

# Monitoring Trends in Female Education in Low and Middle Income Countries: Coherence across Data Sources

Kristen Jeffers  
Albert Esteve

## 1. Introduction

Education is widely recognized as a key instrument of individual agency and progress across various domains of social life. Along with investments in population health, education is viewed as the most important mechanism for development in low- and middle-income countries (LMICs). In places where gender gaps in schooling persist, women's education is considered a crucial vehicle of advancement. As a result, national and international policy agendas include educational expansion and the closing of the gender gaps in education as key objectives. Monitoring of such progress at an international scale and for LMICs requires comparable data across countries and over time. Fortunately, together with age and sex, educational attainment is easily the most frequently collected individual characteristic in surveys, censuses, and population registers worldwide. Similarities in the fundamental structure of education systems around the world facilitates comparisons across countries. Nevertheless, LMICs have limited statistical and administrative resources. They often rely on population censuses for the measurement of educational attainment but censuses are, in the best of cases, conducted every 10 years and access to the individual-level data required to construct comparable measures across time and place is restricted for many countries. As alternatives or complements to censuses, the policy and scholarly community turn to household surveys for more timely and accessible data. The Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS) are the largest scale comparative surveys administered in LMICs. Together with censuses, these sources should be the basis for monitoring trends in education across LMICs. It is not clear, however, how interchangeable estimates provided by these sources are because systematic analyses of comparability are unavailable. Precise coherence analysis is required to investigate consistency in the measurement of education outcomes across and within data sources.

This study responds to this challenge by examining, for the first time, coherence in estimates of educational attainment across the three main repositories of international microdata: IPUMS International, the Demographic and Health Surveys (DHS), and the Multiple Indicator Cluster Surveys (MICS). All censuses, DHS surveys, and MICS surveys include questions about school attendance and educational attainment. Microdata files available from these sources include the variables required to create comparable indicators of educational attainment. The paper examines trends in educational attainment among women age 25 to 29 in 75 countries between 1960 and 2017 during which we have hundreds of observations from multiple sources. This extended abstract justifies the need for an assessment of the coherence of empirical estimates of female educational attainment within and across data sources, describes the data and methodology used in our analysis, and presents preliminary results summarized in a series of macro-level figures. The final paper will include the complete set of results and a discussion of the implications of our main conclusions.

## 2. Background

Education is an important element of human development and economic growth. In addition to its impact on individual agency and social change, education improves the material well-being of individuals and societies. Education provides access to better paying jobs, alleviating poverty and increasing the availability of resources for households. More highly educated individuals tend to live longer, healthier lives than their less educated peers. The children of highly educated parents are less likely to die during the first years of life, are more likely to receive recommended immunizations, and to have better physical and cognitive outcomes. In aggregate, a healthy, highly educated workforce bolsters productivity and innovation, contributing to economic growth. Accordingly, policymakers and development practitioners view access to education as an important opportunity for development in low- and middle-income regions.

Censuses are the main source of data on educational attainment in LMICs. For many countries that lack comprehensive population registers or other administrative databases, census data represent the only source for information on the characteristics of the entire population. Census data serve as the underlying source for global databases on educational attainment like those maintained by the United Nations. Census data also provide the benchmark observations for modelled estimates and projections of educational attainment produced and used by the scholarly community (Barro and Lee 2013; Cohen and Soto 2007; Kc, Barakat, Goujon, Skirbekk, Sanderson, and Lutz 2010; De la Fuente and Doménech 2002; Jordá, and Alonso 2017). Compared to surveys, censuses have clear advantages. Censuses provide universal coverage of the population. They are not affected by sample design effects and usually provide reliable statistics, showing consistent trends over time. Certain

limitations, however, constrain the utility of census data for measuring education. Censuses are conducted infrequently. Most countries conduct a population and housing census every ten years. In low-resource and/or conflict settings, more than ten years may elapse between censuses.

In addition, access to census microdata is limited. Census results published by national statistical offices do not always include the indicators required for monitoring trends in educational expansion across countries. Access to individual-level census microdata is always preferred. Microdata are necessary for disaggregating population-level statistics and indicators by social and demographic characteristics. Because the educational categories identified in census summary tables vary across countries, individual-level microdata are also required to construct comparable measures of educational attainment across countries. Access to complete, original census microdata files is highly restricted in most countries. Fortunately, many countries make anonymized scientific-use samples of census microdata available to the research community. The IPUMS International project simplifies access and use of census microdata from around the world. In 2019, microdata from more than 300 censuses and 98 countries are available to researchers free of charge through IPUMS International. The impressive collection available from IPUMS International notwithstanding, for dozens of countries census microdata are entirely unavailable or too old to be useful for many purposes.

In response to these limitations, the policy and scholarly community use data from household surveys as substitutes for or complements to census data. Data from two global survey programs are frequently used to monitor development progress in LMICs: the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). DHS surveys are nationally representative household surveys that collect data on a number of topics related to population, health, and nutrition. Since 1984, more than 400 DHS surveys have been carried out in over 90 countries in Africa, Latin America, Asia, Oceania, and Eastern Europe. Data on literacy, school attendance, and educational attainment are collected for all household members in every DHS survey. MICS surveys are UNICEF-supported surveys that provide data on the wellbeing of women and children in LMICs. More than 300 surveys have been fielded in 116 countries covering the period 1993 to the present. Most MICS surveys are also nationally representative and collect education information similar to that collected in DHS surveys.

Because data availability varies across countries, assessments of global and regional trends in education rely on data from multiple sources. In theory, comparing educational attainment across countries and data sources should be straightforward. Most education systems are organized around three distinct levels: primary, secondary, and tertiary. Systems within regions and among countries with shared colonial histories are usually quite similar. Moreover, the International Standard Classification of Education (ISCED) provides clear guidelines for measuring educational attainment in censuses and surveys. In practice, accurate comparisons across countries are difficult because the definition of educational levels varies significantly across countries. We presume that scholars considering education at the global level have grappled with comparability across data sources, but none have systematically evaluated these issues across the main sources of population data and made their findings known to the research community.

### 3. Data and Methodology

We use individual-level data from 210 IPUMS census samples, 219 DHS surveys, and 106 MICS surveys to assess coherence in the measurement of education across data sources in 75 LMICs. We include in the analysis all countries with four or more censuses and/or surveys across two or more data sources. The data were collected during the period 1960 to 2017 and cover more than half of the 138 countries designated as low or middle income by the World Bank. A plurality of countries studied are represented in all three data sources: IPUMS, DHS, and MICS (table 1).

**Table 1: Number of analysis countries by data source and world region**

	Africa	Asia	Europe	Latin America / Caribbean	Total
DHS, MICS, IPUMS	17	5	1	3	26
IPUMS, DHS	7	8	–	7	22
IPUMS, MICS	2	4	1	9	16
DHS, MICS	8	2	–	1	11
<b>TOTAL</b>	<b>34</b>	<b>19</b>	<b>2</b>	<b>20</b>	<b>75</b>

From these census and survey samples, we select women age 25 to 29. In many low- and middle-income countries, women trail men in access to and completion of secondary education. Develop agendas focus on confronting barriers to girls' schooling and require accurate information on female educational attainment. Women in this age range represent the population that recently completed formal education, more accurately

reflecting the educational context in the country at the time of the survey compared to older cohorts. Coherence analysis based on men's results yield similar results.

Our analysis focuses on the population completing secondary and higher education. Relatively few people complete secondary or higher education in many low-income countries, and estimates of high educational attainment are based on small subsamples. Furthermore, existing literature suggests secondary education is an important threshold for individual outcomes and development (Grépin and Bharadwaj 2015; Makoka and Masibo 2015; Raghupathy 1996). We generate a dichotomous educational attainment variable that distinguishes women who have completed secondary education or higher from those who have not. For census samples, we recode the IPUMS EDATTAIN variable. DHS survey data were accessed directly from the DHS Program and from IPUMS DHS. For DHS surveys, we use the educational attainment summary variable constructed by the DHS Program during data processing (hv109 or EDSUMM in IPUMS). The MICS datasets do not include a summary educational attainment variable. We construct a summary measure of educational attainment for MICS samples according to each country's educational system using variables indicating the highest education level attended and highest grade completed at level.

We calculate the percentage of women age 25 to 29 who have completed secondary education for each census or survey, yielding 535 country-year observations in total. The mean number of census and survey observations among countries examined is 7.7. We use polynomial regression to examine trends over time in educational attainment within each country. As visualized in figure 1, the relationship between time and educational attainment of the population for the study period is not necessarily linear. Secondary or higher completion has expanded more rapidly in recent decades. By country, we specify a second-degree polynomial (quadratic) regression model to account for the parabolic nature of the relationship, where  $P$  is the proportion of women age 25-29 completing secondary or higher education.

$$\ln(P) = \beta_0 + \beta_1 Year^2 + \beta_2 Year + \varepsilon$$

Each observation of  $P$  is weighted by sample size. By doing this, we give priority to census samples over DHS and MICS surveys. Census microdata samples distributed by IPUMS are systematically drawn from complete-count census databases. Results derived from these census samples are therefore extremely close to those derived from complete-count databases. We use the coefficients from each country-specific regression model to generate predicted values of  $P$ . We then compare observed and predicted values, calculating and summarizing absolute differences between the two and classifying observations based on absolute deviance from the predicted trend.

#### 4. Preliminary Results

Over the last several decades, in all world regions, there has been a general upward trend in the proportion of young women completing secondary education or higher in LMICs. Figure 1 presents by country and world region estimates of the percent of women age 25 to 29 completing upper secondary education or higher for the 75 countries included in our analysis. Each line represents a country, showing change over time based on multiple data sources. The percent of women age 25 to 29 completing secondary education or higher has increased between the earliest census or survey observation and the most recent in nearly every country examined. While the global picture presented in figure 1 shows a generally increasing trend, levels of secondary or higher completion as well as the rate and linearity of change over time vary widely across countries and regions.

Educational attainment is lowest in sub-Saharan Africa. According to surveys conducted since 2010, less than 10 percent of women age 25 to 29 have completed secondary or higher in Burkina Faso, Burundi, Benin, Central African Republic, Chad, Ethiopia, Mali, Mozambique, Niger, Senegal, and Togo. Progress in secondary completion varies across countries within Latin America. At least 50 percent of women age 25 to 29 had completed secondary or higher in 11 of the 17 Latin American countries studied with census or survey observations since 2010. Levels of completion were highest in Cuba (2014) and Trinidad and Tobago (2011), where more than 80 percent of young women complete secondary or higher, and lowest in Haiti (2016) and Guatemala (2014), where less than 25 percent of young women complete secondary or higher. Recent levels of secondary completion vary within Asia as well. Female educational attainment remains low in several of the Southern and Southeast Asian countries studied according to recent surveys. In contrast, educational attainment is high in Central Asia and Eastern Europe: more than 95 percent of women age 25 to 29 complete secondary education or higher in Armenia, Belarus, Kazakhstan, Kyrgyz Republic, and Ukraine according to recent surveys.

The rate of increase in secondary completion in LMICs during the last several decades also varies across countries and regions, as reflected in the variation in the slope of country trend lines in figure 1. Change has been slowest in Africa, where the median annual increase in the percentage of women age 25 to 29 completing secondary or higher among the countries examined is 0.43 percentage points. By comparison, the median annual increase in the percentage of women age 25 to 29 completing secondary or higher is 1.1 percentage points among Latin American countries studied. Trends in educational attainment among young women presented in

**Figure 1: Women age 25-29 completing secondary education or higher (%)**

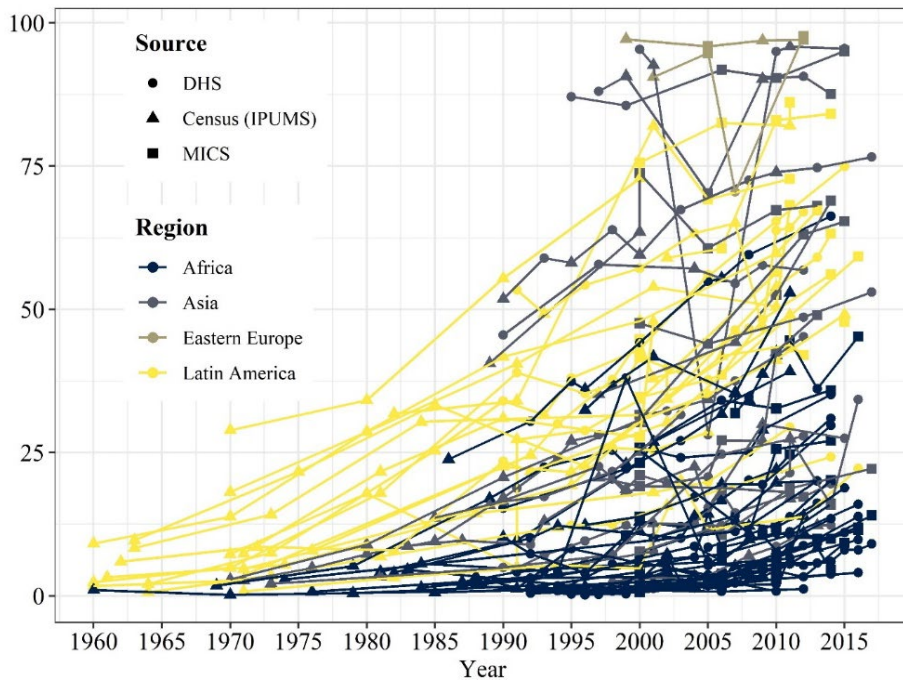
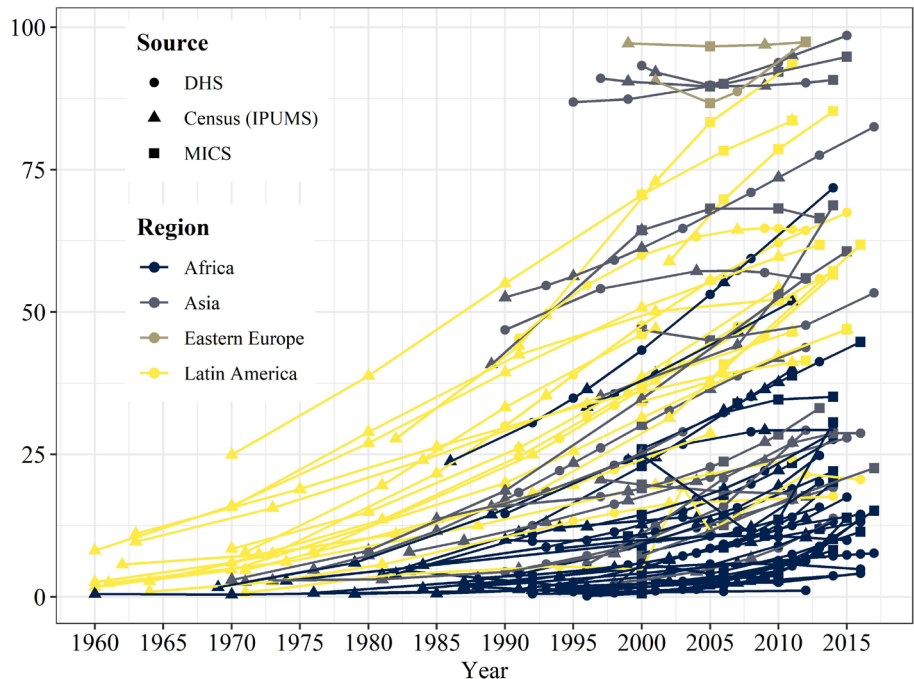


figure 1 are not always linear or monotonic. We do not expect that educational expansion is linear in all countries. In some countries, secondary completion has expanded more rapidly in recent decades, reflected in a J-shaped curve for many countries. In other countries, educational attainment increased rapidly during the MDG-era but has slowed in recent years as countries reach secondary "carrying capacity". This trend is reflected in an inverse J-shaped curve. We do, however, expect that in most countries educational attainment among young women will always be non-decreasing. As visible in figure 1, this is not the case. In 50 of 75 countries examined, at least one point estimate of secondary completion among women age 25

to 29 is lower than the previous point estimate. In 11 countries, we observe a decline of 10 percentage points or more between adjacent point estimates.

**Figure 2: Women age 25-29 completing secondary education or higher (%), predicted values**



To identify which observations diverge from overall country-level trends and to what extent, we use regression coefficients to produce predicted values of the percentage of females age 25 to 29 completing secondary education or higher that align with expected trends (figure 2). The final objective of our analysis is to compare observed and predicted values of educational attainment among young women and assess coherence in its measurement across data sources. Overall, coherence across data sources is strong: the mean absolute difference between observed and predicted values is 2.2. Of 535 census and survey observations, 67.4 percent deviate from expected trends by less than two percentage points. Only 11 percent of observations deviate from expected trends by five or more percentage points. Coherence across data sources is poorest for countries in Latin America. Likewise, MICS-based estimates deviate most from expected trends. The final paper will explore these and other findings.

## References

- Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics*, *104*, 184-198.
- Boyle, E.H., King, M., & Sobek, M. IPUMS-Demographic and Health Surveys: Version 7 [dataset]. Minnesota Population Center and ICF International, 2019. <https://doi.org/10.18128/D080.V7>
- Cohen, D., & Soto, M. (2007). Growth and human capital: good data, good results. *Journal of economic growth*, *12*(1), 51-76.
- De la Fuente, A. & Domenech, R. (2013). Cross-country data on the quantity of schooling: a selective survey and some quality measures. *BBVA working papers*, (13/27).
- Grépin, K. A., & Bharadwaj, P. (2015). Maternal education and child mortality in Zimbabwe. *Journal of health economics*, *44*, 97-117.
- ICF. 2004-2017. Demographic and Health Surveys (various) [Datasets]. Funded by USAID. Rockville, Maryland: ICF [Distributor].
- Jordá, V., & Alonso, J. M. (2017). New estimates on educational attainment using a continuous approach (1970–2010). *World Development*, *90*, 281-293.
- Kc, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W. C., & Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic research*, *22*(15), 383-472.
- Makoka, D., Masibo, P.K. (2015). Is there a threshold level of maternal education sufficient to reduce child undernutrition? Evidence from Malawi, Tanzania and Zimbabwe. *BMC Pediatr* **15**, 96 doi:10.1186/s12887-015-0406-8
- Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. <https://doi.org/10.18128/D020.V7.2>
- Raghupathy, S. (1996). Education and the use of maternal health care in Thailand. *Social science & medicine*, *43*(4), 459-471.
- UNICEF. Multiple Cluster Indicator Surveys (various) [Datasets]. Available: <http://mics.unicef.org/surveys>.