Helen Seitzer
Michael Windzio

Development of Historical Enrolment Trends

Bremen, January 2021
Helen Seitzer, Michael Windzio
Development of Historical Enrolment Trends
SFB 1342 Technical Paper Series, 6
Bremen: SFB 1342, 2021

SFB 1342 Globale Entwicklungsdynamiken von Sozialpolitik /
CRC 1342 Global Dynamics of Social Policy

Postadresse / Postaddress:
Postfach 33 04 40, D - 28334 Bremen

Website:
https://www.socialpolicydynamics.de

[ISSN 2700-0389]

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)
Projektnummer 374666841 – SFB 1342
Development of Historical Enrolment Trends
DEVELOPMENT OF HISTORICAL ENROLMENT TRENDS

Helen Seitzer and Michael Windzio

INDEX

INTRODUCTION ................................................................. 4
DATA ................................................................. 5
METHODS ............................................................. 6
REFERENCES ............................................................ 8
APPENDIX ............................................................. 9
Ensuring universal access to schooling can be considered a crucial part of a welfare state’s responsibilities (Ramirez & Boli, 1987). In fact, universal access to formalized education is often defined as an important goal in establishing a modern welfare state in the first place, and multiple organizations including International Organizations (IOs) are working on ensuring that goal by declaring ‘universal primary education’ as an important target for agreements like the Millennium Development Goals (MDGs). Education is similarly credited to be a crucial determinant of economic development (Hanushek & Woessmann, 2011) and is considered an important tool for providing citizens with equal opportunities and for promoting equality and social mobility (Deary et al., 2005). Enrolment in state-organized formalized schooling is therefore often used to measure a country’s industrial and social standards (Benavot & Riddle, 1988). Consequently, enrolment rates are a recurring measure in comparative studies on welfare states, focusing especially on the development of welfare policy and compliance. For these types of time-series analyses it is crucial to have complete measures for many countries over a long period of time. Historical data on enrolment, however, is often either scarce or ambiguous. There are multiple possibilities for calculating enrolment in the first place, for example, by focusing only on non-repeaters or by separating enrolments according to age group, gender, level or type of education, and so forth. In addition, the basis of calculation is not always sufficiently explained, especially in historical records. Furthermore, there are issues regarding existing data reliability: Depending on the source, national comparative data could be collected through population surveys or reported from statistical office of the respective country, causing suspicion over the accuracy of the data. A source for consistent enrolment statistics is UNESCO’s Institute of Statistics, as they provide enrolment rates from the 1970s onwards for a multitude of countries. Recent changes in their calculation methods reduced the number of available enrolment measures so that there are currently only enrolment rates calculated by two different methods, as opposed to the previously available four measures (which are included here). This measure for enrolment rates combines multiple historical and recent datasets on enrolment into one consistent trendline for up to 192 countries from 1820 to 2018. The strength of this trendline is not the historical accuracy of every single countries’ statistics, but the overall trend, completeness and consistency. In contrast to other datasets focusing on the accuracy of their numbers, this dataset includes correction factors to obtain a consistent trendline representing a probable, complete, trend and rise of enrolment over time.

Previous efforts to consolidate enrolment rates for longer time periods were conducted by Lee and Lee (2016). This dataset combines multiple enrolment measures and previously collected data on schooling to provide a more accurate measure of historical enrolment rates, similar to our approach. However, this dataset only contains 110 countries and provides data every five years. The approach is much more focused on the historical accuracy of the enrolment trend than its completeness. Another collection of historical enrolment rates is Benavot and Riddle’s (1988) dataset for primary enrolment rates. This dataset is also only available for a limited set of countries. Other collections of historical enrolment are embedded in encyclopedias such as Mitchell’s International Historical Statistics (Mitchell, 2007a, 2007b, 2007c). Most historical sources are mostly included in Lee and Lee’s dataset and were therefore not further considered in this calculation. Additionally, measures from statistical institutes like the World Bank or Eurydice for European countries are mostly based on UNESCO’s datasets (UIS, 2019) and were therefore also not considered any further. UNE-
SCO’s data is deemed the most accurate, as the countries report these numbers themselves; however, only UNESCO’s processed data, not the raw data, is available to the public.

Most previous datasets are mainly concerned with the historical accuracy of the data provided. For this dataset, however, we chose a slightly different approach than that of previous efforts: Since most historical records on enrolment can be contested somehow, we wanted to provide a more comprehensive estimation of a possible enrolment trend development. This dataset unites multiple sources on enrolment and therefore provides a trendline, describing how enrolment over time most likely evolved without too much focus on the historical accuracy of single datapoints.

Our estimation includes a total of 219 countries and provides data on enrolment from 1820 to 2018. The significance of this contribution therefore lies not in its historical representativeness but in its comprehensiveness. The strength of the trend as presented here allows for the analysis of school enrolment, including its influence on other welfare policies and vice versa, over a longer time span.

**Data**

Five data sources were included into this trendline, four of which were available from the UNESCO Institute of Statistics as of September 2020. Three measures from UNESCO were included in the dataset: ‘net enrolment rates’, ‘adjusted net enrolment rates’, and current ‘total enrolment rates’. Also included were Lee and Lee’s dataset as well as Benavot and Riddle’s dataset. All these measures defined enrolment differently. While gross enrolment ratios are the standard in most publications, they are disaggregated by level of education but not by age and therefore include all over and under school-age children, skewing the distribution depending on the number of children repeating classes. Lee and Lee define their data as “…adjusted ratios are a modification of the gross enrolment ratios in that they account for the repetition of grades in primary and secondary schools” (Lee & Lee, 2016; 148), while gross enrolment ratios “[are] constructed by dividing the number of all persons enrolled in a given level of schooling by the population of the age group that should be enrolled at that level according to national regulations or customs” (Lee & Lee, 2016; 148).

The UNESCO statistics are calculated slightly differently: “Adjusted net enrolment is the number of pupils of the school-age group for primary education, enrolled either in primary or secondary education, expressed as the percentage of the total population in that age group”. Total net enrolment in contrast is “[The] total number of students of the official age group for a given level of education who are enrolled in any level of education, expressed as a percentage of the corresponding population”. Net enrolment rate on the other hand is “the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age”. Benavot and Riddle’s measures are described as “unadjusted enrolment ratio is found by dividing enrolments at each level by the population of a constant school-age category for all countries” (Benavot & Riddle, 1988; 195). While all these measures sound very similar at first glance, they either include different age groups, do not distinguish between primary and secondary schooling, or do distinguish between the levels of schooling while ignoring age groups. The measures are all highly correlated though and are therefore predictive of the resulting combination trend.
METHODS

All datasets were treated similarly before the main estimation procedure. Missing datapoints had to be interpolated where possible. Lee and Lee’s dataset, for example, has only datapoints for every five years and so its gaps had to be filled. Benavot and Riddle’s dataset is only available for primary education, boys and girls combined, and was therefore only used for this specific series. Data from UNESCO is not available consistently, so if the original series for one country only contained five datapoints or less, the series was excluded.

If there was no data available in 1820, the data for each respective country was set to 0.001 at that time to ensure the interpolation would not go below zero.

To ensure the trendline would not be linear, the interpolation was performed with a logit transformation. First, to ensure values within a 100% range, the original enrolment values were divided by 100. The fractions $\pi$ were then transformed onto the logit scale.

$$y = \ln\left(\frac{\pi}{\pi - 1}\right)$$

These values were then interpolated linearly to complete the time span from 1820 to 2018 and transformed back to their original scale with the inverse logit to get the share or probability $p$.

$$p = \frac{e^y}{1 + e^y}$$

This ensures that the values are kept within range. The values were then multiplied by 100 to restore the original scale. This transformation results in a logistic growth curve of the data instead of a linear or growth curve. As demonstrated in Figures 2.1 to 2.6, the curve is also of a similar shape to previously estimated growth curves for enrolment trends.

The last but most crucial step of the analysis consisted of performing a Principal Component Analysis (PCA) and adding the resulting value to the mean of all values per indicator, i.e. country and year. Factor analyses are used to reduce multidimensional data to a ‘digestible’ format by projecting the covariance matrix of the involved variables onto the first few principal components while maximizing their variance. The results are one or more components representing the eigenvectors of the covariance matrix. This allows the combination of all the different values included and reduces them to one or two components representing the variance between the involved variables. Here, we only included the first resulting factor. This results in a correction of the row-wise values by a general factor to be added to the row-wise mean of all values to smooth out any irregularities and discrepancies of the empirical data. This is displayed in the following equation with $i = \text{country}$, $t = \text{time}$, and $x$ being the different variables included.

$$\text{enrolment trend}_{i,t} = PCA_i + \text{mean}(x_{1;i,t}, x_{2;i,t}, \ldots, x_{n;i,t})_{i,t}$$

If one of the previously collected trend values constantly overestimated the empirical values, this is corrected not only through the mean of the empirical values but also through the added factor. The resulting trendlines can be seen in Figure 1.
Despite this correction, it was impossible to prevent all abnormalities, since the actual data still contained a great number of irregularities. This can be seen in Figures 2 to 7 in the Appendix, which compare the interpolated values with the produced trendline. A total of six trends were estimated, whereby the trends for each gender were estimated both separately and once combined for both, primary and secondary education. The individual trends for each country disaggregated by gender as well as by education level (primary and secondary) can be found in the Appendix as well. Figures 8 to 13 show all six trends for all countries as well as their means and medians, showing the variance of trends between the countries. These Figures clearly show, that our trendline represents a relatively accurate approximate measure for the empirical data. They also show the variability of the produced trendline over the sample. To better understand this variability, the supplementary material available consists of individual trendlines per country. To test the accuracy of the produced trendline we correlated the produced trend with the empirical data, which resulted in correlations ranging from .83 to .99. Overall, this is a very high correlation, as expected.

Table 1 shows descriptive statistics for the six outcome trends for all years and countries combined. The shape of the distribution must be taken into account when inspecting these statistics. The growth curve is relatively slow at first, resulting in the relatively low mean values seen in Figure 1. This table also shows, that not all trends are available for all countries due to the lack of data in the included datasets. Figure 14 in the Appendix therefore shows the data availability for the outcome trendline.

Overall, this dataset provides a good measurement for enrollment trends over a large country sample for a long period of time. Due to the unique combination of existing data, this trendline offers a solution to include enrolment measures into historical country-level
analyses. Unlike other measurements we do not claim to provide absolute accurate data, but rather the most probably trend development in enrolment rates.

REFERENCES


Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary enrolment trend both sexes</td>
<td>192</td>
<td>0</td>
<td>28.2</td>
<td>40.228</td>
<td>38.424</td>
<td>102</td>
</tr>
<tr>
<td>Primary enrolment trend male</td>
<td>194</td>
<td>0</td>
<td>28.0</td>
<td>41.669</td>
<td>39.708</td>
<td>102</td>
</tr>
<tr>
<td>Primary enrolment trend female</td>
<td>194</td>
<td>0</td>
<td>21.4</td>
<td>39.146</td>
<td>39.880</td>
<td>102</td>
</tr>
<tr>
<td>Secondary enrolment trend both sexes</td>
<td>191</td>
<td>0</td>
<td>6.4</td>
<td>25.950</td>
<td>32.581</td>
<td>102</td>
</tr>
<tr>
<td>Secondary enrolment trend male</td>
<td>184</td>
<td>0</td>
<td>6.9</td>
<td>25.794</td>
<td>32.064</td>
<td>102</td>
</tr>
<tr>
<td>Secondary enrolment trend female</td>
<td>184</td>
<td>0</td>
<td>5.4</td>
<td>25.517</td>
<td>32.869</td>
<td>102</td>
</tr>
</tbody>
</table>

Figure 2. Primary enrolment trend female and interpolated empirical data
Figure 3. Primary enrolment trend male and interpolated empirical data

Interpolated enrolment measures
- Primary enrolment male generated
- Primary enrolment male Lee & Lee
- Primary net enrolment male UIS
- Primary adjusted net enrolment male UIS
- Primary total net enrolment male UIS

Figure 4. Primary enrolment trend both sexes and interpolated empirical data

Interpolated enrolment measures
- Primary enrolment both sexes generated
- Primary enrolment Lee & Lee
- Primary net enrolment UIS
- Primary adjusted net enrolment UIS
- Primary total net enrolment UIS
- Primary enrolment both sexes Benavot
Figure 5. Secondary enrolment trend female and interpolated empirical data

Figure 6. Secondary enrolment trend both sexes and interpolated empirical data
Figure 7. Secondary enrolment trend male and interpolated empirical data

Interpolated enrolment measures
- Secondary enrolment both sexes generated
- Secondary enrolment Lee & Lee
- Secondary net enrolment UIS
- Secondary adjusted net enrolment UIS
- Secondary total net enrolment UIS

Figure 8. Primary enrolment trend for females per country

Primary enrolment trend female for all countries
- mean
- median
Figure 9. Primary enrolment trend for males per country

Figure 10. Primary enrolment trend for both genders per country
Figure 11. Secondary enrolment trend for females per country

Figure 12. Secondary enrolment trend for males per country
Figure 13. Secondary enrolment trend for both genders per country
Figure 14. Data Availability
Figure 15. Trends by Country

Measure

- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female
Percentage of enrolled children

Measure
- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female

<table>
<thead>
<tr>
<th>Year</th>
<th>BGD</th>
<th>BGR</th>
<th>BHR</th>
<th>BHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>BLR</th>
<th>BLZ</th>
<th>BMU</th>
<th>BOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>BRA</th>
<th>BRB</th>
<th>BRN</th>
<th>BTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>BWA</th>
<th>CAF</th>
<th>CAN</th>
<th>CHE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The diagram shows the percentage of enrolled children for various countries over different years. The measures include:

- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female

Each country is represented by a different color, and the percentage is displayed on the y-axis, with years on the x-axis. The countries and their corresponding codes are:

- CHL
- CHN
- CIV
- CMR
- COG
- COK
- COL
- COM
- CPV
- CRI
- CUB
- CYM
- CYP
- CZE
- DEU
- DJI
The image shows graphs representing the percentage of enrolled children over years for various countries. Each graph is labeled with the country's name at the top and includes lines for different measures:

- **Primary both sexes**
- **Primary Female**
- **Primary Male**
- **Secondary both sexes**
- **Secondary Male**
- **Secondary Female**

The x-axis represents the years (1850 to 2000), and the y-axis shows the percentage of enrolled children.
Percentage of enrolled children over years for different countries and measures.

Measure:
- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female
Year

Percentage of enrolled children

MDV MEX MHL MKD

MLI MLT MMR MNE

MNG MOZ MRT MSR

MUS MWI MYS NAM

Measure
- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female

Global Dynamics of Social Policy
CEC 1342
Year

Measure
- Primary both sexes
- Primary Female
- Primary Male
- Secondary both sexes
- Secondary Male
- Secondary Female

Percentage of enrolled children

Country:
- VCT
- VEN
- VNM
- VUT
- WSM
- YEM
- ZAF
- ZMB
- ZWE