SCHOOL REPETITION, DROPOUTS AND THE RETURNS TO SCHOOL: THE CASE OF INDONESIA

by

Jere R. Behrman and Anil B. Deolalikar*

October 29, 1990

RP-148

* Williams College, Williamstown, Massachusetts
SCHOOL REPETITION, DROPOUTS AND THE RETURNS TO SCHOOL: THE CASE OF INDONESIA

by

Jere R. Behrman and Anil B. Deolalikar*

October 29, 1990

© 1990-(Jere R. Behrman and Anil B. Deolaikar)

Note: This paper is intended for private circulation and should not be quoted or referred to in publication without the permission of the authors.

*Behrman is the William R. Kenan Jr. Professor of Economics at the University of Pennsylvania, Philadelphia, PA and the Arnold Bernhard Visiting Professor of Economics at Williams College, Williamstown, MA for the 1990-1 academic year. Deolaikar is Associate Professor of Economics at the University of Washington, Seattle, WA.
SCHOOL REPETITION, DROPOUTS AND THE RETURNS TO SCHOOL:
THE CASE OF INDONESIA

by
Jere R. Behrman and Anil B. Deolalikar*

29 October 1990

ABSTRACT

Standard estimates of the rates of return to primary schooling in most developing
countries are high, and have been used to support advocacy of increased investments
in primary schooling. But the standard estimates ignore repetition and dropout
experience. This paper develops a procedure for estimating the impact of repetition
and dropout rates and applies it to Indonesian data. The results are striking,
suggesting that standard procedures overstate substantially the economic returns to
schooling in Indonesia (e.g. by 38 to 78% for primary schooling), distort the pattern
of estimated returns across schooling levels by overestimating especially the returns
to the lower schooling levels, and misrepresent the relative returns to schooling invest-
ments among groups identified by sex, region, rural-urban residence, and age.

*Behrman is the William R. Kenan Jr. Professor of Economics at the University of
Pennsylvania, Philadelphia, PA 19104-6297 U.S.A. and the Arnold Bernhard Visiting Professor of
Economics at Williams College, Williamstown, MA 01267 U.S.A. for the 1990-1 academic year.
Deolalikar is Associate Professor of Economics at the University of Washington, Seattle, WA 98195
U.S.A. The authors are grateful to Martin Godfrey for helpful discussions on this subject and to
Stephen Nickell and other members of the editorial board for helpful suggestions on an earlier draft.
The authors alone are responsible for all interpretations given in this paper.
1. Introduction

Most developing countries have expanded greatly investment in schooling in the past two decades. One index of this expansion is the very considerable growth in enrollment rates as a proportion of the eligible population: for currently low-income countries from 73 and 20% for primary and secondary schooling, respectively, in 1965 to 104 and 37% in 1987 and for current middle-income countries from 92 and 26% in 1965 to 104 and 54% in 1987. For Indonesia, the subject of the present study, comparable percentages are from 72 and 12% in 1965 to 118 and 46% in 1986. The Indonesian increases, thus, are large in comparison with those of other developing countries (also see Behrman 1987, 1988).

A major argument for the expansion of schooling in the developing world is that studies of the rate of return to schooling in terms of wage rates and other labour market outcomes suggest high rates of return to investment in schooling, particularly primary schooling. World Bank (1980), and Psacharopoulos (1985), for example, summarize existing evidence as implying average social rates of return to primary schooling of about 24% and lower, but still considerable, rates of return for secondary and higher education. Available estimates for Indonesia are of a similar order of magnitude.

To obtain such estimates, however, some assumptions must be made about how much time students spend in school in order to attain the grade level or to complete the schooling category that they report. If grade repetition and school dropouts are common (as is the case in much of the developing world), the standard practice of equating the number of grades necessary to complete a given level of schooling with the number of years spent in that level may understate the private and social costs of schooling and thus overstate the private and social returns to schooling. In this study we develop procedures for dealing with and estimating such a bias under alternative assumptions.

---

1. Based on data in World Bank (1990). Percentages over 100% reflect schooling enrollments larger than the population normally relevant for a given schooling level (e.g. ages 6-11 for primary schooling). Such percentages are not uncommon for developing countries with rapidly expanding primary school systems that are attended by numerous students over the normal primary school age.

about the extent of heterogeneity among students. Whether such adjustments are important or not is an empirical question. Therefore we apply these methods to data from Indonesia, the fifth most populous country in the world. We find that, if repetition and dropout rates are ignored, the biases are considerable -- ranging from 38 to 78% for primary schooling. The standard procedure results in considerable overestimates of the rates of return to schooling. Moreover the bias is largest for the lower schooling levels. Thus the advocacy of the World Bank (1980, 1981, 1990), Colclough (1982), Eisemon (1988), Psacharopoulos (1985, 1988), and UNDP (1990) for focusing on primary schooling, to the extent that it is based on the standard estimates, is overstated. Moreover, if repetition and dropout rates vary across schooling levels or among various demographic groups, ignoring such phenomena may cause distortions in our understanding of relative rates of return.

We first present some new estimates of the rates of return to schooling in Indonesia and illustrate how repetition and dropout rates enter into these estimates. Table 1 gives estimates of the private rate of return to schooling, given a Mincerian (1974) interpretation, in Indonesia based on the 1986 Labour Force Survey (SAKERNAS). The first column gives household fixed-effects estimates of a semilog wage function for all 25,555 individuals who report wages in the national sample of about 225,000 individuals at least 10 years of age. The rows give the estimates of the

---

3 It is well-known that there are a number of other problems in interpreting wage-schooling associations to imply that schooling causes the observed wage differences. Some of these problems pertain to the usual failure to control for differences in schooling quality, in price levels, in ability and motivation, and in community characteristics that may affect schooling and wages. See Behrman (1990a,b,c) for reviews of some of these problems and for references to related studies. In this paper we do not focus on such problems (though we attempt to control for many of them by controlling for community and household fixed effects in the estimates in column one of Table 1), but abstract from them in order to concentrate on the bias introduced in the standard treatment of schooling by the failure to incorporate schooling repetition and dropouts.

4 This wage function is based on the subsample of all individuals who report wages. The use of such a subsample raises the possibility of selectivity bias since those who work in wage employment probably are not a random sample. We control for selectivity by controlling for unobserved household fixed effects in these estimates (which also controls for unobserved community effects) since all household members in the sample live in the same community), which other studies (e.g. Heckman and MaCurdy 1980, Pitt and Rosenzweig 1990) argue eliminates selectivity bias. Such a control uses all of the information available in this sample for selectivity corrections. See Behrman and Deolalikar (1990a) for elaboration.
Table 1
Estimated Wage Relations and Private Rates of Returns to Schooling in Indonesia, 1986
(Household Fixed Effects Estimates)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T-Ratio</th>
<th>Implied Rate of Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.044</td>
<td>14.54</td>
<td></td>
</tr>
<tr>
<td>Age squared (X 1000)</td>
<td>-0.400</td>
<td>-9.67</td>
<td></td>
</tr>
</tbody>
</table>

Educational Attainment:

<table>
<thead>
<tr>
<th>Educational Attainment</th>
<th>Estimate</th>
<th>T-Ratio</th>
<th>Implied Rate of Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subprimary</td>
<td>0.160</td>
<td>7.14</td>
<td>W 5.3%</td>
</tr>
<tr>
<td>Primary</td>
<td>0.299</td>
<td>12.42</td>
<td>O 7.6%</td>
</tr>
<tr>
<td>General Junior Secondary</td>
<td>0.683</td>
<td>18.89</td>
<td>M 9.2%</td>
</tr>
<tr>
<td>Vocational Junior Secondary</td>
<td>0.832</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>General Senior Secondary</td>
<td>1.114</td>
<td>33.9</td>
<td>N 9.3%</td>
</tr>
<tr>
<td>Vocational Senior Secondary</td>
<td>1.278</td>
<td>49.69</td>
<td></td>
</tr>
<tr>
<td>Diploma 1,2</td>
<td>1.460</td>
<td>21.71</td>
<td></td>
</tr>
<tr>
<td>Diploma 3</td>
<td>1.507</td>
<td>29.99</td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>1.762</td>
<td>24.78</td>
<td></td>
</tr>
</tbody>
</table>

Whether male

| Whether male                                | 0.062    | 0.91    | 11.7%                  |

Dichotomous Variable for Male Interacted with:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T-Ratio</th>
<th>Implied Rate of Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.018</td>
<td>4.75</td>
<td></td>
</tr>
<tr>
<td>Age squared (X 1000)</td>
<td>-0.193</td>
<td>-3.86</td>
<td></td>
</tr>
<tr>
<td>Subprimary</td>
<td>-0.010</td>
<td>-0.32</td>
<td>M 5.0%</td>
</tr>
<tr>
<td>Primary</td>
<td>0.020</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>General Junior Secondary</td>
<td>-0.115</td>
<td>-.73</td>
<td></td>
</tr>
<tr>
<td>Vocational Junior Secondary</td>
<td>-0.189</td>
<td>-2.73</td>
<td></td>
</tr>
<tr>
<td>General Senior Secondary</td>
<td>-0.295</td>
<td>-7.68</td>
<td></td>
</tr>
<tr>
<td>Vocational Senior Secondary</td>
<td>-0.370</td>
<td>-11.47</td>
<td></td>
</tr>
<tr>
<td>Diploma 1,2</td>
<td>-0.514</td>
<td>-6.32</td>
<td></td>
</tr>
<tr>
<td>Diploma 3</td>
<td>-0.239</td>
<td>-4.07</td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>-0.235</td>
<td>-3.07</td>
<td></td>
</tr>
</tbody>
</table>

R-Squared                                    | 0.320    |         |                        |
F-Ratio                                      | 522.612  |         |                        |
Number of observations                       | 25555    |         |                        |
impact on log wages of having completed the indicated schooling levels, which are the categories given in the Labour Force Survey. In each case there is a test to see if the coefficient estimate for males differs from that for females. The estimates do not differ significantly by sex for subprimary, primary and general junior secondary school completers. However, for vocational junior secondary school completers, and both vocational and general secondary and post-secondary school completers, women receive significantly greater returns to such schooling than do men.\(^5\) The coefficient estimates in this column are estimates of the impact of completing the different schooling levels. They imply that the effect of completing primary schooling as compared to having no primary schooling, for example, is to increase the wage rate by 35% \(( = \exp .299)\).\(^6\)

To obtain the Mincerian private rate of return to spending time in school, we need to adjust these estimates for the time spent in each schooling level. Standard practice is to use six years for primary schooling since there are six grades of primary schooling, three years each for junior and senior secondary schooling since there are three grades in each, etc. Subprimary (less than primary) often is ignored or assumed to be three years --half of the number of grades in primary schooling. Column three gives the estimated Mincerian private rates of return to different schooling levels for males and females, using the standard assumptions about the duration of study at each level. These

---

\(^5\)This does not mean that women receive higher wages than men with equal educational levels. The significant positive additive dichotomous estimate for being male implies that males with no schooling receive on the average wage rates that are about a third (33.7% to be exact since \(\exp .2905 = 1.337\)) higher than females with no schooling (ignoring the age and age\(^2\) interactions with sex, which imply a further wage differential for young males, though the opposite is the case for older males). The apparent higher returns to higher levels of schooling for females than for males only partially offsets the wage advantage of males that is independent of schooling. For further discussion, see Behrman and Deolalikar (1990b).

\(^6\)Sometimes the coefficient itself is interpreted as indicating the percentage increase in the wage since the semilog form implies that \(dw/w = r\) if \(\ln w = r S_p\), where \(w\) is the wage and \(S_p\) is the dichotomous variable for whether or not an individual has completed primary schooling. But this procedure compares the difference in the wage due to primary schooling with the average of the wage with and without primary schooling.
estimates imply private rates of return to different levels of schooling that range from 5% to 11.7%. But the basic point of interest for this study is that these estimates depend directly on the assumptions about the time spent in school based on the respondent's reported completion of a given schooling category. The standard assumption is that a respondent who completed primary school (but not post primary school) spent six years in school. If such respondents on average spend eight years in school owing to repetition and dropouts, the rate of return to those who responded that they completed primary schooling would be overstated by a third.

2. Repetition, Dropouts and the Length of Schooling in Indonesia

Schooling repetition and dropout rates in Indonesia and in many developing countries are nonnegligible. Table 2, for example, gives such rates grade-by-grade for primary and secondary schooling in Indonesia. Repetition rates are as high as 13.9% (grade 1 of primary school) and dropout rates are as high as 8.9% (grade 3 of upper secondary school). Our basic point is quite simple: when survey respondents indicate that they have completed a certain level of schooling, the existence of positive repetition and dropout rates means that on the average they have spent more

---

7 There are the significant differences noted above for females versus males, and a range across schooling-sex combinations of the order of magnitude of one to two (i.e. 11.7/5 = 2.3). But these estimates are relatively low as compared with earlier ones for Indonesia (see note 2), and imply limited variation across schooling levels. Our estimates apparently are lower than others for Indonesia (i) because we control for unobserved household effects that cause upward biases if they are not controlled and if they are correlated with schooling (e.g. if children from better family backgrounds tend to obtain more schooling and to obtain better paying jobs because of their backgrounds), (ii) because we focus on wage rates and not earnings (that confound labor supply effects with the impact of schooling on wage rates), and, at least in comparison with the other estimates based on the 1986 SAKERNAS survey in McMahon and Boediono (1988), because (iii) we use all employees and not just urban employees.

8 These rates are in addition to the even higher dropout rates at the completion of schooling levels: 32.4% of primary completers do not continue in junior secondary school, 22.1% of junior secondary completers do not continue in senior secondary, and an estimated 60% of secondary completers do not continue in post-secondary education. Dropouts at the end of schooling levels, however, do not cause the same problems in estimating the rates of return to schooling as do dropouts within schooling levels given current data collection procedures.
Table 2
Repetition, Dropout, and Progression Rates, by Grades, Indonesia, 1986

<table>
<thead>
<tr>
<th>Grade</th>
<th>Repetition Rate</th>
<th>Dropout Rate</th>
<th>Success Rate</th>
<th>Progression Rate given success</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.139</td>
<td>0.025</td>
<td>0.836</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>0.099</td>
<td>0.027</td>
<td>0.874</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>0.086</td>
<td>0.042</td>
<td>0.872</td>
<td>1.000</td>
</tr>
<tr>
<td>4</td>
<td>0.068</td>
<td>0.053</td>
<td>0.879</td>
<td>1.000</td>
</tr>
<tr>
<td>5</td>
<td>0.049</td>
<td>0.047</td>
<td>0.904</td>
<td>1.000</td>
</tr>
<tr>
<td>6</td>
<td>0.012</td>
<td>0.040</td>
<td>0.948</td>
<td>1.000</td>
</tr>
<tr>
<td>7</td>
<td>0.015</td>
<td>0.060</td>
<td>0.925</td>
<td>0.676</td>
</tr>
<tr>
<td>8</td>
<td>0.016</td>
<td>0.050</td>
<td>0.934</td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td>0.020</td>
<td>0.051</td>
<td>0.929</td>
<td>1.000</td>
</tr>
<tr>
<td>10</td>
<td>0.013</td>
<td>0.082</td>
<td>0.905</td>
<td>0.779</td>
</tr>
<tr>
<td>11</td>
<td>0.016</td>
<td>0.051</td>
<td>0.933</td>
<td>1.000</td>
</tr>
<tr>
<td>12</td>
<td>0.026</td>
<td>0.089</td>
<td>0.885</td>
<td>1.000</td>
</tr>
</tbody>
</table>

years in that level of schooling than a simple count of the grades for that level of schooling indicates.\textsuperscript{9} If one makes the standard assumption that the number of grades of schooling for that level gives the number of years spent in that level of schooling, therefore, one underestimates the average private and social costs of schooling and overestimates the private and social rates of return. For example, those who report that they completed primary schooling in the Indonesian case may have spent more than six years in school for two reasons. First, they may have repeated one or more of the six primary grades. Second, they may have begun, but not completed, junior secondary school (in which case, given the choices given in the survey, completed primary schooling is recorded).\textsuperscript{10} Whether this conceptually quite simple point makes much difference, of course, is an empirical question.

To explore the importance of this question, we need to develop expressions for the expected numbers of years spent in school for individuals who are recorded in the various schooling categories in the survey, dependent on grade-specific repetition and dropout rates.\textsuperscript{11} We do so under two alternative assumptions. First, we assume that everyone who enters a given grade is subject to the same average repetition and dropout rates. Second, we assume that those students who enter the next schooling level have zero values for the repetition and dropout rates at all lower levels and those students who do not enter the next level have appropriately adjusted (i.e. appropriately adjusted so that the average repetition and dropout rates are the same as under the first assumption) repetition

\textsuperscript{9}If students skip grades the opposite phenomenon may be important. We have not been able to find any evidence that the skipping rates are important empirically.

\textsuperscript{10}Many surveys inquire about grades of schooling or years of schooling rather than using categories such as those used by the Indonesian Labour Force survey. If so, they are not likely to have the second problem. Those that inquire about grades of schooling still have the first problem. Those who inquire about years of schooling have neither problem if they in fact receive an answer to that question. However, some nonrandom sampling suggests that individuals often answer the question about years of schooling as if they had been asked about the grade of schooling.

\textsuperscript{11}We do not have data on individual-specific dropout and repetition experience, so we cannot explore the determinants of that experience. Nor do we have data that permit the exploration of the possibility that those with greater ability are less likely to dropout or to repeat grades and are more likely to go to school longer and to obtain higher wages.
and dropout rates at the lower levels. This second alternative probably overstates somewhat the
duration of time in school for those who terminate their schooling at lower levels and understates
somewhat the duration of time in primary school for those who complete higher levels of school. But
we judge that it is a closer approximation to reality than is the first assumption since that assumption
implies, for instance, that the probability of repetition in the first grade is the same 13.9% for
someone who subsequently permanently dropouts from school at a subprimary level as for someone
who eventually obtains an university degree. Our preferred, second assumption implies that the
eventual university graduate does not repeat primary grades.

Assumption 1: Homogeneity -- every one has average repetition and dropout rates: We now
consider how to estimate the expected years spent in school for each schooling level indicated in the
survey under the first assumption. For both assumptions regarding the repetition and dropout rates
for different individuals, we assume that a student who drops out in a given year drops out at the
mid-point of the year, so the time spent in school during the dropout year is 0.5 years.

Let $N_i$ be the number of individuals that enter grade $i$, $r_i$ be the repetition rate for that
group, $d_i$ be the dropout rate for that group, and $s_i$ be the success rate for that group. These last
three rates refer to mutually exclusive and exhaustive categories, so:

\[ r_i + d_i + s_i = 1. \]

The way that the data in Table 3 are defined, for most grades the success rate is the same as the
progression rate in the sense that those who succeed in grade $i$ progress to grade $i+1$. For the grades
that are the terminal years for each schooling level, however, there is an additional attrition in the
sense that not everyone who successfully completes one schooling level enters the next schooling level.
Of those who successfully complete primary school, $p_p = .676$ enter junior secondary school; of those

---

12 To illustrate, we assume that the students who enter secondary schooling never repeat
grades nor dropout of primary school and that all the primary repetition and dropout experience
that is reflected in the average rates is due to the repetition and dropout experience of the
primary students who never enter secondary school.

13 We also assume that dropouts never return to school in both cases, though if they do return
it would not cause any changes in our analysis unless the interruption reduced (due to lack of
continuity) or increased (due to greater maturity) the effectiveness of subsequent schooling. We
have no way of exploring such possibilities with the Indonesian data.
who successfully complete junior secondary school, \( p_j = .779 \) enter senior secondary school; and of those who successfully complete senior secondary school, \( p_s = .400 \) enter post-secondary school. Given these definitions, we now define the expected years spent in school for an individual in each of the schooling categories:

\[
y_{sp} = \sum_{i=1}^{6} d_i \sum_{j=1}^{\infty} (j-0.5) r_{ij}^{j1} + \sum_{i=1}^{6} s_i \sum_{j=1}^{\infty} j r_{ij}^{j1},
\]

\[
y_p = \sum_{i=1}^{6} \sum_{j=1}^{\infty} r_{ij}^{j1} + \left( \frac{N_p}{N_1} \right) \left[ \sum_{i=1}^{9} d_i \sum_{j=1}^{\infty} (j-0.5) r_{ij}^{j1} + \sum_{i=1}^{9} s_i \sum_{j=1}^{\infty} j r_{ij}^{j1} \right],
\]

\[
y_j = \sum_{i=1}^{9} \sum_{j=1}^{\infty} r_{ij}^{j1} + \left( \frac{N_j}{N_1} \right) \left[ \sum_{i=1}^{12} d_i \sum_{j=1}^{\infty} (j-0.5) r_{ij}^{j1} + \sum_{i=1}^{12} s_i \sum_{j=1}^{\infty} j r_{ij}^{j1} \right],
\]

\[
y_s = \sum_{i=1}^{12} \sum_{j=1}^{\infty} r_{ij}^{j1} + \left( \frac{N_s}{N_1} \right) \left[ \sum_{i=1}^{16} d_i \sum_{j=1}^{\infty} (j-0.5) r_{ij}^{j1} + \sum_{i=1}^{16} s_i \sum_{j=1}^{\infty} j r_{ij}^{j1} \right],
\]

\[
y_u = \sum_{i=1}^{16} \sum_{j=1}^{\infty} r_{ij}^{j1}.
\]

where \( y_k \) is the expected years of schooling for the schooling category, with \( k = sp \) for subprimary, \( k = p \) for completed primary, \( k = j \) for completed junior secondary, \( k = s \) for completed secondary, and \( k = u \) for completed university.

Relation (2) gives the expected years spent in school for individuals who start, but do not complete primary school. The first right-side term gives the time spent in the year in which individuals drop out, which may be in any of the first six grades (so the sum is from one through six). Under the assumption that the individual spends half the year in the grade in which he or she drops out, the expression indicates a half year expected in school in the \((j-0.5)\) term if the individual does not repeat that grade before dropping out (so \( j = 1 \)). But individuals may repeat that grade before dropping out, so there is the sum over \( j \). We have written the sum over \( j \) here and below to go from one to infinity since we do not know the maximum repetition of any grade; this is not a matter of consequence since only the first few terms in this expression have an impact even as large
as a tenth of a year. The second right-side term gives the time spent in primary grades in which (perhaps only after repetition) an individual is successful. This sum is from one through five because those in the subprimary category may be successful in each of the first five grades. Because of the possibility of repeating a grade before being successful, the sum over \( j \) again appears. However under the assumption that a successful year and a repeated year each take a year of time, 0.5 is not subtracted from \( j \) in this case as it is for the dropout year in the first right-side term in relation (2).

Relation (3) gives the expected years in school for those who are in the completed primary school category. The first right-side term gives the number of years expected to be spent in primary school given that the individual finishes primary school and given the repetition rates for the primary grades. The entire second right-side term (in brackets) refers to the expected time that primary school graduates who enter but do not complete junior secondary school (and therefore are still classified as primary school completers) spend in junior secondary school. The multiplicative term in front of the brackets refers to the probability that an individual who enters primary school also enters junior secondary school since this second term is relevant only for those who enter junior secondary school. Within the brackets are two terms that give the expected time that someone spends in junior secondary school conditional on entering but not completing junior secondary school and the repetition and dropout rates for the junior secondary school grades. The interpretations of these terms are parallel to the interpretations of the similar terms in relation (2).

Relations (4) - (6) are similar expressions for junior secondary completers, senior secondary completers, and university completers. These are each parallel to the expression for primary school completers in relation (3), except that the sub and superscripts are changed appropriately.

\[ N_i = p_{i-1} s_{i-1} + r_i N_i, \]

where \( p_6 = p_{p}, p_9 = p_{j}, p_{12} = p_{s}, \) and all other \( p_i = 1 \). Therefore,

\[ N_i = N_1 \sum_{j=2}^{i} p_{j-1} s_{j-1} / (1 - r_j). \]

14To evaluate this probability (and others like it below) one only need observe that \( N_i = p_{i-1} s_{i-1} + r_i N_i, \) where \( p_6 = p_p, p_9 = p_j, p_{12} = p_s, \) and all other \( p_i = 1 \). Therefore,

\[ N_i = N_1 \sum_{j=2}^{i} p_{j-1} s_{j-1} / (1 - r_j). \]

15The Indonesian data have two other categories of post-secondary education besides universities: diploma 1-2 and diploma 3. The expressions for these post-secondary categories are identical to that for university except that the number of years required is less (1-2 and 3 respectively).
Of course relation (6) for the expected years in the university does not have a second right-side term since there is no subsequent schooling category in which university graduates may have spent some time but not completed it in contrast to all of the previous schooling levels.\textsuperscript{16} Also we note that the unavailability of data on post-secondary repetition and dropout rates limits our application of relations (5) and (6) for senior secondary school completers and university completers, respectively. This means that we underestimate the expected time for secondary school completers because we do not have a way to represent the time that individuals in this category spend in uncompleted post-secondary schooling. For a parallel reason we also underestimate the expected years spent in post-secondary education. For both of these categories, nevertheless, we will have higher expected years spent than in the standard estimates because of the repetition at earlier schooling levels that is captured in the first right-side term in each expression.

Assumption 2: Heterogeneity -- zero repetition and dropout rates for those who enter the next schooling level: Not all of those who compete a given schooling level enter the next level. In fact the attrition rates are fairly considerable for those who successfully complete a given level but do not enter the next level (p\textsubscript{p}, p\textsubscript{j}, and p\textsubscript{i}). But under this assumption we assume that those who in fact do enter the next level do not repeat grades in the level of interest (and, of course, they do not dropout). That means that all of the repetition and dropouts captured in the average rates in Table 2 are assumed to be experienced by the individuals who enter a given schooling level, but who do not enter the next schooling level. If we let such individuals be represented by single primes, therefore, for subprimary and the subset of primary completers who do not enter junior secondary school the effective repetition, dropout, success, and critical progression rates for the primary grades (i = 1, ..., 6) are:

\begin{equation}
(7) \quad r_i' = r_i \frac{N_i}{N_i - N_j}, \quad s_i' = \frac{s_i N_i - N_j}{N_i - N_j}, \quad d_i' = d_i \frac{N_i}{N_i - N_j}, \quad \text{and} \quad p_p' = 0.
\end{equation}

Let the comparable rates for those who do enter junior secondary school be represented by double primes. Then for i = 1, ..., 6 the rates for this group are:

\begin{equation}
(8) \quad r_i'' = d_i'' = 0 \quad \text{and} \quad s_i'' = p_p'' = 1.
\end{equation}

\textsuperscript{16}The proportion of Indonesians in graduate education is quite small. One could treat time spent in uncompleted graduate education parallel to our treatment for entering but not completely lower schooling levels, but our data do not provide information with which to pursue such a possibility.
The effective rates for the subprimary category are those in relation (7). The effective rates for those who are in the completed primary school category are a weighted average of the rates in relation (7) and in relation (8), with the weights reflecting the proportion of primary completers who do not versus those who do enter junior secondary school.  

Similar adjustments can be made for those who enter junior secondary school but not senior secondary school (by letting \( i = 7,8,9 \) and replacing \( N_7 \) in relation (7) by \( N_{10} \)) and for those who enter senior secondary school but not post-secondary school (by letting \( i = 10,11,12 \) and replacing \( N_7 \) in relation (7) by \( N_{13} \)). To obtain the appropriate rates for each category in the data, weighting of the single and double primes rates has to be undertaken to reflect the fact that some of the individuals in that category entered (but did not complete) the next level. Expressions (2) - (6) then can be used to calculate the expected years of schooling with the repetition and dropout rates at each schooling level dependent upon the category of completed schooling.  

3. Adjusted Expected Years of Schooling in Indonesia

The expressions and calculations discussed in the previous section, while conceptually straightforward, are somewhat messy to apply. As we say at the start of that section, whether such calculations are worth the effort is basically an empirical question. We now turn to that question for the Indonesian data.

For each response category below post-secondary schooling, Table 3 gives the expected years of schooling firstly under the assumption that every one faces the same repetition and dropout rates and secondly under the alternative assumption of differential rates. For primary and junior secondary schooling, the expected years are divided between those spent in attaining exactly (but no more than) the recorded level and those spent in the subsequent (but not completed) schooling level.

---

17 The proportion of respondents indicated as primary completers who enter junior secondary school is \( (N_6 s_s p_p - N_9 s_p)/(N_6 s_s p_p) \).

18 Such weighting is parallel to that described above for primary school completers.

19 The multiplicative weights in front of the second (bracketed) terms in expressions (3)-(5) also have to be changed to represent the probability of entering the next level: \( (N_7 - N_{10})/N_1 \) for relation (3), etc.
Table 3. Expected Years in Various Schooling Levels and Implied Upward Biases in Standard Estimates Given Alternative Treatments of Repetition and Dropout Rates

<table>
<thead>
<tr>
<th>Assumption 1: Homogeneous Rates</th>
<th>Assumption 2: Heterogeneous Rates</th>
<th>Standard Assumption Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Years of Schooling Given Education Reported as:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subprimary</td>
<td>5.46</td>
<td>6.44</td>
</tr>
<tr>
<td>primary</td>
<td>8.29</td>
<td>10.69</td>
</tr>
<tr>
<td>- of which years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spent (unsuccessfully) in junior secondary school</td>
<td>1.11</td>
<td>0.59</td>
</tr>
<tr>
<td>junior secondary</td>
<td>11.14</td>
<td>10.25</td>
</tr>
<tr>
<td>- of which years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spent (unsuccessfully) in senior secondary school</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>senior secondary</td>
<td>13.30</td>
<td>12.20</td>
</tr>
</tbody>
</table>

Implied Upward Bias in Standard Estimates of the Rates of Return to Schooling

| subprimary | 82.0% | 114.7% |
| primary | 38.2% | 78.2% |
| junior secondary | 23.8% | 13.9% |
| senior secondary | 10.8% | 1.7% |

a. These estimates are based on the repetition, dropout, and progression rates in Table 3, with a slight adjustment to assure consistency with the 1986 SAKERNAS proportions of respondents in the various schooling categories.

b. For senior secondary these are lower-bound estimates in that they do not include time spent in uncompleted post-secondary education.
For each case the implied percentage by which the standard method (i.e. assuming that the number of years spent in school equals the number of grades of school) overstates the private and the social rates of return to schooling also is presented (under the assumption that the private and the social costs are proportional to the time spent in school).

The results in Table 3 have five important implications. First, adjusting for repetition and dropout rates makes a considerable difference in the Indonesian case. The standard procedure of not adjusting results in expected overestimates of the rate of return to schooling of 82-114% for the subprimary category, 38-78% for the completed primary category, 14-24% for the junior secondary category, and a lower bound of 2-11% for the senior secondary category.20 Second, the alternative methods of calculating the expected years spent in school for each schooling category make some difference, though the basic story is robust to the choice regarding homogeneity versus heterogeneity in the repetition and dropout rates. Third, how much of the expected excess in years spent for a given schooling level recorded in the survey beyond the number of grades in that level is due to repetition in attaining that level or to beginning (but not completing) the next level varies across the estimates. The percentage of the excess over the standard estimates due to repetition before reaching the recorded level ranges from 26% for junior secondary with the assumption of heterogeneity to 84% for primary also with the assumption of heterogeneity. This means that there would be considerable upward bias in the estimated rates of return to schooling from the standard procedure even if the data was recorded by grades of school instead of the categories used. Fourth, in addition to overstating the levels of the rates of return to schooling, the standard procedure overstates relatively the rates of return to the lower schooling levels. For both sets of estimates, the extent of upward bias declines considerably as the schooling category increases.21 Fifth, disaggregated estimates by sex,

---

20 For the last category these are lower-bound estimates because of the absence of data on post-secondary repetition and dropout rates, so the terms relating to those rates in relations (5) and (6) are ignored in our calculations.

21 Certainly this is the case through the first three schooling categories and quite possibly through the higher levels as well, though this is more speculative since we have lower-bound estimates for the completed secondary and higher levels.
region, and urbanization indicate that the biases differ among groups. For example, the biases tend to be greatest for older females in rural areas of the relatively remote outer islands. The disaggregated estimates suggest that, not only are there the biases discussed at the national level, but the biases differ among groups, so that standard procedures also cause misunderstanding of the differences in the schooling-wage relations among such groups.

4. Conclusions

Repetition of grades and dropping out are widespread phenomena in most developing countries. We have developed a methodology for adjusting standard estimates of rates of return to schooling for these phenomena and have applied it to data from Indonesia. This application suggests that failure to control for these phenomena in this and similar cases causes substantial upward biases in estimated rates of return to schooling, particularly for the lower schooling levels: 82-114% for the subprimary category and 38-78% for the completed primary category. Since the estimated rates of return in such surveys as the World Bank (1980, 1981, 1990), Colclough (1982), Eisemon (1988), Psacharopoulos (1985, 1988) and UNDP (1990) do not in general control for such phenomena, their strong advocacy for primary schooling investments seems to be overstated substantially, ceteris paribus. These biases tend to differ somewhat among subsamples, with somewhat larger values for females, older individuals, and rural and relatively remote areas. Therefore the standard procedures

---

22 We undertook 72 sets of estimates for samples disaggregated by these characteristics. These results are not presented in the paper because of space constraints, but are available upon request from the authors.

23 We have excluded individuals who still are in school from these calculations. This affects primarily the 10-19 age group and, much less, the 20-29 age group. Since at least under the assumption of heterogeneity, those still in school are likely to have the lowest repetition rates at the lower schooling levels, the exclusion of such individuals is likely to bias upward the estimated bias at least for subprimary and primary for the age groups in which there are many individuals still in school. That in fact the estimated biases tend to be higher for subprimary and primary for the 10-19 age group than for the 20-29 age group probably reflects such a phenomenon. If so, this provides some support for the heterogeneity assumption in Section 2.

24 Heckman and Hotz (1986) give a good discussion of some of the problems in making deductions from the comparison of wage or earnings estimates across groups, but they do not note the substantial problem in existing estimates that is discussed here.
tend also to be misleading by overstating the relative returns for individuals in these subsamples.

REFERENCES


UNDP (Mahboub ul Haq, head of research team), 1990, Human Development Report 1990, New York: UNDP.


Department of Economics  
Williams College  
RESEARCH PAPER SERIES

A series of papers written by members of the Department of Economics on topics that do not pertain directly to less developed countries. From time to time documentation for computer programs and other research tools developed by members of the department will be distributed in this series as Special Research Papers and denoted by the letters RPS. A copy of any paper (and/or a reprint of the published version) will be mailed on request. See the order blank for further details.

1984-1990 Research Papers


RP-74** Roger Bolton, "Multiregional Models: Introduction to a


RP-86** Michael McPherson, "Efficiency and Liberty in the Productive Enterprise: Recent Work in the Economics of Work Organization" in Philosophy & Public
Affairs, Vol. 12, No. 4, Fall 1983.


is Optimally Idle", February, 1988.


** Reprint only available
Center for Development Economics  
Department of Economics  
Williams College  
RESEARCH MEMORANDUM SERIES

A series of papers written by members of the Department of Economics on topics pertaining to the economics of less developed countries. A copy of any paper (and/or a reprint of the published version) will be mailed on request. See the order blank for further details.

1984-1990 Research Memorandum

RM-89  

RM-90  

RM-91  

RM-92**  

RM-93*  
Stephen R. Lewis, Jr., Taxation for Development, Oxford University Press, 1984 (available only through the publisher).

RM-94  

RM-95**  

RM-96**  

RM-97  

RM-98  


RM-112  Brian Levy, "Transactions Costs, the Size of Firms and Industrial Policy: Lessons from A Comparative Case Study of the Footwear Industry in Korea and Taiwan,"


* No longer available
** Reprint only available
ORDER BLANK
FOR RESEARCH MEMORANDA AND RESEARCH PAPERS

Please send the indicated number of copies of each paper.

|-------|-------|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|

Place a check mark in appropriate boxes.

Where appropriate, I would prefer the paper _____ reprint _____ version of the papers checked above.

Please send a list of the earlier titles in the Research Memorandum and Research Paper Series and an order blank with which I may request copies._____

My correct mailing address is:

MAIL THIS COMPLETED ORDER BLANK TO THE ADDRESS IN THE TOP LEFT CORNER.