**Theoretical Framework**

I have decided to create an composite indicator to measure and compare education systems between countries.

From my research, I found the following:

* Higher spending per student is associated with better outcomes for student later in life, especially for those from low-income families. This gives access to school resources the potential to reduce intergenerational-poverty. (Jackson et al. 2015)
* Building on this, it has been found that a lack of qualified teachers is a sign of staffing problems (Ingersoll 2001), which leads to poorer student performance.
* Exposure to high staff turnover leads to worse educational outcomes by students, although the effects of this disruption can be reduced through proper management. (Gibbons et al. 2021)
* Gender equality within education is one of the most consistent factors in improving student outcomes. (Campbell 2021)

**Data Selection**

I used three datasets to construct this index:

* A dataset from Kaggle, available [here](#_https%3A%2F%2Fwww.kaggle.com%2Fcode%2Fn), which contained data on enrollment rates, out-of-school rates, completion rates, maths and reading scores, and literacy rates. All of these categories were separated between male and female.
* Education data from the World Bank, which contained education spending as a percentange of each countries’ GDP, split between primary, secondary and tertiary education and the expected years in education for males, females and the average.
* Data from the UNESCO Institute of Statistics, which contained data on the number of compulsary years of education, the number of free years of education, the teacher:pupil ratio of primary and secondary education, the percentage of teachers who are qualified, and the attrition rates for teachers for pre-primary, primary and secondary education.

**Imputation of Missing Data**

Every variable I selected, with the exception of compulsary schooling had at least one missing value. I addressed them as follows:

* For all variables gathered from the UNESCO Institute of Statistics, the World Bank and the Kaggle dataset, if a country had data for a particular variable for any year from 2011 to 2020, but not for 2021, I took an average of the years available instead, as while it would likely not be completely accurate, it would serve as good approximation.
* For the variables related to free schooling, I assumed that no data meant that the respective country had none. I felt this was a reasonable solution as less than 25 countries were missing this information and mostly occured with countries with no formal schooling.
* I decided to drop the variables related to teacher attrition, due to a lack of data. Less than 1/3 of courtries had this data available for any of the previos 10 years.
* For the ratio between qualified teachers to pupils and the percentage of current teachers who are currently qualified, I replaced missing values with the global average.

**Multivariate Analysis**

After carrying out multivariate analysis, some variables were very strongly corrolated with each other. To address this, the following changes were made:

* Principle Component Analysis was used to combine the six completion rate variables. This resulted in one component that explained 95% of the variance.
* Principle Component Analysis was used to combine the out-of-school rate variables. This resulted in three components that explained a combined 95.5% of the variance.
* The expected years of education for males and females were changed into one variable that represented the difference between the two.
* This was repeated for the literacy rate variables.
* The average of the math and reading scores was used instead of the two being separate.

**Normalisation**

**Weighting and Aggregation**

**Links to other Indicators**

**Visualisation of Data**

**References**

Campbell, J.A., (2021). The Moderating Effect of Gender Equality and Other Factors on PISA and Education Policy. Education Sciences [online], 11(1). Available from: <https://www.mdpi.com/2227-7102/11/1/10> [accessed 8th May 2025].

Gibbons, S., Scrutinio, V., Telhaj, S. (2021). *Teacher turnover: Effects, mechanisms and organisational responses* [online]. Available from: <https://www.sciencedirect.com/science/article/pii/S0927537121001147> [accessed 8th May 2025].

Ingeroll, R.M. (2001). *Teacher Turnover and Teacher Shortages: An Organizational Analysis* [online]. Available from: <https://www.gse.upenn.edu/pdf/rmi/TeacherTurnoverTeacherShortages-RMI-Fall-2001.pdf> [accessed 8th May 2025].

Jackson, C.K., Johnson, R.C., Persico, C. (2015). *The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms* [online]. Available from: <https://www.nber.org/system/files/working_papers/w20847/w20847.pdf> [accessed 8th May 2025].