# Applying statistical databases on educational research

DINARA ALIMKHANOVA, ANDREW G. DRYBROUGH

**Introduction**

This review presents information on international and national statistical databases that can be used for the purpose of educational research. Each database is derived from the websites and the authentic website source is provided through a hyperlink. These hyperlinks were accessible and working at the time of writing. Although most of these websites contain a range of types of data on different subject areas (e.g., economic and demographic statistics) each website has at least one section of data or research relating to education, or that might be of interest to the educational researcher. There are also some sites that may be of more specific interest to those involves in educational assessment.

The analysis of each databases will cover two main aspects. First, there will be a short description of the statistical database. This is then followed by a brief analysis of that database that identifies the benefits and/or limitations of it. We begin by describing and analyzing the international education databases, followed by national educational databases of Kazakhstan. We have found these of use in our research and teaching, and hope that they can be of benefit to anyone who is interested in international educational research and educational research in Kazakhstan and the wider Central Asian region.

**International educational statistical databases**

The international data and links to websites can be divided up into three broad categories. First, there are those that are based in international organizations such as the World Bank, UNESCO and OECD. Second, there are those that are more regional in nature, focusing in data from the CIS and EU. Finally, there are those based on the research of various research institutions and organizations that may be global or regional in their scope.

**OECD stat**

OECD.Stat includes data and metadata for OECD countries and selected non-member economies. Particularly, it provides a separate section for the indicators of "Education at a Glance". There are a number of benefits of the OECD stat to educational researchers. To begin with, it is possible to customize the data to the re-searcher's needs. Secondly, there is an information bar for each indicator so it is quite easy to navigate and find specific data. The "Education and Training" tab will most likely be of interest to the education researcher. There is also data on OECD average as well as data for non-OECD nations such as Brazil and India. However there is no specific educational data on Kazakhstan or the Central Asia countries.

**UNESCO Stat**

The United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics (UIS) is the official and trusted source of internationally comparable data on education, science, culture, and communication. Unlike the OECD, it covers most countries in the world. It is also easy to customize your search to focus in on different aspects of education by country, is user friendly, and is benchmarked to the Sustainable Development Goals (SDGs). Some of the data is quite dated though, and there is a lack of data for some countries.

**World Bank data**

The World Bank data provides a listing of available World Bank datasets, including databases, pre-formatted tables, reports, and other resources. It can be customized to your requirements, and has data sets at the world and country level. However, similar to the UNESCO statistical data sets, it may consist of some data that is quite old.

**Interstate Statistical Committee of the Commonwealth of Independent States (CIS)**

The Interstate Statistical Committee of the CIS is the interstate body acting in the framework of the Commonwealth of Independent States for coordinating the activities of national statistical services including educational. Like the previous international data sets, it is relatively easy to customize to your search requirements to specific areas of education, it has comparative data on the CIS countries and it is in English. There are also annual reports, but it is restricted to CIS members only, so the scope of comparison for the same statistical measurement is limited.

## Eurostat

Eurostat is the statistical office of the European Union (EU) whose task is to provide the EU with statistics at a European level that enables comparisons between countries and regions. Your search for specific data can be customized at a country, city or regional level and includes EU averages. It is also very comprehensive, but it only applies to EU members and may consist of some older data (around 2015).

## International Labour Organization

The ILO Department of Statistics is the focal point of the United Nations (UN) on labor statistics. The benefits of this database are that it has a clear information bar, has data at the world level, sub-regional level (e.g., Eastern and Central Asia) and the country level. You should refer to the ILO data explorer:
https://www.ilo.org/shinyapps/bulkexplorer25/?lang=en
Some of the data is quite old though. For any unclear questions you may refer to the ILO library which is quite responsive.

## World values survey (WVS)

The World Values Survey (WVS) is an international research program devoted to the scientific and academic study in the world. The project's goal is to assess which values are stable ad which change over time. It focuses in on the social, political, economic, religious and cultural values of people. It is very up-to-date with the most recent wave of research being from 2017-2022. It is based on a questionnaire sampling approach, but there is limited random sampling, and not all the questions might be applicable to a certain countries.

## Central Asia Barometer

The Central Asia Barometer (CAB) is a regional, independent, non-profit institution for applied social research and analytics to measure the social, economic, and political "atmosphere" in the countries of Central Asia. It is based on a questionnaire sample and its most recent wave is 2020. Its database is in Excel format so is easily transferable to SPSS. However, the questions relating to education are limited.

## Oxus Society

The Oxus Society is a group of Central Asian scholars headquartered in Washington, DC, but part of a global network of experts located around the world. The following two sources at Oxus website may be helpful in social science research. The Central Asia Migration Tracker (CAMT) visualizes the flow of migration from Central Asia from 2016 to 2020, while The Central Asia Protest Tracker tracks and categorizes protests in Central Asia in reports, datasets, and visualizations.

## Sustainable Development Goals (UN) Tracker

The SDG Tracker presents data across all available indicators using official statistics from the UN and other international organizations to track global progress towards the SDGs. The presentation of this website is very clear and easy to navigate as it designed around the 17 SDGs. SDG4 on quality education is likely to be of most interest to educational researchers, but other SDGs such as good health and well-being (SDG3) and gender equality (SDG5) may also be of interest. It includes chart, map, table, source and downloadable data.

## Asian development bank data

The ADB Data Library is a platform that hosts publicly available data from the Asian Development Bank. The Data Library contains macro-economic and social data, financial and corporate data, project-level data, dashboards, and data stories. It provides data at the world level, sub-regional level (e.g., Central Asia) and country level, but some of the data may be a bit old (from around 2015).

## Discuss data

Discuss Data is a new open repository for storing, sharing and discussing research data on Eastern Europe, South Caucasus and Central Asia. It is a new dataset established in 2020 and it includes a range of data sets based around different research projects. This means you will have to browse through the different research topics to find what you need.

**International educational assessment**
Statistical data is also available on a number of international assessments. The OECD also holds the Programme for International Student Assessment (PISA) if you are interested in that specific area of educational research. These are quite extensive data, and you can search for quite specific information on different counties in the OECD. There are also separate reports for the non-OECD countries, including for Kazakhstan. However, it is not always possible to compare data longitudinally for every country as more countries have participated in the test more recently and there is no comparable data on these from previous years when they did not participate in the PISA.

The International English Language Testing System (IELTS) also produces some basic international demographic and test taker performance statistics. These are a lot less extensive compare to the PISA databases, but they provide demographic and performance data on the four English language skills for the two IELTS test (academic and general training) across all the countries that take part in it. However, this data should not be considered to be fully representative of the actual English language proficiency of people of those countries as it is only based on those who did the test.

**National educational statistical databases**
In additional to the wealth of international data available, there are a number of more detailed data sets on different areas of Kazakhstan that may be of interest to the educational researcher. They provide a lot more detailed statistical data for Kazakhstan as a whole, but also for specific regions and cities. These can also be divided up into two basic types. The first are official government sources, and the second are from research institutions and organizations. Most of these area quite general portals, but within each website you can find quite specific educational data. Most also have data presented in English.

**The Bureau of National Statistics**
The Bureau of National Statistics coordinates, manages and implements activities pertaining to the field of state statistics. The benefits of this data set are that its publications are by region and fields and you can request explanations from the Bureau if necessary, but not everything is translated into English.

**Finance center**
The finance center is an institution that ensures the availability of secondary, technical and professional, post-secondary, higher and postgraduate education to citizens of the Republic of Kazakhstan through the development of a system of public-private partnerships in the field of education. Mainly you can refer to the website when conducting document analysis in the field of education financing and if you are interested in examining tuition fees across universities in Kazakhstan.

**'Taldau' online dashboard**
Taldau is information and analytical dashboard of the Bureau of National Statistics. It provides extensive statistical data on Kazakhstan, including a section on education related data on Kazakhstan, much of which is in English. It is easy to navigate and focus on the areas of interest. However, some educational statistics only cover the period 2001-2013.

**National Center for higher education development**
The source provides research on methodological and information analysis support of the implementation of the Bologna process parameters to the higher and postgraduate education system of the Republic of Kazakhstan. You may find data on higher education field relating to the Bologna process and related documents, data on student exchange, international students, Kazakhstani students studying abroad, faculty & staff exchange, international faculty in Kazakhstan  (by country, by universities, sources of funding), and on trilingual education data in higher education.

**Conclusion**

We hope that this brief review of different websites and data bases can provide opportunities for the educational researchers to delve into the large amounts of rich data that are already available to the general public. These include both international and comparative data and more localized Kazakhstani and Central Asian that can be used as part of the background and context of a piece of research, or analyzed in more depth with use of different statistical techniques of data analysis.

As with all secondary sources of information, it is also important to consider who produced the data, how it was collected and for what purpose, as well as consider the currency, credibility, relevance, authority and ac-curacy of them. Inevitably, as an educational researcher you will also use the data for your own purposes, and it is important to consider what that is (e.g., whether to provide background information, or to support a specific argument), and use the data appropriately and ethically.