

Methodology Notes

Author: Nicole Siaw

This article combines data from various publicly available sources to approximate enrolment and dropout rates in Malaysia and Sarawak across education levels and districts. Due to data limitations, several assumptions and estimation techniques were used. The aim is not to provide definitive numbers, but to highlight broad patterns and areas of concern.

Some notes are provided here on the methodology for:

- Enrolment/Dropouts at a National level
- Enrolment/Dropouts at a Sarawak-wide level
- Enrolment/Dropouts at each education district in Sarawak

Broad Methodology

Student progression in Malaysia is typically measured by 3 main indicators:

- **Enrolment Rate**

The percentage of students enrolled in a particular stage of education, calculated as a share of total live births in the corresponding cohort year. For example, the enrolment rate for Primary 1 in 2022 is calculated by dividing the number of students enrolled in Primary 1 by the number of live births recorded seven years earlier, in 2015.

- **Completion Rate**

The proportion of students who complete a given stage of education, relative to the number of students from the same cohort who initially entered that stage. For example, to calculate the completion rate for primary school, the number of students who exit at Primary 6 is compared to the number who enrolled in Primary 1 six years earlier. In some cases, this rate may exceed 100% if students enter the system late, complete school late, or transfer in from other areas.

- **Transition Rate**

The proportion of students who progress from one stage of education to the next, within the same cohort. For instance, the transition rate from Primary 6 to Form 1 compares the number of students from a cohort enrolled in Form 1 to those who were enrolled in Primary 6 the year before.

These indicators are typically published at a national level. It is not always published for the state level, and to my knowledge, not often at the district level. The aim is to calculate the equivalents at a more granular level, down to the administrative district level.

Where available, we always use official figures from published sources. Where not available, we estimate them using reconstructed cohort data based on live birth statistics and enrolment records. Alternatively, we may use proxy indicators to help us understand. Completion rates were available for all calculations seen, so enrolment rates and transition rates are the main calculated indicators.

Once the enrolment rates, completion rates, and transition rates are calculated or obtained, they are converted into the **percentage of the original birth cohort remaining at each education stage**. This provides a unified and intuitive view of how student numbers decline (or persist) as the cohort moves through the education system.

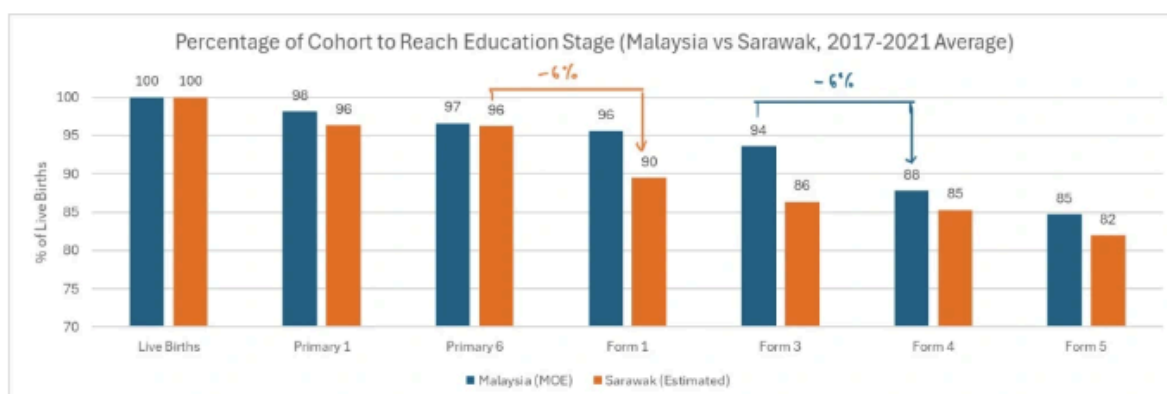
The method begins with **100% of the birth cohort**, and rates are applied sequentially:

- **Enrolment rate** estimates how many enter Primary 1
- Then apply **completion rate** estimates how many reach the end of Primary 6
- Then apply **transition rate** shows how many continue to Form 1
- The process continues stage by stage through to Form 5

This stepwise reduction forms a “remainder curve” of the cohort, showing where the biggest dropouts occur. Whether at school entry, during key transitions, or at completion points. It also enables meaningful comparisons across time, districts, and geographies, even when absolute numbers differ.

Finally, **absolute dropout figures** can be estimated by multiplying the percentage of the cohort lost at each stage with the corresponding **live birth data** for each district or year. This shows the same ‘remainder curve’ but shows the magnitude of the enrolment/dropouts rather than in percentage terms.

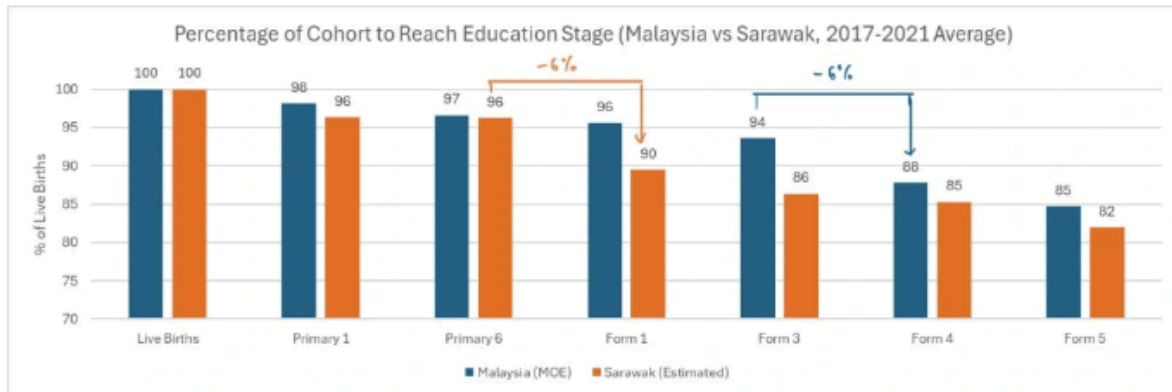
Enrolment/Dropouts at a National Level



Graph showing the percentage of each cohort in Sarawak reaching different education stages compared to the national average from 2017-2021. Source: author's calculations, based on MOE and DOSM data.

At a national level, enrolment rate, completion rate and transition rate data is widely available. We take the average between 2017-2021 to smooth out any idiosyncrasies arising just from a single year.

Enrolment/Dropouts at a State-Wide Level



Graph showing the percentage of each cohort in Sarawak reaching different education stages compared to the national average from 2017-2021. Source: author's calculations, based on MOE and DOSM data.

At the state-level, we begin to need to reconstruct some of this data.

- Enrolment Rate: reconstructed using enrolment and live birth data
- Completion Rate: available from published sources
- Transition Rate: calculated using enrolment rate and completion rate

Enrolment Rates

At the state level, both live birth data and enrolment data are available, though not always in a disaggregated form:

- Live birth data is reported annually for the whole of Sarawak, based on birth registrations from the National Registration Department (JPN), accessible via DOSM.
- Enrolment data is provided as total enrolment figures for the entire primary (Primary 1–6) and secondary (Form 1–5) levels, without breakdowns by year group.

To estimate annual enrolment at Primary 1 and Form 1, we use a simple proportional assumption:

- Primary 1 enrolment \approx total primary enrolment \div 6
- Form 1 enrolment \approx total secondary enrolment \div 5

This provides a rough approximation of annual intake. To smooth out year-to-year fluctuations and minimise estimation error, we calculate **average enrolment rates across a 5-year period (2017–2022)**, capturing broader trends rather than isolated spikes/anomalies.

To match these enrolment figures with the appropriate cohorts, **live birth data is averaged over the years corresponding to the youngest and oldest possible students enrolled during 2017–2022**. For primary school, this means:

- **2005:** birth year of the oldest Primary 6 student in 2017
- **2016:** birth year of the youngest Primary 1 student in 2022

Thus, enrolment rates are calculated using the average number of live births from **2005 to 2016**, aligning with the full range of cohorts present during the 2017–2022 period. A similar logic is applied for secondary school estimates.

Percentage of Students to Reach Form 5

There was a need to estimate the amount of students that were expected to reach Form 5.

To estimate the number of students reaching **Form 5**, the number of students enrolled for the **SPM examination** is used as a proxy. This is a reasonable proxy in the Sarawak context, as the number of private schools is small, and most still follow the national curriculum and sit for SPM. This assumption may not hold in states with more diverse education streams (e.g. technical, religious, or international schools), where students may sit for different examinations.

The **Form 5 attainment rate** is calculated by dividing the number of SPM candidates (2017–2021) by the average number of live births for the corresponding cohort.

Transition Rates

The transition rates between education stages are then derived indirectly using the enrolment rate and school completion rates.

- For the **Primary 6 to Form 1 transition**, the rate is calculated as:

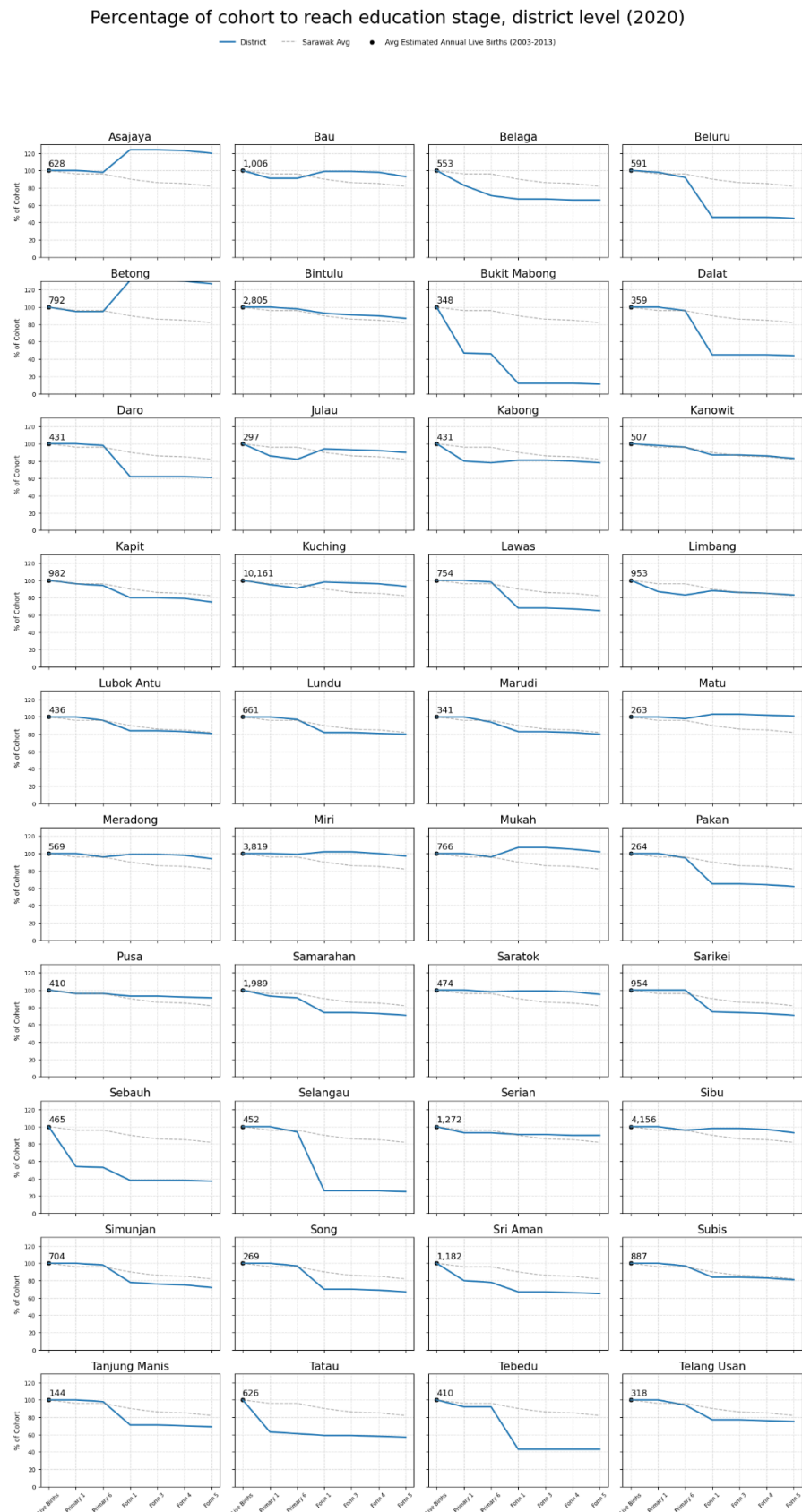
$$\text{Transition Rate} = \text{Form 1 Enrolment Rate} \div (\text{Primary 1 Enrolment Rate} \times \text{Primary Completion Rate})$$

- For the **Form 3 to Form 4 transition**, the formula is:

$$\text{Transition Rate} = (\text{Form 5 Completion Rate} \div \text{Upper Secondary Completion Rate}) \div (\text{Form 1 Enrolment Rate} \times \text{Lower Secondary Completion Rate})$$

These formulas allow us to back-calculate the likelihood of students progressing between key stages, even when direct transition data is unavailable.

Enrolment/Dropouts at each Education District in Sarawak



At the district level, historical data is sparse and more proxies/ inferences need to be made.

- Enrolment Rate: reconstructed using known enrolment and *estimated* live birth data
- Completion Rate: available from published sources, but in different base years
- Transition Rate: calculated using enrolment rate and completion rate

Enrolment Rates

Live Birth Data

Historical live birth data at the district level is incomplete and, where available, not always aligned with current administrative boundaries. Over time, some districts in Sarawak have been subdivided or redrawn, leading to inconsistencies in how populations are recorded.

The following datasets are used in this study:

- 2000–2022: Total live births for Sarawak (overall state-level only)
- 2021–2022: District-level live births based on the current 40 administrative districts
- 2015–2020: District-level live births based on an older boundary with 32 districts
- 2020 Census: Population by age group and district (based on current boundaries)
- 2010 Census: Population by age group and local authority/sub-district

These inconsistencies require careful alignment when matching birth data to enrolment at the district level. The live birth data for 2000-2020, provided in a form that matches our 40 current 40 administrative district structure is estimated as follows:

- **2000-2004:** Used the **proportion of 5–9 year-olds by district** from the **2010 Census**, multiplied by total state-level live births for each year.
- **2005-2009:** Used the **proportion of 0–4 year-olds by district** from the **2010 Census**, multiplied by total state-level live births for each year.
- **2010-2014:** Used the **proportion of 5–9 year-olds by district** from the **2020 Census**, multiplied by total state-level live births for each year.
- **2015-2019:** District-level live births were available for 32 districts. These were **reallocated to the current 40-district structure** based on observed proportions from **2021–2022**.

Larger District	Split District	portion of Split District/Larger District
Betong	Betong	0.386885246
Saratok	Kabong	0.210432191
	Pusa	0.200298063
	Saratok	0.202384501
Bintulu	Bintulu	0.736585366
Tatau	Sebauh	0.112195122
	Tatau	0.151219512
Kapit	Kapit	0.462686567
Song	Bukit Mabong	0.164179104
Belaga	Song	0.119402985
	Belaga	0.253731343
Miri	Miri	0.641221374
Marudi	Beluru	0.099236641
	Subis	0.148854962
	Telang Usan	0.053435115
	Marudi	0.057251908
Mukah	Mukah	0.390243902
Dalat	Tanjung Manis	0.073170732
Daro	Dalat	0.182926829
Matu	Daro	0.219512195
	Matu	0.134146341
Serian	Serian	0.756410256
	Tebedu	0.243589744

- **2021-2022:** Used actual live birth data, already aligned with the current 40-district boundaries.

In the absence of published live birth data by district for those years, this method provides the best available approximation. It allows district-level birth proportions to vary over time, reflecting population shifts observed in different census years, while also aligning with overall state-level trends in annual birth increases or decreases across Sarawak.

Enrolment Numbers

A dataset of all known schools (primary/secondary) in Sarawak, including student and teacher counts with coordinates, was obtained from [MySchoolChildren.com](https://myschoolchildren.com). Total students per district were calculated by matching school coordinates to district boundaries using a Python script. Note that this raw dataset is from Jan 2020, which means it should not include anomalies introduced by Covid in that year.

Calculations

Enrolment rates are calculated by dividing district-level enrolment by corresponding live birth figures, using the method outlined in the sections above.

Post Processing of Enrolment Rates

At the primary level, several rural districts show enrolment rates exceeding 100%. This can be attributed to a combination of factors:

- Underestimation of live births due to limitations in our estimation method
- Delayed or unregistered births, especially in remote areas
- Small cohort sizes, where minor discrepancies lead to large percentage shifts (e.g. 45 unaccounted students in a district with 300 births = 115%)
- Statistical artefacts, where schools near district boundaries serve multiple districts but are recorded under only one, inflating enrolment in one district and undercounting in the other

To partially correct for this, some districts with >100% enrolment and boundary-adjacent schools had their “excess” students reassigned to neighbouring under-enrolled districts. This helps reduce the boundary distortion effect.

For remaining districts, enrolment rates above 100% are retained but noted as likely due to birth data uncertainty or underreporting of live birth data from those districts in general. For consistency with the rest of the analysis, we cap initial enrolment at 100% when constructing cohort progression curves.

Completion Rates

Completion rate data was sourced from DOSM for the years 2021–2023, grouped by PPD (District Education Office) boundaries, rather than administrative districts. It is important to note that these years immediately follow the COVID-19 pandemic, and the data may reflect

some anomalies. For example, many districts report completion rates exceeding 100%, likely due to students graduating later than expected.

While pre-pandemic data would have been preferred to minimise these distortions, it is not publicly available. To reduce the impact of outliers, we use the average completion rate from 2021–2023, while acknowledging this as a limitation

Sarawak is divided into 32 PPD districts, which do not align perfectly with the 40 administrative districts. Some PPDs span multiple administrative districts, and vice versa. To assign completion rates to each administrative district, we use the PPD rate that corresponds to the majority of schools located within that district.

Transition Rates

Transition rates are calculated using the same method described earlier. For the Primary 6 to Form 1 transition, the rate is derived as:

- $\text{Transition Rate} = \text{Form 1 Enrolment Rate} \div (\text{Primary 1 Enrolment Rate} \times \text{Primary Completion Rate})$

Due to data limitations, the Form 3 to Form 4 transition rate could not be calculated at the district level. Instead, we apply the state-wide average transition rate of 99% uniformly across all districts.

Percentage of Students to Reach Form 5

Finally, the percentage of students reaching **Form 5** is estimated by multiplying the following:

- $\text{Form 1 Enrolment Rate} \times \text{Lower Secondary Completion Rate} \times \text{Form 3–4 Transition Rate} \times \text{Upper Secondary Completion Rate}$.

Limitations

While every effort was made to apply reasonable assumptions and minimise error, the results are indicative, not definitive. They should be interpreted as broad patterns rather than precise counts.

Several limitations should be noted:

- **Averaging across years:**
Enrolment and birth data were averaged across 2017–2022 to smooth short-term fluctuations. This assumes stability over time and does not track a single cohort across time, which may mask year-on-year variation.
- **Mismatched base years:**
In some cases, data from different years were used due to limited availability. This may introduce inconsistencies, as district-level demographics and enrolment patterns

change over time.

- **Limited post-2021 data:**

Most of the analysis uses data prior to 2021. While access and reporting may have improved since then, those changes are not fully captured here. The impact of **COVID-19** on education outcomes, particularly delayed graduations and dropout patterns, is only partially addressed.

- **Focus on government school data:**

The analysis relies heavily on publicly available data from the MOE-run government school system. In Sarawak, this is a reasonable approach, as private school enrolment is minimal and most students sit for SPM. However, this assumption may not hold in other regions with more diverse education streams (e.g. religious, technical, or international schools).

These factors mean that findings are most useful for identifying **general trends and disparities**, rather than making precise claims at the individual district or school level.