jenkins, 1992) and in the wider development community (Thomas *et. al.* 1991; Sandiford, 1995). A large literature finds that measures of parental education or literacy have a positive

effect on child height (Strauss and Thomas, 1995) and these results have also been replicated in PNG (Gibson, 1997). However, the children of illiterate parents may still benefit if other household members can provide literacy services (e.g., by reading pamphlets on methods of preventing disease), so these models of child height may have excluded a relevant variable – whether anyone at all in the household can read.³ The results reported here on the size of the externality created when at least one household member can read may help to settle this issue.

2. MEASURES OF LITERACY

Let x_j^h be a variable that indicates the literacy level of the j th adult member of household h in some society. When the usual adult literacy rate is calculated for this society, x_j^h can take just two values: $x_j^h = 1$ if person j in household h is literate, and $x_j^h = 0$ if they are illiterate. If there is a total of m households, and the number of adults in household h is n_h , then the total number of adults in the society is:

$$N = \sum_{h=1}^m n_h$$

and the adult literacy rate, R can be calculated as:

$$R = \frac{\sum_{h=1}^m \sum_{j=1}^{n_h} x_j^h}{N}. \quad (1)$$

The innovation suggested by Basu and Foster (1998) is to let the measure of literacy change when a literate person in a household provides external benefits to illiterate members of the same household. The literacy rate does not recognise this effect because it ignores household structure. The magnitude of these external benefits is given by the parameter \mathbf{a} ($0 < \mathbf{a} < 1$),

which is initially assumed to be independent of the characteristics of household members.⁴

Consequently, the indicator variable for ‘effective’ literacy, \tilde{x}_j^h can now take three values:

$$\tilde{x}_j^h = \begin{cases} 1 & \text{if } x_j^h = 1 \\ \mathbf{a} & \text{if } x_j^h = 0, \text{ and } x_k^h = 1 \text{ for some } k \\ 0 & \text{if } x_k^h = 0 \text{ for every } k . \end{cases}$$

Thus, every *proximate illiterate*, who lives in a household where at least one person is literate, is assigned an effective literacy level \mathbf{a} .⁵ There is a close parallel with the equivalence scale literature, which assigns every individual a relative cost-of-living which depends upon who (in terms of household composition), and how many others, they live with.

Basu and Foster (1998) suggest as an overall measure of effective literacy for the society:

$$* = \frac{\sum_{h=1}^m \sum_{j=1}^{n_h} \tilde{x}_j^h}{N}. \quad (2)$$

This new measure includes the usual literacy rate as a special case because:

$$* = R + \mathbf{a}P \quad (3)$$

where P is the share of the adult population that is proximate illiterate. Thus, if there is no external benefit to an illiterate person from living with a literate person, $\mathbf{a}=0$ and $*$ reduces to the usual literacy rate, R . At the other extreme, if having a literate person in the household gives an illiterate person access to the full range of literacy functionings, such that $\mathbf{a}=1$, then $*=R+P$ and the rate of effective literacy for adults would be equal to $1-I$, where I is the share of the adult population who are *isolated illiterates* (i.e., living in households where all are illiterate). Hence, much depends on the actual value of \mathbf{a} , although Basu and Foster simply use assumed values for \mathbf{a} in their illustration of the new measure of literacy.

Basu and Foster show that L^* satisfies five basic axioms that are intuitive properties for a measure of literacy, while the traditional literacy rate, R satisfies only four of these axioms. The axiom which R fails concerns household splits that create isolated illiterates when part of a household with at least one literate person divides from the rest of the household containing only illiterates. Such a split reduces the effective literacy level L^* because of the decline in the share of the population who are proximate illiterate but has no effect on R . If these household splits are more likely, as a result of changes in the literacy of individual household members, the benefits to illiterates from their proximity to newly-literate adults may be only transitory as the adults who can read may leave to form their own household.⁶ However, little is known about the effect of literacy on the longer-run dynamics of household structure,⁷ so this potential caveat is not explored further in the paper.

(a) Measuring the effective literacy parameter, \mathbf{a}

Let Y be a variable of interest (e.g., the rate of adopting an agricultural innovation) that is affected by some measure of literacy, L and by other characteristics that are represented by the portmanteau variable, X and the random error, u :

$$Y = \mathbf{b}X + \mathbf{g}L + u . \quad (4)$$

If the Basu and Foster measure of effective literacy, L^* is used in this model, then by substituting equation (3) into equation (4), the model can be rewritten as:

$$Y = \mathbf{b}X + \mathbf{g}(R + \mathbf{a}P) + u . \quad (5)$$

If the model is then estimated with both R and P as explanatory variables:

$$Y = \mathbf{b}X + \mathbf{g}R + \mathbf{a}gP + u , \quad (6)$$

and the effective literacy level of a proximate illiterate, \mathbf{a} can be calculated from the ratio of the coefficient on P to the coefficient on R . Of course the estimate of \mathbf{a} may vary somewhat with the choice of the dependent variable, because the external benefits of literacy depend, amongst other things, on where production takes place and whether it is carried out individually.⁸ For example, the estimated \mathbf{a} is unlikely to be very high if Y measures individual wages, because living with someone who can read may not improve an illiterate's workplace productivity. But if Y measures some aspect of home production, such as children's health, the estimated \mathbf{a} is likely to be higher because having one person who can read may raise the productivity of other household members.

3. A 'CONVENTIONAL' ADULT LITERACY PROFILE FOR PAPUA NEW GUINEA

There is a wide range of estimates of the adult literacy rate in Papua New Guinea, with some sources suggesting that it is below 50% and others suggesting that it exceeds 70%.⁹ This uncertainty partly reflects the extraordinary linguistic diversity in the country – there are over 700 language groups in a population of less than five million (Wurm, 1975) – which makes the measurement of literacy especially susceptible to different definitions. The estimates reported here are from a 1996 national household survey (detailed below), which defined literacy according to a respondent's answer to the question: 'Can you read a newspaper?' This definition of literacy is more restrictive than definitions based on reading or writing in the local vernacular (known as *Tok ples* in neo-Melanesian Pidgin) because newspapers are mainly written in the three official languages of PNG (English, Motu, and Pidgin). However, being able to read the official languages is the economically useful skill because these are the languages of agricultural extension bulletins, health information and everyday commerce. The

Tok ples languages are not useful in this regard because it is not feasible to translate technical or commercial information into the hundreds of local vernaculars.

There are large regional disparities in adult literacy rates in PNG (Table 1). Over three-quarters of the adult population in the National Capital District and New Guinea Islands can read, but just over one-third of the adults in the Highlands can read. Across the whole country, approximately 52% of adults can read a newspaper. However, about 29% of the adult population are proximate illiterate, and thus may have access to a portion of the benefits conveyed by literacy. The proportion of proximate illiterates is greatest in the Highlands, so there may be rather smaller regional disparities in effective literacy levels, depending on how great are the benefits of having a literate person in the household. However, even at the extreme of a proximate illiterate being equivalent to a literate ($\alpha=1$), the results in the “isolated illiterates” column of Table 1 suggest that substantial regional variation in effective literacy rates would still remain.

(Table 1 about here)

The gender gap between male and female literacy rates varies widely between regions but shows the same pattern as the adult literacy rate and the isolated illiteracy rate. Female literacy rates are almost on par with male rates in the National Capital District and New Guinea Islands, but are less than two-thirds as high as rates for males in the Highlands and North Coast, which are the two regions with the lowest overall literacy rates. Differences between male and female literacy rates will be returned to below, in the discussion of a gender-adjusted measure of effective literacy.

4. MEASURING INTRAHOUSEHOLD EXTERNALITIES: METHODS, DATA AND ESTIMATION RESULTS

(a) The model

The empirical model is concerned with the relationship between adult literacy and the height of young children (0-5 years old). Height is considered a good measure of long-run nutritional status, reflecting the accumulated effect of past illnesses and periods of inadequate food intake. Adult literacy is likely to affect height by improving the ability of family caregivers to access information about health and nutrition.¹⁰ Ideally, the adult literacy measure should be lagged somewhat, because child height reflects the accumulated effect of nutrition and illness over a number of years. But over a five year period (the maximum age in the sample) literacy rates are unlikely to change much, so current values should be good indicators of the relevant lagged values, which are unavailable because surveys of literacy occur only infrequently in PNG.¹¹

The model in equation (6) used to estimate \mathbf{a} is based on the literacy *rate*, R and the *proportion* of proximate illiterates, P but it does not specify at what level of aggregation these rates should be calculated. Usually, literacy rates are calculated at community or state level, and indeed this is the level used by Basu and Foster (1998) in the illustration of their effective literacy measure. But the phenomenon studied here – the spillover benefits from literates to illiterates – is assumed to take place mainly within households. One practical issue that affects the choice of aggregation levels is that with an average of only three adults per household in PNG, the household may not be a large enough unit to allow much variation in the calculated values of either R or P – there will tend to be mass points at 0, 0.25, 0.33, 0.5, 0.66, 0.75 and 1.0.¹² Moreover, the sum of R and P will equal 1 for all households with at least one literate adult and 0 for households with no literate adults, so collinearity problems may result.

Similarly, if the *number* of literates and proximate illiterates were used these would sum to the total number of adults for households where at least one person can read and so would be collinear with measures of household composition.

In light of the practical and conceptual issues involved, the model is estimated at both community level and household level, with a comparison of the two sets of estimates serving as an indicator of the robustness of the findings. Emphasis is placed on the specification that predicts the average height of children in a community because this corresponds to the usual level at which literacy rates are calculated and also circumvents some problems that afflict the empirical modelling of children's height. In particular, the model is not affected if households simultaneously choose consumption expenditures along with some of the inputs into the production of child height (which makes household expenditures endogenous in models of individual children's height). The average income level in each community is unlikely to be determined jointly with inputs into children's health, because these averages are formed over the decisions of many households who have different tastes and exogenous constraints.

Therefore, the first model estimated is of the average height of children in a community, given the adult literacy rate and the proportion of proximate illiterate adults in that community. If literate household members do not provide external benefits to illiterate members, P should not influence the height of children and \mathbf{a} will be estimated to be zero. But if there are intrahousehold externalities, $\mathbf{a} \neq 0$ and both R and P will be relevant explanatory variables. The same model is then estimated at household level, with results reported in Appendix 1.

Child height is systematically (and non-linearly) related to age and sex, so these characteristics need to be controlled for when comparing the average height of children in each community.

The approach used here follows Thomas and Strauss (1992) and relates each child's height to the height of a well-nourished United States child of the same gender and age. Thus, the dependent variable in the model is (the logarithm of) the community-level average of children's height-for-age as a percentage of the median height in the reference population.

Even though the age of children is controlled for in the dependent variable, it is also common to include age as an explanatory variable in anthropometric analyses in developing countries (Thomas and Strauss, 1992). This allows for the fact that the height of children in developing countries may fall further below the US median as growth retardation worsens in the early years of life (Sahn and Alderman, 1997). The model therefore includes the variable *AGE*, measured as the average, in months, over all children whose heights were measured in each community. The variable *MALE*, which equals one for boys and zero for girls, is also included in case the pattern of growth retardation differs between the sexes (although Gibson and Rozelle (1998) suggest that it does not differ in PNG).

Household economic resources are another important determinant of child growth (Sahn, 1994). The real value of expenditure per adult equivalent (*PCX*) is used in the model to measure household resources.¹³ It is also possible that, in terms of child health, there are diminishing returns to increased household expenditures as budgets shift towards luxury goods. Therefore, a quadratic in *PCX* is included in the model.

Access to public health facilities, clean water and sanitation are also important determinants of child height in developing countries (Lavy *et. al.*, 1996). All of these forms of public infrastructure are readily accessible and in reasonable condition in urban areas of PNG but are

less easily accessible and of more variable quality in rural areas. Therefore, a dummy variable, *URBAN* is included in the model to control for these effects.

The final variable included in the model is a more country-specific factor. Previous analyses of child growth patterns in PNG have discovered a paradox – children in the highlands are shorter (implying poorer health and nutrition) but heavier (implying better health and nutrition) than children in the lowlands (Heywood, Singleton and Ross, 1988). There is no consensus yet as to whether these altitudinal differences in the relationship between weight and height of children reflect dietary, environmental or genetic factors (Smith, *et. al.*, 1993). To deal with the unusual growth pattern of highlands children, some previous anthropometric analyses in PNG have abandoned the use of external growth standards and formed internal standards separately for the highlands and lowlands regions. A similar effect can be achieved by the use of different intercept variables to allow for the different average heights of highlands and lowlands children. Thus, the variable *HIGHLANDS*, which equals one if the community is in the highlands and zero otherwise, is included in the model. Therefore, the specification to be estimated is:

$$\ln(HAM)=f(R, P, AGE, MALE, PCX, PCX^2, URBAN, HIGHLANDS).$$

(b) Data

Data used in this paper come from the 1996 Papua New Guinea Household Survey (PNGHS), which was a complex sample survey of living standards carried out as part of a World Bank-sponsored poverty assessment. The sample was drawn from the capital city, Port Moresby, plus 80 other towns and rural villages, giving a total of 81 communities. The selected communities were obtained in a stratified manner from the enumeration areas of the 1990 Census, and came from all provinces except one.¹⁴ In Port Moresby, 240 households were

interviewed, with six households randomly selected from each of 40 ‘clusters’ (approximately one city block). In all other areas, 12 households were randomly selected per community. In both parts of the sample, information was obtained on the literacy and education levels of household members, the heights and weights of young children, and household expenditures.¹⁵ With at least 12 randomly selected households per community, reliable estimates of R and P can be calculated. To account for the unequal sampling rates between the capital city and the rest of the country, and for the change in population since the 1990 Census, a set of sampling weights were derived to allow the results reported to be representative of Papua New Guinea in 1996.

On average, the height of young children in Papua New Guinea is only 93% of the median height in the U.S. for the same age and sex (Table 2). A common indicator for stunting, suggested by Waterlow *et. al.* (1977), is height that is less than 90% of the U.S. median. According to this indicator, 29% of the young children in PNG are stunted. Table 2 also reports the means and standard deviations of the other variables used in the analysis.

(Table 2 about here)

(c) Estimation results

Initial estimation of the model suggested that differences in the gender composition of children in each community did not affect the average height ($p=0.62$), so to save space, results reported are for the version of the model with the variable *MALE* excluded.¹⁶ The insignificance of the variable *MALE* is consistent with earlier findings in PNG (using individual-level data) that there is no gender difference in the growth retardation of young children (Gibson and Rozelle, 1998).

The adult literacy rate, R is a highly significant predictor ($p=0.001$) of the average height of children in a community, controlling for age, household economic resources, access to health infrastructure and regional effects (Table 3). The coefficient on R suggests that a 10 percentage point increase in the adult literacy rate in a community, holding other variables at their mean, would raise the average height of young children, relative to the U.S. median height-for-age, by 0.6 percentage points. Although this may seem like a small increase, if the height of all young children in PNG could be raised by this percentage, the total number of stunted children would be reduced by 12%.

(Table 3 about here)

It is apparent from Table 3 that the adult literacy rate is not the best measure of literacy for predicting child height, because the proportion of proximate illiterates in the community, P is also a relevant predictor ($p=0.049$). This suggests that the Basu and Foster (1998) measure of effective literacy, R^* is superior to the usual measure of the adult literacy rate, given that

$R^* = R + aP$. Further evidence for the superiority of R^* is given by the reduction in the adjusted- R^2 from 0.470 to 0.448 when P is excluded from the model.

The results in Table 3 suggest that the usual measure of literacy ignores important intrahousehold externalities, such as when a literate person interprets disease-prevention guidelines on behalf of the illiterate parents of a young child. How large is this externality? According to equation (6), the ratio of the coefficient on P to the coefficient on R implies an ‘effective’ literacy level of a proximate illiterate (a) of 0.76. This is rather higher than the values of a assumed by Basu and Foster in the illustration of the R^* measure (0.25 and 0.5). However, as mentioned above, a is likely to be high in certain types of home production, such

as child health, because having one person in the household who can read may raise the productivity of all other household members. Further estimates of equation (6) in different settings are needed before any overall view on the appropriate value of α can be formed.

It is also apparent from Table 3 that the full effect on child height of any increase in the adult literacy rate will depend on how that extra literacy is distributed. If the newly-literate adults live in households where someone could already read, the impact is likely to be smaller than if they live in households where previously everyone was illiterate. A numerical example can illustrate this point: In PNG, a 10 percentage point increase in R , properly distributed, would be sufficient to eliminate all isolated illiteracy, so that $R+P=100\%$.¹⁷ Given initial values of $R\approx 50\%$ and $P\approx 30\%$, this would mean an increase of 10 percentage points in both R and P . Therefore, the combined effect on (the logarithm of) average standardised child height would be $0.1\times(0.062+0.047)$, which corresponds to raising the average height-for-age of children by just over one percentage point. If the height of all young children in PNG increased by this proportion, the total number of stunted children would fall by 20%. But if the newly-literate adults all lived in households where someone could already read, the 10 percentage point increase in R implies an offsetting decrease in P , so that the combined effect on average height would be $0.1\times(0.062-0.047)$, giving a reduction in the number of stunted children of only 1%. This result depends partly on the assumption that the effective literacy level of proximate illiterates is achieved with just one literate adult in the household so the addition of further literates has no effect.

Although the other variables are included in the model mainly so that inferences about R and P do not suffer omitted variables bias, it is worthwhile briefly discussing their coefficient

estimates. The average standardised height of young children falls with increases in their age, which is common with the pattern found in other developing countries. The average height increases with average per capita expenditure levels in the community, although possibly at a declining rate because the coefficient on squared expenditure is statistically significant at the $p=0.066$ level. This result is interesting because with the exception of the results reported by Thomas and Strauss (1992) for their 'illiterate mothers' sub-sample, most studies of child height in developing countries have not found significant quadratic terms in per capita expenditures. It also appears that the average height of young children in PNG is higher in the urban communities, which tend to have better public infrastructure, although this effect is imprecisely measured. Finally, in common with other research in PNG, the model suggests that *ceteris paribus* the height of children is lower in the highlands.¹⁸

When the model is estimated at household level, the results are very similar to the community level estimates (see Appendix Table 1). Both R and P are statistically significant predictors of the standardised height of individual children, with the significance of P providing support to the Basu and Foster measure of effective literacy. The magnitude of the coefficients on both R and P are lower than in the community level estimates, but the ratio of the coefficients, which indicates the magnitude of \mathbf{a} according to equation (6), is higher. Specifically, the effective literacy level of a proximate illiterate is estimated as 0.86, with a standard error of 0.22. The hypothesis that $\mathbf{a}=0.76$ (the estimate from community level data) is not rejected ($p<0.63$), so the similarity of the estimates from community level and household level data is reassuring. The estimated coefficients on the control variables in the household level regression also have similar signs and statistical significance levels to the community level estimates, except that quadratic on per capita expenditures becomes statistically insignificant while the urban dummy variable becomes significant.

5. A GENDER-ADJUSTED MEASURE OF EFFECTIVE LITERACY

Thus far, it has been assumed that the external benefits from having a literate member of the household are independent of the characteristics of household members. However, some research suggests that at least one characteristic – gender – may affect the size of any external benefits of literacy. For example, infant and child mortality appears to be influenced more by maternal literacy (or education) than by paternal literacy (Caldwell, 1979; Mensch, *et. al.*, 1986).¹⁹ To allow for gendered externalities, Basu and Foster (1998) distinguish between two types of proximate illiterates: *m-proximate illiterates* are those living in households with at least one literate male and no literate females, and *f-proximate illiterates* are those in households containing at least one literate female. With this distinction the indicator variable for effective literacy now takes four values: zero for isolated illiterates, \mathbf{a}_m for each *m*-proximate illiterate, \mathbf{a}_f for each *f*-proximate illiterate, and one for each literate, and it is expected that $0 < \mathbf{a}_m < \mathbf{a}_f < 1$. The gender-adjusted measure of effective literacy is a generalisation of equation (3):

$$** = R + \mathbf{a}_m P_m + \mathbf{a}_f P_f \quad (7)$$

where P_m and P_f are the share of *m*-proximate and *f*-proximate illiterates in the population.

Empirical estimates of \mathbf{a}_m and \mathbf{a}_f can be made in the same way that \mathbf{a} is estimated – by including P_m , P_f , and R in a model that predicts some outcome that depends on literacy skills. Hence the differential between \mathbf{a}_m and \mathbf{a}_f will vary with what is being predicted, although it should be possible to form an overall view of whether it is better to have access to a literate female or a literate male after estimates are made in several contexts. To see whether there is a

gender differential in the current setting, the model has been revised by disaggregating the proximate illiterate rate in each community into its male and female components, P_m and P_f .

(Table 4 about here)

The results from the model of standardised child height suggest that Basu and Foster (1998) are correct in expecting a larger positive externality when illiterates have access to a literate female than when they have access to a literate male (Table 4). In particular, the coefficients on P_f and P_m imply that $\mathbf{a}_f = 0.82$ and $\mathbf{a}_m = 0.74$. However, once the proximate illiteracy rate is disaggregated into its m -proximate and f -proximate components, the point estimates become imprecisely measured (especially the coefficient on P_f). Therefore, the null hypothesis that $\mathbf{a}_m = \mathbf{a}_f$ is not rejected ($p=0.937$). Consequently, the calculation of the gender-adjusted ** measure is left as a topic for further research.

6. SUMMARY AND IMPLICATIONS

Usual measures of literacy ignore the positive, intrahousehold, externality created when a literate household member provides literacy services to illiterate members. Basu and Foster (1998) have developed a new measure of literacy that takes account of this externality by assigning an effective literacy level, \mathbf{a} to each proximate illiterate who lives in a household where at least one person is literate. An estimate of $\mathbf{a} = 0.76$ has been made in this paper, using a method based on an analogy with the literature on measuring economies of household size. More specifically, the estimate of \mathbf{a} comes from a regression model where both the adult literacy rate, R and the proportion of proximate illiterate adults, P in a community influences children's height-for-age in Papua New Guinea. If there were no intrahousehold externalities

created by literacy, the proportion of proximate illiterates would have no role to play in explaining children's height.

A number of practical implications follow from these results. Perhaps most importantly, the ranking of Papua New Guinea regions changes, once the adult literacy rate, R is replaced with the measure of effective literacy, R^* . This is apparent from a comparison of Table 1 and Table 5 which shows that the South Coast region has a higher adult literacy rate than the North Coast, but a lower level of effective literacy. The reason for this is that even though a lower proportion of adults are literate in the North Coast, they are more evenly distributed amongst households, so that the effective literacy level exceeds that of the South Coast region. Thus, the rankings of regions for the PNG *Human Development Report* may change if the measure of effective literacy was used instead of the adult literacy rate (although the other three regions do not change their rank). This re-ranking could make some difference to the regional targeting of programs that aim to improve human development. One can also infer from this result that the human development index rankings of countries might change if R^* was used as the measure of literacy.

(Table 5 about here)

A second implication of finding external benefits (at least in terms of child height) when literate household members provide literacy services to illiterate members is that it focuses greater attention on the group of isolated illiterates. People in households where everyone is illiterate may be the group most in need of interventions designed to improve literacy. There are a number of research questions that need to be answered about isolated illiterates, including what their characteristics are so that they can be better targeted and whether they

currently get literacy services outside of the household (e.g., from extended family, clan, pastors, public servants).

The results reported here also have implications for the econometric modelling of child height (and weight). It is usual in this literature for measures of parental literacy and/or education to be included as explanatory variables (Strauss and Thomas, 1995). However, these variables do not capture all sources of literacy services within the household because the children of illiterate parents may still benefit, in terms of better health and nutrition, if there are other household members who are literate. Thus it may be worthwhile for these models of child height to include an additional variable – whether anyone at all within the household is literate.

(Appendix Table 1 about here)

ACKNOWLEDGEMENTS

Data used in this paper were originally collected as part of a World Bank poverty assessment for Papua New Guinea, for which financial support from the governments of Australia (TF-032753), Japan (TF-029460), and New Zealand (TF-033936) is gratefully acknowledged. All views in this paper are those of the author and should not be attributed to the World Bank. I am grateful to Kathi McKay, John Tressler and two anonymous referees for helpful comments.

REFERENCES

- Basu, K. and Foster, J. (1998) On measuring literacy. *Economic Journal* **108** (Nov), 1733-1749.
- Caldwell, J. (1979) Education as a factor in mortality decline: an examination of Nigerian data. *Population Studies* **33** (3), 395-413.
- Fallon, J. (1992) *The Papua New Guinea Economy: Prospects for Recovery, Reform and Sustained Growth*. Australian International Development Assistance Bureau, Canberra.
- Foster, A. and Rosenzweig, M. (1998) Household public goods, household division, and rural economic growth. Brown University, Providence.
- Gannicott, K. and Avalos, B. (1994) *Women's Education and Economic Development in Melanesia*. National Centre for Development Studies, Canberra.
- Gibson, J. (1997) Women's education and child growth in urban Papua New Guinea. *Papua New Guinea Journal of Education* **33** (1), 29-31.
- Gibson, J. and Rozelle, S. (1998) Results of the household survey component of the 1996 poverty assessment for Papua New Guinea. Population and Human Resources Division, The World Bank, Washington DC.
- Green, S., Rich, T. and Nesman, E. (1985) Beyond individual literacy: the role of shared literacy for innovation in Guatemala. *Human Organization* **44** (4), 313-321.

Gould, D. (1995) Human capital, trade, and economic growth. *Weltwirtschaftliches Archiv* **131** (3), 425-445.

Haddad, L. (1999) The income earned by women: impacts on welfare outcomes. *Agricultural Economics* **20** (2), 135-141.

Heywood, P., Singleton, N. and Ross, J. (1988) Nutritional status of young children: the 1982/82 National Nutrition Survey. *Papua New Guinea Medical Journal* **33** (2), 91-101.

Jenkins, C. (1992) Issues in the promotion of improved health and nutrition in Papua New Guinea. In *Proceedings of the Papua New Guinea National Nutrition Policy Workshop*, pp. 177-186. Institute of National Affairs, Port Moresby.

Lanjouw, P. and Ravallion, M. (1995) Poverty and household size. *Economic Journal* **105** (Nov), 1415-1434.

Lavy, V., Strauss, J., Thomas, D. and de Vreyer, P. (1996) Quality of health care, survival and health outcomes in Ghana. *Journal of Health Economics* **15** (3), 333-357.

Lin, S. (1997) Education and economic development: evidence from China. *Comparative Economic Studies* **39** (3), 66-85.

McElroy, M. (1997) The policy implications of family bargaining and marriage markets. In *Intrahousehold Resource Allocation in Developing Countries: Methods, Models and Policies*, ed. L. Haddad, J. Hoddinott, and H. Alderman, pp. 53-74. Johns Hopkins University Press, Baltimore.

Mensch, B., Lentzner, H. and Preston, S. (1986) *Socio-economic Differentials in Child Mortality in Developing Countries*. United Nations, New York.

Mitch, D. (1984) Underinvestment in literacy? The potential contribution of government involvement in elementary education to economic growth in nineteenth century England. *Journal of Economic History* **44** (2), 557-566.

Noorbakhsh, F. (1998) A modified human development index. *World Development* **26** (3), 517-528.

Sahn, D. (1994) The contribution of income to improved nutrition in Côte d'Ivoire. *Journal of African Economies* **3** (1), 29-61.

Sahn, D. and Alderman, H. (1997) On the determinants of nutrition in Mozambique: the importance of age-specific effects. *World Development* **25** (4), 577-588.

Sandiford, P. (1995) The impact of women's literacy on child health and its interaction with access to health services. *Population Studies* **49** (1), 5-17.

- Smith, T., Earland, J., Bhatia, K., Heywood, P., and Singleton, N. (1993) Linear growth of children in Papua New Guinea in relation to dietary, environmental and genetic factors. *Ecology of Food and Nutrition* **31** (1), 1-25.
- Strauss, J. and Thomas, D. (1995) Human resources: empirical modeling of household and family decisions. In *Handbook of Development Economics, Volume III*, ed. J. Behrman and T.N. Srinivasan, pp. 1883-2023. Elsevier, Amsterdam.
- Thomas, D. and Strauss, J. (1992) Prices, infrastructure, household characteristics and child height, *Journal of Development Economics* **39** (2), 301-331.
- Thomas, D., Strauss, J and Henriques, M. (1991) How does mother's education affect child height? *Journal of Human Resources* **26** (2), 183-211.
- Waterlow, J., Buzina, R., Keller, W., Lane, U., Nichaman, M. and Tanner, J. (1977) The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of ten years. *Bulletin of the World Health Organization* **55**, 489-498.
- Wurm, S. (1975) *New Guinea Languages and Language Studies I: Papuan Languages and the New Guinea Linguistic Scene*. Australian National University, Canberra.
- Yala, C. and Levantis, T. (1998) Recent economic developments in Papua New Guinea: a continuing drought in development. *Pacific Economic Bulletin* **13** (2), 1-16.

Table 1. *Adult literacy in Papua New Guinea, 1996*

Region	Literates	Proximate Illiterates	Isolated Illiterates	Gender Gap in Literacy
	<i>% of adult population</i>			
National Capital District	85.8	12.1	2.1	5.3
South Coast (Papuan)	59.0	23.4	17.6	15.8
Highlands	34.6	37.7	27.7	17.8
North Coast (Momase)	56.3	28.6	15.1	25.1
New Guinea Islands	77.7	15.2	7.1	5.8
Papua New Guinea	51.9	29.1	19.0	18.2

Note:

Estimates are for the population age 15 years and above.

The Gender Gap is the male adult literacy rate minus the female adult literacy rate.

Table 2. *Data definitions and description*

Variable	Unit	Mean ^a	Standard deviation ^a
Standardised height of children ^b	Percent	93.0	3.5
Adult literacy rate (R)	[0,1]	0.503	0.262
Proximate illiteracy rate (P)	[0,1]	0.283	0.160
Male-proximate illiteracy rate (P_m)	[0,1]	0.240	0.133
Female-proximate illiteracy rate (P_f)	[0,1]	0.043	0.061
Proportion of male children	[0,1]	0.529	0.177
Age of children	Months	28.6	5.6
Expenditure per equivalent person ^c	K'000/year ^d	1.042	0.870
Urban community (=1, otherwise=0)	0/1	0.134	0.342
Highlands region (=1, otherwise=0)	0/1	0.390	0.491

^aWeighted estimates, with weights reflecting the number of households represented by each observation.

^bPercentage of the median height for children in the United States of the same age and gender.

^cChildren age 0-6 years count as 0.5 equivalent persons, all others count as 1.0.

^dIn national average prices, where the value of the regional poverty line is used as the spatial price deflator and K1.3=US\$1 in 1996.

Table 3. *The effect of adult literacy, proximate illiteracy and other household and community characteristics on the average standardised height of young children^a*

Variable	Beta	Standard Error	Significance
Adult literacy rate (<i>R</i>)	0.062	0.018	0.001
Proximate illiteracy rate (<i>P</i>)	0.047	0.024	0.049
Age of children	-0.001	0.001	0.067
Expenditure per equivalent person	0.024	0.012	0.045
Squared expenditure per equiv. person	-0.003	0.002	0.066
Urban community (=1, otherwise=0)	0.016	0.013	0.215
Highlands region (=1, otherwise=0)	-0.016	0.008	0.064
Intercept	4.503	0.018	0.000

$N=81$, Adjusted $R^2=0.470$, $F_{(7,73)}=11.13$ ($p=0.000$)

^aDependent variable is the logarithm of the average standardised height of young children (percentage of the median height of a child in the United States of the same age and sex) in each community. Results are estimated by OLS and are weighted by the number of households that each observation represents.

Table 4. *The gender-disaggregated effects of proximate illiteracy on the average standardised height of young children^a*

Variable	Beta	Standard Error	Significance
Adult literacy rate (R)	0.062	0.018	0.001
Male-proximate illiteracy rate (P_m)	0.046	0.027	0.089
Female-proximate illiteracy rate (P_f)	0.051	0.058	0.379
Age of children	-0.001	0.001	0.070
Expenditure per equivalent person	0.024	0.012	0.047
Squared expenditure per equiv. person	-0.003	0.002	0.069
Urban community (=1, otherwise=0)	0.016	0.013	0.218
Highlands region (=1, otherwise=0)	-0.016	0.009	0.068
Intercept	4.504	0.018	0.000

$N=81$, Adjusted $R^2=0.463$, $F_{(8,72)}=9.61$ ($p=0.000$)

^aSee notes to Table 3.

Table 5. *Effective adult literacy rates in Papua New Guinea, 1996*

Region	*
National Capital District	95.0
South Coast (Papua)	76.7
Highlands	63.2
North Coast	78.0
New Guinea Islands	89.3
Papua New Guinea	74.0

Note:

Estimates are for the population age 15 years and above.

* is the measure of literacy defined by Basu and Foster (1998), using $\alpha=0.76$.

Appendix Table 1. *Household level estimates of the effect of adult literacy, proximate illiteracy and other characteristics on the standardised height of young children^a*

Variable	Beta	Standard Error	Significance
Adult literacy rate (<i>R</i>)	0.038	0.008	0.000
Proximate illiteracy rate (<i>P</i>)	0.033	0.011	0.004
Age of children	-0.001	0.000	0.000
Expenditure per equivalent person	0.011	0.006	0.056
Squared expenditure per equiv. Person	-0.000	0.000	0.241
Urban community (=1, otherwise=0)	0.029	0.007	0.000
Highlands region (=1, otherwise=0)	-0.019	0.008	0.015
Intercept	4.522	0.010	0.000

$N=968$, Adjusted $R^2=0.185$, $F_{(7,960)}=31.02$ ($p=0.000$)

^aDependent variable is the logarithm of the standardised height of young children (percentage of the median height of a child in the United States of the same age and sex). Results are estimated by OLS and are weighted by the number of households that each observation represents.

NOTES

¹ The Human Development Index gives a one-third weight to life expectancy, a one-third weight to adjusted average income, a two-ninths weight to the adult literacy rate, and a one-ninth weight to an educational enrolment index (Noorbakhsh, 1998).

² There is no need to limit these external effects to the household. A person who can read may provide literacy services to people in their wider social network, including friends and kin who live in other households. But as a practical matter, more information is available on household membership than on social network membership.

³ These reading services could also be provided by people outside the household but it would be difficult for econometric models of child height to include a variable measuring this effect.

⁴ The extension to the case where \mathbf{a} varies between males and females is discussed below.

⁵ Basu and Foster assume that the effective literacy level of proximate illiterates is achieved by having just one literate person in the household, with the addition of further literates having no effect on \mathbf{a} .

⁶ I am grateful to an anonymous referee for suggesting this point.

⁷ In rural India, Foster and Rosenzweig (1998) find that intrahousehold inequality in schooling raises the probability of household division (controlling for the maximum schooling level in the household) while increases in maximum schooling level for a given mean and variance within

the household result in decreases in household division. No results are available from this study using literacy variables.

⁸ There is a close parallel with measures of household size economies, which may vary for different goods (e.g., food versus heat) and depending upon whether the welfare of children or of adults is of concern (Lanjouw and Ravallion, 1995).

⁹ The *Human Development Report* gives an adult literacy rate of 72.2 percent for PNG in 1995, which is consistent with the estimates of adult male and adult female *illiteracy* for the same year in the *World Development Report*. However, Gannicott and Avalos (1994) report male and female literacy rates for 1990 that imply an overall adult literacy rate of only 52 percent, while Fallon (1992) reports an adult literacy rate for 1989 of only 46 percent.

¹⁰ To the extent that literacy influences employment and wages, it may also help to raise household incomes which allows extra spending on food and health care for children. This effect is controlled for by including per capita expenditures in the model.

¹¹ The rate of change in literacy in PNG is difficult to determine because the results from the most recent Census (1990) cannot be compared with the previous 1980 Census, which did not include any questions on literacy.

¹² Shapiro-Wilk tests show that the normality of R and P cannot be rejected when they are defined at community level (z -statistics of 0.51 and 1.22) but is conclusively rejected when

they are defined at household level (z -statistics of 11.1 and 11.9). This non-normality is due to the mass points that result when R and P are calculated at household level.

¹³ The adult equivalence scale counts children aged 0-6 years as one-half of an adult and everyone else as an adult. This scale is based on estimates of child costs, made using the Engel and Rothbarth methods, and on a comparison of the dietary requirements of adults and children of various age groups. Details are provided by Gibson and Rozelle (1998).

¹⁴ The Bougainville crisis prevented the Census from being conducted in North Solomons province in 1990, so there was no information available for the selection of communities, even if the safety of interview teams could have been guaranteed in 1996.

¹⁵ The expenditure data include the imputed value of own-production, net gifts received, and food stock changes, and also include estimates of the value of services provided by durable goods and owner-occupied dwellings.

¹⁶ Results from this initial regression are available from the author.

¹⁷ The basis of this is as follows: a 10 percentage point increase in R would make almost 300,000 adults literate, while there are only 200,000 households where all adults are illiterate.

¹⁸ Removing the *HIGHLANDS* dummy variable from the model affects the size and statistical significance of a number of variables, especially PCX and its square, which become

insignificantly different from zero. There is also a small fall in the size of the coefficient on the proximate illiteracy rate, from 0.047 to 0.040 and the t -statistic on P falls from 2.00 to 1.78.

¹⁹ There is less consensus that child height is more sensitive to maternal literacy (or education) than it is to paternal literacy: Thomas *et. al.* (1991) find that mother's education has more effect on child height than does father's education in one Brazilian survey but Thomas and Strauss (1992) do not find this pattern in another Brazilian survey and neither do Lavy *et. al.* (1996) in a Ghanaian survey.