"An Introduction to the Analysis of Student Enrollment and Flows Statistics"

By

Birger Fredriksen

Education and Employment Division
Population and Human Resources Department
The World Bank

June 1991
## Contents

<table>
<thead>
<tr>
<th>I. Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>II. Enrollment Ratios</td>
<td>2</td>
</tr>
<tr>
<td>A. Definitions</td>
<td>3</td>
</tr>
<tr>
<td>B. Data problems</td>
<td>10</td>
</tr>
<tr>
<td>C. Interpretation</td>
<td>12</td>
</tr>
<tr>
<td>D. Examples</td>
<td>17</td>
</tr>
<tr>
<td>III. Importance of Analyzing Student Flows</td>
<td>21</td>
</tr>
<tr>
<td>IV. Analyzing Admission to a Cycle of Education</td>
<td>30</td>
</tr>
<tr>
<td>A. Apparent intake rates</td>
<td>31</td>
</tr>
<tr>
<td>B. Age-specific intake rates</td>
<td>33</td>
</tr>
<tr>
<td>V. Analyzing the Flow of Pupils Between Two School Years</td>
<td>37</td>
</tr>
<tr>
<td>A. Student flow models</td>
<td>37</td>
</tr>
<tr>
<td>B. Data requirements</td>
<td>38</td>
</tr>
<tr>
<td>C. Calculations of flow rates</td>
<td>43</td>
</tr>
<tr>
<td>VI. Effect on Flow Rates of Common Errors in Data</td>
<td>46</td>
</tr>
<tr>
<td>A. Over-reporting of grade 1 enrollment</td>
<td>47</td>
</tr>
<tr>
<td>B. Under-reporting of the number of repeaters</td>
<td>50</td>
</tr>
<tr>
<td>C. Yearly variation in the coverage of data</td>
<td>53</td>
</tr>
<tr>
<td>D. Incorrect definitions of the flows</td>
<td>53</td>
</tr>
<tr>
<td>VII. Cohort Flow Analysis</td>
<td>57</td>
</tr>
<tr>
<td>A. Cohort reconstruction</td>
<td>59</td>
</tr>
<tr>
<td>B. Cohort flow indicators</td>
<td>65</td>
</tr>
<tr>
<td>C. Further comments on the assumptions behind the reconstructed cohort method</td>
<td>76</td>
</tr>
<tr>
<td>VIII. The Apparent Cohort Method</td>
<td>79</td>
</tr>
<tr>
<td>References</td>
<td>82</td>
</tr>
<tr>
<td>Annex 1</td>
<td>85</td>
</tr>
</tbody>
</table>
Tables

Table 2.1 Zaire, enrollment and population by age (1983) .......... 10
Table 2.2 Level and age-specific enrollment ratios for selected countries, both sexes (1980-86) .......... 15
Table 2.3 Illustration of impact of dropout on gross enrollment ratio .......... 20
Table 3.1 Enrollment ratios (including and excluding repeaters in the first and final grades of primary education, 1980 and 1987 (percentages) .......... 24
Table 3.2 Percentage repeaters in primary education, both sexes (latest year available between 1984-89) .......... 26
Table 3.3 Percentage repeaters in secondary education, both sexes (latest year available between 1984-89) .......... 26
Table 3.4 Changes in the percentage of repeaters in primary education by region, for 1965, 1975 and 1985 .......... 27
Table 3.5 Percentage of repeaters in primary education by sex (latest year available between 1989-1984) .......... 27
Table 4.1 Enrollment ratios in first grade and intake rates for selected countries (males) .......... 32
Table 4.2 Simulations of apparent intake rates .......... 34
Table 7.1 Enrollment and repeaters by grade in primary education, Burkina Faso, both sexes .......... 59
Table 7.2 Rates of promotion, repetition and dropout by grade in primary education for 1986-87, Burkina Faso, both sexes .......... 60

Figures

Figure 3.1 Proportion of cohort reaching last grade of primary education around 1965 and 1985 .......... 28
Figure 6.1 Flows between Grade 1 and Grade 2 in school years $t$ and $t+1$ .......... 49
Figure 6.2 What happens to pupils entering Grade $g$ of primary education? .......... 55
Figure 7.1 Flows out of Grade 1 (absolute numbers) .......... 60
Figure 7.2 Flows out of Grade 1 (relative numbers) .......... 61
Figure 7.3 Hypothetical flow of the 1986 cohort through primary education in Burkina Faso .......... 63
AN INTRODUCTION TO THE ANALYSIS OF
STUDENT ENROLLMENT AND FLOWS STATISTICS

I. Introduction

World Bank education appraisal and sector reports commonly use a few standard indicators to describe the level of access, enrollment coverage and internal efficiency of the education system studied. Most of these indicators are rates or ratios describing the extent to which those eligible for admission to a given cycle of education are actually admitted and, once admitted, how they progress through the cycle. For example, admission rates are normally computed to assess access to the first grade of a given education cycle, while different types of enrollment ratios are used to appraise the overall enrollment coverage of the cycle. Progress from grade to grade through the cycle is normally described by means of promotion, repetition and dropout rates estimated separately for each grade.

Employing this type of indicators together with other information, World Bank appraisal and sector reports routinely identify low and/or unequal access and high repetition and dropout as major concerns, and most Bank projects support policies and investments designed to address these concerns. To be successful, the development of such project elements needs to be based on a solid diagnosis of the problem. Frequently, errors in available statistical data and incorrect interpretation of these simple indicators may seriously affect their usefulness as a basis for designing appropriate project interventions.

On the above background, the main purpose of this paper, and the seminars for which it has been prepared as background material, is to review the usefulness and limitations of the above type of indicators. More specifically, the purpose is to review:
(a) definitions, interpretations, data needs, and pitfalls in using enrollment ratios and student flow rates (intake, promotion, repetition, dropout and graduation rates); and

(b) methods using such rates to derive commonly used indicators of internal efficiency of school systems, (e.g., percentage of cohort completing the education cycle, relative contributions of repetition and dropout to overall wastage, and number of school years required to complete a cycle of education).

The paper focuses on various practical problems encountered in the above type of analysis in countries where both the reliability and availability of data on student flows leave much to be desired. It attempts to point out how an adequate statistical analysis of repetition and dropout can help to correctly diagnose the problem and, in turn, design policies and investments to address it. The paper also gives examples of how common pitfalls in this type of analysis along with errors in available education statistics may adversely impact project design.

II. Enrollment ratios

Enrollment ratios are the most commonly used indicator for assessing the extent to which a country's children of school-age are enrolled at school. The purpose of this section is to describe the enrollment ratios most commonly used, to explain what data are required to compute them, and to discuss pitfalls in their use. The discussion will be oriented towards addressing practical problems staff may encounter in computing and interpreting enrollment ratios when using such indicators in sector and appraisal reports.

Although different types of enrollment ratios exist, they all relate the number of pupils enrolled in a given age group or level of education to the size of a given reference population. Hence, enrollment ratios are relative measures of numbers enrolled as compared to a population age-group, and the interpretation of such ratios depends on the pupils and reference population
considered. For example, enrollment ratios may be calculated by level of education or without regard to level, they may take into account the age of the pupils or disregard the age. Furthermore, in each of these cases, enrollment ratios may be calculated separately for different sub-populations, e.g., for boys and girls, part-time and full-time students, private and public schools, urban and rural areas, and ethnic groups.

We shall discuss below the enrollment ratios most commonly used in Bank documents. There is no clear terminology in this area. The terms used below are believed to be the ones most commonly employed.

A. Definitions

The least refined enrollment ratio is the overall enrollment ratio, also referred to as the crude or general enrollment ratio. This ratio is defined as the total enrollment at all levels of education divided by the population in the age-group which corresponds to the normal age range for attending primary, secondary and higher education, for example, the age group 6-24 years.

This ratio gives an approximate idea of the extent of enrollment among the population of school age. It was commonly used, for example, for international comparison of education development before population estimates by single years of age or by five-year age groups became available. It is, however, not an adequate measure for detailed studies of enrollment trends, and is of little use for the type of sector description required in sector and appraisal reports. For example, overall enrollment ratios do not provide any

---

1/ It should be recalled that the general availability of data and projections of population by age for developing countries, especially for Sub-Saharan Africa, is of a fairly recent date. For example, about half of the countries in Sub-Saharan Africa conducted their first population census in the 1970s. An illustration of how little was known about the population situation in Africa at the time of independence in the early 1960s is provided by the well-known "Addis Ababa Plan" for development of education in Africa (adopted in May 1961 by 35 African States). The targets for universal primary education in Africa by 1980 included in this plan imply that the population of school age would grow by about 5% during the twenty-year period 1960-1980. In reality, this population age-group grew by about 95%, see Fredriksen (1983a).
information on the level at which pupils are enrolled, or on their age. Therefore, such ratios are of little use to planners and policy makers in establishing an adequate sector diagnosis. This is why we normally use level-specific or age-specific enrollment ratios.

The level enrollment ratio is probably the most commonly used measure of enrollment coverage. Level enrollment ratios are normally calculated separately for primary, secondary and higher education. Typically, sector and appraisal reports quote two such ratios:

(a) the gross enrollment ratio (GER); and
(b) the net enrollment ratio (NER).

The gross level enrollment ratio relates total enrollment at a given level of education (e.g., primary education) to the population which, according to national regulations, should be enrolled at this level. While classification of students by level of education is normally straightforward, the definition of the age-range for the population group to use may pose some problems. The lower limit for this range should be the official admission age to the level considered, while the upper limit is found by adding as many single ages as there are grades in the cycle. This works well for primary education. For example, for a country where children may start school at the age of six and where this cycle has six grades (this is the most common admission age and duration of primary education in developing countries), the age-group 6-11 should be used.

For general secondary education, the lower limit of the age-group is normally the age at which a pupil who started primary education at the official age, and who does not repeat, is ready to enter secondary school. In the above case, this would be at age 12 and, in a case where general secondary education comprises six grades (the most common situation), the age-group 12-17 would be used.

For higher education, the situation is more complex. First, while all primary and secondary students follow programs organized into a given number of grades, this is not the case for higher education where the duration of study may vary by field of study, and where there is not a standard set of grades that
all students have to complete -- e.g., some students are enrolled in programs
not leading to a first degree, others are enrolled in programs leading to a lower
degree, and yet others are in post-graduate programs. Second, there is normally
no legal entry age to higher education, and there may be a wide age-range among
new entrants. Since the main purpose of level enrollment ratios is to give an
indication of the coverage of a given level of education as compared to the
population "at risk", the guiding principle should be that the number of single
ages included in the population age-group should correspond as closely as
possible to the average duration of higher education in the country concerned.
The five-year age-group 20-24 years is very commonly used. This would imply an
average duration of five years.

Bearing in mind the above discussion, for a country with six grades
of primary education and for which the official admission age is six, the gross
enrollment ratio for primary education (GERP) would be computed as:

\[
(1) \quad \text{GERP} = \frac{\text{Total enrollment in primary education}}{\text{Total number of children aged 6-11 years.}}
\]

Normally, the ratio derived from this computation is multiplied by 100 to be
expressed as a percentage.

Tabulations showing gross level enrollment ratios for countries with
differing educational structures often employ the term adjusted enrollment ratio.
This means that the age range of the population used in the denominator of the
enrollment ratio for a given level has been adjusted for each country to the
admission age and prescribed duration of this level. Thus, while the population
aged 6-11 would be used for a country where the official admission age is six
and the primary cycle comprises six grades, the population aged 7-14 would be
used for a country where the official admission age is seven and the primary
cycle comprises eight grades.

Before estimates of population data by single years of age became
available on an international scale, unadjusted enrollment ratios were used.
These were obtained by employing the same age-range, comprising one or two five-
year age-groups, for all countries. For example, the age-group 5-14 years was commonly used for primary education. Unadjusted ratios may have some interest for international comparison since they compare the enrollment at a given level to the same yardstick (same age-group) for all countries. However, such ratios are of little interest for country-specific work since they are poor indicators of the coverage of the national system as compared to the population age-group eligible for enrollment at this level. For example, it is difficult to interpret a gross enrollment ratio of, say, 70% calculated on the basis of the 5-14 age-group for a country that has six grades of primary education. Does this country have sufficient capacity to enroll all children of primary school age? This question cannot be answered on the basis of this ratio since all children aged 5-14 years are not expected to be enrolled in primary education.

The net enrollment ratio differs from the gross enrollment ratio in that the numerator includes only those pupils who are within the prescribed age for the level of education considered. For a country where the official age of admission to primary education is six and the duration of primary education is six years, the denominator would include the population aged 6-11 years for both the gross and net enrollment ratios for primary education. The numerator of the gross enrollment ratios would, however, include all pupils in primary education, while the net enrollment ratio would include only those belonging to the age group 6-11. i.e.,

\[
(2) \quad \text{NERP} = \frac{\text{Primary school pupils aged 6-11 years}}{\text{Total number of children aged 6-11 years}}.
\]

\[
\text{2/ For example, in 1987, the gross enrollment ratio for primary education was 68% in Mozambique and 67% in Rwanda. However, while the official entry age for both countries was seven years, the former country had only five grades of primary education while the latter had eight. This means that the ratio for Mozambique was calculated with reference to the age groups 7-11 years while that for Rwanda was calculated with reference to the age groups 7-14 years. Thus, despite the similar enrollment ratio, it is clear that the respective "effort" made in primary education by these two countries is larger in Rwanda than in Mozambique.}
\]
Note that while the gross enrollment ratio can exceed 100% because of repetition and because of admission of children below or above the official admission age, the net enrollment ratio has 100% as its maximum value. The difference between the gross and net ratios for a given level is an indicator of the importance of under- and over-aged pupils enrolled at this level. We shall discuss this more thoroughly below.

A third ratio sometimes used in sector and appraisal reports is the age-specific enrollment ratio. While the net and gross enrollment ratios both refer to the enrollment at a given level of education, this ratio relates the enrollment of a given age or age group, regardless of the level at which the pupils are enrolled, to the population of the same age. For example, for children aged x years, the age-specific enrollment ratio (ASER) is defined as:

\[
\text{ASER} = \frac{\text{Pupils aged } x \text{ years enrolled at any level of education}}{\text{Total number of children aged } x \text{ years}}
\]

This ratio may hence be calculated for any single age (e.g., the 10 year-olds) or for a given age-group (e.g., the age-group 6-11 years). In both cases, the problem of under- and over-aged pupils is eliminated since the numerator and denominator refer to the same age or age group. The ratio is a purely demographic measure, distinguishing those who are out of school from those in school, regardless of the level and type of education in which they are enrolled.

It should be noted that, when studying yearly trends in age-specific enrollment ratios covering several ages, changes in the age-structure of the population may affect such ratios even in a case where all enrollment ratios for the single age comprising this age-group remain unchanged. This is because any enrollment ratio covering several ages is a weighted average of the enrollment ratios by single year of age for the ages covered, the weights being the population in each single age. If the weights change (i.e., the population structure), then the average enrollment ratio will also change.
Thus, a shift towards a more youthful population can increase an age-specific enrollment ratio calculated for an age-group, by placing more persons in the typical enrollment ages, even if all the enrollment ratios by single years of age remain unchanged. Likewise, a decline in the birth rate leading to a shift towards an "older" population might lead to a decrease in an age-specific enrollment ratio calculated for an age group, by giving higher relative weights to the ages having a relatively low school participation. In particular, the average may decline even if all the enrollment ratios by single years of age were to increase. Such shifts have taken place, for example, in many European countries during the last three decades as a result of the post World War II "baby boom." Similar shifts are presently taking place in some developing countries that have experienced a relatively sharp decline in the population of primary school age for periods in the recent past (e.g., China).

When such changes in population age-structure takes place, only enrollment ratios by single years of age give a correct account of the change in enrollment over time. It is often desirable, however, to work with some aggregated measure of enrollment and not with enrollment ratios by single years of age. Such a measure is obtained by calculating standardized enrollment ratios. The purpose of such ratios is to eliminate the effect of differences in the age-structure of the population groups for which the enrollment ratios are calculated and compared. In addition to being used for study of changes in enrollment over time, standardized enrollment ratios may hence be used for comparing enrollment coverage of different sub-populations that have different age-structure, e.g., between the population of different regions, urban and rural areas, or ethnic groups.

A standardized enrollment ratio is obtained by computing the weighted average of the enrollment ratios by single years of age for a given population, using as weights the age distribution of a chosen standard population. Thus, if we wanted to compare the enrollment ratios of urban and rural areas for a given age group, this could be done by computing the weighted average of the enrollment ratios by single years of age for the rural population, using as weights the age distribution of the urban population. This would show what the enrollment ratio for this age group would have been in rural areas if the age
structure of the rural population had been the same as that of the urban population. This standardized ratio for rural areas may then be compared with the enrollment ratio for urban areas. Any difference between these two ratios would be caused by differences in the enrollment pattern and not by differences in the age structure of the two populations.

Although standardized enrollment ratios are normally not used in Bank sector and appraisal reports, they may be more relevant in the future as declining fertility rates cause marked waves in the size of the population of school age.

Of the enrollment ratios defined above, the gross enrollment ratio is definitely the one most used in Bank reports. As will be discussed under point C., the main reason is that enrollment ratios are normally used to assess enrollment coverage and capacity at a given level of education. Since age-specific enrollment ratios do not distinguish enrollment by level, they can not be used for this purpose. Similarly, since net enrollment ratios exclude all enrollment outside the official age-group, they under-estimate, often to a considerable degree, the enrollment capacity of a given level of education. Thus, gross enrollment ratios, although they are very crude measures of enrollment coverage and capacity, are still the ratios most commonly used.

Table 2.1 provides an illustration of how to calculate the gross, net and age-specific enrollment ratios for primary and general secondary education for Zaire.

---

2/ Use of standardized enrollment ratios are further discussed in Fredriksen (1975) and Thonstad (1980).
Table 2.1: Zambia, Enrollment and Population by Age (1983)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Female Male</td>
<td>Total Female Male</td>
<td>Total Female Male</td>
</tr>
<tr>
<td>5</td>
<td>130329 55739 74590</td>
<td>0 0 0</td>
<td>918000 444000 470000</td>
</tr>
<tr>
<td>6</td>
<td>414840 177173 237087</td>
<td>0 0 0</td>
<td>573500 236000 447000</td>
</tr>
<tr>
<td>7</td>
<td>685998 246948 339140</td>
<td>0 0 0</td>
<td>520000 312000 423000</td>
</tr>
<tr>
<td>8</td>
<td>665161 284758 380885</td>
<td>0 0 0</td>
<td>780000 387000 401000</td>
</tr>
<tr>
<td>9</td>
<td>781325 334242 447283</td>
<td>0 0 0</td>
<td>665000 332000 333000</td>
</tr>
<tr>
<td>10</td>
<td>595790 255596 337190</td>
<td>0 0 0</td>
<td>448000 234000 214000</td>
</tr>
<tr>
<td>11</td>
<td>521316 222959 338357</td>
<td>9867 241 6186</td>
<td>720000 359000 361000</td>
</tr>
<tr>
<td>12</td>
<td>609457 162191 227266</td>
<td>187928 42726 115024</td>
<td>606000 346000 350000</td>
</tr>
<tr>
<td>13</td>
<td>287695 127405 170290</td>
<td>355650 100306 253340</td>
<td>586000 332000 334000</td>
</tr>
<tr>
<td>14</td>
<td>162911 66674 96237</td>
<td>497048 13149 346899</td>
<td>420000 204000 216000</td>
</tr>
<tr>
<td>15</td>
<td>69819 29860 39959</td>
<td>374430 108242 266168</td>
<td>620000 312000 312000</td>
</tr>
<tr>
<td>16</td>
<td>32682 13934 18648</td>
<td>279747 79348 200399</td>
<td>604000 302000 302000</td>
</tr>
<tr>
<td>17</td>
<td>4654 1980 2664</td>
<td>200129 58764 143362</td>
<td>586000 293000 293000</td>
</tr>
<tr>
<td>18</td>
<td>1396 597 799</td>
<td>142025 40284 101741</td>
<td>568000 285000 284000</td>
</tr>
<tr>
<td>19</td>
<td>0 0 0</td>
<td>81772 23194 58578</td>
<td>551000 276000 275000</td>
</tr>
<tr>
<td>20</td>
<td>0 0 0</td>
<td>45189 12817 223768</td>
<td>531000 257000 254000</td>
</tr>
<tr>
<td>20+</td>
<td>0 0 0</td>
<td>21516 6103 15415</td>
<td>284000 284000 284000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Female Male</td>
<td>Total Female Male</td>
<td>Total Female Male</td>
</tr>
<tr>
<td></td>
<td>464613 1990710 2659303</td>
<td>2151900 610374 1841526</td>
<td></td>
</tr>
</tbody>
</table>

B. Data problems

Enrollment ratios suffer from the weaknesses in education and population statistics generally found in developing countries. Some problems with the quality of the education statistics are reviewed in Annex 1 to this report. The impact on admission, promotion, repetition and dropout rates of some common data problems (e.g., over-reporting, incomplete data, use of incorrect definitions) will be discussed in Section VI. The discussion below is limited to mentioning a few problems specific to the calculation of enrollment ratios.

First, it is often difficult to obtain population and education data that are consistent in the sense that they refer to the same point in time, have
the same coverage, and are classified (defined) in the same manner. For example, population estimates normally refer to the middle of the calendar year while the enrollment data normally refer to the beginning of the school year. However, the error introduced by such differences normally is small, depending on the rate of population growth and when the school year starts.4/.

Another problem with respect to population data is that the statistics available from national sources normally are disaggregated into five-year age groups, and do not provide the single-year-of-age breakdown required to be able to estimate the population aged, e.g., 6-11 years. The normal procedure is to estimate data by single years of age by applying an interpolation formula to split five-year age-groups into single ages. While different formulas exist, the one most commonly used in probably the Sprague Multipliers.5/

Bank staff would often use population estimates and projections established by the World Bank. These may differ from those prepared by the country. They may also differ from those used in computing the enrollment ratios published by UNESCO, which are based on population estimates and projections prepared by the United Nations. As a matter of fact, the differences frequently found in enrollment ratios quoted in national, World Bank and UNESCO documents often stem from use of different population data. Another reason for differences between different sources is use of different reference age-groups for, e.g., the population of primary school age. In short, the lack of consistency often found between different sources publishing enrollment ratios is often due to use of different population data.

The problem of obtaining enrollment and population data classified in the same manner is often more serious when calculating enrollment ratios for different sub-populations. For example, in many countries, calculation of

4/ For example, using mid-year population estimates in a case where the enrollment data refer to around November would normally mean an underestimation of the population by less than 1%.

5/ These can be found in any standard textbook on demography. For a discussion of various types of interpolation techniques, see U.S. Department of Commerce, Bureau of the Census (1971).
enrollment ratios separately for urban and rural areas is difficult because, while population data normally are broken down in this way, enrollment data normally refer to administrative units covering both urban and rural areas.

Another example is that children counted in one administrative area may attend school in another area. For instance, children from rural areas with poor educational infrastructure may move to live with friends or family residing in urban areas and attend school there. Such pupils would be included in the enrollment count for urban areas and in the population figures for rural areas. Hence, this would cause under-estimation of the enrollment ratios for rural residents and over-estimation of the enrollment ratios for urban residents. A similar problem is caused by schools that draw their enrollment from several administrative areas, e.g. private schools and public schools recruiting nationwide.

In conclusion, care should be taken when using enrollment ratios calculated for different sub-populations as a basis for assessing differences in enrollment coverage between these populations.

C. Interpretation

Level enrollment ratios may be interpreted as indicators of two similar, yet different, aspects of education development. Frequently, users do not distinguish between these two interpretations. As we shall see below, both are approximations that should be used with care.

First, level enrollment ratios are often used as indicators of enrollment coverage at a given level of education as compared to the age-group which, according to national regulations, should be enrolled at that level. For example, in a case where the gross enrollment ratio for primary education is, say, 70%, this is often taken to indicate that "70% of the children of primary school age are enrolled in primary education". Similarly, if the ratio is about

---

Note that the definition of what constitutes "rural" and "urban" areas generally differs between countries and over time for the same country.
100%, this is generally taken to indicate that the country has attained universal primary education, i.e., it has universal enrollment "coverage" for this age-group.

The second interpretation is one of enrollment capacity at a given level of education as compared to the age-group which, according to national regulations, is eligible for enrollment at that level. Thus, in the above example, we would say that "the enrollment capacity of primary schools corresponds to 70% of the population of primary school age". Or, if the ratio is 100%, that the country's primary schools have sufficient capacity to enroll everybody of primary school age. However, we are not claiming that everybody of primary school age is actually enrolled at school, i.e., there is not necessarily universal coverage. Some of this capacity may be used on repeaters, and some children of primary school age may not be enrolled at all, either because they never started school, or because they dropped out.

In countries where all new entrants start school at the official admission age, where there is no or negligible repetition, where attendance is compulsory, and where all schools are used to the maximum of their capacity (however defined), the two interpretations would coincide. Furthermore, the difference between the net and the gross enrollment ratios for a given level would be small. This is approximately the case in most industrialized countries (though full use of all capacity may not always be the case). For such countries, either ratio would, for primary and secondary education, provide reasonable indications of both enrollment coverage and capacity at these two levels, relative to the size of the respective population age groups.

More care needs be taken when interpreting enrollment ratios for developing countries. First, there are problems related to the reliability of the enrollment and population statistics, confer the above discussion. Second, the data on enrollment refer to the number of children registered at school at the beginning of a given school-year. As attendance normally is not compulsory, the number that actually attend school on a regular basis is considerably lower in many countries. Third, many developing countries have a wide age-spread in the enrollment at any given level, caused by early-and late entrants, and by
repetition. Therefore, the difference between the gross and net level enrollment ratios is often considerable.

The above factors complicate the interpretation of net and gross enrollment ratios as indicators of coverage and capacity. The net enrollment ratio is often a poor indicator of both aspects since it excludes all enrollment outside the official age range. For example, assume a school system for which the official age-range for primary education is 6-11 years. The net enrollment ratio for such a system will not include pupils who are five years old, or pupils who entered at the age of seven and who completed the cycle without repetition (i.e., at the age of 12). In an assessment of coverage or capacity of primary schools as compared to the population of primary school age, such pupils should be included in the enrollment count since both are receiving primary education. The fact that they did not enter school exactly at the official age is less important than the fact that they are enrolled and that existing capacity is sufficient to "cover" them. Research shows that children mature and learn at different speeds. Therefore, although too wide an age-spread within a given grade may pose disciplinary and pedagogical difficulties, to insist -- in a developing country setting -- that all children should enter school at the same age, is unwise. Another example would be a case where all children start school but at the age of seven rather than at the age of six (the official age of admission). If there is no dropout, the net enrollment ratio would be, say, around 85% although everybody can be -- and are -- "covered" by existing capacity.

Use of the net enrollment ratio as an indicator of enrollment capacity is even more arguable in systems where, in addition to early and late entrants, there is a "high" level of repetition. As this ratio excludes all pupils who, whether or not they entered at the official admission age, fall outside the

For example, in Lesotho where the official admission age to primary education is six years, in 1990, only 25% of those enrolled in Grade 1 belonged to that age group. About 7% of the pupils were less than six years old and 13% were ten years or older, including a few 15 and 16 year-olds. As a total for primary education, about 35% of those enrolled did not belong to the primary school age group.
official age-group because they have repeated once or more, it underestimates the real effort the country is making at the level of education considered. The fact that part of this effort is used (wasted?) on repeaters does not change the fact that capacity is available for more children than suggested by the net enrollment ratio. As illustrated in Table 2.2, the proportion of enrollment outside the official age group may be quite high in many countries (e.g., exceed 30% for primary education and even more for secondary education). However, as discussed below, used in conjunction with gross enrollment ratios, net ratios do provide very useful information on the extent of enrollment outside the official age-group.

### Table 2.2: Level and Age-Specific Enrollment Ratios for Selected Countries, Both Sexes (1980-86)*

<table>
<thead>
<tr>
<th></th>
<th>PRIMARY EDUCATION</th>
<th>Enrollment Ratio for 6-11 age group</th>
<th>SECONDARY EDUCATION</th>
<th>Enrollment Ratio for 12-17 age group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gross Enrol</td>
<td>Net Enrol.</td>
<td>6-11</td>
<td>Gross Enrol</td>
</tr>
<tr>
<td>AFRICA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CENTRAL AFR REP 1980</td>
<td>71.6%</td>
<td>56.9%</td>
<td>56.9%</td>
<td>13.7%</td>
</tr>
<tr>
<td>CENTRAL AFR REP 1984</td>
<td>78.2%</td>
<td>64.1%</td>
<td>64.1%</td>
<td>15.4%</td>
</tr>
<tr>
<td>CENTRAL AFR REP 1986</td>
<td>65.3%</td>
<td>53.4%</td>
<td>58.3%</td>
<td>13.3%</td>
</tr>
<tr>
<td>ZAIRE 1980</td>
<td>94.4%</td>
<td>73.6%</td>
<td>73.6%</td>
<td>33.5%</td>
</tr>
<tr>
<td>ZAIRE 1983</td>
<td>95.3%</td>
<td>73.6%</td>
<td>73.6%</td>
<td>34.7%</td>
</tr>
<tr>
<td>ZAIRE 1986</td>
<td>76.0%</td>
<td>58.3%</td>
<td>58.3%</td>
<td>13.3%</td>
</tr>
<tr>
<td>ASIA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOREA REP OF 1980</td>
<td>109.9%</td>
<td>104.3%</td>
<td>105.4</td>
<td>76.3%</td>
</tr>
<tr>
<td>KOREA REP OF 1983</td>
<td>100.9%</td>
<td>96.9%</td>
<td>97.5%</td>
<td>86.7%</td>
</tr>
<tr>
<td>KOREA REP OF 1986</td>
<td>98.5%</td>
<td>86.4%</td>
<td>97.1%</td>
<td>89.2%</td>
</tr>
<tr>
<td>LAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEXICO 1980</td>
<td>115.5%</td>
<td>92.6%</td>
<td>92.6%</td>
<td>45.6%</td>
</tr>
<tr>
<td>MEXICO 1983</td>
<td>117.8%</td>
<td>96.5%</td>
<td>96.5%</td>
<td>52.6%</td>
</tr>
<tr>
<td>MEXICO 1986</td>
<td>119.2%</td>
<td>102.0%</td>
<td>102.0</td>
<td>53.1%</td>
</tr>
<tr>
<td>URUGUAY 1980</td>
<td>106.5%</td>
<td>83.0%</td>
<td>83.0%</td>
<td>80.2%</td>
</tr>
<tr>
<td>URUGUAY 1983</td>
<td>108.4%</td>
<td>87.6%</td>
<td>87.6%</td>
<td>87.3%</td>
</tr>
<tr>
<td>URUGUAY 1986</td>
<td>109.9%</td>
<td>82.0%</td>
<td>82.0%</td>
<td>71.8%</td>
</tr>
<tr>
<td>EMENA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYRIA 1980</td>
<td>101.6%</td>
<td>91.2%</td>
<td>91.2%</td>
<td>47.2%</td>
</tr>
<tr>
<td>SYRIA 1983</td>
<td>105.1%</td>
<td>93.6%</td>
<td>93.6%</td>
<td>56.1%</td>
</tr>
<tr>
<td>SYRIA 1986</td>
<td>110.8%</td>
<td>99.6%</td>
<td>99.6%</td>
<td>60.5%</td>
</tr>
<tr>
<td>TUNISIA 1980</td>
<td>103.1%</td>
<td>93.0%</td>
<td>93.0%</td>
<td>27.3%</td>
</tr>
<tr>
<td>TUNISIA 1983</td>
<td>112.2%</td>
<td>91.7%</td>
<td>91.7%</td>
<td>34.2%</td>
</tr>
<tr>
<td>TUNISIA 1986</td>
<td>117.1%</td>
<td>95.1%</td>
<td>95.1%</td>
<td>38.9%</td>
</tr>
<tr>
<td>BELGIUM 1980</td>
<td>101.3%</td>
<td>95.0%</td>
<td>95.0%</td>
<td>87.6%</td>
</tr>
<tr>
<td>BELGIUM 1983</td>
<td>95.7%</td>
<td>90.3%</td>
<td>90.3%</td>
<td>91.0%</td>
</tr>
<tr>
<td>BELGIUM 1986</td>
<td>95.5%</td>
<td>91.4%</td>
<td>91.4%</td>
<td>94.2%</td>
</tr>
</tbody>
</table>

* All these countries have six grades of primary education and the official admission age is six years.

** The Central African Republic and Tunisia have seven grades of general secondary education.

The six remaining countries have six grades.

** Age group 12-18 for Central African Republic and Tunisia.
Interpretation of the gross enrollment ratios presents a different set of problems. Such ratios give an ambiguous and often overstated indication of enrollment coverage at a given level of education as compared to the age-group corresponding to that level since the numerator includes pupils who do not belong to the official age-group. It should be noted that it is enrollment of overaged pupils caused by repetition, and not by late entrance, that leads to over-estimation of coverage. As long as a child enters school only once, whether or not it enters at the official age, it still will be counted once only. However, even in cases where there is overaged pupils due to repetition, gross enrollment ratios do provide a very useful indicator of available enrollment capacity as compared to the size of this age-group. The fact that part of this capacity is used on repetition, and thus not necessarily efficiently used, does not detr from the fact that it exists.

To illustrate the above discussion, assume a country for which the gross enrollment ratio for primary education is 100% and the net enrollment ratio is 75%. The former ratio indicates that the capacity of primary schools in this country is about sufficient to cater to the primary school age-group. However, the net ratio shows that 25% of this capacity is used on pupils outside this age-group. If there is no repetition in the system, then we can conclude that the coverage is about universal, but many children do not enter at the official admission age. However, if there is known to be significant repetition, then the only conclusions we can draw on the basis of these two ratios are that: (a) the capacity of primary schools would have been

---

8/This statement does not apply in systems where there is a shift in the age-distribution of new entrants towards the official admission age and where, in addition to rapid growth in intake of children of official admission age, children from older age-groups are also admitted. However, as will be shown in the discussion of intake rates later in this paper, in such cases the number of potential late entrants will soon be phased out.

9/ Naturally, we can not, on the basis of this information, say anything about whether or not all existing capacity is actually used, or whether existing capacity is sufficient to cater adequately to those enrolled. As will be discussed later, enrollment ratio say nothing about aspects such as capacity utilization, quality, admission, internal efficiency, or output of a school system.
sufficient to enroll everybody, had there been no repetition, and (b) 25% of the pupils enrolled are outside the official age-group, but we do not know the extent to which this is due to early/late entrance or repetition. The interpretation is further complicated if there is dropout, confer the example given in connection with Table 2.3 below.

D. Examples

The problems associated with the interpretation of enrollment ratios can be illustrated by studying a few concrete examples. Table 2.2 provides gross and net enrollment ratios for primary and secondary education as well as age-specific enrollment ratios for the populations of, respectively, primary and secondary school age. The data cover three years during the 1980s for eight countries selected to illustrate different situations in different Bank regions. The different situations illustrated by these data will be discussed in detail at the seminar. The comments here will be limited to emphasizing that: (a) to benefit maximally from the information available, the three types of enrollment ratios should be used in conjunction, and (b) to understand why enrollment ratios change over time, one needs to examine changes in intake, repetition and dropout rates.

Take as an example the data provided for Zaire. In 1980, the gross enrollment ratio for primary education was 94.4%, while the net ratio was 72.2%. Thus, in that year, the enrollment capacity of primary schools corresponded to 94.4% of the population of primary school age. Comparison of the net and gross enrollment ratios shows that a large proportion (23.5%) of this capacity was used to enroll pupils outside the primary school age-group. However, on the basis of the information provided by these two ratios it is not possible to establish
the extent to which enrollment outside the official age-group is caused by, respectively, repetition or early/late entrance.

Continuing the above example, we note that the age-specific enrollment ratio for the age-group 6-11 for Zaire in 1980 is slightly larger than the net enrollment ratio for primary education. Thus, some of the children aged 6-11 years who are enrolled at school are not in primary education. This is verified by the (1983) figures given in Table 2.1 which show some pupils aged 11 years are enrolled in general secondary education.

As regards secondary education, we note that the difference between the gross and the net enrollment ratios for Zaire is relatively small -- 14.3% are outside the official age-group. This may seem low in view of the high share of overaged pupils in primary education, and the fact that 8% of all secondary school pupils in 1980 were repeaters, including a repetition rate of 18% in the final grade of the secondary cycle. The comparatively small difference is explained by the fact that about 85% of all secondary pupils in Zaire were in Grades 7 and 8, and only 15% in Grades 9-12.

Finally, we note the marked difference between the net enrollment ratio for secondary education and the age-specific ratio for the age-group 12-17 years. This is mainly explained by the fact that a large share of the pupils who are 12-17 years old are enrolled in primary education (34% in 1983, cf. Table 2.1).

Other aspects of the data provided in the table will be discussed at the seminar, e.g., the large difference in the ratios for secondary education

10/ About 19% of all primary school pupils in Zaire in 1980 were repeaters. Thus, most of the primary school pupils who are not of primary school age, are overaged because of repetition.

11/ Concentration of secondary enrollment in the lower grades of the cycle is an important factor explaining the comparatively small difference between gross and net enrollment ratios for this level of education for other countries included in the table as well. The results would have been quite different if we had computed enrollment ratios separately for the lower and upper cycles of secondary education.
for Zaire for 1983 as compared to 1980 and 1986, the marked decline between 1983 and 1986 in the ratios for primary education for the Central African Republic and Zaire, the comparatively larger change in the net than in the gross enrollment ratio for primary education for Mexico and Uruguay between 1983 and 1986, etc.

It should be clear from this example that, while very useful, enrollment ratios are quite crude indicators. They do not purport to provide more than a first rough quantitative indication of the enrollment in a cycle of education as compared to the population age-group which, according to national regulations, could be enrolled in this cycle. In particular, they do not purport to say anything about access to, internal efficiency of, or output from the cycle. In fact, it is important to be aware of that enrollment ratios of about the same magnitude may hide quite different situations in this respect. In other words, when designing a project, countries having similar levels of primary school enrollment ratios may require quite different project interventions although, on the surface -- judging on the basis of enrollment ratios -- the situation appears quite similar. We shall end this section by providing an example illustrating this point.

Consider for example, two countries -- Country A and Country B -- which both have six grades of primary education, where admission is at the age of six years, and where everybody who enters school starts at this age. To simplify, assume that each country has 100 children in each of the single ages 6, 7, 8, 9, 10 and 11 years and, hence, 600 pupils of primary school age. Assume further that in Country A, 60% of the children of admission age have, over the last several years, entered school, and that there in this country is no repetition or dropout. Under these assumptions, total enrollment equals 360 pupils and the gross, net and age-specific enrollment ratios for the age-group 6-11 years as well as for each single age are all 60%.

In Country B, all six-year olds have been entering school over the past several years. There is no repetition, but heavy dropout which, to simplify the example, is assumed to amount to 16 pupils from each grade annually. This means that there are 100 pupils in Grade 1 (aged six years since everybody starts at
that age and there is no repetition), 84 pupils in Grade 2 (aged seven years), and so on until Grade 6 where the enrollment is 20 pupils all aged eleven years. Thus, also Country B has 360 pupils enrolled, and the gross, net and age-specific enrollment ratio for the age group 6-11 years are all 60%. However, the age-specific enrollment ratio for each single age ranges from 100% for the six year-olds to 20% for the eleven year-olds. The example is summarized in Table 2.3.

Table 2.3: Illustration of Impact of Dropout on Gross Enrollment Ratio

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
<th>Grade</th>
<th>Country A</th>
<th>Country B</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>100</td>
<td>1</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>2</td>
<td>60</td>
<td>84</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>3</td>
<td>60</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>4</td>
<td>60</td>
<td>52</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>5</td>
<td>60</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>100</td>
<td>6</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>Totals</td>
<td>600</td>
<td></td>
<td>360</td>
<td>360</td>
</tr>
</tbody>
</table>

The table shows that, although both countries have 60% enrollment, they are very different in terms of access, internal efficiency and output. In Country A, only 60% of the children enter school, but internal efficiency is very high for those who are admitted and 60% of the age group complete successfully the cycle. In Country B, everybody enters Grade 1, but the retention capacity is very poor and only 20% complete the cycle, i.e., one-third of the completion rate in Country A.

The above example could be further amplified by adding the possibility of repetition. However, it should suffice to illustrate the importance of supplementing data on enrollment ratios by information on the student flows into, through and out of the system since these flows determine the enrollment ratio.
III. Importance of Analyzing Student Flows

Section II emphasized that, while useful, enrollment ratios are rather crude indicators of enrollment coverage and capacity and that, to gain a better understanding of what actually goes on within a cycle of education, we need to study the flow of students into, through, and out of the cycle.

Student flow analyses are a must for all sector and appraisal reports that recommend or support policies and investments to improve access to, internal efficiency of, and output from a cycle of education. It is true that the indicators normally used to assess access, internal efficiency and output also are rather crude, and they need to be supplemented with more qualitative information in order to allow an adequate diagnosis to be established. However, information on student flows, i.e., on admission, promotion, repetition, dropout, and graduation is a very good starting point for understanding the functioning of a cycle of education. Furthermore, in many cases, this is the most detailed statistical information that can be hoped for. Therefore, it is important that staff know how to draw the maximum benefit from this information. The main purpose of the remainder of this paper is to help in this regard. Before discussing, in the following sections, the methods and statistical requirements for student flow analyses, we shall, in this section, highlight the magnitude of repetition and dropout throughout the developing world.

High levels of repetition and dropout generally indicate major weaknesses in an education system. Therefore, these two phenomena warrant serious attention by Bank staff who design education lending operations. Achievement of many of the objectives promoted by Bank assistance in areas such as quality improvement, greater equity, and more effective resource use, would necessarily mean reduction in repetition and dropout. Take as an example the goal for development of primary education\(^{12}\) adapted at the 1990 World Conference on Education for All.

\(^{12}\) In addition to these goals for primary education, the "Framework for Action" calls for expansion of early childhood education, adult literacy programs, and programs that provide individuals and families with the knowledge, skills and values required for better living, and sound and sustainable development.
of Education for All (WCEFA). The "Framework for Action to meet Basic Learning Needs" adapted by the WCEFA calls for:13/ 

(a) universal access to, and completion of, primary education (or whatever higher level of education is considered "basic") by the year 2000; 
(b) improvement in learning achievement such that an agreed percentage of an appropriate age cohort (e.g. 80% of 14 year-olds) attains or surpasses a defined level of necessary learning achievement.

To achieve the above targets, (i) all children must enter school and (ii) at least 80% must be retained until they have attained the "level of necessary learning achievement". It is clear that, in many countries, neither of these two goals can be attained without a major reduction in repetition and dropout.

Frequently the quality of the education statistics seriously hampers an adequate analysis of the extent to which these two goals have been achieved. For example, as regards point (i), in many developing countries, the level of accuracy of the data available on new entrants is not sufficient to allow a valid assessment of these countries' admission capacity14/, e.g., it is not possible on the basis of available official statistics to assess whether the share of children who have never entered school is, say 10% or 20%. This can be a serious handicap when designing lending operations for countries which are approaching universal primary education. Should the operation help expand available capacity or should it not? If yes, where? In particular, lack of accurate data on new admission may inversely affect our ability to address effectively equity problems in provision, since the last 5-10 percent of children to be enrolled often belong to disadvantaged population groups.

With respect to point (ii) above, if for the sake of illustration, we assume that the minimum achievement required is literacy, completion of Grade

13/ Cf. Inter-Agency Commission, WCEFA, p. 53.

14/ Confer discussion of common errors in data in Section VI.
4 is often used as a proxy for the minimum level of retention required. In 1985, about 65% of those who entered primary schools in developing countries completed that grade. However, many would argue that completion of Grade 4 is rarely sufficient to ensure, let alone retain literacy, and that pupils need to complete the primary cycle. Nevertheless, irrespective of how the target for "necessary learning achievement" is defined, it is clear that to achieve the WCEFA targets will require a major reduction in dropout. Because dropout is caused by a complex set of factors determining both the supply of and demand for education, to reduce dropout is a much more difficult task than to increase admission.

Table 3.1 illustrates the impact of dropout on enrollment and the importance of the share of present enrollment capacity used by repeaters. Column 1 shows that already in 1980, the existing enrollment capacity in Grade 1 corresponded approximately to the number of children of official entry age in Africa, and exceeded it by 28% in Asia and by as much as 86% in Latin America. However, a considerable share of this capacity was used for repeaters (25% in Latin America, 18% in Africa, and 16% in Asia), and a high proportion of those entering Grade 1 dropped out prior to the final grade of the cycle (45% in Latin America, 40% in Africa, and 35% in Asia excluding China). As a result, enrollment in the final grade of the cycle, excluding repeaters, corresponded to only 47% of the relevant age-group in Africa, 53% in Asia (excluding China) and 64% in Latin America (column 7).

---

15/ The very high ratio in columns (1) - (4) for Latin America is largely caused by Brazil for which evidence suggests that Grade 1 intake is artificially high due to a combination of double-counting of new entrants over time and underestimation of the level of repetition, see Schiefelbein and Grossi (1981). Severe under-estimation of the level of repetition is a common problem in Latin America, see Cuadra (1989). Similar data problems exist for India and Bangladesh, see Fredriksen (1983a).
Table 3.1: Enrollment ratios (including and excluding repeaters) in the first and final grades of primary education, 1980 and 1987. (Percentages)

<table>
<thead>
<tr>
<th></th>
<th>Enrollment Ratio Grade 1 Including repeaters</th>
<th>Enrollment Ratio Final Grade Including repeaters</th>
<th></th>
<th>Enrollment Ratio Grade 1 Excluding repeaters</th>
<th>Enrollment Ratio Final Grade Excluding repeaters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>103</td>
<td>89</td>
<td>84</td>
<td>72</td>
<td>58</td>
</tr>
<tr>
<td>Asia incl. China</td>
<td>128</td>
<td>132</td>
<td>95</td>
<td>101</td>
<td>75</td>
</tr>
<tr>
<td>Asia excl. China</td>
<td>118</td>
<td>124</td>
<td>98</td>
<td>103</td>
<td>61</td>
</tr>
<tr>
<td>Latin America</td>
<td>186</td>
<td>177</td>
<td>140</td>
<td>138</td>
<td>68</td>
</tr>
</tbody>
</table>

Note: The ratios relate a continent's enrollment in, respectively, Grade 1 and the final grade of primary education (including or excluding repeaters) to the population corresponding to this grade. The ratios for Asia exclude Japan.

Source: Calculations made by the author based on data supplied by the UNESCO Office of Statistics.

The corresponding figures for 1987 show the capacity for new admission to Grade 1 remained well above the size of the population of admission age in Latin America, and improved to above 100% in Asia (excluding China). The marked decline for Africa is very disturbing. The 1980s is the first period since independence that the rate of growth in new intake to Grade 1 (0.8% as an annual average for 1980-87) has been well below that of the continent's population of entry age (3.0% annually). The capacity of the final grade of the cycle remained unchanged during this seven-year period in Africa and Latin America, and improved markedly in Asia.

The development in new intake to Grade 1 resulted in a decline in the gross enrollment ratio for primary education for Africa during the period 1980-87 (from 79% to 75%), and in increases for Latin America (from 105% to 108%) and Asia (from 97% to 105%).

16/ This fact by itself shows that there are serious data problems for this region since it is not possible to maintain a net intake rate, i.e., one that excludes repeaters, of about 140% for a seven-year period. The causes for these errors are discussed in the references given in the previous note.
Tables 3.2 and 3.3 illustrate the overall magnitude of repetition in primary and general secondary education. The tables show an amazing variety in the percentage of repeaters for countries at similar levels of economic and educational development. It is also quite striking to note the differences in Africa between former English dependencies on the one hand, and French and Portuguese dependencies on the other. For example, the latter group includes all the 14 countries that had more than 25% repeaters in primary education and all the 21 countries that had more than 13% repeaters in general secondary education. The factors causing such differences will be further explored at the seminar for which this paper has been prepared. However, the tables clearly suggest that, in addition to factors such as low student achievement, repetition is also heavily influenced by promotion practices adopted by different systems of education.

Table 3.4 illustrates the persistence of the problem of repetition in primary education since the mid 1960s. It also shows some disturbing features with respect to the development of repetition in Sub-Saharan Africa. For example, while only eleven of the 48 non-African countries for which data were available for at least one of the two ten-year periods covered experienced increased repetition up to 1985, this was the case for as many as 20 of the 30 Sub-Saharan African countries covered. Furthermore, the 13 of these 30 countries that had the highest level of repetition in 1985 (ranging from 26.9% to 40.9%) saw their level of repetition increase. All these 13 countries were francophone or lusophone.

---

17/ The data shown are derived mainly from UNESCO (1984), (1988), (1989), and (1990).
### Table 3.2: Percentage Repeaters in Primary Education, Both Sexes
(Last year available between 1984-89)

<table>
<thead>
<tr>
<th>Country</th>
<th>Below 5%</th>
<th>5% - 9.9%</th>
<th>10% - 14.9%</th>
<th>15% - 19.9%</th>
<th>20% - 24.9%</th>
<th>25% - 29.9%</th>
<th>30% and over</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AFRICA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIMBABWE</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOTSWANA</td>
<td>6.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWANDA</td>
<td>10.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIGER</td>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAWI</td>
<td>21.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONGO</td>
<td>30.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>1.8</td>
<td>SOLOMON IS 7.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIRIBATI</td>
<td>2.0</td>
<td>CHINA     7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THAILAND</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LAÇ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAMAICA</td>
<td>3.3</td>
<td>GUYANA    6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRINIDAD &amp; TOBAGO</td>
<td>3.5</td>
<td>EL SALVADOR 7.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUATEMALA</td>
<td>4.3</td>
<td>PARAGUAY  8.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EMENA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREECE</td>
<td>0.1</td>
<td>UNITED AR EM 5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYPRUS</td>
<td>0.4</td>
<td>JORDAN    5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>YUGOSLAVIA</strong></td>
<td>1.7</td>
<td>AFGHANISTAN 5.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLAND</td>
<td>2.3</td>
<td>KUWAIT    6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALTA</td>
<td>2.6</td>
<td>ALGERIA   6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGYPT</td>
<td>2.6</td>
<td>SYRIA     7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUNGARY</td>
<td>2.9</td>
<td>TURKEY    7.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Source:</strong> Unesco Office of Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.3: Percentage Repeaters in Secondary Education, Both Sexes
(Last year available between 1984-89)

<table>
<thead>
<tr>
<th>Country</th>
<th>Below 5%</th>
<th>5% - 9.9%</th>
<th>10% - 14.9%</th>
<th>15% - 19.9%</th>
<th>20% - 24.9%</th>
<th>25% - 29.9%</th>
<th>30% and over</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AFRICA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIMBABWE</td>
<td>0.0</td>
<td>LÆSOTHO   6.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOTSWANA</td>
<td>0.1</td>
<td>DJIBOUTI  9.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAWI</td>
<td>1.6</td>
<td>SWAZILAND 9.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAMBIA</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAMBIA</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWANDA</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOREA</td>
<td>0.0</td>
<td>BHUTAN    5.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIRIBATI</td>
<td>1.1</td>
<td>SRI LANKA 6.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THAILAND</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LAÇ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL SALVADOR</td>
<td>0.6</td>
<td>NICARAGUA 5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAMAICA</td>
<td>1.5</td>
<td>HAITI     8.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRINIDAD &amp; TOBAGO</td>
<td>2.3</td>
<td>COSTA RICA 9.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EMENA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLAND</td>
<td>1.4</td>
<td>UNITED AR E 7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYPRUS</td>
<td>1.9</td>
<td>JORDAN    7.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALTA</td>
<td>2.9</td>
<td>ALGERIA   13.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFGHANISTAN</td>
<td>4.1</td>
<td>QATAR     13.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREECE</td>
<td>4.8</td>
<td>SAUDI AR 14.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Source:</strong> Unesco Office of Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Changes in the Percentage of Repeaters in Primary Education by Region, For 1965, 1975 and 1985

<table>
<thead>
<tr>
<th>Region</th>
<th>1965</th>
<th>1975</th>
<th>1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFRICA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUINEA-BISSAU</td>
<td>41.6%</td>
<td>40.9%</td>
<td>34.8%</td>
</tr>
<tr>
<td>TOGO</td>
<td>33.9%</td>
<td>28.6%</td>
<td>34.8%</td>
</tr>
<tr>
<td>MADAGASCAR</td>
<td>28.9%</td>
<td>24.4%</td>
<td>33.2%</td>
</tr>
<tr>
<td>GABON</td>
<td>34.0%</td>
<td>32.4%</td>
<td>12.8%</td>
</tr>
<tr>
<td>CONGO</td>
<td>26.5%</td>
<td>25.1%</td>
<td>29.9%</td>
</tr>
<tr>
<td>MALI</td>
<td>25.9%</td>
<td>22.9%</td>
<td>29.6%</td>
</tr>
<tr>
<td>CENTRAL AFR R</td>
<td>28.6%</td>
<td>35.2%</td>
<td>18.9%</td>
</tr>
<tr>
<td>CAMEROON</td>
<td>25.1%</td>
<td>29.3%</td>
<td>19.0%</td>
</tr>
<tr>
<td>COTE D'IVOIRE</td>
<td>20.9%</td>
<td>28.3%</td>
<td>17.7%</td>
</tr>
<tr>
<td>BENIN</td>
<td>21.9%</td>
<td>21.2%</td>
<td>25.9%</td>
</tr>
<tr>
<td>CHAD</td>
<td>27.1%</td>
<td>36.7%</td>
<td>23.6%</td>
</tr>
<tr>
<td>LESOTHO</td>
<td>6.0%</td>
<td>22.6%</td>
<td>15.5%</td>
</tr>
<tr>
<td>ZAIRE</td>
<td>22.9%</td>
<td>11.1%</td>
<td>18.8%</td>
</tr>
<tr>
<td>MAURITANIA</td>
<td>14.8%</td>
<td>18.3%</td>
<td>10.4%</td>
</tr>
<tr>
<td>MALAWI</td>
<td>16.5%</td>
<td>17.8%</td>
<td>12.0%</td>
</tr>
<tr>
<td>BURUNDI</td>
<td>21.9%</td>
<td>25.6%</td>
<td>17.7%</td>
</tr>
<tr>
<td>GAMBIA</td>
<td>13.5%</td>
<td>10.4%</td>
<td>17.4%</td>
</tr>
<tr>
<td>SENEGAL</td>
<td>18.1%</td>
<td>14.9%</td>
<td>16.3%</td>
</tr>
<tr>
<td>BURKINA-FASO</td>
<td>17.9%</td>
<td>17.5%</td>
<td>15.9%</td>
</tr>
<tr>
<td>NIGER</td>
<td>23.2%</td>
<td>13.1%</td>
<td>15.0%</td>
</tr>
<tr>
<td>UGANDA</td>
<td>10.2%</td>
<td>14.1%</td>
<td>27.0%</td>
</tr>
<tr>
<td>KENYA</td>
<td>5.1%</td>
<td>5.2%</td>
<td>13.1%</td>
</tr>
<tr>
<td>RWANDA</td>
<td>21.5%</td>
<td>16.9%</td>
<td>12.1%</td>
</tr>
<tr>
<td>BOTSWANA</td>
<td>0.3%</td>
<td>2.5%</td>
<td>5.7%</td>
</tr>
<tr>
<td>GHANA</td>
<td>3.0%</td>
<td>2.3%</td>
<td>2.1%</td>
</tr>
<tr>
<td>ZAMBIA</td>
<td>2.1%</td>
<td>2.1%</td>
<td>2.1%</td>
</tr>
<tr>
<td>TANZANIA</td>
<td>0.5%</td>
<td>1.0%</td>
<td>12.5%</td>
</tr>
<tr>
<td>SUDAN</td>
<td>2.6%</td>
<td>2.7%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 3.5: Percentage of Repeaters in Primary Education by Sex, (Latest year available between 1989 - 1984)

<table>
<thead>
<tr>
<th>Region</th>
<th>Male</th>
<th>Female</th>
<th>Difference (1)-(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFRICA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAMBIA</td>
<td>2.2%</td>
<td>1.9%</td>
<td>0.3%</td>
</tr>
<tr>
<td>BOTSWANA</td>
<td>0.5%</td>
<td>4.9%</td>
<td>4.4%</td>
</tr>
<tr>
<td>BENIN</td>
<td>25.9%</td>
<td>26.8%</td>
<td>-0.9%</td>
</tr>
<tr>
<td>GUINEA</td>
<td>18.9%</td>
<td>14.6%</td>
<td>4.3%</td>
</tr>
<tr>
<td>PARAGUAY</td>
<td>21.9%</td>
<td>19.0%</td>
<td>2.9%</td>
</tr>
<tr>
<td>CHILE</td>
<td>10.4%</td>
<td>12.5%</td>
<td>-2.1%</td>
</tr>
<tr>
<td>CHAD</td>
<td>23.7%</td>
<td>21.9%</td>
<td>1.8%</td>
</tr>
<tr>
<td>TUNISIA</td>
<td>3.3%</td>
<td>3.5%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>SENEGAL</td>
<td>13.2%</td>
<td>15.0%</td>
<td>-1.8%</td>
</tr>
<tr>
<td>MALAWI</td>
<td>21.4%</td>
<td>20.8%</td>
<td>0.6%</td>
</tr>
<tr>
<td>NIGER</td>
<td>23.8%</td>
<td>29.0%</td>
<td>-5.2%</td>
</tr>
<tr>
<td>SUDAN</td>
<td>2.6%</td>
<td>2.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AFRICA &amp; AR.</td>
<td>34.6%</td>
<td>34.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td>TRINIDAD</td>
<td>3.5%</td>
<td>3.1%</td>
<td>0.4%</td>
</tr>
<tr>
<td>TOGO</td>
<td>30.2%</td>
<td>31.6%</td>
<td>-1.4%</td>
</tr>
<tr>
<td>MADAGASCAR</td>
<td>35.7%</td>
<td>33.4%</td>
<td>2.3%</td>
</tr>
<tr>
<td>CENTRAL AFR</td>
<td>34.0%</td>
<td>33.7%</td>
<td>0.3%</td>
</tr>
<tr>
<td>CHAD</td>
<td>31.4%</td>
<td>34.6%</td>
<td>-3.2%</td>
</tr>
<tr>
<td>BAHRAIN</td>
<td>8.9%</td>
<td>9.3%</td>
<td>-0.4%</td>
</tr>
<tr>
<td>SPAIN R.</td>
<td>15.3%</td>
<td>15.1%</td>
<td>0.2%</td>
</tr>
<tr>
<td>JERUSALEM</td>
<td>16.8%</td>
<td>17.4%</td>
<td>-0.6%</td>
</tr>
<tr>
<td>ASIA</td>
<td>1.6%</td>
<td>2.0%</td>
<td>-0.4%</td>
</tr>
</tbody>
</table>

Source: UNESCO, Office of Statistics
Figure 3.1: Proportion of Cohort Reaching Last Grade of Primary Education Around 1965 and 1985

Country

Korea
Greece
Yugoslavia
Hong Kong
Jordan
Egypt
Syria
Algeria
Singapore
Cote d'Ivoire
Botswana
Burundi
Uganda
Syria
Algeria
Burundi
Uganda

Between 1975 and 1985, the median level of repetition in primary education increased from 16.7% to 18.6% in Sub-Saharan Africa, while it declined from 12.5% to 10.6% in Latin America\(^{18}\), and from 10.8 to 7.5% in EMENA. In short, although repetition is a serious problem in all regions, it is increasing in Africa, i.e., in the region that already at the outset had the highest level of repetition, and that today has the lowest level of admission and coverage of primary education. Considering the present resource-constrained situation in most of these countries, these trends are of great concern and suggest that the development of measures to address this problem adequately needs to be given more attention in bank sector work and lending operations. This is especially so for francophone and hispanophone countries.

Table 3.5 illustrates the fact that, with the exception of Africa, the propensity to repeat in primary education is higher for boys than for girls in practically all countries outside Sub-Saharan Africa. Of the 33 Sub-Saharan countries covered, the level was higher for boys than for girls in 19 countries.

Finally, Figure 3.1 illustrates the development in survival to the final grade of primary education for cohorts entering school in 1965 and 1985. The survival rates have been derived using the Reconstructed Cohort Method discussed in Section VII. The figure shows that 35 of the 42 countries covered improved the holding power of their primary schools over this twenty-year period. However, the figure also shows that five of the seven countries that experienced increased dropout during this period already had low survival rates in 1965, and that -- despite the improvement registered for most countries -- even in 1985 more than 20% of the cohort dropped out prior to the final grade of the cycle.

\(^{18}\)As discussed elsewhere in this paper, many studies indicate that the level of repetition in Latin America is grossly underestimated. However, also these studies suggest that the level has been declining.
IV. Analyzing Admission to a Cycle of Education

One important aspect of assessing the adequacy of the existing enrollment capacity to a cycle of education is the estimation of past trends in new entrants to this cycle, and to project how future admission would need to develop in order to reach specific targets for access. Because of the particular problem of assessing admission capacity in primary education in countries with a high share of late entrants, the discussion below will focus on this level of education. However, the methods discussed are, in principle, applicable for any cycle of education.

In countries with universal primary education, where practically all new entrants start school at the official entry age, the problem is relatively simple: the number of new entrants can normally be expected to equal the population in the admission age group. The number of new entrants for future years may consequently be derived directly from the population projections.

The matter is more complex in developing countries since attendance of primary school generally is not compulsory and because new entrants usually come from several age groups. In addition, the data available on new entrants often suffer from over-reporting, either because there is a general over-reporting of Grade 1 enrollment, or because the data available on enrollment in this grade do not distinguish correctly between repeaters and new entrants. These data problems will be discussed Section VI. The present discussion will be limited to illustrating (a) some of the methodological problems arising when estimating and projecting new entrants for countries for which new admission is drawn from several age groups; and (b) the fact that there are, for some countries, quite severe problems in the data available on new entrants.

The problems discussed below are of particular relevance when designing project interventions for the many developing countries which now are approaching universal intake to primary education. For these countries, it is important to be able to estimate to a fairly accurate degree, the extent to which the existing enrollment capacity is sufficient to cater to the population of admission age and, if not sufficient, how much and in which geographical areas of the country
additional capacity should be provided. In a country's effort to reach universal primary education, the last 5-10% of the population is normally the most difficult to reach, and proper targeting of interventions must be based on an accurate analysis of the current situation.

A. Apparent intake rates

In a country where all new entrants start school at the same age, the number of new entrants to the educational system in a given year depends upon the number of children in the admission age group and the proportion of this age group which enters school. Assuming that children start school at the age of six, this may be expressed as:

\[ N_t = e_t P_6 \]

where

- \( N_t \) = Number of new entrants in year \( t \)
- \( e_t \) = Intake rate in year \( t \)
- \( P_6 \) = Population aged 6 years in year \( t \)

It is an apparent or gross intake rate since many of the new entrants may be older or younger than the official entry age. This rate is also referred to as an admission rate. For countries where the intake rate equals 1, the number of new entrants is equal to the population aged six years. However, difficulties arise when formula (4) is used for assessing the admission capacity of countries where new admission comes from several age groups, which is the case in most developing countries. In such cases, the intake rate may exceed 1 during a certain period, and it is not evident what its upper limit will be or for how many years it will exceed 1.

To illustrate this point, Table 4.1 provides data on (a) the Grade 1 enrollment ratio (i.e., new intake plus repeaters divided on the population of admission age), and (b) the apparent intake rate for a few selected countries and years covering the period 1965 until the latest year for which data were available after 1985. The data refer to boys only.

---

\(^{10}\) The problem is similar to that experienced for the gross enrollment ratio for primary education which may exceed 100% due to enrollment of over-aged pupils.
The enrollment ratios for Grade 1 have been included since data on repeaters were not available in all cases to permit estimation of new entrants and, hence, intake rates.

The main purpose of this table is to illustrate that (a) apparent intake rate in many cases exceed 100\% by a considerable margin; and (b) as will be shown in the below discussion, for many of these countries there are clearly major errors in the data available since it is theoretically not possible for an apparent intake rate to remain way above 100\% for period extending over 10-
15 years as in the case of, e.g., Lesotho, Zaire, Brazil, Colombia, Ecuador, Syria, Indonesia and the Philippines.

Although the data available on repetition are too sparse to allow a thorough analysis of this aspect for many populous countries, the very high enrollment ratios for Grade 1 for most such countries suggest that world-wide estimates of the intake capacity to primary education is severely overestimated. For countries where overestimation of Grade 1 enrollment or new intake seems likely, more attention needs to be given in Bank sector work to understanding better what goes on in the first grade of the cycle -- how many children enter, how many drop out, when and for what reason, how many repeat and for what reason. We shall revert to these factors when discussing more fully, data errors in Section VI.

B. Age-specific Intake Rates

Disregarding data errors, the main factor causing apparent intake rates to exceed 100% is rapid increase over time in new admission in cases where new entrants are recruited from several age groups. The effect is reinforced if at the same time there is a shift in the age distribution of new entrants towards younger ages. This is illustrated in Table 4.2 below for a hypothetical country where the new entrants are assumed to come from two age groups only, and where the population in these two age groups, for the sake of simplifying the example, is assumed to be constant during the period considered. The effects of mortality, furthermore, are ignored.
Table 4.2: Simulations of Apparent Intake Rates

<table>
<thead>
<tr>
<th>Age</th>
<th>1980</th>
<th>1981</th>
<th>1982</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td>% New entrants</td>
<td>Intake rate*</td>
</tr>
<tr>
<td>6 years</td>
<td>100</td>
<td>40%</td>
<td>40</td>
</tr>
<tr>
<td>7 years</td>
<td>100</td>
<td>50%</td>
<td>50</td>
</tr>
</tbody>
</table>

* This rate has been obtained by dividing the total number of new entrants age 6 and 7 years by the population aged 6 years.

The example starts out by assuming that 40% of the six year-olds and 50% of the seven year-olds start school in 1980, leading to an intake rate of 0.9 if the population aged six year-olds is used as the basis for the calculation. In 1981, there is an increase in the intake from both ages, leading to an intake rate of 1.15. In 1982, the increase in enrollment of six year-olds continues, while the new enrollment of seven year-olds decreases to 40%. As 60% of this cohort was already enrolled in 1981, it would not be possible to enroll more than 40% since we assume that a child can only enter school once. The increase in the enrollment of six year-olds continues until 1984, where it reaches 100% of the age group, while the decrease in the new enrollment of seven year-olds reaches zero in 1985, at which time the intake rate is 1.00.

In this example, the intake rate reached 1.20 as its maximum value and remained higher than 1.0 for four years. It is obvious that the upper limit and the development over time would have been different if the intake rates for each of the two age groups had developed differently, and even more so if the new entrants had come from more than two age groups. In reality, the age distribution of new entrants changes less rapidly than assumed in the above example and new entrants are drawn from more than two single ages. While these two factors would extend the period for which the apparent intake rate would
exceed unity, they would also limit the maximum value the ratio could reach. If data are correctly reported, the ratio should only in exceptional cases, exceed 1.10, and then only for a very short period.

A satisfactory analysis of this problem requires data on new entrants by age for consecutive years. Such data will permit us to study the intake rate by single years of age and not only the general intake rate as defined above. If the new entrants come from several age groups, this may be expressed as:

\( N_t = e_{6}^t P_{6}^t + e_{7}^t P_{7}^t + \ldots + e_{6+x}^t P_{6+x}^t \)

where \((6+x)\) is the oldest age group from which new entrants are drawn and \(e_{6}^t, \ldots, e_{6+x}^t\) are the percentages of each age group starting school in year \(t\).

Combining equations (4) and (5), the apparent intake rate \(e^t\) may be expressed as:

\( e^t = e_{6}^t + e_{7}^t \frac{P_{7}^t}{P_{6}^t} + e_{8}^t \frac{P_{8}^t}{P_{6}^t} + \ldots + e_{6+x}^t \frac{P_{6+x}^t}{P_{6}^t} \)

It is clear from this equation that \(e^t\) may exceed 1. It will remain stable over time as long as \(e_{6}^t, \ldots, e_{6+x}^t\) and the age structure of the school population remain unchanged. Built on past data on the age distribution of new entrants, the trend of the intake rates by single years of age may be studied and projected. Note, however, that even in this case, upper limits have to be fixed each year for each of the intake rates by single years of age since they are inter-related over time. This is the sort of limit which had to be applied in the example given above where the intake rate for seven year-olds in 1982 could at the maximum be 40% since 60% of the cohort were already enrolled at school the previous year. In a more general manner, this is illustrated in the table below where \(P_{6}^t, P_{7}^t, \ldots, P_{6+x}^t\) refer to children belonging to the same cohort.
The sum of the intake rates $e_6^t + e_{t+1}^t + \ldots + e_{t+x}^t$ can therefore, at maximum, reach 1, since these rates refer to the same cohort of children. Thus, if $e_6^t$ increases rapidly over time, $e_{t+1}^t, \ldots, e_{t+x}^t$ will decrease in the future. Consequently, although the apparent intake rate defined by (4) during a certain stage of development of primary education may exceed 1, in the long run it will decrease again to 1 and the number of new entrants will be determined by the population in the official admission age.

One possible way of ensuring that the sum of the intake rates for one cohort does not exceed 1, would be to project each intake rate by single years of age separately under the constraints that:

$$E^t = e_6^t + e_{t+1}^t + \ldots + e_{t+x}^t \leq 1$$

where $E^t$ is the intake rate for the cohort being six year-olds in year $t$.

The above analysis ignores mortality. If this aspect is taken into account, the inequality sign in (7) will be effective, as $E^t$ will be less than unity.

We shall not pursue this discussion further here. Readers who would like to study methods of analyzing and projecting new entrants to primary education or to other levels of education may consult Thonstad (1980).
V. Analyzing the Flow of Pupils Between Two School Years

A. Student Flow Models

In analyzing past enrollment trends, as well as for projecting the number of pupils enrolled in future years, educational planners in most countries have found it useful to employ a set of models normally referred to as student flow models. There exists a great variety of such models, ranging from the extremely limited, covering only one type of education within a given cycle of education, to the complex models covering the whole educational system of a country.20/ The models differ also a great deal in the degree of detail and sophistication of their treatment of each part of the system.

However, in spite of this considerable variation in coverage and detail, the logic behind each of these models is generally very similar as they purport to describe the flow of pupils into, through and out of the educational system or some part thereof. In other words, student flow models take explicit account of the fact that, of the pupils enrolled in a given grade in a given school year, some are promotees from the grade below in the previous year, others are repeaters from the same grade, and yet others may be new entrants (i.e., pupils who have never been to school before) or transfers from other schools. The models further take into account that, in the following school year, the pupils may repeat the grade, be promoted to the next grade, leave the school (i.e., drop out, graduate, or die) or be transferred to another school.

In short, flow models simulate the flow of pupils into, through, and out of the school system. Their fundamental objective is to provide an explicit, logical, and integrated framework into which to fit available data describing the flow of pupils through the system. When such a framework has been developed,

20/ Thonstad (1980) provides an extensive review of flow models. Many such models are used by Bank staff. A set of models developed by Manuel Zymelman has been used frequently in the Africa Region. This is particularly the case for a model designed to simulate the impact on primary education of different budgetary and cost saving measures, see Zymelman (1987).
it may be used to project, i.e., to demonstrate the implications for enrollment at future dates of explicit assumptions about the future developments of the propensities of children to enter school, as well as to be promoted, to repeat, drop out, transfer, or graduate once they are enrolled. The models may also be used to reconstruct the flows of a given cohort of children through a cycle of education. This type of analysis answers the following questions: What proportion of the children entering a cycle of education in a given year is likely to complete the cycle? How many will drop out from different grades? How many "student-years" will need to be invested in this cohort during its stay in the system? And so on. This type of analysis is useful in evaluating the effect of repetition and dropout on the internal efficiency of and output from school systems.

This paper will be limited to discussing the use of the most commonly used flow model for cohort reconstruction. Prior to developing the model, we need, however, to define the parameters describing the flow of pupils through a cycle of education and to show how these parameters or flow rates may be estimated on the basis of the education statistics normally available in developing countries. We shall start by discussing the latter.

B. Data Requirements

Until the mid-1960s, few developing countries collected the minimum data needed for quantifying the parameters of flow models. The data available were almost exclusively stock data, e.g., the number of pupils enrolled in each grade at the beginning of given school years. To estimate the parameters of flow models, flow statistics are needed, i.e., statistics which describe the movement of pupils between the beginning of two consecutive school years. Although the availability of flow statistics has improved during the last two decades, few countries collect information on all the key pupils flows which take place. However, as we shall demonstrate below, since collection and processing of data are costly activities, it is not necessary to collect data on all flows, provided that we have reliable information on some key flow and stock data. In particular, the majority of developing countries now collect data on enrollment
39

and repeaters by grade. Under certain assumptions to be described below, these data permit us to derive the number of new entrants, promotees and dropouts.

To illustrate this aspect, let us study a hypothetical country which has the following (stock) statistics on enrollments in Grades 1 and 2 at the beginning of two consecutive years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Enrollment</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 1</td>
<td>Grade 2</td>
</tr>
<tr>
<td>1989</td>
<td>463,700</td>
<td>410,300</td>
</tr>
<tr>
<td>1990</td>
<td>496,800</td>
<td>424,370</td>
</tr>
</tbody>
</table>

We now ask: In 1989, where were the 496,800 pupils who, in 1990, were enrolled in Grade 1? Obviously, most of them were not enrolled in school and are thus new entrants in 1990. But if repetition is permitted, then some of the pupils enrolled in Grade 1 in 1990 are likely to have been enrolled in this grade also in 1989. Without further information, it is not possible to divide these 496,800 pupils into new entrants and repeaters.

Furthermore, where were the 424,370 Grade 2 pupils of 1990 in the previous year? Again, some of them are repeating this grade. But the largest part was most probably in Grade 1 and has hence been promoted to Grade 2. Again, without further information, it is not possible to know how many are repeaters and how many are promotees.

The relevance of being able to answer these questions in a sector or appraisal report should be obvious. For example, if we assume that the above enrollment data refer to a country where the number of children of admission age in 1990 was 490,000, we would not be able to decide on the basis of this information whether or not to support capacity expansion beyond that required to cater to, e.g., population growth or reduction in class size. Furthermore, should the project support policies/investments to reduce repetition and/or dropout? Again, these data do not permit us to assess whether or not repetition and/or dropout are issues that warrant attention.
In addition to data on enrollment by grade, most developing countries are now also collecting data on repeaters by grade. By combining these two sets of data, it is possible -- under certain assumptions -- to estimate the number of new entrants, promotees and dropouts. To illustrate this, assume that in the above example in 1990 there were 51,900 repeaters enrolled in Grade 1 and 38,570 repeaters enrolled in Grade 2. The number of new entrants to Grade 1 in 1990 can now be derived by subtracting Grade 1 repeaters from Grade 1 enrollment that year, i.e., 496,800 - 51,900 = 444,900. Thus, if the population of admission age is 490,000, the intake rate in 1990 was 90.8%, and there seem to be room for increasing the enrollment capacity in Grade 1. However, before we draw this conclusion, we need to examine factors such as: (i) whether the propensity to repeat Grade 1 should/can be reduced; (ii) whether available capacity can be used more efficiently, e.g., through use of multigrade or double-shift teaching; and (iii) what the geographical distribution is of available capacity and of non-enrolled children.

Next, the number of pupils enrolled in Grade 1 in 1989 who are enrolled in Grade 2 in 1990 can be derived if we assume that there are no new entrants (for example, transfers from abroad) to Grade 2. It equals Grade 2 enrollment in 1990 minus Grade 2 repeaters that year, i.e., 424,370 - 38,570 = 385,800. Finally, if we disregard transfers to and from other countries, the number of 1989 pupils who have dropped out (including dropout due to mortality) can be derived. It equals Grade 1 enrollment in 1989 minus those who repeated this grade in 1990 minus the pupils who were promoted to Grade 2 in 1990, i.e., 463,700 - 51,900 - 385,800 = 26,000.

In this way, based on data on enrollments in Grades 1 and 2 in two consecutive years and on repeaters for the latter of these two years, one can derive the number of promotees and dropouts from Grade 1. The same approach may be used to derive these flows for all the remaining grades of the cycle, provided the above type of data on enrollment and repeaters are available for all grades. For the final grade of the cycle, we need, in addition, data on the number of graduates.
The above procedure is the one generally followed when deriving the enrollment flows between two subsequent grades between two successive school years. Thus, by collecting data on one flow (repeaters) one can derive the three other main flows (new entrants, promotees and dropouts). The approach is based on the assumption that all new entrants start school in Grade 1 and that the only two in-flows to grades other than Grade 1 are promotees and repeaters. Furthermore, at the end of the school year, pupils may be promoted to the following grade (graduate if this is the final grade of the cycle), may repeat the same grade, or may drop out.

Clearly there may be other flows, for example:

(a) There may be new entrants to grades other than Grade 1.
(b) There may be re-entrants to all grades caused by pupils who interrupt temporarily their schooling.
(c) In some cases, pupils may be allowed to skip grades and go, for example, directly from Grade 2 to Grade 4. Inversely, some pupils may be sent down to a lower grade.
(d) There may be international migration, i.e., some pupils may emigrate while some new pupils may be gained through immigration.
(e) In cases where data refer to sub-systems within a country, generally there are transfers between different sub-systems (e.g., between public and private schools, regions and urban and rural areas).

If any of the above flows exists, it would obviously be desirable to take them directly into account. The stumbling block is that statistics on all these flows are rarely available in developing countries. As the collection of such statistics would be costly, it is important to know to what extent the existence of flows of the types listed above will actually distort the estimates of the promotion, repetition, graduation, and dropout rates. Such a distortion would be due to the fact that flows which are not taken explicitly into account must, if they exist, necessarily be included among one or more of the flows which are taken into account. For example, if some pupils transfer to another school system (e.g., to private schools in case our data cover public schools only), they will be counted among the dropouts if transfer is not taken
explicitly into account. To count someone who transfers as a dropout is not serious in a case where we are concerned only with projecting enrollment and graduates. As long as a pupil who transferred is not enrolled in "our" system, and hence does not need to be provided for, it may not be important to know why he/she left. But for someone interested in studying the extent of dropout, it is, of course, crucial to distinguish real dropout from transfers or other causes for leaving the system (e.g., emigration).

In general, we may say that although the four flows (a - d) might be important for certain grades in some countries, generally they are negligible and may thus be disregarded without risk of significantly affecting the flow rates. To collect the statistics required to take these flows into account would be quite costly compared to the benefits derived in terms of improved precision in the flow analysis. It would, therefore, normally be better for developing countries to spend the limited resources available for the collection of education statistics on other more high-priority areas.

Transfer between sub-systems is, however, a flow component that needs to be taken into account when transition rates are calculated separately for each sub-system. However, provided they are correctly reported, global national statistics are not affected by transfers within a country since they cancel out when data for all sub-systems are aggregated. This being said, available flow statistics, in many countries, are limited to covering the public system. In some of these systems, there may be considerable transfer of students from private schools in all grades of the cycle, sometimes leading to negative dropout rates. This happens, for example, in many countries where there are religious schools providing partial or full-fledged primary education (e.g., Koranic schools or medrassa schools in West African countries). To ignore such transfers may significantly affect our conclusions with respect to the internal efficiency of different sub-systems. For example, if there is a net transfer of students from private to public schools, to ignore this would, other things equal, lead to under-estimation of dropout in public schools and over-estimation of dropout in private schools.

21/ For example, emigration is an important factor in some small countries.
C. Calculations of Flow Rates

In the illustration given above, the numbers of promtees and dropouts were estimated from data on enrollment and repeaters. This example may now be taken one step further and we may compute the proportion of pupils enrolled in Grade 1 in 1989 that, in the following year, repeated this grade, were enrolled in Grade 2 as promtees, or dropped out. These proportions are normally called transition or flow rates.22/

For instance, by calculating the proportion of Grade 1 pupils in 1989 who repeated that grade the following year, one obtains a flow rate which is called the repetition rate for Grade 1 in 1989:

\[
\frac{51,900}{463,700} = 0.112
\]

Similarly, the promotion rate for Grade 1 in 1989 is computed as the proportion of Grade 1 enrollment in 1989 promoted to Grade 2 in 1990:

\[
\frac{385,800}{463,700} = 0.832
\]

Finally, the dropout rate is the proportion of Grade 1 enrollment in 1989 which was not enrolled in school in the beginning of 1990:

\[
\frac{26,000}{463,700} = 0.056
\]

Obviously, the sum of these three rates must equal unit. Thus, if two rates are known, the third can always be derived. Therefore, to avoid rounding problems in cases where the three rates computed in the above manners do not add up to unity, it is preferable to derive the dropout rate as a residual.

It should be emphasized that the flow rates for a given grade describes the changes which take place in the enrollment in this grade between the beginning of two consecutive school years. Thus, we must distinguish between,

\[22/\] Sometimes the term "transition rates" is used only to denote the rates which describe transition between different cycles of education, e.g., from the final grade of primary education to the first grade of secondary education.
e.g., the repetition rate for a given grade and the percentage of pupils in this grade who are repeaters. The former is a flow rate describing a change in the stock between the beginning of two school years, the latter shows the composition of this stock at the beginning of a given school year. For example, in the above illustration, the repetition rate for Grade 1 in 1989 is 11.2% while the percentage of repeaters in Grade 1 in 1990/91 is 10.4%. Furthermore, as will be discussed in Section VII, it is important to distinguish between promotion rates and cohort survival rates. The former are flow rates, showing the share of pupils enrolled in a given grade in a given year who are enrolled in the following grade in the following school year. The latter shows the proportion of those enrolled in a given grade who eventually reach the following grade, i.e., after all repeaters have either dropped out or have been promoted.

Note that by convention, the year used to denote a given flow rate refers to the first of the two school years between which the rate describes the flows. Thus, in the above example, the rates calculated refer to 1989. In cases where the school year does not coincide with the calendar year, it is usual to use notations such as school year 1989/90. This would be the case where the school year starts in e.g., September and ends next June. For ease of reference, we shall not, in the following, use the double notation.

It is useful for the subsequent discussion to define the above rates in mathematical terms. The following symbols are normally used for this purpose.

The repetition rate for Grade $g$ in year $t$ =

\[
\frac{\text{number of pupils repeating Grade } g \text{ in year } t+1}{\text{total number of pupils in Grade } g \text{ in year } t}
\]

or in symbols: \[ (8) \quad r_g^t = \frac{R_g^{t+1}}{E_g^t} \]
The promotion rate for Grade $g$ in year $t$ =

\[
\frac{\text{number of promotees enrolled in Grade } g+1 \text{ in year } t+1}{\text{total number of pupils in Grade } g \text{ in year } t}
\]

or in symbols: \( p_{g}^{t+1} \frac{E_{g+1}^{t+1}}{E_{g}^{t}} \)

The dropout rate for Grade $g$ in year $t$ =

\[
\frac{\text{number of pupils dropping out from Grade } g \text{ in year } t}{\text{total number of pupils in Grade } g \text{ in year } t}
\]

or in symbols: \( d_{g}^{t} \frac{D_{g}^{t}}{E_{g}^{t}} \)

For the final grade of the cycle, we may calculate the graduation rate defined as:

\[
\frac{\text{number of pupils graduating at the end of year } t}{\text{total number of pupils in the final grade in year } t}
\]

or in symbols: \( g_{m}^{t} \frac{G_{m}^{t}}{E_{m}^{t}} \)

(The subscript $m$ designates the final grade of the cycle)

As already observed, the sum of the rates describing the flows out of a given grade between the beginning of two successive school years equals unity:

\( p_{g}^{t} + r_{g}^{t} + d_{g}^{t} = 1 \)

For the final grade of the cycle, the graduation rate replaces the promotion rate in this formula. To sum up, based on data on enrollment by grade for two consecutive school years and data on repeaters by grade for the last of these two years, we may, under certain assumptions, estimate intake, promotion, repetition, and dropout rates. We shall discuss in Section VII how, on the basis of these rates, we may derive indicators not only of annual flows, but of what happens to the cohort during the whole duration of its stay in the cycle.
However, first we shall discuss below how errors in available data may affect the flow rates estimated in the way explained above and, in turn, the cohort flow indicators derived in Section VIII.

VI. Effect on Flow Rates of Common Errors in Data

As explained, the omission of flows such as migration and re-entrants will, if they exist, affect the estimates of the flow rates derived in the way defined above. But even if none of these other flows takes place, the accuracy of the estimated promotion, repetition and dropout rates will still depend on the quality of the data available on enrollment and repeaters by grade. The accuracy of these data often leaves much to be desired. This section reviews how four categories of errors in data common to many countries will affect the estimates of the flow rates. These errors are:

(a) over-reporting of Grade 1 enrollment;
(b) under-reporting of the number of repeaters;
(c) yearly variation in coverage of data; and
(d) incorrect definitions of the flows.

The reason for discussing these errors in some detail is that some of them, and particularly those mentioned in (a) and (b) above, may significantly affect the estimates of repetition, dropout, and new entrants. In turn, this affects our ability to design project interventions to effectively address problems in the sector. For example, as will be shown below, over-reporting of Grade 1 enrollment leads to over-estimation of enrollment, new intake and dropout. This may lead to design of a project that gives too little emphasis to provision of additional capacity and too much to reducing dropout. Similarly, under-reporting of repetition leads to over-estimation of intake and dropout. Again, this may lead to design of a project that does not adequately address the need to improve quality to reduce repetition, or the need to provide additional capacity to cater to non-enrolled children.
A. Over-reporting of Grade 1 Enrollment

Schools reporting enrollment figures exceeding actual enrollment is not uncommon in developing countries. One reason is that schools sometimes have a financial incentive to overstate enrollment: for example, the grant received or the number of teachers paid by the government may be a direct function of the number of pupils enrolled.\textsuperscript{23}

The fact that over-reporting may be serious in some cases is illustrated by the following quotation from an Indian report referring to primary education:

"... the existing data on enrollment suffer from over-reporting; a certain proportion of the total enrollment is bogus in the sense that it includes children whose names are shown on the registers as enrolled but who, in fact, do not attend schools so that the daily average attendance is often too small in relation to the overall enrollment. There is evidence to show that this over-reporting is large (it varies from 8 to 47 percent) and that it is larger in those States which are backward in elementary education."\textsuperscript{24}

Discussion with national specialists indicates that most of this over-reporting takes place in Grade 1 and that a major reason for exaggerating the actual enrollment is to obtain additional teachers. Over-reporting of Grade 1 enrollment is also found in other countries, e.g., in Bangladesh.\textsuperscript{25}

A second and different type of over-reporting occurs in countries where parents have incentives to register their children in school at the beginning of the school year, but where a large number of those registered do

\textsuperscript{23} Incentives for over-reporting of enrollment may also exist in industrialized countries. For example, in countries where the school-age population is declining, the resulting decline in enrollment may not be fully reported since this would lead to cut in resources. In one European country, it has been reported that small schools over-reported enrollment because the headmaster would have to teach if enrollment declined below a certain minimum. This error in available enrollment statistics was discovered when the statistics were compared with those on the population of school age.

\textsuperscript{24} See Government of India (1977), page 5.

\textsuperscript{25} See Khan (1980).
not attend school or attend only for a brief period. If the enrollment statistics in such cases refer to, for example, the first week of the school year, a number of children who, for learning purposes, are not enrolled, will be included in the enrollment figure. Contrary to the first example above, in this case there is usually no deliberate over-reporting on the part of the school.

Figure 6.1 illustrates the flows between Grades 1 and 2 between two subsequent school-years. As regards the effect of over-reporting of Grade 1 enrollment on the flow rates for this grade, we note that if the enrollment and repeaters in Grade 2 are both correctly reported, then the estimated number of promotees from Grade 1 will not be affected since this estimate is based on the data for Grade 2. However, if the enrollment in Grade 1 is overstated, then the promotion rate will be under-estimated since the denominator in equation (9) will be over-stated. The same conclusion holds true for the repetition rate in Grade 1 if the number of repeaters in this grade is correctly reported. In the case of the number of repeaters also being overstated, the extent to which the repetition rate is over- or under-estimated depends on the relative magnitudes of the over-reporting of Grade 1 enrollment and repeaters, respectively [see equation (8)].

Finally, in the case where Grade 1 enrollment is over-reported but where Grade 1 repeaters are correctly reported, and where the number of promotees to Grade 2 is correctly estimated, the number of dropouts from Grade 1 will be over-estimated. Similarly, if Grade 1 enrollment is over-estimated, and Grade 1 repeaters are correctly estimated, the number of new entrants will be over-estimated since new admission is estimated as the difference between enrollment and repeaters. Because they are derived as residuals, the estimates of new admission and dropout are very sensitive to data errors of this type. Comparatively small over-reporting of Grade 1 enrollment may cause considerable distortion in the estimate of these two flows, and consequently in our diagnosis of priority project interventions.
To illustrate this, suppose that the "true" enrollment in Grade 1 is 100 pupils while the number reported is 110, i.e., an over-reporting of 10%, and that the number of promotoes and repeaters originating from these 100 pupils is correctly estimated as 75 and 15 respectively. This means that 85 new students entered Grade 1 and that 10 dropped out. Using the incorrect enrollment figure, we estimate that 20 pupils have dropped out and that 95 entered. The estimated promotion, repetition and dropout rates will be 68%, 14% and 18% respectively, as compared to the "true" rates of 75%, 15%, and 10%. Thus, accounting for errors in the data, the magnitude of the dropout--while still a
major concern that needs to be addressed in project design—is reduced by about half. Similarly, the real level of admission is considerably lower than indicated by available statistics.

We shall discuss further the implications of this type of errors in data after having discussed impact of under-reporting of repetition.

B. Under-reporting of the Number of Repeaters

This type of error in the data available appears to be common to many developing countries.\textsuperscript{26} In particular, this error is often found in the data available for Grade 1 of primary education where pupils who have already attended school are frequently recorded as new entrants. In some cases, it is caused by pupils dropping out from one school, or who are obliged to leave after having repeated the maximum number of times permitted, and who present themselves as new entrants at another school. In other cases, it is caused by registration as new entrants of children who were already registered as such the previous school year, but who dropped out after a short period. In yet other cases, pupils who completed the year unsuccessfully may, because of inadequacy of school registers, be counted as new entrants, rather than repeaters. Finally, because of rules limiting the level of repetition, teachers may deliberately under-report the actual level of repetition, to stay within such limits. In all these cases, the same children are counted more than once as new entrants, implying an over-reporting of the number of new entrants and under-reporting of repetition.

The teachers' task of distinguishing between new entrants and repeaters is complicated by the inadequacy of available school records. For example, Schiefelbein and Grossi (1980) found that:

\textsuperscript{26} Several studies suggest that this type of error is particularly common in Latin America, see for example, Schiefelbein (1975), AID (1977), Schiefelbein and Grossi (1980), Schiefelbein, et. al. (1989), and Cuadra (1989). While all these studies refer to Latin America, evidence suggests that under-reporting of repetition is quite common in many countries in Africa and Asia as well.
"Latin American education system does not keep a master record where the whole educational history of each child is recorded. Good records exist in some schools, but these records do not follow the students when they change school. Thus, students transferring from one school to another will usually be considered as non-repeaters. Students leaving school mid-year due to illness or in anticipation of failure are usually counted as dropouts in that year, and as newcomers if they register in the following year rather than as repeaters. To count the true number of repeaters would require that sophisticated records are kept in each school and, furthermore, that the teachers have time to use and maintain them. However, field work carried out in Paraguay showed that it is too lengthy to get the information from official school files and there is no administrative support for such work. Therefore, teachers tend to shortcut the work as they have exams to grade and other tasks to perform. The total enrollment was easily obtained for each course from the 'Libro de Clases,' while repeaters were counted by the teacher asking the students to raise their hands if they were repeaters," (op.cit. p.24).

The magnitude of this type of error may be considerable in some countries. For example, the study cited above estimates that the actual repetition rate for Grade 1 in Latin American countries may be more than double the size of that derived from official statistics. This makes a significant difference for countries for which even the official rate for Grade 1 generally exceeds 20%. Schiefelbein, et. al. (1989) estimates that the real repetition rate in Grade 1 in Latin America was almost 50% in Grade 1 in 1980 and declined to about 40% in 1987. Cuadra (1991) estimates that the real rate of Grade 1 repetition in Honduras varied between 50.1% and 57.6% during the period 1978 - 1983 as compared to between 26.1% and 28.0% as shown in official statistics.

While under-reporting of repetition is most well-known (because most studied) in Latin America, the problem is common in other regions as well. For example, a recent exploratory study suggests that Grade 1 repetition is quite high in China. Data on repeaters are not collected in China, and the general presumption has been that repetition is fairly low since the level of failure at the end of each school year typically ranges from seven to ten percent. However, estimates derived by comparing time-series data available on enrollment by age and grade with data on the population of primary school age, suggest that the repetition rate in Grade 1 was around 36% in 1986 and 39% in 1987. The study indicates that a temporary dropout during the school year is a major factor in
explaining the difference between failure and repetition rates, see Schiefelbein (1990).

As regards the effect on the flow rates of this type of error, we note that as long as total enrollment, as well as the number of repeaters in Grade 2 are correctly reported, the number of promotees from Grade 1 to Grade 2 will not be affected. However, if the enrollment in Grade 1 is correctly reported and the number of promotees from Grade 1 is correctly estimated, to register pupils who are actually Grade 1 repeaters as new entrants means: (i) under-estimating the number of repeaters in this grade; (ii) over-estimation of the number of dropouts from this grade; and (iii) over-estimation of the number of new entrants.

The above discussion referred to Grade 1. If there is a general tendency for repeaters to be registered as newcomers (i.e. promotees) in other grades also, this will result in a general over-estimation of the number of promotees and under-estimation of the number of repeaters for these grades. Whether the number of dropouts will be over- or under-estimated will depend on the relative magnitude of the error in the repetition figures for subsequent grades. For the final grade of the cycle, under-estimation of repetition will result in over-estimation of dropout if we assume that the number of graduates is correctly reported.

The magnitude of the errors in the intake, promotion, repetition, and dropout rates for Grade 1 caused by the two types of errors in the data for Grade 1 discussed above (i.e., over-reporting of enrollment and under-reporting of repetition) are considerable. As these errors are present in data for some populous countries (e.g., China, Bangladesh, Brazil, India), they affect in a significant manner world-wide estimates of non-enrolled children, repetition and dropout. (See the discussion in connection with Table 4.1.)
C. Yearly Variation in the Coverage of Data

In some developing countries, the statistics published do not always cover all the schools since some schools do not supply the required information. While some countries make efforts to estimate data for the non-responding units based, for example, on past data, others publish the incomplete data. If the number of pupils enrolled in non-responding schools varies from one year to the next, this will lead to distortion in the estimated flow rates. Assume, for example, that the data available for year \( t \) are complete while those available for year \( t+1 \) are incomplete. Disregarding other types of errors, this implies that the number of promotees and repeaters will be under-estimated and the number of dropouts over-estimated. If, further, the data for school year \( t+2 \) are complete, this will imply that some of the promotees and repeaters that year were not included in the enrollment the previous year, leading to over-estimation of the promotion and repetition rates and under-estimation of the dropout rates, which may even be negative in some cases.\(^\text{21}^\)

The large annual fluctuations in flow rates observed for some countries are partly caused by yearly variations in the coverage of data. This is particularly the case for the dropout rates since they are determined residually. Note that while the types of errors discussed under points A and B above are mainly expected to affect the flow rates for the first grade of primary education, incomplete data will distort the rates for all grades.

D. Incorrect Definitions of the Flows

The pupil flow models most commonly used are based on the definitions of promotees, repeaters, graduates, and dropouts given in Section V. In cases where these flows are defined otherwise, this will usually affect the interpretation of the results derived from such models. This point may be

\(^{21}\) Negative dropout rates, arising from this and other types of errors in the available statistics, are quite common. Sometimes even promotion rates may exceed 100%. For example, in Zambia, the promotion rates exceeded 100% in four grades in the school year 1979/80 in which dropout rates were negative in the same four grades, see Fredriksen (1984), p. 57.
illustrated by the flow diagram of Figure 6.2, which shows what may happen to the pupils entering Grade g in a given school year in a cycle of education where attendance is not strictly compulsory. Some pupils may drop out during the school year (the intra-year dropouts), and the rest complete the school year (the completers), with or without passing the final exam or assessment. The intra-year dropouts may either drop out permanently, or repeat the grade the following school year. Some of the completers pass the requirements (the passers), and some fail to do so (the failures). Among the passers, some proceed to the next grade, some repeat the same grade the following school year and some leave the school system. The failures may either drop out or repeat the grade the following year.

All the flows included in Figure 6.2 are compatible with the more aggregated flows of promotees, repeaters, dropouts, and graduates defined in Section V.28/ However, as shown in the figure, each of these flows comprises several groups of pupils. For example, in cases where intra-year dropout does take place, the total number of repeaters recorded in a given year may be split into three different groups, i.e., repeaters among intra-year dropouts, repeaters among failures, and repeaters among the pupils who passed the grade examination. In some countries, lack of capacity in the subsequent grade is one important reason for passers to repeat. Hence, "repeaters" as defined in flow analysis are not synonymous with "failures" since some of the latter may drop out while some of the intra-year dropouts and some of the "passers" may join the remaining failures in repeating. Likewise, not all those who, at the end of the school year, are granted permission to go on to the next grade will actually be found in this grade the following school year. Some may, for various reasons, drop out or repeat.

28/ Note, however, that while flow models normally assume that all dropouts take place at the end of the school year, Figure 6.2 allows for dropout during the school year.
Figure 6.2: What Happens to Pupils Entering Grade $g$ of Primary Education?

- Initial enrollment in Grade $g$
  - Intra-year dropouts
    - Permanent intra-year dropouts
    - Intra-year dropouts repeating Grade $g$ next school year
  - Failures dropping out
    - Failures repeating Grade $g$ next school year
  - Passers leaving the primary school system
    - Passers repeating Grade $g$ next school year
    - Passers proceeding to next grade

* If Grade $g$ is the last grade of a cycle, these pupils are graduates leaving primary education. Otherwise, they are dropouts.

Figure 6.2 may now be used to illustrate two common types of errors in the flow rates caused by erroneous definitions of the basic flows. Firstly, the data on enrollment and repeaters by grade, from which the flow rates are derived, should refer to the beginning of the school year. The common practice in developing countries is that the pupil count refers to data a month or so after the official start of the school year. In this way, one avoids inclusion in the enrollment figure of pupils who attend school for only a very brief period. There are, nevertheless, considerable differences between countries as regards the date to which the enrollment figures refer. There may even be significant differences between different regions (states, provinces) within a given country. One reason for such differences is that the school year, due, for example to climate, may start at different dates in different regions.  

It is obvious that, in countries where school attendance is not strictly compulsory and where some pupils drop out during the school year, both enrollment and dropout figures will be affected by the date at which the enrollment count takes place. The later in the school year the pupils are counted, the lower will be the estimated dropout since the pupils who leave school prior to the date of the count will not be included in the enrollment figure.  

As a second example of the use of incorrect definitions which may affect the estimates of the flow rates, we mention that some countries report as repeaters pupils who fail the examination at the end of a given school year.  

29/ For example, in the Sudan, the data on enrollment refer to a date close to the beginning of the school year in the Northern Regions while they refer to a date close to the end of the school year in the Southern Region. The reason for this difference is that, while the data are collected at approximately the same date in all regions, the school year starts earlier in the South than in the North, see Fredriksen (1982).  

30/ As examples of countries for which the enrollment figures refer to the end of the school year, we mention some of the former English dependencies in the Caribbean, see Caribbean Community Secretariat (1978), p. 56; and Critchlow (1979), p. 41.
year. As shown in Figure 6.2, not all "failures" return to repeat at the beginning of the following school year, and not all those registered as repeaters are failures, since some intra-year dropouts as well as some passers may also repeat. Obviously, in cases where there is a significant difference between the number of failures at the end of a given school year and the number of repeaters at the beginning of the following school year, it is important to ensure that the figures used in deriving the repetition rates refer to repeaters as defined in the context of pupil flow models. For Bank staff designing projects, it is even more important to understand why such differences occur so that factors causing intra-year dropout leading to repetition or causing repetition among passers (e.g., because of incomplete schools in rural areas) can be addressed in the project.

VII. Cohort Flow Analysis

We have discussed above how to estimate promotion, repetition and dropout rates from education statistics commonly available in developing countries. We shall now see how we may use these rates to estimate the flow of a cohort of pupils through a cycle of education.

The most reliable way of studying the progress of a group of pupils through a cycle of education is on the basis of an individualized data system where each pupil is given a code number and can be followed throughout his/her school career. This method is often referred to in literature as the true

\footnote{For example, Cuadra (1989) reports that in Honduras, many students who repeat were not recorded as repeaters because they had not failed the end of the year examination. More generally, those who complete the school questionnaires often do not understand the meaning of the term "repeater." For example, this was the case for 20% of the headteachers in Somalia, see Chapman et. al. (1990).}

\footnote{To illustrate that all three groups of repeaters may be important in some countries, we mention that of the 131,914 repeaters in primary education in El Salvador in 1970, it has been estimated that 28% were passers who repeated, 46% were failures who repeated, and 26% were intra-year dropouts who repeated, see AID (1977), p. 71-72. The study referred to above for China (see Schiefelbein 1990) suggests that also in that country, failure at exams is not the main cause of repetition.}
To establish and operate an individualized data system is, however, very costly and only a few industrialized countries have introduced it so far.

A more inexpensive approach would be to introduce a "cohort coding system" whereby all pupils in a cohort experiencing the same educational events receive the same code number. The data thus collected year by year would permit the analysis of some aspects of the flow of this cohort through a cycle of education. However, even this approach is quite complex and has not yet been successfully implemented in any country. Both of these methods are further discussed in another background paper prepared for this seminar, see Cuadra (1991).

For want of data collected by means of the above two approaches, educational statisticians and planners have developed various simplified methods which, under certain hypotheses, permit us to estimate the progression of a cohort through a cycle of education. The Reconstructed Cohort Method is the one most frequently used when data are available allowing the calculation of promotion, repetition, and dropout rates. The method implies using the most commonly used student flow model—the Grade Transition Model—to project the flow of a given cohort of children through a cycle of education. This is the method used to derive indicators such as "percentage of cohort completing primary education" or "number of pupil-years needed to produce one graduate." Both of these indicators are frequently quoted in sector and appraisal reports.

In cases where data on repeaters are not available, a more approximate method, generally referred to as the Apparent Cohort Method, may be used. While the Reconstructed Cohort Method purports to describe in terms of promotion, repetition, dropout, and graduation the way in which a cohort of pupils progresses through a cycle of education, the ambition of the Apparent Cohort Method is limited to estimating the proportion of the cohort which reaches successive grades. The difference between the two approaches will be explained further in Section VIII.

33/ By the term "cohort" we shall mean, in this report, "a group of children who start Grade 1 of primary education in a given year."
A. **Cohort Reconstruction**

We shall use the flow rates presented in Table 7.2 to illustrate how we may employ the Reconstructed Cohort Method to estimate the flow through primary education in Burkina Faso of the cohort enrolled in Grade 1 in school year 1986-87. The rates have been computed on the basis of official statistics using the method presented in Section V. These statistics are shown in Table 7.1. Let us start by examining the flow between the two school years 1986 and 1987. The enrollment in a specific grade in a specific year may be written inside a rectangle. For example, in Burkina Faso in 1986, the total enrollment (boys and girls) in Grade 1 of primary schools was:

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>88,345</td>
</tr>
</tbody>
</table>

### Table 7.1: Enrollment and Repeaters by Grade in Primary Education

**Burkina Faso, Both Sexes**

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>88,345</td>
<td>80,347</td>
<td>74,493</td>
<td>51,877</td>
<td>44,186</td>
<td>51,166</td>
</tr>
<tr>
<td>of which repeaters from 1986</td>
<td>10,782</td>
<td>9,768</td>
<td>11,159</td>
<td>8,636</td>
<td>7,138</td>
<td>20,514</td>
</tr>
<tr>
<td>1987</td>
<td>83,681</td>
<td>85,747</td>
<td>76,527</td>
<td>65,582</td>
<td>45,335</td>
<td>55,035</td>
</tr>
</tbody>
</table>

Source: UNESCO Office of Statistics
Table 7.2: Rates of Promotion, Repetition, and Dropout by Grade in Primary Education for 1986-87, Burkina Faso. Both Sexes

<table>
<thead>
<tr>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promotion rate</td>
<td>0.860</td>
<td>0.813</td>
<td>0.764</td>
<td>0.737</td>
<td>0.781</td>
</tr>
<tr>
<td>Repetition rate</td>
<td>0.122</td>
<td>0.122</td>
<td>0.150</td>
<td>0.166</td>
<td>0.162</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>0.018</td>
<td>0.065</td>
<td>0.086</td>
<td>0.097</td>
<td>0.057</td>
</tr>
</tbody>
</table>

* As no data were available on graduation, this rate includes all pupils not repeating Grade 6, including those who drop out.

We now ask: where were the 88,345 pupils the following school year, i.e., in 1987? According to the assumptions of the Grade Transition Model, the pupils may have either been promoted to Grade 2, repeated Grade 1, or dropped out. To illustrate these flows, we shall adopt the following notation:

![Diagram](https://via.placeholder.com/150)

Table 7.1 shows that of the 88,345 pupils enrolled in Grade 1 in 1986, 10,782 repeated this grade in 1987. Based on this table, one also finds that 85,747 - 9,768 = 75,979 pupils were promoted to Grade 2 and that 88,345 - 10,782 - 75,979 = 1,584 dropped out. The flows out of Grade 1 may be written as:

![Figure 7.1: Flows Out of Grade 1 (Absolute Numbers)](https://via.placeholder.com/150)
Thus in 1987, we find that $75,979 + 10,782 = 86,761$ of the 88,345 pupils are still enrolled in school, 10,782 are enrolled in Grade 1 (repeaters) and 75,979 in Grade 2 (promotees).

Next, where would the 86,761 pupils enrolled in Grades 1 and 2 in 1987 be in 1988? This may be computed, if the promotion, repetition, and dropout rates for Grades 1 and 2 in 1987 are known. Before computing these flows we note that it is easier to work with relative than with absolute numbers. For example, instead of starting with 88,345 pupils in Grade 1 and saying that 10,782 repeated, 75,979 were promoted and 1,584 dropped out, one may as well base the calculations on the promotion, repetition, and dropout rates for Grade 1. In this example, these three rates are given in Table 7.1 as 0.860, 0.122, and 0.018, respectively. This means that out of every 1,000 pupils, 860 would be promoted, 122 would repeat, and 18 would drop out. In addition to simplifying the calculations when we work with large numbers, this approach also facilitates the interpretation of the results since every figure computed may be read as "per thousand." Thus, following this approach, Figure 7.1 may be expressed as:

**Figure 7.2: Flows Out of Grade 1 (Relative Numbers)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>1,000</td>
</tr>
</tbody>
</table>

We are now ready to illustrate the flow through primary education of this cohort of 1,000 pupils. To simplify the example, we shall assume that the rates shown in Table 7.2 will remain constant over the whole period members of this cohort are enrolled. Of course, in practice flow rates do change over time and these changes should be taken into account. The technique is, however, the same whether the rates shown remain stable or change. In addition, when
preparing a sector report, it is interesting to ask what the implications would be of maintaining constant the flow rates observed for the latest year for which data are available. The results, thus derived, give a good indication of whether or not efforts should be made to change these rates in the future.

Figure 7.3 shows the extension of Figure 7.2 to cover all grades and years pupils from the 1986/87 cohort are enrolled in school. The procedure followed in drawing up this diagram is easily understood. Of 1,000 pupils enrolled in Grade 1 in 1986, we have already shown that, based on the flow rates of Table 7.2, 122 repeated this grade in 1987, 860 were promoted to Grade 2, and 18 dropped out. These figures are entered in the diagram against their respective arrows. By applying the promotion, repetition, and dropout rates of Grade 2, we find further that of the 860 pupils enrolled in this grade in 1987 (i.e., the pupils promoted to Grade 1 in 1986), 860 x 0.813 = 699 were promoted to Grade 3 in 1988 while 860 x 0.122 = 105 repeated Grade 2 in this latter year. In order to obtain the figure 210 in the rectangular box (a figure showing how many of the original 1,000 pupils would be in Grade 2 in 1988 after having repeated once either Grade 1 or Grade 2), we must add to these 105 pupils those who are promoted to Grade 2 after having repeated once Grade 1, i.e., 122 x 0.860 = 105. (Naturally, it is a coincidence that both figures equal 105.)

All the other figures in the flow diagram may be obtained in the same way employing for each year and grade the promotion, repetition, and dropout rates given in Table 7.2. The assumptions on which this reconstruction is based may be summarized as follows:
Figure 7.3: Hypothetical Flow of the 1986 Cohort through
Primary Education in Burkina Faso

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>1000</td>
<td>880</td>
<td>66</td>
<td>18</td>
<td>109</td>
<td>12</td>
</tr>
<tr>
<td>1987</td>
<td>122</td>
<td>108</td>
<td>106</td>
<td>12</td>
<td>110</td>
<td>109</td>
</tr>
<tr>
<td>1988</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>106</td>
<td>34</td>
</tr>
<tr>
<td>1989</td>
<td>2</td>
<td>39</td>
<td>22</td>
<td>275</td>
<td>554</td>
<td>22</td>
</tr>
<tr>
<td>1990</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>72</td>
<td>99</td>
<td>22</td>
</tr>
<tr>
<td>1991</td>
<td>1</td>
<td>17</td>
<td>13</td>
<td>109</td>
<td>554</td>
<td>109</td>
</tr>
<tr>
<td>1992</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>78</td>
<td>367</td>
<td>205</td>
</tr>
<tr>
<td>1993</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>97</td>
<td>236</td>
<td>141</td>
</tr>
<tr>
<td>1994</td>
<td>2</td>
<td>13</td>
<td>10</td>
<td>62</td>
<td>85</td>
<td>65</td>
</tr>
<tr>
<td>1995</td>
<td>0</td>
<td>2</td>
<td>26</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>1996</td>
<td>27</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>1997</td>
<td>11</td>
<td>7</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1998</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1999</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Cohort Members Still Enrolled: 1000

Output (No of Pupils):
- 1940

Waves Input
- Per completing pupil: $2.3$

Input/Output: $1.8$

<table>
<thead>
<tr>
<th>Grades</th>
<th>Prom</th>
<th>Rep</th>
<th>Drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86.0</td>
<td>12.2</td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>61.3</td>
<td>12.2</td>
<td>6.6</td>
</tr>
<tr>
<td>3</td>
<td>76.4</td>
<td>16.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>75.7</td>
<td>16.8</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>76.1</td>
<td>16.2</td>
<td>0.7</td>
</tr>
<tr>
<td>6</td>
<td>68.9</td>
<td>40.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>
(a) We want to estimate the implications for the progression through primary education of the pupils enrolled in Grade 1 in Burkina Faso in 1986 of maintaining constant the promotion, repetition and dropout rates observed that year. As we are thus reconstructing the flows of one given cohort, there are no new entrants to this cycle after 1986. Nor does any of the other flows listed under points (a) to (e) in Section V.B. take place (e.g., re-entrants, new entrants to grades other than Grade 1, migration, and transfers). For the sake of analytical convenience and clarity, we have based the calculation on a hypothetical cohort of 1,000 new entrants instead of on the number of pupils actually enrolled in Grade 1 in that year (i.e., 88,345 pupils). As explained above, this does not affect the interpretation of the results.

(b) The promotion, repetition, and dropout rates are assumed to remain constant over the entire period members of this cohort are enrolled in school and they are those calculated for the starting year, i.e., 1986, and

(c) The reconstruction assumes that the same promotion, repetition, and dropout rates apply to all pupils enrolled in a given grade, regardless of whether or not they have previously repeated this or preceding grades once or more. This assumption, usually referred to as the hypothesis of homogeneous behavior, will be discussed further under point C below. One implication of this assumption is that, as shown in Figure 7.3, in the case of Burkina Faso, a pupil may repeat as many as nine times during primary education. This is seen from the number of boxes for Grade 6 below the top diagonal of the diagram which includes pupils who have not repeated. This is, of course, a theoretical maximum and only a few repeat more than four times. However, as will be discussed later, this does raise questions as regards the validity of some of the indicators derived by means of this method.
In the above example, data were not available on the number of pupils who actually graduated successfully from primary education at the end of 1986. We have therefore assumed that there is no dropout from Grade 6. Under this assumption, 674 of the original 1,000 students will eventually successfully complete Grades 1-6. Of these 674 students, only 184 are estimated to graduate without repetition (in 1991), 208 will graduate in 1992 after repeating once, 141 in 1993 after repeating twice, 77 in 1994 after repeating three times, and so on until the last student graduates in year 2000 after having spent 15 years in this six-year cycle, i.e., after having repeated nine times. This large number of repetition results mainly from the very high level of repetition in Grade 6. Prior to that grade, only 26 students, i.e., 2.6% repeated for the fourth time. (This figure is found by adding the figures in the fifth diagonal of boxes in Figure 7.3. All pupils in the first diagonal of boxes were promoted without repetition, all in the second diagonal of boxes have repeated once, and so on.) Nevertheless, as already indicated, the fact that the method implies that pupils can repeat almost without limit is one of the weaknesses of the hypothesis of homogeneous behavior referred to above.

Finally, the row of boxes at the foot of the flow diagrams, and the column of boxes and the table in the right hand corner will be explained below.

B. Cohort Flow Indicators

Apart from providing a good visual illustration of how a cohort of students progresses through a cycle of education, the main purpose of constructing the flow chart in Figure 7.3 is that this permits us to derive a number of indicators describing this progression. We may distinguish between four groups of such cohort flow indicators:

(a) Indicators of retention (or survival) by grade;
(b) Indicators of retention by years spent in the cycle;
(c) Indicators of duration of study;
(d) Indicators or wastage in education.

We shall briefly explain how these indicators may be derived from Figure 7.3.
(a) **Indicators of retention by grade**: Three such indicators may be derived from Figure 7.3.

1. **Survival by grade**: This indicator is given by the figures included in the row of boxes shown at the foot of the flow diagram. For example, the figure included in the box for Grade 2 shows that of the 1,000 pupils starting Grade 1 in 1986, 980 will *eventually* reach Grade 2. This figure is found by adding the number of pupils promoted from Grade 1 to Grade 2 in each of the years pupils from this cohort remain enrolled in the cycle, i.e., $860 + 105 + 13 + 2 = 980$.

   This indicator illustrates a very important point. While the flow rates as shown in Table 7.2 describe the promotion, repetition, and dropout *between two subsequent school years*, the cohort flow indicators are not related to any particular school year. They show what *eventually* happens to the cohort after all those who repeat have finally either been promoted or dropped out. For example, we have calculated that 860 pupils (or 86% of the cohort) are promoted from Grade 1 to Grade 2 between the two school years 1986 and 1987. However, when it is taken into account that some of the 122 pupils who repeat will be promoted in later years, we find that 980 or 98% of the cohort eventually reach Grade 2. The difference between, respectively, the promotion and cohort survival rates is particularly large for Grade 6, i.e., 59.9% as compared to 100%.

   This indicator of survival by grade is one of the two cohort flow indicators most commonly quoted in sector and appraisal reports. (The other is the input/output ratio or, expressed in a different manner, the average number of school years spent by graduate, see below.) It is a very useful indicator because it gives a summary description of the impact of the repetition and dropout rates on the capacity of a school system to retain pupils enrolled until the end of the cycle. It can also be shown that this is a robust indicator in the sense that it is not affected by the assumption that pupils' previous school history does not affect their propensity to repeat, be promoted or drop out (see point C below).
2. **Dropout by grade:** This indicator can also be derived from the row of boxes shown at the foot of the flow diagram. The number of dropouts between two subsequent grades is found either by adding the dropouts shown in the flow diagram or, more easily, as the difference between the number of pupils surviving to each of these two grades. For example, the number of dropouts between Grades 3 and 4 is computed as $60 + 24 + 6 + 1 = 91$ or as $909 - 818 = 91$. Similarly, the total number of dropouts prior to, e.g., Grade 4 is computed either as $18 + 2 + 56 + 13 + 2 + 60 + 24 + 6 + 1 = 182$ or as $1,000 - 818 = 182$.

3. **Number of graduates from the final grade:** The flow diagram shows that, under the assumption that there is no dropout from Grade 6 (see explanation above), 674 of the 1,000 pupils will eventually graduate from the cycle.

(b) **Indicators of retention by years spent in the cycle:**

4. **Enrollment by years of study:** This indicator of survival by years of study is found by adding the figures in each of the boxes of the flow diagram for the year in question. For example, for 1988, we find that $15 + 210 + 699 - 924$ pupils of this cohort are still enrolled. This indicator is summarized in the column of boxes at the right hand side of the flow diagram. The last year in which members of this cohort would be enrolled is year 2,000 (1 pupil in Grade 6).

5. **Dropouts by years of study:** The number of dropouts by years of study is found either by adding the dropouts from each grade for the relevant year or, if Indicator 4 has already been calculated, as the difference between the enrollment in two subsequent years. For example, in 1988 we find that $13 + 60 = 73$ pupils dropped out. Alternatively, this figure could have been found as enrollment in 1988 minus enrollment in 1989, i.e., $924 - 851 = 73$. The last dropout from this hypothetical cohort would take place in 1999 when one pupil drops out from Grade 6 after having spent 14 years at school, i.e., he/she has repeated eight times.
6. **Graduates by years of study:** The diagram shows that 184 pupils would graduate in 1991 after six years of study (i.e., without repetition), 208 would graduate in 1992 (one repetition), 141 in 1993 (two repetitions), and so on.

(c) **Indicators of duration of study:**

7. **Average duration of study for graduates:** Figure 7.3 shows that 184 pupils would graduate after 6 years of study, 208 after 7 years, and so on. The total number of pupil-years\(^{24}\) spent at school by the 674 graduates is hence:

\[
(184 \times 6) + (208 \times 7) + (141 \times 8) + (77 \times 9) + (37 \times 10) + (16 \times 11) + (7 \times 12) + (2 \times 13) + (1 \times 14) + (1 \times 15) = 5,066 \text{ pupil-years}.
\]

Dividing this total by the number of graduates gives average study time for graduates:

\[
\frac{5,066}{674} = 7.5 \text{ years}
\]

8. **Average duration of study for dropouts:** Figure 7.3 shows that of those dropping out, 18 spent one year in school, 2 + 56 = 58 spent two years, 13 + 60 = 73 spent three years, and so on. Adding all these figures, we find that the total number of pupil-years used by dropouts equals:

\[
(18 \times 1) + (58 \times 2) + (73 \times 3) + (77 \times 4) + (57 \times 5) + (27 \times 6) + (10 \times 7) + (3 \times 8) + (2 \times 9) + (1 \times 10) = 1,230 \text{ pupil-years}.
\]

Dividing this by the total number of dropouts gives average study time for dropouts:

\[
\frac{1,230}{326} = 3.8 \text{ years}
\]

9. **Average duration of study for the cohort:** The average duration of study for all cohort members is found by dividing all pupil-years used by graduates and dropouts by the size of the cohort:

\[^{24}\text{One pupil spending one year in a grade is said to have used one pupil-year.}\]
Average length of study by cohort: \[
\frac{5,066 + 1,230}{1,000} = 6.3 \text{ years}
\]

(d) **Indicators of wastage**: We complete this brief survey of cohort flow indicators by presenting three indicators commonly used to estimate the amount of the resources used by a cohort that is spent by pupils repeating grades and by pupils dropping out. Resources used on repetition and dropout is commonly referred to as education wastage.

The Reconstructed Cohort Method defines wastage in education as the number of pupil-years spent on repetition and dropout. It is recognized that this concept of wastage constitutes a rather crude view as regards the benefits pupils may draw from the education received before dropping out or from spending an extra year in a grade. It is not the purpose of this introduction to the use of cohort flow indicators to fully discuss this aspect. Nevertheless, some factors that need to be considered when interpreting the indicators presented below need to be mentioned.

The extent to which dropout should be regarded as wastage will depend on factors such as the structure and objectives of the educational system studied, the educational attainment achieved by those dropping out, and from which level of education the dropout takes place. As regards primary education, the main objective, at least for the first few grades, is to achieve literacy. Those who leave school without having completed the cycle are not likely to have strengthened basic literacy and numeracy to the point where it becomes resistant to forgetting. To the extent such pupils relapse into illiteracy, one may argue that the resources they have consumed are, at best, inefficiently used and, at worst, completely wasted. This definition of "waste" implies that the resources used on dropouts have an alternative use which would somehow be more "efficient"; they could, for example, have been spent on other pupils or could simply represent savings which could be used for meeting other needs of society.

For pupils who drop out from grades towards the end of the primary cycle after having spent, say, five to seven years at school, the argument of complete waste implied by the Reconstructed Cohort Method seems even more
questionable. On the one hand, these dropouts use a high number of pupil-years, thus contributing significantly to the level of wastage. On the other hand, to the extent they have acquired some of the skills the system sets out to teach them, these dropouts also represent some output.

Practically all countries aim at providing all children with some basic schooling. Thus, even if one could devise some sort of test for distinguishing the new entrants who would drop out from those who would complete the cycle, to apply this type of selection for admission into primary education would generally not be acceptable. It is, therefore, useful to stress again that the main purpose of the wastage indicators is to throw light on the effects of repetition and dropout on pupil flows. In the case of "high" levels of wastage, this information is an indication of a country needing to study the factors causing this waste in order to better design interventions to address them. Similarly, for Bank Staff, high wastage indicates a need for the Bank, through its policy advice and lending operations, to help countries design and implement measures to better meet their learning objectives and, thereby, using their education resources more efficiently.

For secondary education, the relationship between "dropout" and "waste" is even more complex than for primary education. On the one hand, the fact of not completing a stage of secondary education is all the more striking since entry to this level in most developing countries is based on selection. On the other hand, one may argue that pupils will benefit in various ways from a few years of secondary schooling regardless of whether or not they complete the cycle successfully.

The extent to which repetition should be regarded as wastage is also a moot point. The supporters of repetition usually claim that it serves two major purposes: to remedy inadequate achievement and to aid pupils who are judged to be emotionally immature. Conversely, to promote pupils automatically is claimed to lower academic standards, to destroy pupils' incentive to learn and teachers' motivation to teach, and to create pedagogical problems in the class by increasing the ability range within each grade. Thus, according to this view, repetition does not represent a waste of resources. On the contrary,
it is considered as a necessary and appropriate investment to assist pupils with academic or adjustment difficulties.

The proponents of automatic promotion reject the pedagogical assumptions on which a repetition policy is based. Firstly, supported by a number of studies which conclude that schooling has little independent impact on achievement, they claim that out-of-school variables such as family background and personal pupil characteristics are equally, or possibly, more important in determining pupils' progress in school. Thus, to provide a slow learner with another year of the same program is not necessarily the best way of improving this pupil's achievement. Secondly, they contest the validity of the tests generally used to determine who should be promoted and who should repeat. Thirdly, they claim that it is far from obvious that repetition of the same grade will improve the achievement of slow learners more than being promoted to the next grade. Lastly, to stigmatize pupils as "failures" may have a negative effect on their incentive to learn, as well as on their self-respect.

In short, the views on the value of repetition held by the proponents of automatic promotion correspond quite closely to that implied by the Reconstructed Cohort Method, i.e., repetition is largely a waste of resources that could alternatively be used, e.g., to provide access to pupils not enrolled, or to improve the quality of instruction for those enrolled.

As should be clear from the above discussion, the factors leading to failure and consequent repetition are complex. The causes differ among countries and across grades within each country, and it would be wrong to believe that their negative effects on achievement could be eliminated simply by one administrative stroke of the pen whereby repetition was abolished in
favor of automatic promotion. The real issue is not to design measures that merely ensure no repetition, but measures that address the factors that cause repetition. The fact that some of the few developing countries which had introduced automatic promotion during the last two decades have reintroduced the possibility of repetition (e.g., Costa Rica, Egypt, Singapore and Venezuela) might possibly be taken as an indication that automatic promotion alone is not sufficient.

We shall not pursue this discussion further. It is sufficient for our purpose to reiterate that the sole purpose of computing the type of indicators of wastage used by the Reconstructed Cohort Method is to throw light on the effect of repetition and dropout on pupil flows. "High" levels of wastage simply indicate a need to study the factors (internal and external to the educational system) which make pupils leave school prematurely or repeat grades. But, as already stressed, "low" wastage as measured by this method does not necessarily indicate that wastage is low if output was measured otherwise, e.g., in terms of pupils' achievement. It simply indicates that there is little repetition and dropout in the system.

To summarize, whatever opinion one may have regarding the benefits derived from attending school less than the complete cycle by pupils dropping out, or regarding the value of repetition, it is still important for educational planners and policymakers to know the magnitude and pattern by grade of these two events. In spite of their crudeness, the indicators presented below do provide interesting information in this regard.

---

35/ For example, a major cause for Grade 1 repetition in many countries is simply that the pupils insufficiently know the language of instruction and therefore need more than one year to learn both the language and the content of the curriculum. In other countries, Grade 1 curriculum is just too demanding (often more demanding than in industrialized countries). Repetition in Grade 6 is, on the other hand, often caused by insufficient capacity in secondary education.

36/ For a further discussion of the effects of repetition, readers are referred to Haddad (1979). The impact of repetition on costs is discussed in a separate document prepared for this seminar series, see Cuadra and Fredriksen (1991).
10. **Input-output ratio**: This is an indicator of the "efficiency" with which a school system produces a given number of graduates. Efficiency is a concept which refers to the relationship between the **inputs** into a system (be the system, e.g., agricultural, industrial, or educational), and the **outputs** from that system (be they wheat, vehicles, or educated individuals). An activity is said to be "efficient" if maximum output is being obtained from given inputs, or if a given output is being obtained with the minimum possible inputs. Inputs and outputs have somehow to be valued so that they may be aggregated; and usually prices are used to perform this valuing function. The problems of measuring efficiency in education, however, are considerable. They stem mainly from difficulties in measuring educational output, as well as from quantifying the relationship between inputs and outputs. How educational output is to be measured depends, of course, on the nature of the objectives of the educational system. Depending on the analytical and philosophical viewpoint adopted, the objectives may differ considerably.

Efficiency indicators derived by means of the Reconstructed Cohort Method do not have the ambition of quantifying the efficiency by which a given learning output is imparted. As for all the other indicators derived on the basis of this method, the "efficiency" indicators only purport to give another facet of the impact of repetition and dropout on the flow of a cohort of pupils through a cycle of education. To this end, the method defines "output" and "input" in a very simplistic manner. The output of a given cycle of education is simply considered to be the number of pupils who successfully complete this cycle (the graduates). This is naturally a rather restricted definition since the dropouts, no doubt, have acquired some of the skills which the system has set out to teach them. Furthermore, in addition to the number of graduates, their achievement level should be taken directly into account. Nevertheless, this way of measuring output still gives us some useful insights into the functioning of an educational system.

Educational **inputs** comprise buildings, teachers, textbooks, etc., which may all be aggregated financially in terms of expenditures per pupil-year. The inputs also include the **income foregone** by students while studying. However, in the context of the Reconstructed Cohort Method, the input indicator
most commonly employed is the number of pupil-years used by the cohort. Naturally, the amount of input expressed in monetary terms is related to the number of pupil-years used.

The table in the upper right hand corner of Figure 7.3 shows the number of pupil-years spent in each grade by the cohort in the above example. The total for a given grade is obtained by adding the figures inside the boxes of the diagram referring to that grade. For instance, a total of 1,139 pupil-years was spent in Grade 1, i.e., 1,000 in 1986, 122 in 1987 by pupils repeating once, 15 in 1988 by pupils repeating twice, and two pupil-years in 1989 by pupils repeating for the third time. Considering all six grades, the total number of pupil-years spent by this cohort is 6,296. This represents the total input. The "output," i.e., the number of successful completers of this six-year cycle is 674. The average number of pupil-years per successful completer was therefore:

\[
\frac{6,296}{674} = 9.34 \text{ pupil-years}
\]

Naturally, had there been no repetition or dropout, the average number of pupil-years would have been six. Thus, one often finds in sector and appraisal reports statements such as: "The internal efficiency of the system is low; it requires, on average, x number of pupil-years to complete the six-year cycle," (in this case, 9.34 years, or some 56% more than in the "optimal" situation).

It should be noted that this indicator is very sensitive to the exam results in the final grade of the cycle. In this example, we have assumed that there is no dropout from that grade. In many countries, the completion or graduation rates for primary and secondary education are very low and often affected by the intake capacity in the first grade of the subsequent cycle. This is one reason for the often very high average number of pupil-years quoted in some reports. To avoid the often arbitrary impact of capacity limitations in the subsequent cycle on completion rates, it is often advisable to use the number of pupils reaching the final grade of the cycle as a proxy for the number of completers rather than the actual number of graduates.
Another way of expressing the above indicator of efficiency is the following. In the above example, the cohort actually used 6,296 pupil-years. Had there been no repetition, the 674 graduates would have needed 674 x 6 = 4,044 pupil-years to complete the cycle. By dividing the number of years actually spent by this cohort by this optimal number, one obtains an indicator of efficiency generally referred to as the input-output ratio, which in this case is: \[ \frac{6,296}{4,044} = 1.56 \]

In a "perfectly efficient" system, i.e., one in which there is no repetition or dropout, this ratio would equal 1.00. Thus, this cohort used 1.56 times the input required in a system without repetition or dropout.

Another way of expressing this indicator is in terms of "the percentage of the total number of pupil-years wasted." In the above case, 6,296 - 4,044 = 2,252 pupil-years wasted, i.e., 35.8% of the total.

11. Proportion of total wastage spent on repetition and dropout, respectively: This indicator splits total wastage on the two events causing waste, i.e., repetition and dropout. In the above example, a total number of 2,252 pupil-years was spent on repetition and dropout. Using the figures for dropout by grade given in the row of boxes at the foot of Figure 7.3, we may calculate the number of pupil-years wasted on dropout (note that this calculation excludes wastage due to repetition among pupils who later drop out): \[ (20 \times 1) + (71 \times 2) + (91 \times 3) + (95 \times 4) + (49 \times 5) = 1,060 \text{ pupil-years} \]. Thus, \[ \frac{1,060}{2,252} \times 100\% = 47\% \text{ of the pupil-years wasted by this cohort was due to dropout.} \]

Consequently, 53% of the total wastage was due to repetition. Some of this wastage was due to repetition by pupils who eventually graduated (see below), while some was due to repetition by pupils who later dropped out.

12. Proportion of total wastage used by graduates and dropouts, respectively: Within the framework of this method, wastage by graduates can only be due to repetition. The number of pupil-years spent on repetition by graduates may be derived directly from Figure 7.3 which shows that 208 pupils graduated after having repeated once, 141 after having repeated twice, 77 after
having repeated three times, and so on. Thus, the total number of pupil-years spent on repetition by graduates equals:

\[ (208 \times 1) + (141 \times 2) + (77 \times 3) + (37 \times 4) + (16 \times 5) + (7 \times 6) + (2 \times 8) + (1 \times 9) + (1 \times 10) = 1,026 \text{ pupil-years}. \]

This means that \( \frac{1,026}{2,252} \times 100\% = 46\% \) of all wastage was accounted for by graduates repeating grades. The remaining part, i.e., 54% was accounted for by dropouts. This figure is higher than the one calculated above (47%) because it includes pupil-years spent on repetition by pupils who dropped out.

C. **Further Comments on the Assumptions Behind the Reconstructed Cohort Method.**

How well the cohort flow indicators presented above describe the way in which a cohort actually progresses through a cycle of education depends on the realism of the assumptions on which the Grade Transition Method is based, and on the quality of the statistical data available for estimating the flow rates. Possible effects of some common errors in data have already been highlighted in Section VI. In this section, we shall say a few words about how cohort flow indicators may be affected by violations of the main assumption behind the Grade Transition Model, i.e., the hypothesis of homogeneous behavior of all pupils enrolled in a given grade.

The validity of the hypothesis of homogeneous behavior has been questioned in the literature on the use of flow models for cohort reconstruction. Some authors have pointed out that pupils' behavior with respect to promotion, repetition, and dropout is likely to be affected by previous repetitions and that models disregarding this fact may, therefore, give unsatisfactory results, see for example, Thonstad (1980) and OECD (1967). This would imply that, for example, the flow rates depend not only on the grade in which pupils are enrolled now, but also on the grade in which they were enrolled in the previous year. Even the pupils earlier school history may influence their transition rates. For example, pupils who have already repeated Grade 1 and Grade 2 may have a different propensity to repeat Grade 5 than pupils who have not repeated at all so far, or who have repeated Grade 3 only.
One aspect of the hypothesis of homogeneous behavior which has been of particular concern to users of the Grade Transition Model for cohort reconstruction is the fact that this assumption implies that pupils may repeat grades without any limits. This is obviously not realistic. Some authors applying the flow models to industrialized countries have pointed out that since the repetition rates are normally low, the error made is probably very small. Many developing countries, however, show quite high levels of repetition and for these countries, the method clearly implies an unrealistically high number of repetition in each grade. For example, a repetition rate of 20% means that a fraction 0.04 repeats twice, a fraction 0.008 repeats three times, etc.

The discrepancy between reality and the assumption of unlimited repetition is particularly unsatisfactory in cases where countries are known to apply clearly defined limits to the number of repetitions permitted. This has led some to question the validity of indicators derived on the basis of the Grade Transition Model under such circumstances.

It is not possible here to enter into a detailed and technical discussion of this question. This would imply constructing models which take restrictions on repetition explicitly into account and then comparing indicators derived from such models to the corresponding indicators derived on the basis of the Grade Transition Model. However, as this is an important issue, we shall summarize the main conclusions of a study conducted to throw light on the effect of the most common types of regulations of grade repetition found throughout the world. The study aims at answering the following question:27/

"Consider a cycle of education for which cohort flow indicators have been derived by means of the Grade Transition Model. To what extent will these estimates describe correctly the flows which actually take place if the data on which they are based have been generated, not by a system where there is no restriction on repetition, but by one where such restrictions both exist and are applied in practice? In other words, how would indicators, derived on the basis of the Grade Transition Model, differ from the corresponding indicators derived by means of the "true model," i.e., the model which takes explicitly into account existing regulations governing repetition?"

The results presented in the study show that while some of the 12 indicators listed in the previous section are rather insensitive to restriction on repetition, others may be affected quite significantly. The former group includes the indicators used for quantifying average study time, as well as the level of dropout, graduation, and wastage associated with the whole cohort, i.e., Indicators 1, 2, 3, 7, 10, and 11 above. For these six indicators, the Grade Transition Model may thus be expected to yield satisfactory results even for systems in which the number of repetitions permitted is limited.

As the above six indicators are those which are probably most commonly used for analyzing cohort flows and wastage, this conclusion also has a bearing on the data collection priorities in the field of education statistics. For example, it is likely that the returns to any investment aimed at improving the quality of this group of indicators would, in most countries, be higher if these resources were used to improve the accuracy of the statistics collected on total enrollment and repeaters by grade, rather than to collect more detailed flow statistics permitting the use of models which are more sophisticated than the Grade Transition Model.

The indicators which may be quite significantly affected by the rules governing repetition are those which refer to enrollment, dropout, and graduation in any specific year originating from a given cohort (e.g., Indicators 4, 5, and 6) as well as indicators referring to sub-groups of pupils belonging to this cohort (e.g., Indicators 8, 9, and 12). It follows from this that care should be taken when the Grade Transition Model is used to study, for example, yearly teacher requirement or yearly supply of educated labor associated with a given cohort.

To summarize, we may say that despite the fact that the Grade Transition Model assumes homogeneous behavior and thus disregards restriction on grade repetition, this model yields acceptable results for the indicators normally used in sector and appraisal reports even in cases where this assumption is incorrect. This conclusion lends rather strong support to continued use of this quite simple model for deriving this type of indicators. The explanation for this is that the method is based on the average flow rate
for each grade and describes correctly what happens, on average, to the cohort. However, as flow rates are not available for different groups of pupils (e.g., dropout, graduates) to the extent the rates for the various groups differ, the results derived for each of these groups (based on the average rates) will also differ.

VIII. The Apparent Cohort Method

The cohort reconstruction presented above was based on the Grade Transition Model. We recall that use of this model requires data on enrollment and repeaters by grade.

When data on enrollment, but not on repeaters, are available, the Grade Retention Model is sometimes used to derive indicators of survival by grade. This approach is normally referred to as the Apparent Cohort Method. It is based on comparing enrollment in successive grades in successive school years, the hypotheses being that the decline in enrollment between two successive grades in two subsequent school years represents dropouts. For example, Table 7.1 gives the enrollment by grade in primary education in Burkina Faso for the two school years 1986 and 1987. Applying the Apparent Cohort Method, we estimate the following survival rate between Grades 1 and 2:

\[
\text{Survival between Grades 1 and 2: } \frac{85,747}{88,345} = 0.971
\]

Thus, according to this method, 97.1% of the 1986 cohort would reach Grade 2, i.e., 2.9% would drop out prior to this grade. Based on data on enrollment in Grade 3 in 1988, one could estimate the survival of the 1986 cohort to that
grade, and so on, for the survival to subsequent grades if we have data for later school years.38/

Obviously, as the Apparent Cohort Method ignores repetition, the extent to which its estimates of cohort survival will correspond to the actual level of survival depends, *inter-alia*, on the level of the repetition rates. It can also be shown that in cases where repetition takes place, even estimates based on time-series data on enrollment are affected by the rate of growth in new entrants. If there is no repetition, then the Grade Transition Model will be identical to the Grade Retention Model and the two models will give the same estimates of cohort survival.

Further, it can be shown that the Apparent Cohort Method gives reasonably exact estimates of dropout between grades and survival if the repetition rates are of very similar magnitude, see Fredriksen (1983), Chapter 5. This explains why the dropout between Grades 1 and 2 as estimated by this method in the above case (2.9%) is fairly similar to the dropout rate of 2% estimated by the Reconstructed Cohort Method. However, the method will generally *over-estimate* dropout between two grades if the repetition rate in the first of the two grades is larger than in the second. The Apparent Cohort Method will generally *under-estimate* dropout between two grades in a case where the repetition rate in the latter of these two grades is larger than in the

---

38/ In some cases, such survival rates are even calculated on the data for one *single year*. For example, employing the data for 1986 given in Table 2, this approach would estimate the survival until the final grade of primary education as:

\[
\text{Enrollment in Grade 6 in 1986} = \frac{51,166}{88,345} = 0.579
\]

This would imply that only 57.9% of those enrolled in Grade 1 in 1986 would reach Grade 6, as compared to 67.4% given by the Reconstructed Cohort Method, see Figure 7.3. This application of the Apparent Cohort Method on data for one year only is *not* to be recommended. In particular, it should not be employed for developing countries where the enrollment in primary education generally is growing rapidly. Under such conditions, this method will grossly *over-estimate* dropout. This is so because the enrollment in Grade 6 in a given year does not depend on Grade 1 enrollment in the same year, but on Grade 1 enrollment six years earlier, as well as on the repetition and dropout rates in all the grades of the cycle throughout this six-year period.
former. If the difference between the rates is large, the estimated dropout may be negative. For example, using this method for estimating the dropout between Grades 5 and 6 in the case of Burkina Faso, using the data in Table 7.1, gives a negative dropout rate of -25%.

To summarize, the Apparent Cohort Method based on time-series data, can be used to estimate dropout and survival by grade in cases where one knows that there is no repetition. This method may also be used when the repetition rates are quite low, provided that their magnitude does not vary much between grades. In cases where repetition rates are relatively high and vary between grades, the Apparent Cohort Method may over- or under-estimate the level of dropout considerably\(^2\).

It should be noted that the Apparent Cohort Method can, at best, be used to estimate the three indicators of retention by grade presented above (i.e., Indicators 1, 2, and 3). None of the remaining 9 indicators can be estimated. As the level of repetition is not taken into account by this method, it can naturally not provide estimates of enrollment by years of study, average duration of studies or level of wastage.

Finally, it is important to realize that the estimates of dropout and survival by grade derived on the basis of the Apparent Cohort Method are not comparable to the dropout and promotion rates defined by equations (9) and (10) in Section V. Rather, the former two measures should be compared with cohort flow Indicators 1 and 2 as explained above. The promotion and repetition rates describe annual changes in enrollment stocks while estimates of cohort survival rates refer to what eventually happens to cohort members.

\(^2\) This aspect is discussed in further detail in Fredriksen (1983) and Cuadra (1991).


-----. "First Grade Repetition Levels in China," UNESCO - OREALC, Santiago, August 1990.


improving the effectiveness and efficiency of educational systems entails, among other things, strengthening the managerial and analytical capacities of ministries of education. A condition to improving these capacities is that policy makers, educational planners and managers increase the use of quantitative data for decision making, planning, monitoring and evaluation. This would, in turn, require that developing countries establish management information systems that provide reliable, relevant and timely information about the operation of schools, the progress of students within the system and the quality of education.

One of the main reasons why educational administrators and policy makers in developing countries do not use sufficiently quantitative data to support day to day operations, or for strategic and medium term planning, is that available education statistics are of poor quality, outdated or unreliable. Therefore, a condition to strengthen the managerial and analytical capacities of ministries of education in such countries is to improve the quality of educational statistics.

The major sources of error effecting educational statistics can be grouped in two major categories: (I) problems originating at the data collection level, and (II) problems arising at the data processing level. This annex discusses key factors causing these two types of problems and provides examples of how these factors affect data quality in some developing countries.

---

This annex was prepared by Ernesto Cuadra and is based on the article by D. W. Chapman and R. A. Boothroyd, "Threats to Data Quality in Developing Country Settings," Comparative Education Review. Vol. 32 No. 4, 1988.
I. ERRORS ORIGINATING FROM DATA COLLECTION

Errors in data collection result from problems created due to: (1) lack of consensus about the meaning of key concepts or because the concepts are not well understood by respondents; (2) instruments for data collection which are poorly designed; and (3) inaccuracies introduced by poor reporting of school level data.

I.1. Lack of Consensus About Data Definitions. It is important that officials responsible for data collection realize that terms used in statistical questionnaires may not have the same meaning for school administrators and teachers as they do for policy makers and analyst working at the central level. In some cases, school level people may not understand correctly what type of information they have been asked to supply.

For example, concepts such as repeaters or dropout may not mean the same to school administrators and to the statisticians who compile national statistics. In Honduras, for example, many students repeating a grade were not recorded as such because they had not failed the end of the year examination the year before, although they were still enrolled in the same grade as in the previous year. Furthermore, children who left Grade 1 before the end of the academic year were considered to be repeaters when they returned to the first grade the following year\(^2\). A study conducted in Somalia revealed that 20% of headteachers did not understand what a repeater was\(^3\).

I.2. Low Reliability or Validity of Data Collection Instrument. Data collection instruments in developing countries are often poorly designed because they either overload respondents with questions about low priority information or because the wording and type of questions included do not reflect what happens

---


at the school level.

(a) Poorly designed instruments. There are two main problems with questionnaires sent by ministries of education to schools. First they are often too long as they request some information that is seldom used or even tabulated. Second, instruments are changed from year to year, creating confusions at the school level with respect to what is the relevant information that they need to collect and keep. A study conducted in the Yemen Arab Republic found that, due to lack of a standardized data collection system, schools generally collect more information than what they were asked to report yearly to the Ministry of Education. Since schools lacked a standardized book-keeping system, headmasters were forced to design their own data collection form according to her/his understanding and experience\(^5\). In Nepal an evaluation of data flow from schools to the Ministry of Education, commissioned by the government, found that the kind of information collected varied from school to school due, primarily, to a lack of an official form for data collection. In most of the schools the only information collected at the beginning of the school year was the name of the student, and very few schools kept records on the progress of the students, e.g. on repetition and promotion. Consequently, when the MOE requested this information, teachers and headmasters reported based on their memory\(^5\).

In Pakistan LeBlanc and Valdes (1991)\(^5\) found a total of 123 different data collection instruments which were used by different government offices to collect information from schools in the provinces of Balochistan and North West Frontier. These instruments basically collected the same information but used different definitions of some key concepts such as headmasters and teachers duties and responsibilities, shifts, repeaters and dropouts. As a

\(^4\) Cuadra, E., M. Anderson, S. Moreland, and F. Dall. Female Access to Primary Education: Trends, Policies and Strategies. Harvard University, Project BRIDGES, 1988

\(^5\) ibid.

consequence of this duplication of effort and multiple definitions, teachers and headmasters were confused and discouraged, which lead to poor quality of data and inconsistencies between the information reported by different government agencies.

(b) Use of inadequate questions or use of questions in inappropriate settings. Many times those responsible for designing the data collection instruments are not fully familiar with the day to day operation of the system. For example, in Egypt, a large number of schools operate in rented facilities. When the Minister needed to assess the cost implications of this practice, no information of how many schools were operating in rented building was available because, due to lack of understanding of this practice, the question about the status of the building included in the annual school census did not capture the complexities of this arrangement. This problem is compounded by a practice observed in many developing countries where the questionnaires used to collect school level data are designed based on instruments used by other countries, or on samples of questionnaires provided by international organizations such as UNESCO. This practice often creates misunderstandings due to inaccurate translations, or use of questions that are not adequate to local realities. For example, in nomadic societies such as Somalia it does not make much sense to ask schools to report on number of dropouts since students move in and out of school as their communities move. Furthermore, a survey showed that 36% of headteachers in this country did not understand at least one question in the MOE school census questionnaire.1/

I.3. Errors in Reporting School Level Data. This type of errors can be grouped in three categories: (a) failures to report data; (b) intentional reporting of false data and; (c) unintentional reporting of inaccurate data.

(a) Failures to report data. Schools might not report data due to communication problems, lack of incentives, lack of clarity about who is responsible for reporting, or simple because the data is unavailable. For example, in countries with automatic promotion (such as Egypt and Costa Rica in

---

the 1980s) school principals and teachers do not count, nor report, repetition in grades where automatic promotion is supposed to be enforced. However, visits to schools in both countries showed that repetition does exist in all grades. Another example is provided by a study conducted in Somalia where more than 10% of headteachers reported that the data requested by the MOE were not available. Of particular difficulty were data on repeaters by grade\(^{2/}\). A third example is provided by the provinces of Balochistan and NWFP in Pakistan. As headmasters and headteachers are overloaded with information requests from different agencies within the government, they prioritize which agency to respond to and which one not to respond to. Generally, agencies that control resources, such as the Finance Department or the Members of the Provincial Assembly, receive high priority and immediate response, as oppose to the Statistic Cell which receives a very low priority and usually has to request the information more than once.

(b) Intentional reporting of false data. This is a frequent source of errors because reporting officials believe that schools or districts will benefit from providing distorted information. Over-reporting of enrollment is the most common problem found in this area and can be attributed to the fact that educational resources — such as teachers and instructional materials — often are allocated to schools based on the level of reported enrollment. However, under-reporting also occurs, either because of poor communications (Somalia, for example) or because there are unintended incentives embedded in the operation of the system. An example of the latter is provided by a Liberian study which found that some local authorities and headteachers under-reported enrollment so they could keep part of the money collected from students' registration fees at the district or school level, largely to cover operating expenses\(^{2/}\).

Over-reporting of enrollment seems to be a special serious problem in countries where access to primary education is still problematic or in

---

\(^{2/}\) Ibid.

countries that have experienced an explosive growth of enrollment. For example a study on data quality in Yemen, where the gross enrollment ratio for primary education went from 44% in 1980 to 92% in 1988, found that enrollment reports submitted to the MOE by a sample of schools from the capital city exceeded by 56% the enrollment recorded in the school files\textsuperscript{10}. A different form of inflation of enrollment occurs in countries where parents are penalized if their children are not registered at schools but suffer no sanction if they do not attend. This appears to be the case in China where a recent Bank mission found a 30% difference between school attendance and enrollment registration in the first grade in rural areas\textsuperscript{11}.

Another example of reporting errors of enrollment is found in Pakistan where LeBlanc and Valdes\textsuperscript{12} found that the only province which, in 1987, reported enrollment by age and grade to the central government was the province of Balochistan. Although this information was not collected directly from schools, two officers working in the statistic cell of the Directorate of Education produced the report based on their "experience". For example, the age of students in each grade was estimated based on the "legal" entrance age to first grade. As a consequence of this, the majority of the student enrolled in, for example, Grade 3 were reported to be seven years old because the official entrance age to first grade is five years. Some overage students were reported in each grade based on the fact that some students enter late. The reason why these officers took time to produce this report was a belief that all other provinces would send this information to the central government and that their department would look bad and could suffer financially if they did not provide such data.

Not only data on enrollments suffer from false reporting. Data on repetition are often subject to the same type of error. For example, in Honduras, Article 144 of the law that created the system of promotion and evaluation

\begin{itemize}
\item \textsuperscript{10} Chapman, D. W. and Boothroyd, R. A. (1988).
\item \textsuperscript{11} China. Provincial Education Planning and Finance Sector Study, January 1990.
\item \textsuperscript{12} LeBlanc, T. and J. Valdes. (1991).
\end{itemize}
(Sistema de Evaluación y Promoción Controlada) states that, if more than 15% of the students of any given grade fail to be promoted, then the teacher will be held accountable for the failures and the Principal and Supervisor will study the reasons for the teachers "incompetence" in order to overcome his or her limitations. This rule explains, in part, why 15% of the student reported to have been promoted from Grade 1 to Grade 2 in 1983 were still enrolled as repeaters in Grade 1 the following year.\(^{13}\)

(c) Unintentional reporting of inaccurate data. Unintentional reporting of inaccurate data occurs either because the reporting agencies do not understand the real meaning of the information that they are asked to report, or because they lack time to collect the information properly and rely on their memory. For example, in countries such as the Philippines where headmasters report the number of dropout, many transfer students are reported as dropout because parents do not get a transfer certificate when they move their children to a new school and for schools it is almost impossible not to accept a new students even in cases where he or she does not bring the transfer certificate. On the other hand, many dropouts are not reported at all because either headmasters assume that they moved to a new school and, consequently, report them as transfer or because parents, after getting the transfer certificate, withdraw their children from school. Another example of this problem is what takes place in Honduras where teachers, as a matter of practice, fill the annual school census at home, relying more on their own memory than on official documents when reporting number of repeaters. This in part explains why there was only 36% agreement between the number of repeaters reported by a sample of schools to the MOE and what the school attendance record showed.\(^{14}\) Something similar happened in Paraguay where teachers reported first grade repeaters based on students' being asked to raise their hand if they were repeating. Many students --specially in Grade 1-- did not even understand what a repeater was.

\(^{13}\) Cuadra E. (1988).

II. ERRORS ORIGINATING FROM DATA PROCESSING

A second general threat to data quality in developing countries derives from errors resulting from data processing. This type of problems occur in cases where officials responsible for processing information lack clear guidance on how to handle situations where data are incomplete, or where there are no systematic procedures for double checking results when data is aggregated, or where educational statistics are handle by personnel who is not familiar with the operation of the system, or because there is little understanding of the value of having consistent and unique coding schemes.

II.1. Errors due to Treatment of Missing Data. Missing data occurs mainly due to poor communication. In both Guatemala and El Salvador there were wide fluctuation in enrollment data in the early 1980s due to partial coverage resulting from the civil war that affected those countries. Many times supervisors could not reach schools because they were located inside a war zone. Since they did not have clear instructions about what to do in these cases, some reported the enrollment of those schools based on pass trends to find out, a year later when they came back to the zone, that often schools were closed because the population had relocated. In Somalia and Nepal, due to transportation problems, it is difficult for the MOE officials to know if schools not reporting data to the MOE do so because they are closed, or because they have not received the questionnaire, or simple because they do not want to respond.

II.2. Errors in Transferring and Summarizing Data. Generally school level data is collected at the classroom level first, then is aggregated at the school level. From the school it usually goes to a district or department where it is aggregated again to be sent to the state or provincial authorities. Before the information is sent to the central office of the Ministry of Education it is again aggregated. Each time the information is aggregated, the risk of involuntary errors increases. In Yemen, Chapman and Boothroyd (1988) found discrepancies of 25% between the enrollment records sent from a sample of 35 primary schools near the capital to the Ministry of Education, and the summary reports produced by the Ministry. They discovered that "data for only 19 of 35 classrooms (54%) were correctly transferred" (p.421) and, for one classroom, the
II.3. Errors in Producing Statistical Reports and Analysis. This problem occurs when those in charge of data analysis and reporting are not fully familiar with the day to day operation of the system. For example, in Pakistan, two government agencies reported different enrollment of girls in primary education because one of the agencies, unaware of the fact that many girls in rural areas attend boys schools, reported girls enrolled in boys schools as boys enrollment.

II.4. Failure to Use Common Coding Schemes. A classical coding problem is the assignment of different codes to the same unit. Sometimes this happens because different offices are responsible for collecting different kind of information. This seems to be the case in Jordan where the statistics unit of the MOE collects information on schools characteristics and student enrollment, and the examination unit collects information on examination results. These two units have their own coding systems for schools with the results that it is almost impossible to match achievement scores with school characteristics because schools have different identification codes in each subsystem. Although it is possible to recode schools and then match them based on the new code, this is a time consuming activity that must be done manually since the only common element shared by the two systems is the name of the school, which is not only not unique to each school but also admits different spelling. A variation of this problem occurred in Egypt in the early 1980s where the same unit received different codes in two consecutive years. This situation developed when a new governorate was created by splitting a large governorate into two smaller units. The new governorate was given the same code as the old one which, in turn, received a new code. This change made it difficult to compare historic data specially for people unaware of the changes, who happened to be the majority of people working at the MOE except for the few people working at the statistics department.