The Global Longitudinal University Enrollment Dataset (GLUED)

Elizabeth Buckner

Abstract

Existing cross-national data on higher education participation has many limitations: It is collected primarily at the national level, combines short-cycle tertiary education with university enrollments, and lacks disaggregated data on the private sector before 1998. The Global Longitudinal University Enrollment Dataset (GLUED) seeks to address these limitations. The dataset compiles and estimates institution-level enrollment data on universities worldwide from 1950 to 2020. This article discusses the advantages that GLUED offers researchers of higher education.

How large was the private university (i.e., ISCED 6+) sector globally in the 1970s? And has the recent growth of private higher education occurred primarily in new private universities or through expanding enrollments in existing institutions? These are just some of the questions that scholars of private higher education are raising to better understand the growth of private higher education worldwide and its consequences for higher education landscapes, policies, and students.

However, until now, it has been difficult for scholars of higher education to put recent growth in perspective relative to earlier eras. This is due to a lack of data on private higher education before the year 2000. This article discusses the many problems with data on university enrollments and introduces a new dataset that seeks to advance the study of universities cross-nationally, over time.

The Lack of Data on Higher Education Enrollments by Sector

Existing data for studying private higher education cross-nationally has many limitations. UNESCO's Institute for Statistics (UIS) began collecting data on the private share of tertiary enrollments worldwide in 1998. However, in 1998, data was only available for 14 countries. By 2005, data on the private share of tertiary enrollments was available for 72 countries. UIS data peaks in 2015, at which point an estimate is available for 127 countries, after which data becomes scarcer again, most likely due to natural lags in reporting. By 2020, estimates are only available for 54 countries. Another issue with UIS data is that it depends on national or federal government reporting, which means that there are some countries for which data is never available, including my own country, Canada.

Moreover, UNESCO definitions for "public" and "private" do not always correspond with national definitions or popular understandings of public and private higher education. The consequences of this are apparent for countries such as the United Kingdom, where private universities are listed as comprising 100 percent of all enrollments in all years.

As a result, scholars interested in more accurate estimates of private higher education that better align to popular understandings and government policy turn to data sources such as the Program on Private Higher Education (PROPHE). The PROPHE website offers detailed country estimates based on UIS, as well as government sources and country-level experts in cases when government data is not available. However, one limitation with PROPHE is that, as of now, data is only available publicly for 2010, which makes it difficult to track long-term trends.

A third limitation of existing datasets is that available data on private higher education combines short-cycle tertiary education (i.e., ISCED 5) with universities (i.e., ISCED 6+), meaning institutions that offer bachelor's and/or advanced degrees. Yet, decades of scholarship in higher education has documented the distinctive characteristics of universities, which are much more likely to be research oriented, participate in global rankings, and which tend to enjoy more academic freedom. For those of us interested specifically in universities as a social institution, the conflation of short-cycle tertiary institutions with universities is a major concern.

The issues related to private higher education are shared by researchers of higher education more generally: Most of the publicly available data on higher education is calculated at the national level, combines short-cycle enrollments with university enrollments, and has only been widely available in recent years.

Introducing GLUED–Beta Version

Clearly, there is a need for better data on sector-specific enrollments in universities over time. The Global Longitudinal University Enrollment Dataset (GLUED) is a new dataset that seeks to do this. The three main advantages of GLUED are that it represents a near-census of universities (ISCED 6+) worldwide; that data is collected at the institutional level, which will allow researchers to investigate organizational-level phenomena; and that for many institutions, raw and estimated enrollment data is available starting in 1950.

Constructing GLUED has taken many years. The data was collected at the Ontario Institute for Studies in Education (OISE) at the University of Toronto and generously funded by a grant from the Social Science and Humanities Research Council (SSHRC) of Canada. Between 2018 and 2020, a team of research assistants manually entered enrollment data on thousands of universities from hard-copy volumes of both the Europa World of Learning and the International Association of Universities (IAU)'s World Higher Education Database. This data was supplemented with data from the Integrated Postsecondary Education Data System (IPEDS) for the United States, recent enrollment data from the digital version of the World Higher Education Database hosted by IAU, and student enrollment data scraped from Wikipedia. All data was extensively checked and cleaned. Missing data for prior years was estimated based on institution-specific growth rates or institutional characteristics, and calibrated to align to published sources at the global and country levels.

GLUED specifies the name and institutional characteristics, including estimated student enrollment data at five-year intervals, of 15,263 institutions in 185 countries and territories.

GLUED specifies the name and institutional characteristics, including estimated student enrollment data at five-year intervals, of 15,263 institutions in 185 countries and territories. A second country-level data, with sector-specific enrollment data totaled to the country level, is also available.

GLUED also includes a number of other useful variables on institutional characteristics, merged from various sources, including sector (i.e., public/private), founding year, and whether the institution is PhD granting or not. The university sector is self-declared—meaning whether it is public or private—in the original sources by the institution, based on institutional control or ownership, and GLUED adopts that definition. The dataset does not distinguish between for-profit or nonprofit universities, nor does it capture whether a university is religious or not, but this is something that future versions of the dataset may be able to add.

Finally, one of the exciting features of GLUED is that it also includes the geographic location (i.e., GPS coordinates) of each university, scraped from Google Maps. This data can be used to map universities geographically in the dataset and visualize the growth of universities worldwide, and may be of use to researchers for a wide range of purposes.

The dataset is currently in its beta version, during which dataset cleaning and missing value estimation are finalized. The dataset will be published in early fall 2022 on the University of Toronto's Dataverse, and will be accessible through Borealis, the Canadian Dataverse Repository, at the following permalink: https://doi.org/10.5683/SP3/P0D1KE.

Elizabeth Buckner is an assistant professor of higher education at the Ontario Institute for Studies in Education (OISE), University of Toronto, Canada. Email: elizabeth.buckner@utoronto.ca.