Modeling methodology and assumptions in HIPE-Trend based model

Baseline Data

The EPDC projection model estimates pupil volume based on several key pieces of input information, including the gross intake rate (GIR), and the repetition, dropout, and transition rates. Available historical values on these key indicators are used to estimate trends across time, which are then extrapolated for the period of the projections, following the methods described below.

The baseline data for education indicators is based on information published by the UNESCO Institute for Statistics (UIS). Historical GIR values are calculated based on the number of pupils and repeaters in the first grade of primary for the period of 2000-2013 and age specific population data published by the United Nations Population Division (UNPD). The repetition, dropout, and transition rates for the years 2000-2013 were calculated by EPDC when possible based on the number of pupils and repeaters in each year, by grade, provided by UIS.

Flow rates are calculated by grade. In instances where no repetition or dropout rates could be calculated across 2000-2013 data, average rates for the level (primary or lower secondary) are used as the starting values for the projections. When no rates could be calculated for the entire level, no projections are generated by the model.

Progress on Key Education Outcome Indicators

Modeling Assumptions and Definitions

A. ***Input Indicators.*** This section describes the inputs, or driving indicators, that are used to project values for the entire set of outcome indicators. These include gross intake, transition, repetition and dropout rates. The basic assumption is that if the indicator was moving in the right direction (towards full enrollment, full completion), and the trends were found to be statistically significant, they were used to project future enrollment. If the trends have been in the undesirable direction, or not statistically significant, trends observed in historical data available for other countries have been used instead.

1. The latest available ***gross intake rate (GIR)*** was used as a starting value for pupil projections. GIR progress projections are calculated based on the following equations:

$$1.a)\ GIR_{\Delta} = \propto\ +\ \beta(GIR_{t-1})$$

where the yearly change GIR_{Δ} is calculated via a regression with a constant $\propto$ (the average change), and a variable rate of change β based on the previous year's level of GIR (GIR_{t-1}).

$$1.b.)\ GIR_t = GIR_{t-1} + GIR_{\Delta}$$

where GIR at year t equals GIR in previous year (year *t-1*) plus a yearly change GIR_{Δ}.

a. The GIR is projected using an extrapolation of a regression model (equation 1.*a.* above), where the annual change in GIR is dependent on the GIR value in the previous year (regression with available historical data). The estimated annual change is then added to the previous year's GIR value to project the following year's GIR (equation 1.*b.* above). The model projects increasing trends are for countries with historical intake rates below 100%, and decreasing trends for countries with intake rates above 100%. The shape of the projected trends is curvilinear, such that annual change diminishes as the value of the GIR approaches 100. See Figure 6 for an illustration of the GIR trends in this model. The regression equation, based on all available data for low income and middle income countries, sets the trends to converge around 115.

b. If the GIR most recent value was more than 100%, it is projected to decrease using the equation shown in Figure 6. However, if the most recent value of GIR was higher than 120%, it is kept constant for 5 years before the equation is applied. As historical data shows, many countries with particularly high gross entry ratios have experienced a rapid increase in enrollment to grade 1 in the recent years. It is reasonable to expect that in these countries a decrease in gross entry rates will not be immediate, but rather will initially stabilize before beginning on a gradual downward trajectory.

2. ***Dropout and repetition rate*** trends are based on the average calculated rates of change for low income and middle income countries, determined by the country's most recent historical value.

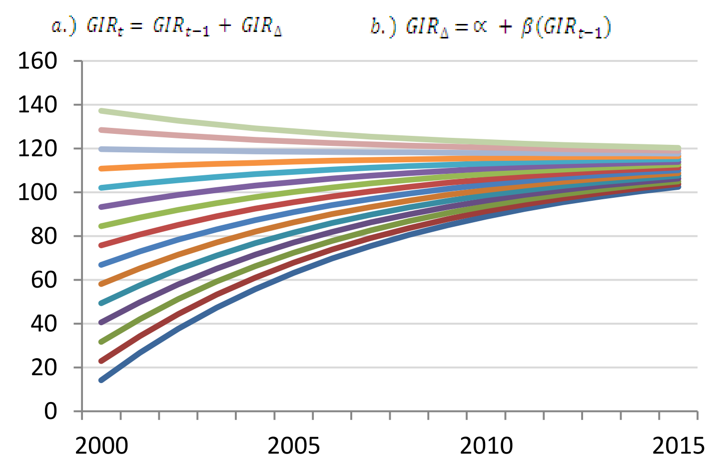

Figure 1 Projected Gross Intake Rate

Repetition and dropout projections are based on the following equation:

$RD_t = \propto T + \delta RD_{mr} + \gamma TRD_{mr}$; where repetition or dropout at time t is determined by the most recent historical value of repetition or dropout RD_{mr}, the year of the projection T and an average rate of change $\propto$; and where the country-specific rate of change varies, based on the level of the most recent value RD_{mr}.

a. Based on historical trends, we assume that the rate of decline for repetition and dropout varies based on the level of these indicators at the latest observed point in time. Thus, we project that countries with higher repetition and dropout will decline at a faster (but constant) rate than countries with historically low levels of repetition and/ or dropout. See Figure 7 for an illustration of the repetition and dropout trends projected by the model.

b. Dropout and repetition are projected individually for each grade of primary and lower secondary.

c. Dropout and repetition rates are projected from the most recent year's rate for historical data to decline linearly at the country's historical slope until it reaches the minimum value IF that country showed a declining rate, AND the correlation of the historical values of repetition and dropout with time is strong (i.e. correlation coefficient stronger than 0.7).

d. If a country lacks data to estimate its own slope of decline, or if the historical slope is positive (i.e. repetition and/or dropout have increased in the past ten years), dropout and repetition are set to decline linearly by the average rate of other low income and middle income countries, using the equation above, with regression coefficients based on all available values[1].

$$RD_t = \propto T + \delta RD_{mr} + \gamma TRD_{mr}$$

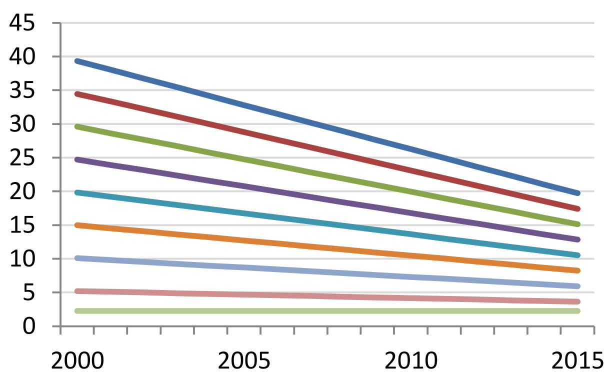

Figure 2 Projected Repetition and Dropout Rates

e. The minimum possible values for repetition and dropout are set at the 25th percentile of available repetition and dropout rates, separately for males, females, primary, and lower secondary.

f. For countries with high GIR (above 100%), AND a growing trend in historical GIR rates, **first and second grade repetition rate** is set constant at the most recent available value. This is done to offset the deflation in the number of pupils in these grades, which would result from the combination of the projected decline in GIR imposed by the model settings (declining trend for countries above 100%), and the projected decline in repetition and dropout.

g. Regardless of the direction of historical trends in gross intake rates, repetition rates are also kept constant for some period of time where intake is particularly high. As mentioned above, gross intake rates higher than 120% at the most recent year of available data are kept constant in the model for 5 years. Based on an assumption that repetition in early grades is not likely to decrease when entry rates remain at such high levels, repetition rates in the first and second grades are set constant for the same years as intake. In addition, particularly high

[1] An exception is made for countries where the latest known rate of dropout is below 8% for primary grades, and 10% for lower secondary grade: the general equation projects increasing dropout rates, based on historical data. Therefore, a separate slope of 0.19 per year was set for these cases, calculated using a time series regression on a subset of countries that showed decreasing dropout at these low levels.

gross intake rates may be linked to underreported repetition, i.e. treating repeaters as new entrants in the calculation of intake rates. In order to offset the deflation in the number of pupils in those cases, repetition rates in the first two grades of primary school are kept constant for additional 5 years after high intake rates start decreasing.

3. The ***transition rate*** is projected it to grow exponentially until it reaches 100%. The growth is assumed either at a historical rate for that country, IF it was a positive trend, or at the average historical growth rate of countries that had progressed in the right direction, whichever is faster[2].

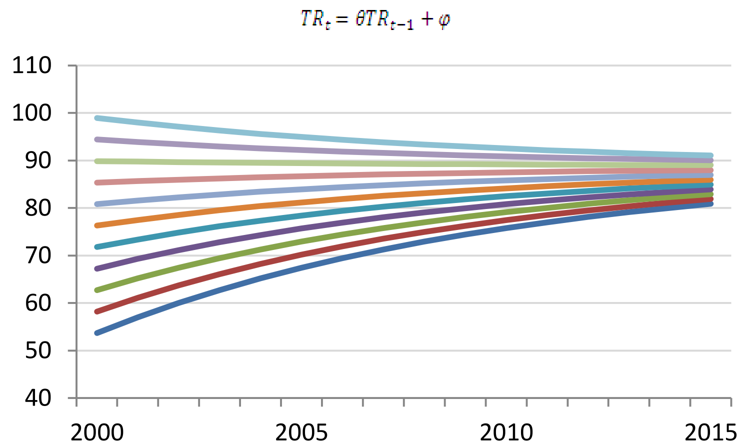

Figure 3 Projected Transition Rate

Transition trends are set to converge to 100%. The equation for the transition projections is the following:

a) $TR_t = TR_{t-1} + TR_\Delta$

where transition at time t TR_t equals the previous year's level TR_{t-1}, plus yearly change TR_Δ, which is estimated as:

b) $TR_\Delta = \propto + \theta(100 - TR_{t-1})$

where the estimated change in transition rate is determined by the difference between the previous year's value and 100, the optimal rate, multiplied by coefficient θ, calculated using a regression with historical data. If transition in the previous year equals 100, TR_Δ cancels out, and the change equals zero.

B. ***Output Indicators.*** The milestones for these indicators are based on the expected trends in the input indicators, as outlined above.

1. ***Number of public pupils*** is sum of the pupil projections for the grades officially designated in primary or lower secondary, less the estimated number of pupils enrolled in private schools.
2. ***Number of new entrants*** is the GIR times the population of official school entry age.
3. ***Gross enrollment rates (GER)*** are estimated based on pupil projections (see above), as a proportion of the age-appropriate population group for each school level (UNPD).
4. ***Completion rate*** is the proportion of non-repeating students in the last grade of primary over the population of the official age appropriate for the last grade of primary.

[2] This choice was made based on the observation of trends: transition rates were changing at an extremely slow pace, and applying a common set of coefficients imposed a faster, albeit still very gradual, increase.

Projection coefficients and resulting equations in EPDC HIPE 8.0 (August 2013)

Driver Rate	Formula	Equation, Male enrollment	Equation, Female enrollment
Gross intake rate	$GIR_{\Delta} = \propto + \beta(GIR_{t-1})$	$GIR_{\Delta} = 14.20 - 0.12\,(GIR_{t-1})$	$GIR_{\Delta} = 12.87 - 0.11(GIR_{t-1})$
Repetition- primary	$R_t = \propto T + \delta R_{mr} + \gamma TR_{mr}$	$R_t = .067 * T + .975 * R_{mr} - 0.034 * TR_{mr}$	$R_t = .070 * T + .977 * R_{mr} - 0.034 * TR_{mr}$
Repetition-lower secondary	$R_t = \propto T + \delta R_{mr} + \gamma TR_{mr}$	$R_t = .141 * T + 1.01 * R_{mr} - 0.025 * TR_{mr}$	$R_t = .147 * T + 1.05 * R_{mr} - 0.033 * TR_{mr}$
Dropout- primary	$D_t = \propto T + \delta D_{mr} + \gamma TD_{mr}$	$D_t = .225 * T + .845 * D_{mr} - .027 * TD_{mr}$	$D_t = .21 * T + .854 * D_{mr} - .03 * TD_{mr}$
Dropout- lower secondary	$D_t = \propto T + \delta D_{mr} + \gamma TD_{mr}$	$D_t = .385 * T + .929 * D_{mr} - .052 * TD_{mr}$	$D_t = .254 * T + .856 * D_{mr} - .038 * TD_{mr}$
Transition	$TR_{\Delta} = \propto + \theta(100 - TR_{t-1})$	$TR_{\Delta} = -1.063 + .095 * (100 - TR_{t-1})$	$TR_{\Delta} = -.703 + .073 * (100 - TR_{t-1})$