Module 5: Data Analysis and Results Communication

Module overview – objectives, topics and learning outcomes

Many countries in the region have data on different aspects of education collected through annual school censuses, periodic censuses and household surveys.

Collected data has little relevance until it is transformed into meaningful information and converted into a body of knowledge on particular issues. One of the challenges countries in the Asia-Pacific region are facing is that data is not processed appropriately so it can be effectively used in decision-making.

There is also the issue of not disseminating data to wider audiences, i.e. parents, to inform them about and involve them in progress and the required assistance in education. On top of this, all data requires a thorough disaggregation to identify populations that have a tendency to get left behind.

The dissemination of comprehensive, reliable and timely data is also viewed as an issue of transparency which is crucial for building trust in education systems and this must be undertaken in a way that is credible, both inside the country and internationally.

The modern education statistical system should assume the role of an information

provider and knowledge manager with data going in and out, rather than the traditional role of merely collecting data and keeping ownership of it.

This module consists of two main parts: one focusing on data analysis; and one focusing on data communication.

The data analysis section will present the essentials for understanding education data processing. The data communication section will encapsulate a range of communication products and what to look out for when creating them.

The following topics are covered in this module:

- Purpose of data analyses and their types;
- What constitutes a good indicator and how to distinguish between different types of indicators;
- The basic steps in data analysis;
- Choosing the right data presentation method for strategic communication with various audiences.

After completing the module, learners will have acquired the following learning outcomes:

- An ability to comprehend data analysis requirements;
- Knowledge of how to create and select appropriate education indicators;
- An understanding of how to choose the right data presentation method for strategic communication.

1 The Purpose of Data Analysis

Data analysis is the process of systematically applying statistical, or logical techniques to describe and illustrate, condense, recap and evaluate data. Analysis is the procedure to make broad generalizations by identifying trends and situations (phenomena) among present information.

In principle, we want to achieve an effective monitoring and evaluation of the education sector. We know that we have a wealth of data at our hands coming from education management information systems and household surveys. This data now must feed back into the planning of the education sector to serve as our evidence to base planning on. In short, the purpose of data analysis in education serves to:

1. Assess progress in the achievement of national and global targets;
2. Identify remaining shortcomings with regard to access, quality and equity;
3. Review policies, strategies, actions and success stories to identify bottlenecks;
4. Use the results to formulate better policies and plans to achieve aspired outcomes.

With the 21st century being characterized as the 'information age', we often have too much information available that can obfuscate its meaning and the ability to make a clear decision.

Due to the lack of analytical and technological capacities, the available information, especially raw data, is often not transformed into meaningful information to serve decision-making processes positively. Data analysis, therefore, functions as a production of meaningful information to make evidence-based decisions.

Meaningful information as a result of proper data analysis allows for:

- Tracking progress towards a set target;
- Effective decision-making for planning and managing at the national, sub-national and local levels;
- Predicting sector trends and performances.

Depending on the type of data we have at hand, it is important to decide between two different types of data analysis as each generate different information:

- Quantitative data analysis;
- Qualitative data analysis.

1.1 Quantitative and qualitative data analysis

Quantitative analysis

Quantitative data analysis can be understood as explaining situations by means of numerical data. In quantitative analysis, we collect numerical data and analyse it using mathematic methods (in particular statistics). In order to be able to use mathematic methods, our data has to be in a numerical form[1].

As quantitative analysis is about collecting numerical data, the following four specific phenomena are best suited to analyse quantitatively:

a. Questions that demand a quantitative answer, such as: 'How many students choose to study social science at higher education?' Or: 'How many maths teachers do we need? Or: How many have we got in our school/district?'

b. Comparisons of numerical values, for example change prior, during, or past a time period, or numerical characteristics of individuals/social groups, such as: 'Are the numbers of students in our university rising or falling?' 'Is learning achievement going up or down?'

c. Understanding the state of something, or other or to identify factors for the situation or the change, e.g., factors which predict the recruitment of maths teachers. What factors are related to changes in student achievement over time?

d. The study which needs testing of hypotheses – e.g. whether there is a relationship between a pupil's achievement and their self-esteem and social background. By looking at the theory, a possible hypothesis to test, would be that a lower social class background leads to low self-esteem, which would in turn be related to low achievement. Quantitative analysis can test this kind of model.

The essence of quantitative analysis is to confirm an assumption, or hypothesis by identifying patterns among a larger sample from a population and this approach is useful in policymaking and planning.

Qualitative Analysis

There are many definitions of qualitative analysis as there are books on the subject. The key word to differentiate qualitative analysis from quantitative analysis is 'exploration'. Qualitative analysers are interested in exploring an observed phenomenon to understand the meaning that people have constructed. That is, how people make sense of their world and the experiences they have in the world.

[1] For more on quantitative analysis, read: Aliaga, M. & Gunderson, B., 2002: Interactive statistics, New Jersey, Prentice Hall.

As Johnson and Christensen (2004) state, qualitative analysis involves working with data that is non-numerical in nature and does not indicate an order, hierarchy or rank[2]. Symbology is an example. Social scientists apply a form of observing and interpretive sociology; that means, to adopt a point of view from the individual's perspective to understand beliefs, values or behaviours by means of participant observation, or case studies, which result in a narrative, descriptive account of a setting, or practice.

In other words, qualitative analysis involves analysing written and spoken language, audio-visual imagery, or in short symbolism, as perceived by the subject of study[3]. Numbers can be analysed qualitatively, if they represent symbolism.

Qualitative analysis studies are mainly useful in:

- Identifying and exploring individual issues or cases;
- Describing complex processes;
- Explaining processes, such as motives, values, and causes of behaviour.

1.2 Qualitative versus quantitative analysis purposes exemplified

The UIS and UNICEF Global Initiative on Out-of-School Children (OOSC) serves as a good example in portraying the use of **both** research methods in analysis of out-of-school children[4]. The evaluation of the initiative studied the problem of OOSC to quantitatively capture who is excluded, where, and how many, as well as to qualitatively explore the barriers and causes of exclusion.

Quantitative example

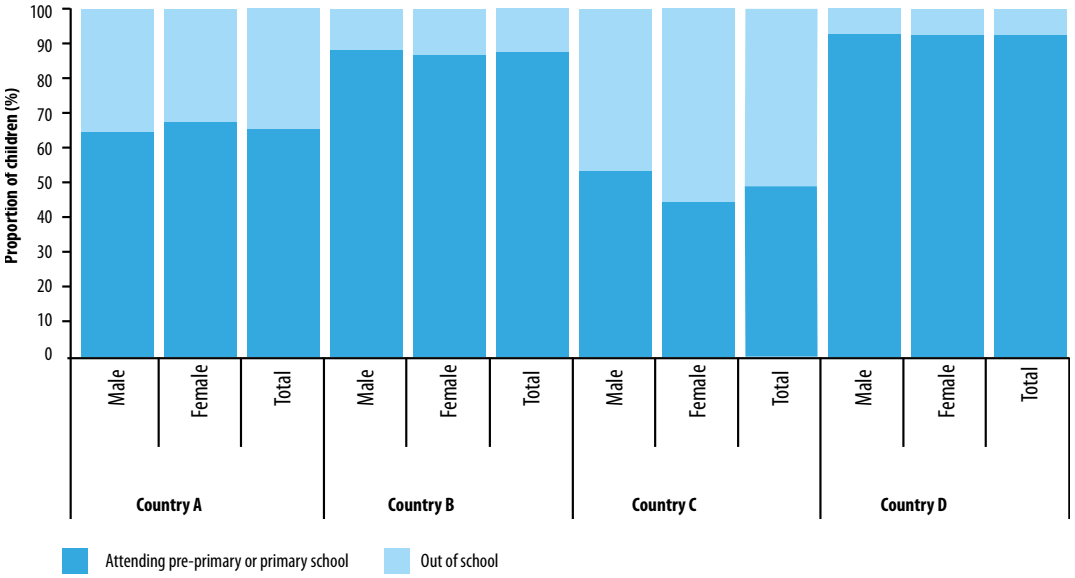
The numbers on the out-of-school population can be collected and analysed to describe the situation at the national and sub-national levels, disaggregated by gender, location, wealth, ethnicity, language, disability and other factors. As seen in Figure 1 below, we can identify the distribution of OOSC by country and gender. We can recognize that differences exist from country to country in terms of the total out-of-school population size, as well as a greater difference by gender in Country C. While, with this example, we can identify the size of a population and take first measures to target the specific population, the question remains why these populations are out-of-school in the first place.

[2] Johnson, R. B. and Christensen, L. B., 2004: Educational Research: Quantitative, Qualitative, and Mixed Approaches. Boston, MA: Allyn and Bacon.

[3] For more on qualitative research, see: Elwell, Frank W., 2009: Macrosociology: The Study of Sociocultural Systems (Chapter 8 on 'Verstehen: The Sociology of Max Weber'), Edwin Mellen Pr; Nkwi P, Nyamongo I, Ryan G. 2001: Field Research into Social Issues: Methodological Guidelines. Washington, DC, UNESCO.

[4] For more practical guidance on the monitoring and evaluation of OOSC, see the operational manual created by UIS and UNICEF, access: http://allinschool.org/wp-content/uploads/2015/12/F_UNICEF1017_OOSCI_manual-web.pdf

Figure 1: Example on the proportion of pre-primary age children in pre-primary or primary school, by sex



Qualitative example

A qualitative report on the situation of the children who are not in school presents us with an example on how we can identify underlying causes. Interviews and focus group discussions took place to collect views and the perceptions of the different stakeholders in schools to gather in-depth data on why children were out-of-school; what they were doing; and to take into account OOSC's family background. These stakeholders were comprised of the children themselves, their parents, the school heads and teachers.

Let's read the following excerpts.

Figure 2: Qualitative study excerpts on the situation of OOSC

3.2.1.4 Lack of safety and security in school

School children interviewed in the study area revealed that bullying, name calling and corporal punishments, made the school environment unsafe. The majority of interviewed OOSC reported that corporal punishment was a contributing factor to their failure to attend school. Other school children interviewed also reported
that excessive corporal punishment made the school environment scary to the extent that some children left school. However, in focus group discussions, village leaders had divided views on the use of corporal punishment in schools, with some arguing that corporal punishment was suitable for naughty children, while others argued against corporal punishment, asserting that excessive application led to poor school attendance. The majority of the village leaders reported the need for more controlled corporal punishment and alternative means to correct children.

3.3.1.1 Approved school contributions

3.3.1.1.1 Primary school

The study revealed that education regulations do not impose any contributions on the parents of primary school children because primary education is entirely free. However, school committees approved some contributions to support running of school activities. Head teachers reported that some of the contributions were paid in cash while others were covered in kind or in labour.

They further reported that contributions included cash for processing and printing midterm and terminal tests, speed tests, preparing and administering mock examinations for Standard IV and VII children, school badges, desks, and school meals. Very few schools reported charging cash for school uniforms, but parents were free to get uniforms elsewhere depending on the unit cost. All the contributions in cash were paid directly through school bank accounts. Teachers interviewed were able to explain cost per parent in school, but it was hard to make generalisations as the contributions appeared to vary from one school to another and from one region to another.

Source: UNICEF and UIS, 2018: Tanzania Qualitative Study Report. Global Initiative on Out-of-School Children.

From a qualitative analysis of the situation, based on experiences and interpretations, by the stakeholders involved, it was revealed that corporal punishment contributes to children not being willing to attend school, as well as some teachers not knowing how to exercise alternative strategies to deal with difficult behaviour.

It was also found that despite national regulations on keeping education free, schools take the liberty to request cash contributions from parents to supplement the running of school activities.

The exploration of these issues brought to light factors that – in order to capture them in numbers – first had to verbalized. From this perspective, qualitative analysis often precedes quantitative analysis as it provides topics to be monitored and acted upon. But it also works the other way around; when quantitative analysis shows discrepancies that cannot be explained without conducting exploration, then qualitative analysis is the way forward.

2 Education Indicators

As we already know, Sustainable Development Goal (SDG) 4 has put a strong emphasis on monitoring certain education aspects in order to evaluate progress towards achieving the Education 2030 Agenda. These aspects have been agreed upon by countries worldwide and resulted in the ten targets of SDG 4. We could say, developing the ten targets has been the result of qualitative data analysis by exploring each and every country's perception on crucial education problems that require solving. To track the situation for each of these ten targets, this requires being able to express the particular situation in some form of a numerical measure that tells us whether a situation has changed for the worse, or for the better. We call this an indicator.

An indicator can be used for different circumstances as different indicators can have different functions. Table 1 attempts to provide a brief overview of the functions of indicators, distinguished by monitoring, evaluation and also assessment.

Table 1: Monitoring, evaluation and assessment using education indicators[5]

FUNCTIONS OF INDICATORS			
	MONITORING	**EVALUATION**	**ASSESSMENT**
Purpose:	During implementation. Formative – continuously drawing lessons and insights to adjust implementation.	After implementation. Summative – evaluating how policies/programmes have been implemented.	During and after implementation. Can be both formative and summative to review outcome/impact.
Aimed at:	Operations and management.	Policy-making and leadership.	Curriculum designers, teachers, course providers.
Focus on:	Account of what and how things have been done and immediate results and lessons.	Effectiveness of policies and strategies vis-à-vis goals and targets.	Learning objectives, teaching-learning methods and materials, outcomes and impact.

[5] Monitoring can take place regularly during policy implementation to check progress and to identify issues and lessons to adjust the implementation processes. Evaluation is carried out after the completion of specific stages of implementation; to reflect on what has and has not been accomplished; and the factors and constraints. Assessment occurs when we review the outcomes and impacts of curriculum design and various teaching/learning processes and materials.

Information regarding:	Inputs and process.	Outcomes and impact.	Learning achievements per objectives, teaching-learning processes.
Indicators on:	Resource inputs, access and participation, efficiency, quality of delivery.	School capacities; graduates; learning achievement, socio-economic changes.	Learning results according to teaching-learning methods and materials

Note: Evaluation indicators may overlap with monitoring and assessment indicators.

Source: Compare UNESCO, 2011: Systemic Monitoring of Education for All. Training Modules for Asia-Pacific. Bangkok, UNESCO, **access:** http://uis.unesco.org/sites/default/files/documents/systematic-monitoring-of-education-for-all-training-modules-for-asia-pacific-en.pdf

2.1 What is an indicator?

An indicator is used to show – or indicate – the state of a situation or condition, expressed as a numerical (for example, 90 per cent) or categorical value (e.g. 'yes' vs 'no'). In the context of monitoring and evaluating education, an indicator is a quantitative metric that provides information, for example, to track performance, measure achievement and determine accountability. It is important to note that a quantitative metric can be used to provide data on the quality of an activity, project or programme.

Indicators help to identify problems and issues, which leads to setting targets and developing strategies to reach those targets. Indicators also serve to monitor progress towards achieving, (or failing to achieve) a goal when comparing the same indicator for the same condition over time.

Indicators also provide an early warning by drawing the attention of policy makers and planners to underperformance in time so they can address systemic issues by implementing strategies to improve a system. Indicators are an essential tool that can describe the characteristics, the effectiveness, the equity, or the trends of a particular aspect of the education system. They also provide an objective representation to evaluate progress and outcomes.

In short, an indicator:

- Enables managing the monitoring of delivering education services;
- Enables objective judging on aspects of the functioning of the education system;
- Enables measuring changes in the education system over time.

For more on indicators, see:

An Introduction to Indicators[6]

An example of using education indicators

Understanding the situation of an education system requires more than a simple counting of the number of schools, students and teachers. Let's take a look at the table for country X.

Table 2: Example of education indicators

	2013	2014	2015	2016	2017
School-age children enrolled	325,781	336,043	346,024	356,508	367,061
Percentage change year-by-year		3.15%	2.97%	3.03%	2.96%
Total school-age population in country X	458,201	468,679	478,594	490,383	501,449
Net enrolment rate (percentage)	**71.1**	**71.7**	**72.3**	**72.7**	**73.2**

Here the 'school-age children enrolled' and the 'total school-age population' are raw data, whereas the 'net enrolment rate' and the 'percentage change year-by-year' serve as indicators.

To explain, this table shows that the number of enrolments increased from 325,781 to 367,061 between 2013 and 2017. This is a total increase of 41,280 students over a period of four years.

Year-by-year, this represents an average annual increase of about 10,000 students, or around three per cent. These figures may indicate steady, positive growth in enrolment. However, because we notice fluctuations with a downtrend (3.15 per cent to 2.96 per cent), the following years would need observation to confirm whether this trend remains.

To calculate the 'net enrolment rate', we divide 'school-age children enrolled' by the 'total school-age population', multiplied by 100 to convert it to a percentage figure. Now we have created the indicator that tells us how many of all school-age children actually enrol in education (expressed in per cent). Therefore, we can observe steady progress in enrolling children over time.

[6] UNAIDS, 2010: Monitoring and Evaluation Fundamentals. Geneva, UNAIDS, access: http://www.unaids.org/sites/default/files/sub_landing/files/8_2-Intro-to-IndicatorsFMEF.pdf

2.2 The essence of a good indicator

Creating an indicator can confront us with the dilemma of setting a standard that accounts as progress. A simple illustration is setting a minimum criterion for learning outcomes.

If the standard for what accounts as a positive learning outcome is set too high, achieving the standard will be beyond the capability of the majority of pupils we are trying to educate.

In consequence, the indicator will show that most fail to reach the set standard – without the ability to diagnose the cause. Conversely, if the standard is set too low, all will pass over the criterion level and the indicator fails to measure increased learning outcomes.

Indicators must be designed in a way to capture differences in performance[7]. A good indicator should be clear and concise. It should focus on a single issue that provides relevant information on a situation; particularly information that provides the strategic insight required for effective planning and sound decision-making[8].

Good indicators should accurately measure what they claim to measure. If it is not feasible to collect data for an indicator, or data that can be collected is/are not meaningful, the indicator will have little or no utility[9].

Furthermore, to be meaningful across systems and countries, an indicator needs to reach a certain degree of consensus on its usefulness and validity.

A good indicator has to be:

RELEVANT: While it is difficult for indicators alone to fully capture the vision behind the proposed targets, indicators should ideally reflect the most critical policy themes in the targets. Across all SDG 4 targets, emphasis has been placed on measuring equity.

ALIGNED: The construct to be measured must be valid and reliable relative to the targets. This means that the indicator must have the same meaning and significance in all settings, ideally measured by a similar question, or item. Measuring constructs that vary across settings poses a challenge for global tracking. It may be possible to measure some elements globally, while others may be best measured at the national, or regional level, with flexibility to adapt constructs to local contexts.

FEASIBLE: Global tracking is most effective when data is collected on a regular basis (though not necessarily annually), and all, or nearly all countries routinely collect data in a similar manner. Infrequent or low coverage of data constrains the ability to track changes

[7] For further discussion on the above topic see: Lewin, K. M., & Sabates, R. (2012, March 23). Who gets what? Has improved access to basic education been pro-poor in sub-Saharan Africa? International Journal of Educational Development, 32(4), 517-528; Lewin, K. M. (2005, July). Taking targets to task: Planning post primary education. International Journal of Educational Development, 25(4), 408-422; Technical Advisory Group of the Education For All Steering Committee. (2014). Towards indicators for a post-2015 education framework. Montreal, Canada: UNESCO Institute for Statistics.

[8] UNAIDS, 2010: An Introduction to Indicators. Geneva, UNAIDS, access: http://www.unaids.org/sites/default/files/sub_landing/files/8_2-Intro-to-IndicatorsFMEF.pdf

[9] ibid.

over time. Collecting data over time must also be feasible and cost-effective.

COMMUNICABLE: The indicators must be easily understood by a wide audience and lend themselves to the development of a clear narrative regarding progress towards the goal and its targets. The indicators for education should facilitate clear and transparent reporting about the objectives and achievements at each stage of implementation.

INTERPRETABLE: The indicator values and their changes over time must be easily understood to ensure correct interpretation in education sector planning. Limitations of an indicator should also be visible to prevent misinterpretation.

2.3 Defining education system and education service indicators

Education indicators can be derived by looking at education from two different perspectives:

a. Education as a **SYSTEM**; and
b. Education as a **SERVICE**

2.3.1 Education indicators from a system's perspective

For purposes of monitoring education as a system, education indicators (and indicators in general) are broadly classified into input, process, output, outcome and impact indicators. Input and process indicators are used for monitoring whether appropriate education policies have been issued and adequate resource inputs have been allocated and implemented. Output, outcome and impact indicators are used to evaluate the results, effectiveness and progress of education policies and their implementation.

INPUT INDICATORS: In an education system, input indicators f**ocus on the human, financial and material resources that have been assembled and channelled into educational activities**. Such resource inputs are used to organize the provision of educational services in order to create intermediate outputs such as classes and learning activities. Examples of education input indicators include the percentage of the government budget allocated to education, pupil-teacher ratios, pupil-classroom ratios, the percentage of pupils without textbooks, etc.

PROCESS INDICATORS: In an education system, process indicators show **how the resource inputs discussed above have been utilized to deliver educational services**. These indicators show what actually happened in the classroom and during teaching/learning processes. The students' attendance rate at school, average number of class hours they participated in as a percentage of official class hours and the frequency of use of teaching/learning materials are some examples of process indicators. Other process indicators may include repetition rates and drop-out rates.

OUTPUT, OUTCOME AND IMPACT INDICATORS: These measure **the result of a particular set of interventions** (policies, learning topics, financial aid, etc.). Improvement in these types of indicators should determine the success of the interventions. Output and outcomes should relate to specific objectives, but there may be different levels of objectives. A distinction between output, outcome and impact should be made in education to distinguish the level of results[10].

> **OUTPUT INDICATORS:** These describe **the direct result of the products, or services delivered in education**. Output indicators can provide measures of efficiency and describe the relationship between investments in an education activity and its result[11]. Output indicators are seen as a more direct 'outcome' of schooling, directly measured, with for example, achievement tests and/or the number of students who have graduated, potentially further distinguished by the score with which they graduated[12]. They typically describe total numbers.
>
> **OUTCOME INDICATORS:** Outcome indicators are used to evaluate the **end results of all the educational inputs and processes**. Outcome indicators may be used to evaluate the degree of access to educational services and the degree of satisfaction with the services received. Examples of outcome indicators include intake rates, enrolment ratios, completion rates and GPI (Gender Parity Index).

IMPACT INDICATORS: Impact indicators show the effect implemented activities in education have on the situation of individuals, families, communities, the nation and society as a whole. For example, earnings, employment, contribution to productivity, improved health, decreased crime and other non-monetary long-term outcomes associated with the availability and completion of a quality education. A specific example is the adult literacy rate. It shows the proportion of the adult population who have learnt to read, write and comprehend written text and who can continue to learn using written words. Other impact indicators include those that measure the effect of increased knowledge and skills, emotional development and the impact of changes in students' values, attitudes and behaviour on their family, community, society and the nation.

---

[10] Parsons, Jim, Caitlin Gokey, Monica Thornton, 2013: Indicators of Inputs, Activities, Outputs, Outcomes and Impacts in Security and Justice Programming. London, UKaid, access: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/304626/Indicators.pdf; Vos, Rob, 1996: Educational Indicators: What's To Be Measured? Indes Working Papers Series I-1. Washington, D.C., Indes Working Papers, access: https://pdfs.semanticscholar.org/3ff9/7df813e86dbaf55544e389c0c6c582757005.pdf

[11] Parsons, Jim, Caitlin Gokey, and Monica Thornton, 2013: Indicators of Inputs, Activities, Outputs, Outcomes and Impacts in Security and Justice Programming. London, UKaid, access: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/304626/Indicators.pdf

[12] Scheerens, Jaap, 2011: Measuring educational quality by means of indicators. In: Perspectives on educational quality: illustrative outcomes on primary and secondary schooling in the Netherlands, pp 35-50. Dordrecht, Springer, access: https://www.springer.com/cda/content/document/cda_downloaddocument/9789400709256-c2.pdf?SGWID=0-0-45-1132975-p174098969; UNESCO, 1998: Monitoring progress towards education for all: a methodological guidebook. Paris, UNESCO, access: https://unesdoc.unesco.org/ark:/48223/pf0000112816?posInSet=6&queryId=9c8838b6-bacd-49d9-978d-f40e6f65adb7 (page 3).

Box 1: Outputs, outcomes or impacts?[13]

There is often confusion about the differences between outputs (products and services) and outcomes (the short- and medium-term benefits that those products, or services deliver). One way to distinguish between outputs and outcomes is to consider whether the indicator describes effectiveness.

If yes, we are talking about an outcome indicator. For example, installing computers in every classroom (as resource inputs) has led to an increase in the use of interactive presentations in teaching (the output). However, this does not offer any indication of whether the new technology has improved the teaching effectiveness yet (the outcome).

The actual short-to-medium term effect may be an increase in ICT skills among the students. Going one step further, the potential long-term impact might include an increase in students enrolling in technical education – be it in Technical and Vocational Education and Training (TVET) or university.

The following figure shows the contents of an indicator framework based on an Input-Process-Output/outcome (IPO) model.

Figure 3: Indicator framework-based Input-Process-Outcome model

<table>
<tr><th colspan="6">CONTEXT</th></tr>
<tr><td colspan="3">Demographic, social and economic context of education</td><td colspan="3">I.e.: Education Attainment, Adult literacy rates</td></tr>
<tr><th colspan="2">INPUTS</th><th colspan="2">PROCESSES</th><th colspan="2">OUTCOMES</th></tr>
<tr><td>Financial and human resources invested in basic education.</td><td>I.e: Education expenditure per student.</td><td>Access, participation, progression, transition from lower to upper levels.
Learning environments.</td><td>I.e: Overall participation in formal and non-formal education.
I.e: Total intended instructional time for pupils by level of education; instructional time per subject</td><td>Achievements.</td><td>I.e: Mathematics, science, reading achievements of students in a given grade or level of education.</td></tr>
<tr><td colspan="2">Examples</td><td colspan="2">Examples</td><td colspan="2">Examples</td></tr>
<tr><td colspan="2">Proportion of GDP spent on education.
Education expenditure per student.
Proportion of public/private investment in education.
School-level financial and material resources (Building facilities, computer equipment, water access, etc.).
Teachers working conditions (salaries, working times, training).</td><td colspan="2">GIR/NIR, GER/NER, ANER.
Promotion, dropout and repetition rates.
Out-of-school rate by level of education.
Pupil-teacher-ratios by level of education.
Average class size.
Proportion of trained/ qualified teachers.
Instruction time per subject.</td><td colspan="2">Graduation rate.
Completion rate.
Survival rate to the last grade of basic education.
Percentage of students achieving functional minimum competencies in mathematics, reading, science.</td></tr>
</table>

[13] With Inspiration from: Parsons, Jim, Caitlin Gokey, Monica Thornton, 2013: Indicators of Inputs, Activities, Outputs, Outcomes and Impacts in Security and Justice Programming. London, UKaid, access: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/304626/Indicators.pdf (page 15)

2.3.2 Education indicators from a service's perspective

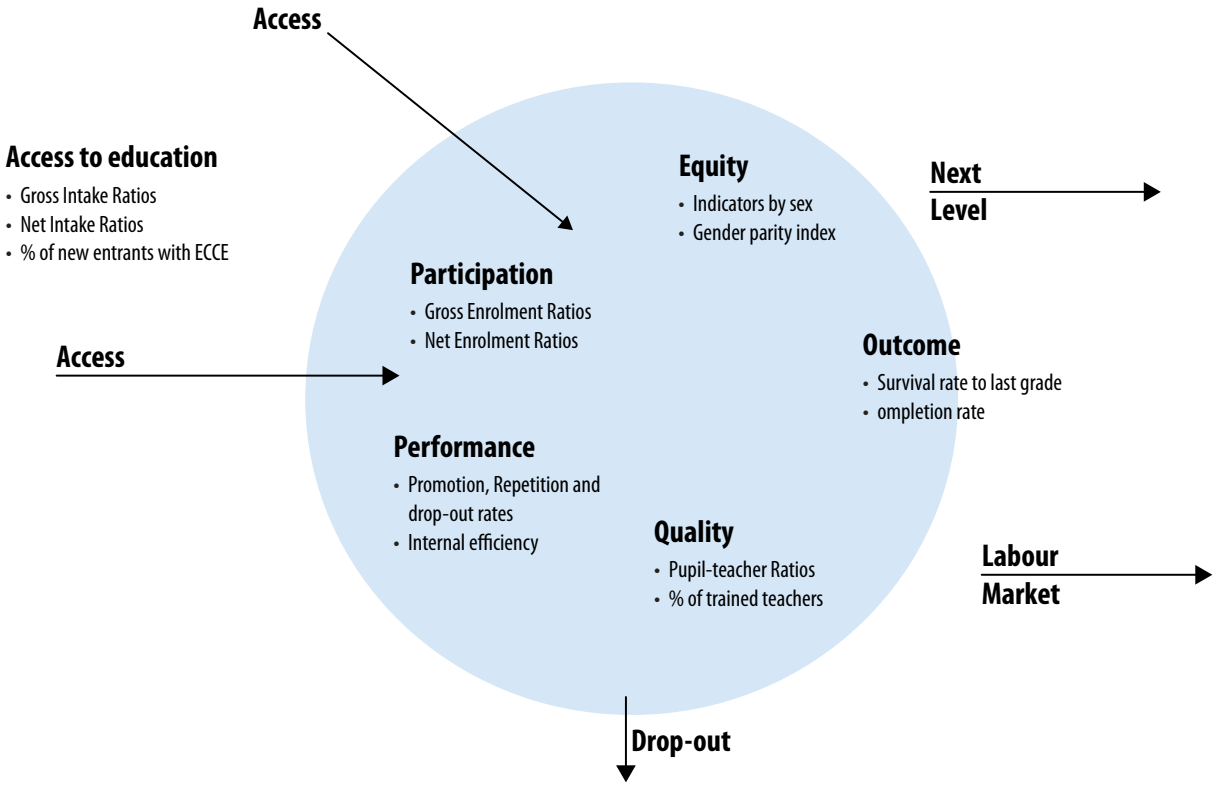
To monitor education as a service, education indicators can be distinguished by the dimension of a part of education that is meant to be measured. We identify dimensions of education service when thinking of questions, for example:

- *Do all children have access to education?*
- *Are all students able to actively participate in education?*
- *Do all students benefit from a good quality education?*
- *Are all students treated with equality?*
- *Is the management of education efficient and effective?*
- *Are the outcomes of education relevant and satisfactory?*

In consequence, and for the sake of structuring the information, education service indicators are organized as indicators on:

- School characteristics, environment and facilities;
- Access and participation;
- Retention and progress within the education system;
- Teaching and learning resources;
- Teaching-learning processes;
- Quality of education;
- Equity;
- Learning achievement and outcomes;
- The impact of education.

Figure 4: Dimensions of education service indicators



Source: UNESCO, 2011: Systemic Monitoring of Education for All. Training Modules for Asia-Pacific. Bangkok, UNESCO, Access: http://uis.unesco.org/sites/default/files/documents/systematic-monitoring-of-education-for-all-training-modules-for-asia-pacific-en.pdf

2.3.3 Selecting an indicator

We can use indicators to monitor the overall situation, to examine specific aspects and concerns and to consider the implications of decisions. Often more than one indicator is required to explain a certain issue. For example, we monitor participation in school through the use of a number of ratios, for instance gross, net and age-specific enrolment ratios.

Sometimes, one indicator can be used to depict several issues. For example, we use the pupil-teacher ratio to check whether there are enough teachers for the number of students. However, this indicator can also be used as a proxy for the quality of education as a lower pupil-teacher ratio will indicate a teacher can dedicate more attention to each pupil – provided all children are actually attending school.

Knowing how to select the most appropriate indicator for a specific purpose is important, because if the wrong indicators are chosen, they can lead to confusing, misleading or even contradictory conclusions.

When we choose indicators, we need to ensure they are based on the right concept,

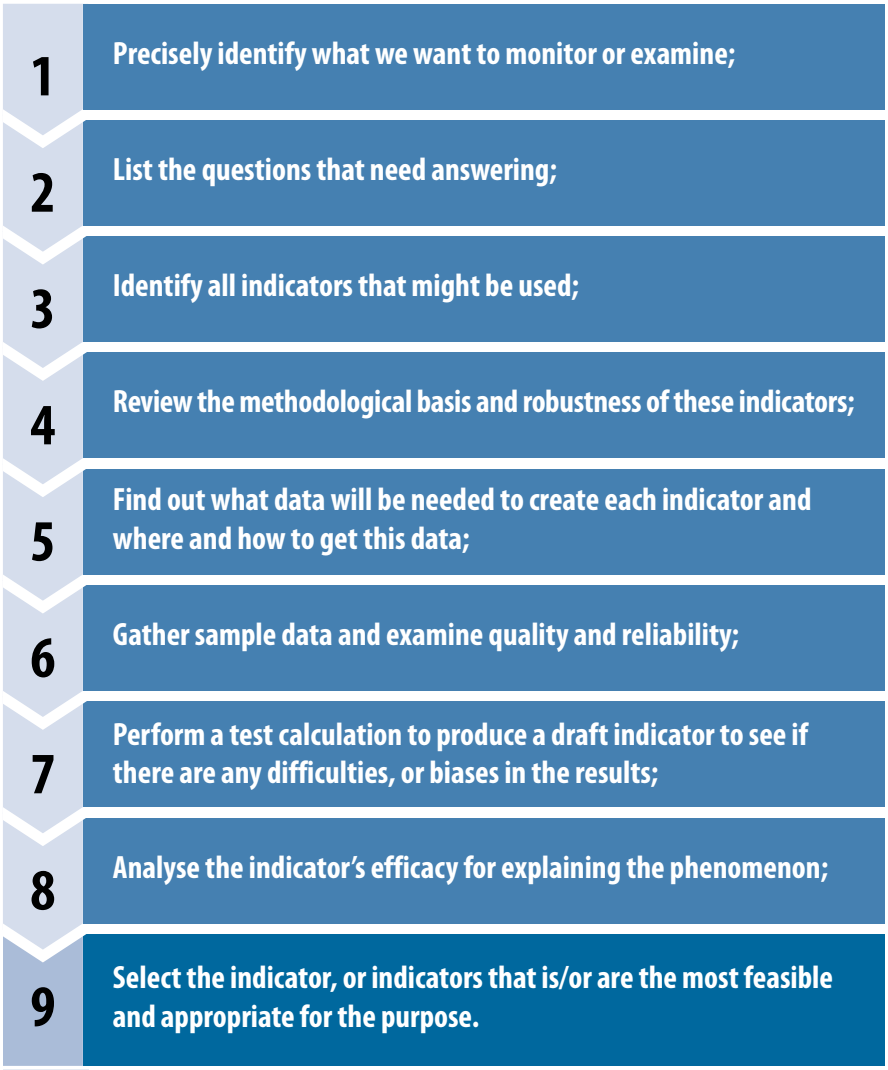
definition, analytical objective, data type, data source and angle of interpretation.

Think in terms of:

- *Is the indicator relevant in its concept to measure the issue looked at?*
- *Has the indicator a clearly defined purpose and limitation?*
- *Is data available and is the available date reliable?*
- *Can the indicator be rigorously calculated?*
- *Can the indicator be easily presented and interpreted?*

The identified steps below should aid in selecting an indicator practically:

Figure 5: Guiding steps to select education indicators



Source: UNICEF and UIS, 2018: Tanzania Qualitative Study Report. Global Initiative on Out-of-School Children.

After selecting a suitable indicator, it is recommended to capture the information needed to generate the indicator, as well as its purpose, application and limitations. Collecting this information will allow us to create an indicator document that provides other users of the indicator the information to replicate the indicator. We can also call this the metadata.

Elements to keep in mind when creating an indicator metadata catalogue

The development of an indicator document on education and SDG 4 helps organize and identify the information that is required to generate indicators. It guides the data collectors, producers, analyst and users. Such a document should include the following[14]:

- **DEFINITION**
- **PURPOSE**
- **METHOD AND FORMULA OF CALCULATION**
- **REQUIRED DATA**
- **POSSIBLE DATA SOURCES**
- **DISAGGREGATION FOR ANALYSIS OF DISPARITIES**
- **INTERPRETATION**
- **QUALITY STANDARDS**
- **LIMITATIONS AND CONSTRAINTS**

If you have not seen it already, the UIS document; 'Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030' serves the exact purpose of an indicator document and allows countries to apply the exact same standards and procedures, thus, making indicators comparable.

See the following figure for an example and follow it to download the document:

Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030[15]

[14] Note: You can find the latest details for each of the SDG 4 indicators at: http://uis.unesco.org/en/glossary

[15] UIS, 2018: Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030, access: http://uis.unesco.org/sites/default/files/documents/sdg4-metadata-global-thematic-indicators-en.pdf

Figure 6: Example of what should be included in an indicator document

4.1.5 Out-of-school rate (primary education, lower secondary education, upper secondary education)

Definition:

Children and young people in the official age range for the given level of education who are not enrolled in primary, secondary or higher levels of education. Children and young people who are enrolled in pre-primary education are considered to be out of school.

Purpose:

To identify the size of the population in the official age range for the given level of education who are not enrolled in school in order that they can be better targeted and appropriate policies can be put in place to ensure they have access to education.

Calculation method:

The number of students of the official age for the given level of education enrolled in primary, secondary or higher levels of education is subtracted from the total population of the same age.

$$OSR_n = \frac{SAP_n - \sum_{i=1}^{8} E_{i,AGn}}{SAP_n}$$

where

OSR_n = out-of-school rate for children and young people of the official age for level n of education

SAP_n = population of the official age for level n of education

$E_{ni,AGn}$ = enrolment in ISCED level i of children and young people of the official age for level n of education

Interpretation:

The higher the number of out-of-school children and adolescents, the greater the need to focus on improving access to education. Some children have never been in school or may not eventually enrol as late entrants. Other children may have initially enrolled but dropped out before reaching the intended age of completion of the given level. When disaggregated by sex, location and other characteristics, this Indicator can identify excluded population groups.

Type of data source:

Administrative data, household surveys.

Disaggregation:

By age or age-group and sex (administrative data); by age or age-group and sex, location, and socioeconomic status (household surveys) and others as available.

Data required:

Enrolment by single year of age in each level of education, population estimates by single year of age and data on the structure (entrance age and duration) of each level of education.

Data sources:

Administrative data from schools or household survey data on enrolment by single year of age; population censuses and surveys for population estimates by single year of age (if using administrative data on enrolment); administrative data from ministries of education on the structure (entrance age and duration) of the education system.

Limitations and comments:

Inconsistencies between enrolment and population data from different sources may result in inaccurate estimates of out-of-school children and adolescents. Data from household surveys conducted late in the school year where ages are recorded at the enumeration date may result in over-estimates.

Source: UIS, 2018: Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030, Montreal, UIS, **access:** http://uis.unesco.org/sites/default/files/documents/sdg4-metadata-global-thematic-indicators-en.pdf

3 Conducting Statistical Data Analysis

Monitoring SDG 4 follows the same principle and requires dissecting the SDG 4 indicators according to the different concepts they address (learning, participation, completion, teachers, resource inputs and educational processes, etc.) and analysing them by:

- Characteristics of distribution and/or patterns;
- Differences, disparities and imbalances;
- Changes over time and space;
- Progress and shortfalls against targets and plans.

This dissection helps identify causes and actions to address pertaining issues. Data analysis for monitoring Education 2030 goes beyond covering spatial and temporal data by further disaggregating for age, sex, wealth, disability, ethnicity, mother tongue/first language, second/other language(s) and migration, or refugee status.

When we conduct an analysis of statistical data, we must at all times keep in mind our possibilities for disaggregation, as only then we can capture the remaining issues to tackle in education. Let's take a look at the steps for our data analysis in the next section.

3.1 Five steps for data analysis

Step 1: Objectives of data analysis

Before initiating data analysis, we must spell out clearly the objectives of the data analysis, such as what questions should the data analysis answer? For example, should data analysis examine potential gaps in learning achievements, while distinguishing between rural and urban areas? Should it take a stock of completion of basic education? Should it compare the progress between various social groups? Or should it provide analysis on whether there are enough human and technical resources to implement a programme/programmes effectively?

Step 2: Selection of Indicators

After having a clear purpose of the data analysis with the relevant questions that we seek to provide answers for, the second step is to select indicators for the analysis. The

selected indicators should be able to answer directly the objective in question. The data needed for the indictor must also be available, or initiative must be taken to collect the data. Example indicators include the completion rate, or the survival rate, to analyse the performance of the education system under scrutiny.

Step 3: Processing the data

Data processing is the next bigger step in data analysis which encompasses the extraction and merging of raw data from available data sources and, if required, to correct erroneous data. While preparing the data, one should thoroughly check for completeness and coverage (e.g. data for all the areas and groups, for all types of public, private and other types of education). Processing data means to prepare all the data from different sources for one singular database with which to calculate the desired indicator(s).

Box 2: What if data is missing?

Data coming from various sources – especially administrative data, based on school censuses, or other methods – would invariably see some degree of data missing for some of the variables. If data is missing and not properly indicated, or substituted, this can cause the analysis to be biased. The question is: How should we deal with missing data?

1. Recover the missing values

The obvious solution at hand is to contact the participants and ask them to fill out the missing values. For in-person studies, having an additional check for missing values before the participant leaves helps. So, if there is data missing for some of the school, we need to contact the participants so that the missing data can be filled.

2. Substitute the missing values

Imputation is a technique to replace missing values with substitute values. There are different types of imputation techniques (see the annex of this module). To explore some of the techniques to substitute missing values, this Module's annex will provide an overview.

3. Delete the missing values

Also called 'list-wise deletion', this method deletes all data from any participant with missing values. If the sample is large enough, one may delete data without substantial loss in a statistical meaning. However, we must be certain that the values are missing at random and that you are not inadvertently removing a specific group of participants for which there might be reasons why no data exists.

Note: For more information, see the technical note at the end of this module.

Step 4: Producing summary statistics

Once data has been processed, summary statistics can be created to describe the data; to identify patterns for trends and issues; or plainly to describe a situation. This can also be called 'descriptive statistics' because these statistics 'describe' the data we are looking at. We can produce these statistics by:

- Transforming raw data to indicators, e.g. from total numbers to percentages, rates and ratios (this includes anonymizing data);
- Sorting and grouping data by common social, economic and demographic characteristics;
- Presenting the indicators in tables, charts and flow text to enable easy interpretation.

These statistics provide summary indicators on the average, range, median, mode and standard deviation. These summary statistics help to understand a situation to carry out decisions.

As required, every indicator can be disaggregated by many aspects, such as male versus female; rural versus urban; districts, by year; by language, etc. This will allow us to analyse the indicator in many ways.

Box 3: A common misunderstanding on averages

A common problem in dealing with quantitative data is the misinterpretation and reporting of 'averages'. Averages have been reported as means, medians and modes. However, there are important differences:

1. The MEAN: is a mathematically derived average of values, by adding up all values and dividing them through the total number of values. The mean is the most commonly used average. Note that most of the time when people talk about the average, they refer to the mean.
   - i. The mean is a helpful measure of centrality when we want to determine, for instance, the average time spent in years to complete a full cycle of basic education, including repetition and potential intermitting dropout. If the mean (time spent in years) results are too high than the foreseen years for basic education, the mean can indicate a problem.
   - ii. The mean can also help determine a comparison of different population groups for males and females regarding their average respective years spent in education. The same can also be done for other disaggregation.
2. The MEDIAN: is the middle value among all values; the median represents the value where half of all values are below and half of all values are above; it is significant to note that in order to determine the median, all the values must be sorted first (otherwise the highest, or lowest value might end up in the middle).

i. The median can be relevant when a sample has very high (or very low) outliers. Income is a classic example. For instance, determining the average teacher salary at a school can be distorted by individual high salaries, even though the majority of teachers earn a low salary. Using the median (the salary value in the middle of all salary values) can report a more accurate average than the mean.

ii. Determining the average age among primary school students may be distorted when a number of students are much older than most primary students (due to late school entrance and repetition). In such a case, the median can report a more accurate age average.

3. The MODE: is a frequency-based conclusion of all values, or a measure of frequency; it is the value that occurs most often; that means in reverse, if no value in the list is repeated, then there is no mode for the list.

i. The mode can be relevant when we want to determine which school has the most incidents of bullying, corporal punishment, harassment, violence and/or sexual discrimination. Knowing the mode (the school with the most incidents) would serve as an indication were urgent intervention has to be prioritized.

ii. The mode can be relevant when we want to determine which school has participated the most often in the annual UNESCO-Japan Prize on Education for Sustainable Development. Knowing which school participated the most often in the competition would serve as an indication that this school might have a special curriculum and/or pedagogy in this area.

An analysis depending on each of these averages results in different information for different responses in practice and since this mistake is common but easily avoided it is therefore mentioned.

Step 5: Analysing the statistics

Summarizing descriptive statistics in education provides us with an initial image of the situation at the school, district, country, or even global level. However, at times it is necessary to analyse the retrieved information deeper regarding differences among and relationships between variables of populations. Going for more complex analysis allows us to test hypotheses and generalize results of a population as a whole.

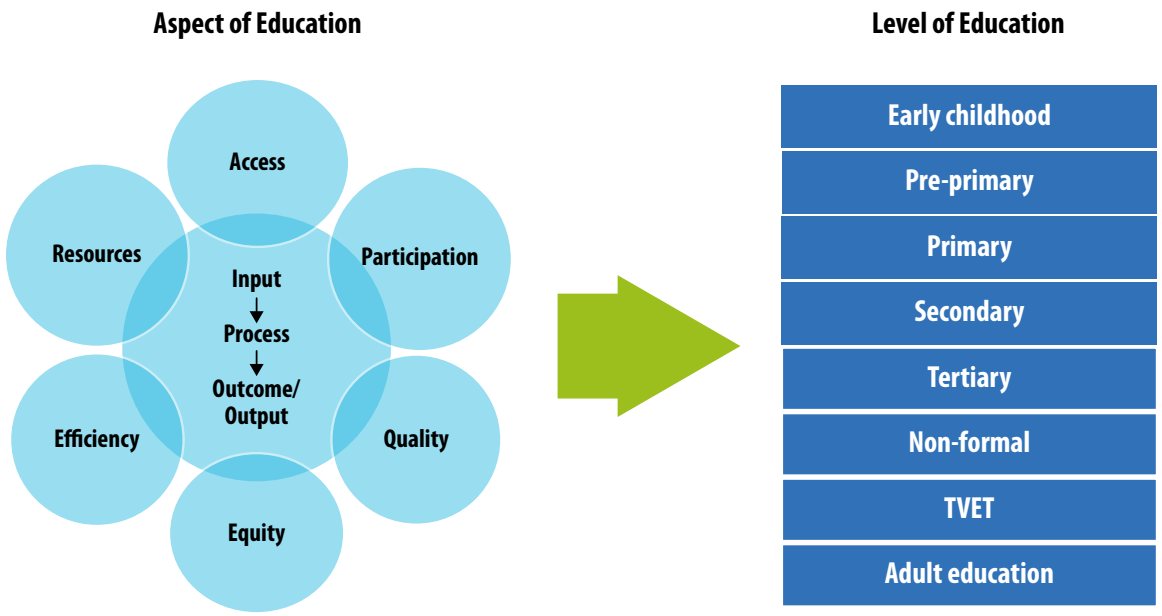
There are several analysis types, some of which are widely known. Table 3 identifies some of the most important data analysis types. We will find explanations in the next section to provide an understanding of their purposes.

Table 3: Different types of statistics used for different types of analysis

ANALYSIS	STATISTICAL TOOL EXAMPLES	PURPOSE AND PERIODICITY
Progress analysis	Time series, percentage point increase, annual growth, etc.	Analysis of trends, of changes overtime.
Gap analysis	Distance from target, population comparison.	Assessing the differences in performance, or results.
Relationship analysis	Correlation (coefficient), regression.	Analysis of whether or not variables influence one another.
Forecasting analysis	Regression, moving average.	Prediction of a future value/ result/outcome. (Can be based on trends).
Multivariate analysis	Causal models.	Analysis of multiple variables at a time.
Equity Analysis	Employs the above analysis techniques, including distributions, parity indices, ranges, percentiles.	Looking into different population groups.

All these analysis types essentially serve the purpose of monitoring the different aspects of education and for all levels of education. It is important to keep in mind that each level of education needs to be evaluated for each of the aspects of education (see Figure 7 for a mind map).

Figure 7: A mind map to remember that all aspects of education should be considered for each level of education



3.2 Statistical analysis tools

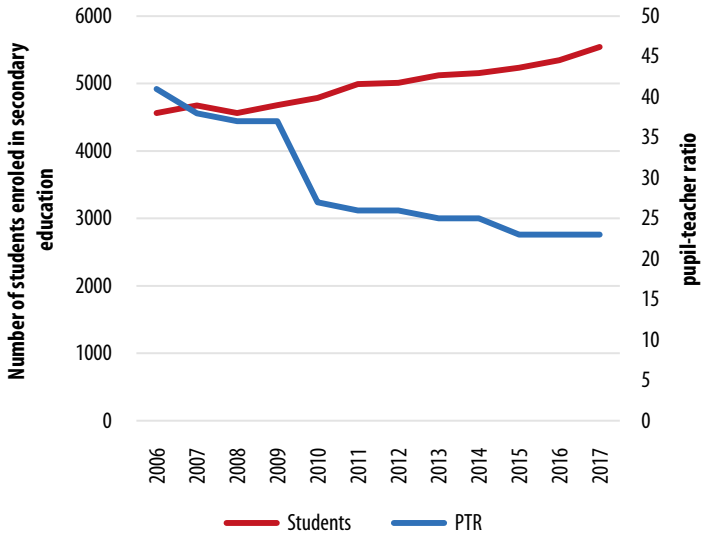
The following section introduces the aforementioned types of analysis in order to familiarize learners with the information these types of analysis can add to the monitoring of education. The explanations, however, are not meant to teach the mathematical procedures to create these types of analysis. Professional statisticians will be capable of creating these types of analysis.

3.2.1 Progress Analysis

Time series

In essence, the time series analysis tells us about the development of an observable phenomenon expressed by its indicator value over time. Meaning, this analysis compares values over time, as long as the definition of the data and the collection methods were consistent. We can create summary tables and charts to visualize the development, ordered by time (chronological order). In Figure 8, we can identify how the number of students and pupil to teacher ratio changes over the years.

Figure 8: Example of time series analysis for number of students; number of teachers; pupil-to-teacher ratio (PTR)



Year	Students	PTR
2006	4563	41
2007	4675	38
2008	4563	37
2009	4682	37
2010	4786	27
2011	4992	26
2012	5012	26
2013	5123	25
2014	5156	25
2015	5234	23
2016	5345	23
2017	5543	23

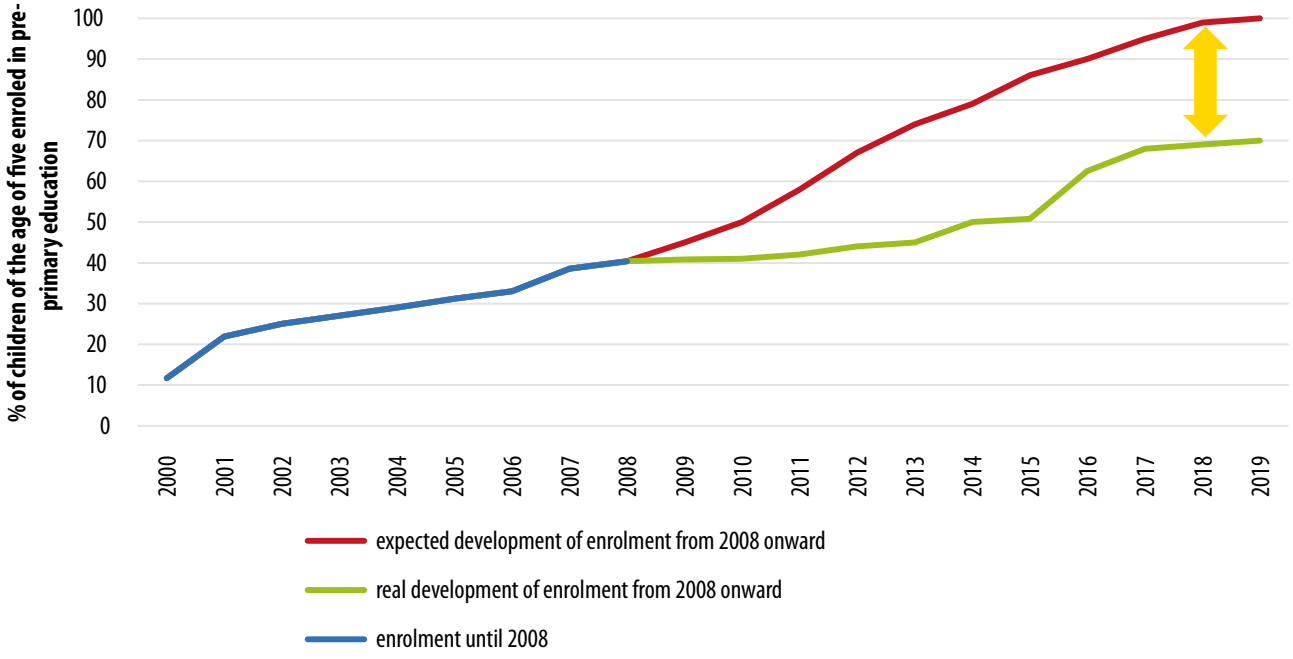
Note: Hypothetical data

3.2.2 Gap analysis

Distance from target

Gap analysis in education can serve different purposes. It can analyse the gap between a current and an expected result. Gap analysis can take advantage of the time series analysis by comparing expected development with the actual development over time. For example, it is expected that 100 per cent of children of the age of five have enrolled in pre-primary schools by 2019. In reality, 70 per cent have enrolled (Figure 9). The distance between two indicator values at a given time is the gap.

Figure 9: Fictional example of a gap analysis for enrolment in pre-primary education



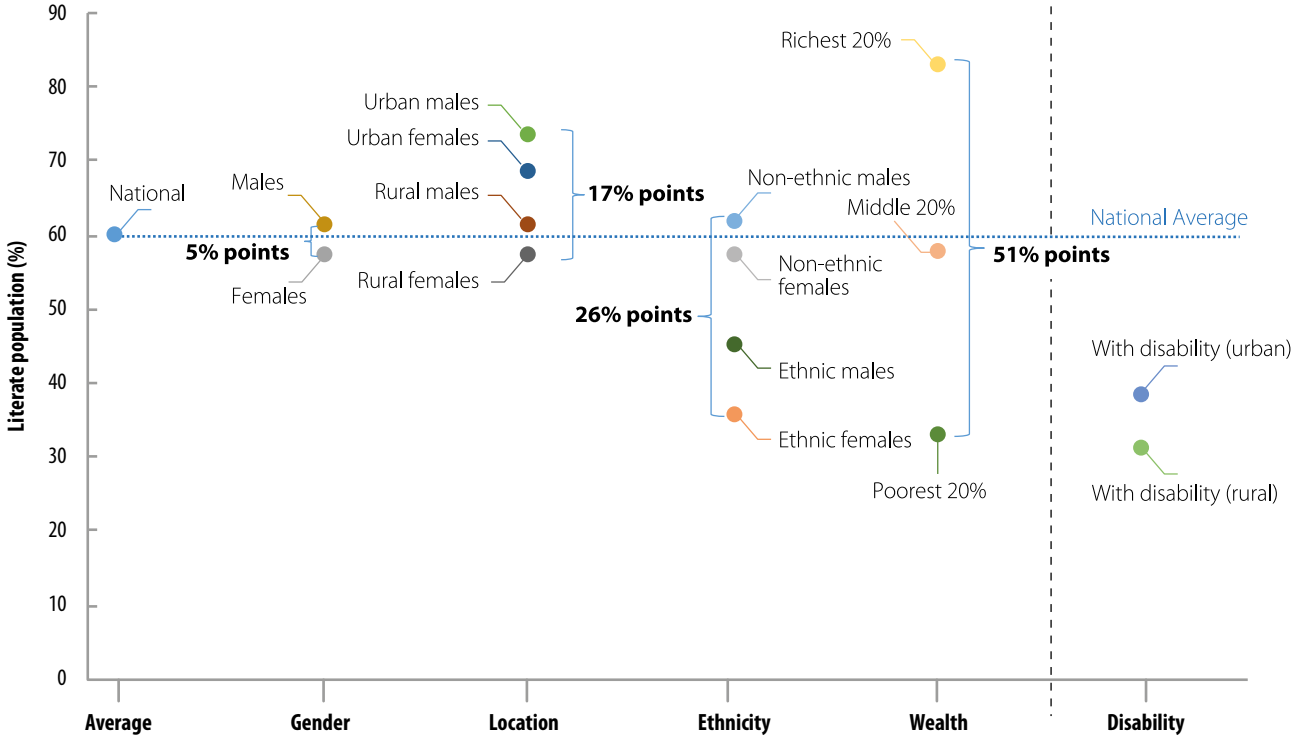
Note: Hypothetical data

Population comparison

Gap analysis can also serve to compare different groups with one another at a given time to identify which group shows higher or lower characteristics, tendencies or achievements, depending on the indicator in question.

In the following example, gaps can be identified for social groups by gender, location, ethnicity, wealth and disability. Furthermore, within these groups' location and gender difference helps further identify gaps.

Figure 10: Literacy rate comparison for a population distinguished by gender, location, ethnicity, wealth and disability



Source: UNESCO, 2018: Paving the Road to Education. A Target-by-Target Analysis of SDG 4 for Asia and the Pacific. Bangkok, UNESCO.

3.2.3 Relationship analysis

Regression

Regression analysis is a statistical technique to examine the relationship between variables. Usually, we seek to ascertain whether a variable has an effect upon another variable.

Regression analysis is used to estimate the strength and the direction of the relationship between two (or more) variables. Thereby, it is distinguished between a dependent variable Y (the variable that is meant to change depending on the change of the other variable) and the independent variable X (the variable believed to influence the other variable).

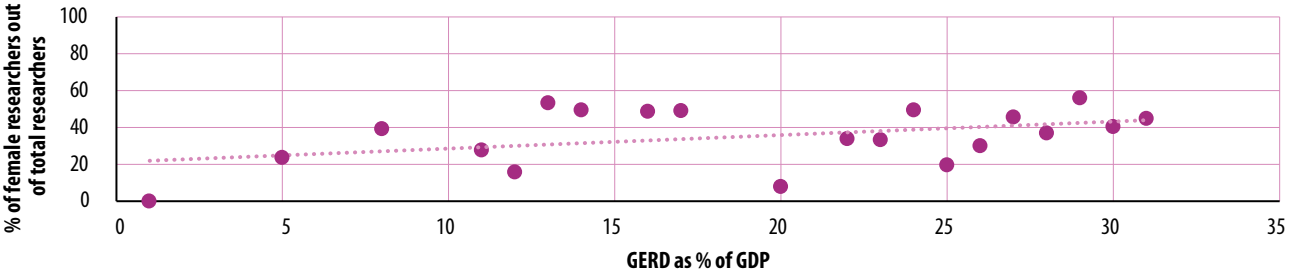
The two basic types of regression analysis are:

- Simple regression analysis: This is used to estimate the relationship between one dependent variable and one independent variable. For example, we can test whether increasing tuition fees will lead to decreasing enrolment numbers; or whether an increase in teachers will lead to better learning achievements.

- Multiple regression analysis: This is used to estimate the relationship between a dependent variable and two or more independent variables. For example, the relationship between a teacher's salary; the years of teacher training; and the class size (student head count); and how these impact on learning achievements. Multiple regression analysis introduces several complexities, but may produce more realistic results than a simple regression analysis.

Regression analysis is useful when we have to identify the impact of a unit change in the independent variable on the dependent variable. The regression types can be further distinguished by being linear or non-linear. The latter requires a more complex mathematical modelling to describe a noticeable pattern in your data (for example a U-shape of the data in your scatter plot, or a circular scatter of your data). Note that regression analysis is also important for forecasting.

Figure 11: Simple linear regression showing how expenditure on research and development increases the percentage of female researchers among the total of researchers in Asia-Pacific (each dot represents a country)



Source: UNESCO, 2019: International Day of Women in Science. Factsheet. Bangkok, UNESCO.

Correlation

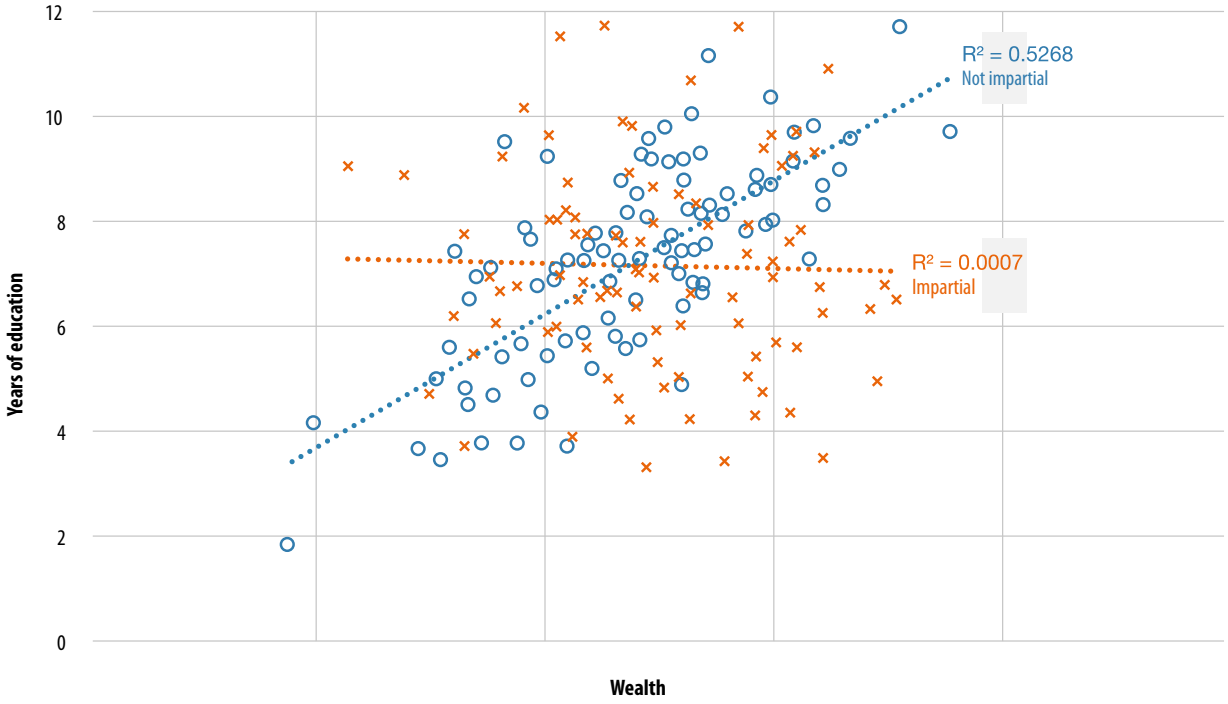
Correlation, or coefficient analysis is used to study the strength of a relationship between two numeric continuous variables (e.g. age, test scores, household income). This analysis is useful to quantify the degree to which two variables are related. Depending upon the values measured, this can be expressed as a positive or negative correlation.

- Positive correlation means that one variable *increases* with the other.
- Negative correlation means that one variable *decreases* when the other *increases*.

The correlation is expressed as between +1 (highest positive correlation) and -1 (highest negative correlation). At a value of 0, there is no correlation. Values closer to zero highlight a weaker/poorer correlation.

Note, that it is often misunderstood that correlation indicates cause and effect. This is not the case because other variables that are not considered in this analysis (e.g. environmental factors) may have impacted on the results.

Figure 12: Impartiality of education with respect to wealth



Note: Hypothetical data.

Source: UIS, fhi 360, Oxford Policy Management and Research for Equitable Access and Learning (REAL), 2018: Handbook on Measuring Equity in Education.

3.2.4 Forecasting

Regression forecasting

Forecasting relies typically on linear regression analysis to help predict future values from past values (e.g. enrolment numbers, budgetary spending). Statistical analysis often uses a linear relationship to predict the average value *Y* for a given value *X*.

A forecasting analysis rests on the assumption that the present circumstances will continue to operate in the future, or at a different place. Forecasting can therefore apply to time and also space.

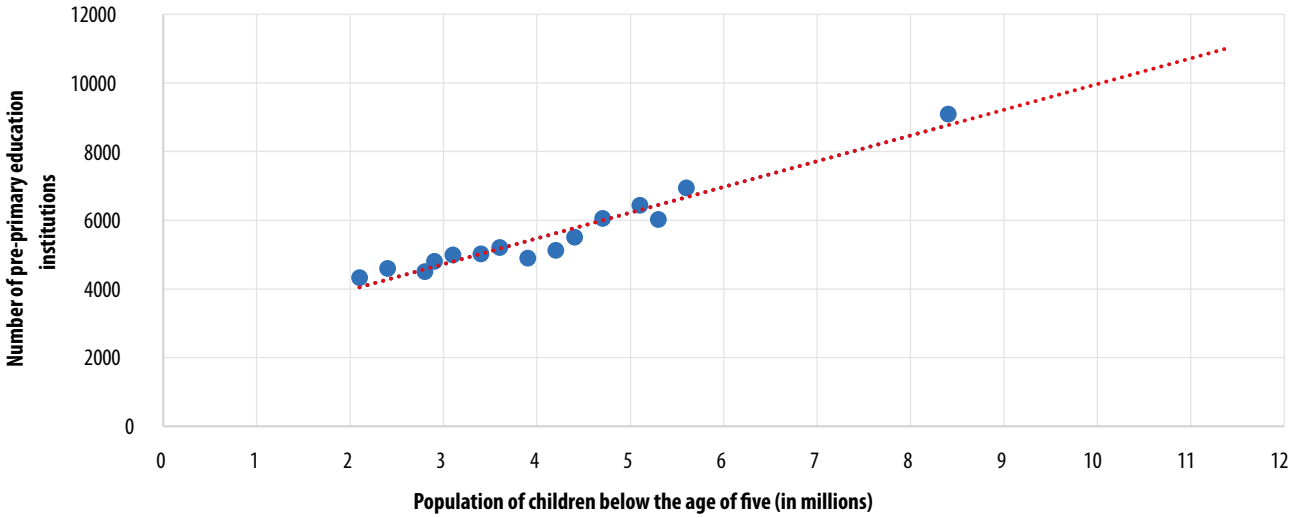
The assumption that the same circumstances apply is often valid when forecasting short-term results, but it falls short when creating medium to long-term forecasts. The further out we attempt to forecast, the less certain we become of the forecast.

Equally, while we may forecast assumptions relating to a location within a country (e.g. number of schools predicted in a district), the same prediction may likely not hold true for a different country due to different circumstances in the other country.

The following are fictional examples with which we can forecast the average number

of pre-primary education institutions in a given district (example 1), or the number of pre-primary intuitions over time (example 2), judging by the known number of children below the age of five and the known number of pre-primary institutions.

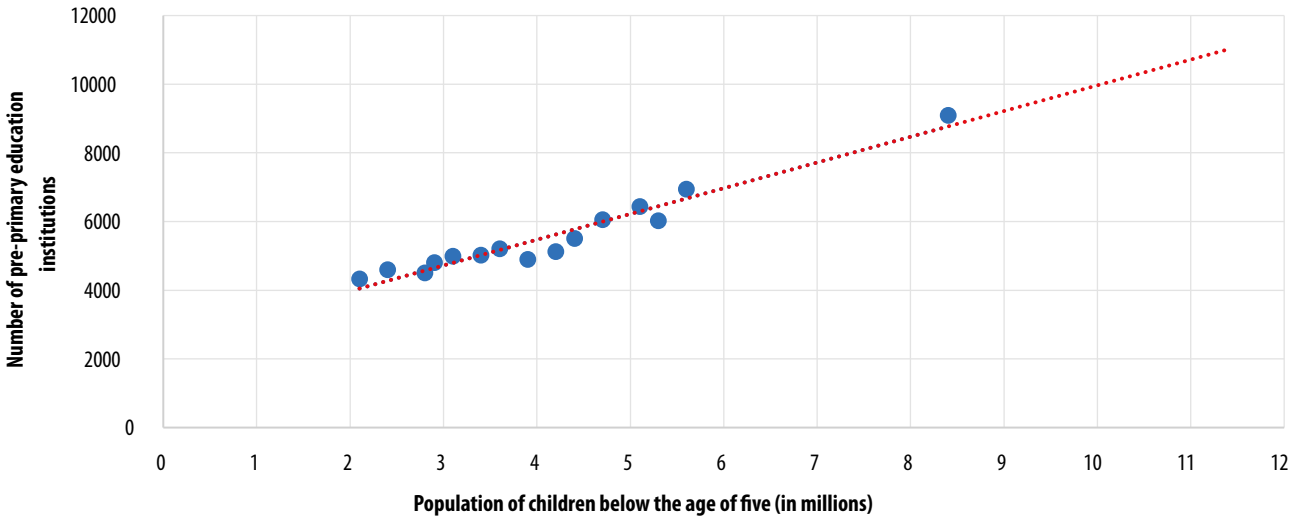
Figure 13: Example 1 of forecasting the number of pre-primary education institutions based on district information



Note: Hypothetical data

The value of the presumed dependent variable Y can be calculated by the function describing formula 'y = 749.2x + 2475.2.' If the population of a district were to increase to 11 million children under the age of five, we can expect to find about 10,716 pre-primary education institutions.

Figure 14: Example 2 of forecasting the number of pre-primary education institutions based on annual statistics



Note: Hypothetical data

The value of the presumed dependent variable Y can be calculated by the function describing formula 'y = 40.543x + 4890.5.' If the total population of children below the age of five were to increase to six million over time, we can expect to find about 5,134 pre-primary education institutions.

3.2.5 Multivariate analysis

Causal model

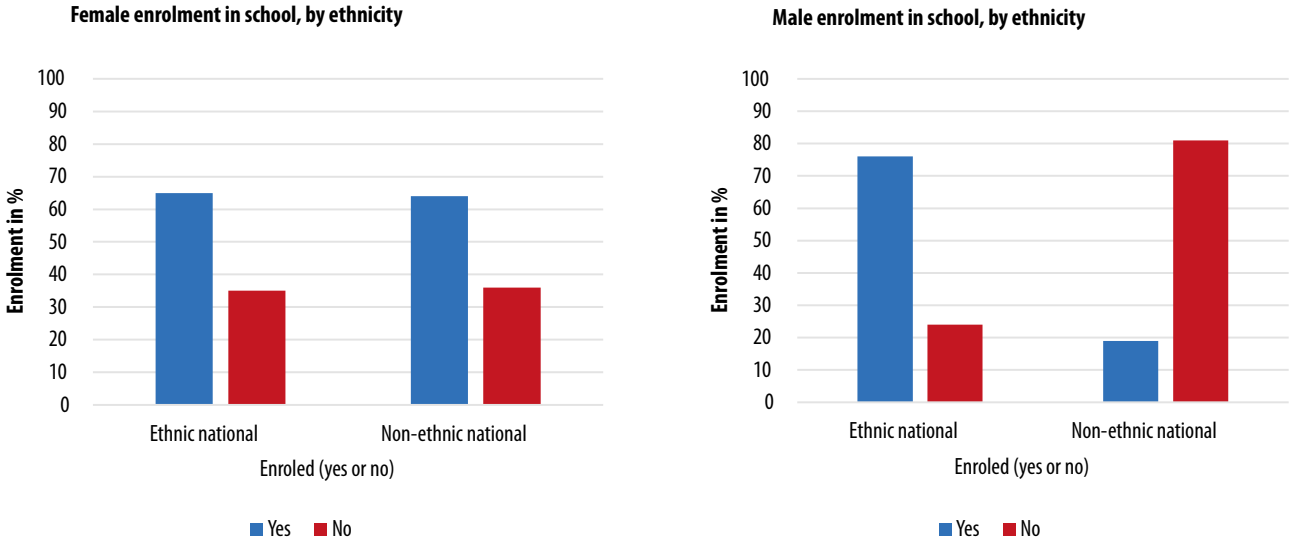
There are a variety of complex mathematical methods of analysis to test whether and how several variables influence one another. For the purposes of monitoring education (unlike research in education), mathematical modelling of multidimensional variables is rarely, if ever needed. There is, however, a technique that is appropriate in exploring, explaining and presenting multivariate analysis.

CROSS TABULATION of two variables differentiated by a third variable is a simple way of analysing whether or not a third variable influences a relationship between two variables. This method requires having identified a previous relationship between two variables, for example, through a preceding correlation analysis.

For example, if we are controlling for the effect of gender on the relationship between enrolment in school and ethnicity, we will produce two partial tables, one for male respondents and another one for the female respondents. While we can never be absolute certain about an underlying cause in multivariate analysis, it will allow us to assume a suggested relationship.

Figure 15: Fictional example of a cross tabulation of three variables and their representation

Gender, by ethnicity status / Enrollment status		Females			Males		
		Ethnic national	Non-ethnic national	Total	Ethnic national	Non-ethnic national	Total
Enrolled	Yes (%)	65	64	64.5	76	19	48.7
	No (%)	35	36	35.5	24	81	51.3
	Total (%)	100	100	100	100	100	100



Note: Hypothetical data

Ideally, analysing multiple variables, we are enabled to identify at least tentative causal relationships. In the example above, a causal relationship is suggested with the variable gender in relation to ethnicity and enrolment.

Certain causal effect can be interpreted as follows:

IRRELEVANT EFFECT: When the anticipated result shows no difference for a population disaggregated by two (or more) variables the causal effect is irrelevant.

CONCURRENT EFFECT: When the effect remains the same for a population disaggregated by two (or more) variables, the causal effect is concurrent (e.g. boys have higher enrolment than girls in both urban and rural areas; gender and location have a concurrent causal effect).

CONDITIONAL/INTERACTION EFFECT: When the effect for one population remarkably changes in the face of a second (or more) disaggregation, we speak of a conditional, or

interaction effect (e.g. the difference for boys is magnified with a higher enrolment in urban areas; for girls in the rural areas).

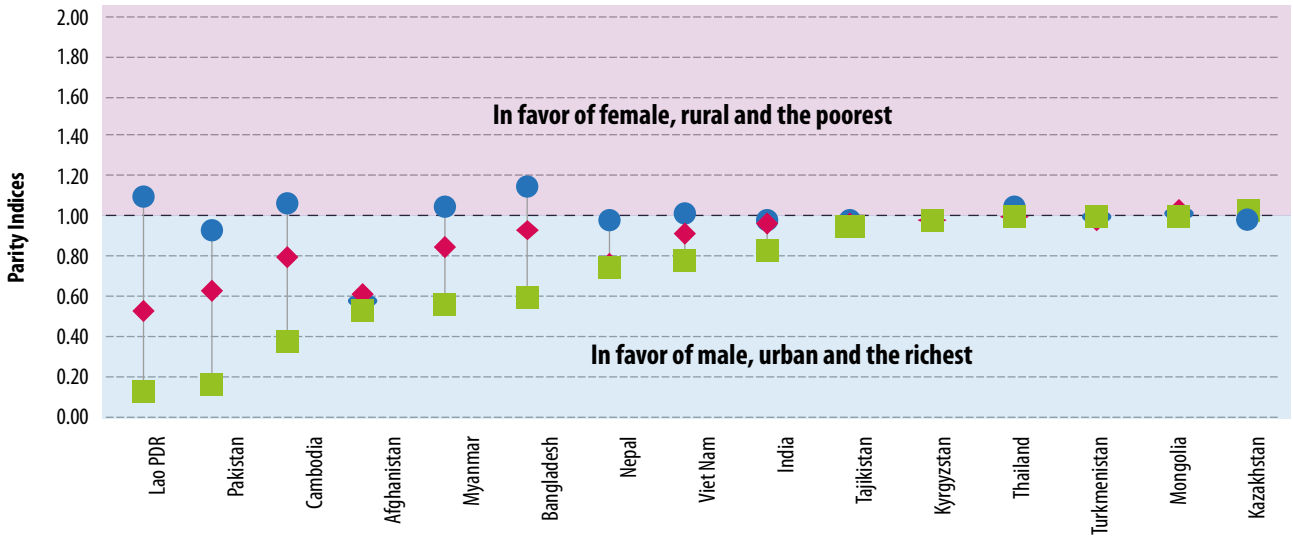
INTERVENING/COMMON EFFECT: When a population, already distinguished by one variable (e.g. gender) shows differences only for another variable (e.g. location), we speak of an intervening or common effect (e.g. the location intervening on the effect of gender).

3.2.6 Equity Analysis

Equity analysis by itself is not a statistical technique but a focus area that employs previous statistical methods to highlight equality, or inequality between opposite populations. Equity itself is a concept, and it can be analysed as such by different means. The above example chart on literacy has already shown one method to analyse equity – or the missing thereof – between groups.

A common approach in monitoring education, as well as social development, is comparing **parity indices** by gender, wealth, location and also migrant, disability and ethnicity status, among others. The parity index is limited to a range between '0' and '2'. The wider the distance from one, either above or below, the greater the disparity between the two values. A gender parity index compares females and males; a location parity index compares rural and urban locations and a wealth parity index compares the poorest 10 per cent of a population and the richest 10 per cent of a population. A value of less than '1' represents disparity in favour of the category in the denominator (males; urban; richest). A value greater than '1' represents disparity in favour of the numerator category (females; rural; poorest), as seen in the example below.

Figure 16: Net attendance rate parity indices for gender (blue), location (red) and wealth (green) at the lower secondary level, 2016



Source: UNESCO, 2018: Paving the Road to Education. A Target-by-Target Analysis of SDG 4 for Asia and the Pacific. Bangkok, UNESCO.

Analysing equity can be approached from several more angles: meritocracy, minimum standards, impartiality, equality of condition and/or equality of distribution. Analysing equity, in the end, depends on the actual measure.

For the purpose of this training manual, the different equity angles will find a brief description. At the end, we will provide you with a link to the Handbook on Measuring Equity in Education for further reading.

MERITOCRACY means that educational opportunities are distributed on the basis of merit, which is commonly applied in the education system (in particular higher education). Examinations are applied to measure merit, although they cannot serve to evaluate a student's real ability.

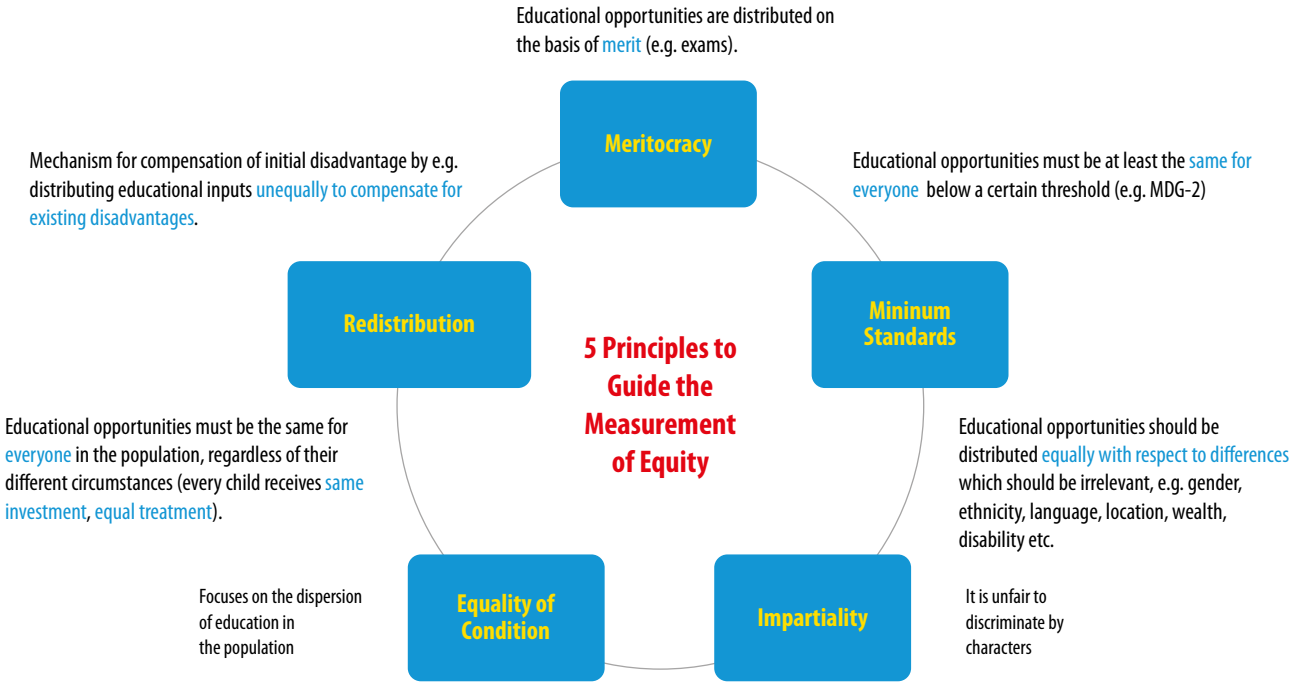
MINIMUM STANDARDS look at whether or not a condition is true for groups, such as being enrolled, or to have completed primary education. This measure is usually taken when an agreement was made on a least condition that must be fulfilled by everyone (e.g. access to free education).

IMPARTIALITY quantifies a relationship between an education indicator and one (or more) measures of circumstance (e.g. wealth, gender, ethnicity, etc.). Perfect impartiality is given when any statistical relationship with education is absent. Impartiality can be analysed with aforementioned tools of gap, ratio, correlation, regression, standard deviation and other tools.

- Ratios are an appropriate and easy way of showing differences between groups, when at least one variable is ordinal or binary.
- When at least two continuous variables are to be compared (e.g. years of education and wealth), correlation or regression analyses are typically applied.

Measuring the **equality of condition** looks at the distribution of a continuous variable across all persons. One visualization option for this method is the cumulative distribution function. This method serves to determine the share of a population for a given value. It is also applicable for test scores, years of education, or access to certain resources for a population.

Figure 17: Overview of equity concepts



Source: Adapted from UIS, 2018: Handbook on Measuring Equity in Education, **access:** http://uis.unesco.org/sites/default/files/documents/handbook-measuring-equity-education-2018-en.pdf

For more information on equity analysis, see:

Handbook on Measuring Equity in Education[16]

For analysis specifically for children with disabilities, see:

Students with Disabilities, Learning Difficulties and Disadvantages, Statistics and Indicators[17]

[16] UIS, fhi360, Oxford Policy Management and Research for Equitable Access and Learning (REAL), 2018: Handbook on Measuring Equity in Education. Montreal, UNESCO Institute for Statistics, access: http://uis.unesco.org/sites/default/files/documents/handbook-measuring-equity-education-2018-en.pdf

[17] Robson, Colin (2005) Students with disabilities, learning difficulties and disadvantages: statistics and indicators. OECD: Organisation for Economic Cooperation and Development, France. Access: http://eprints.hud.ac.uk/id/eprint/464/1/RobsonStudents.pdf

Box 4: Ethical code of conduct in research and analysis

The analysis and reporting of findings must always avoid exaggeration, or plain misrepresentation of data. Analysis and reporting must be executed with honesty and rigour, as otherwise wrong information bears counterproductive national planning.

- Reports should always draw attention to the limitations of the results that have been analysed with regards to reliability and in applicability.
- The significance of results must not be exaggerated, nor misrepresented.
- Data must not be fabricated, falsified nor intentionally misrepresented to allow for concluding with desired recommendations.
- Analysis must be reported fully without omission of significant data, disclosing details of undergirding analytical methods which might bear upon interpretations of any findings.
- Concepts, procedures and results must be presented in sufficient detail to allow others to understand and interpret the present information equally.
- The misuse of findings and misunderstanding of their scope and limitations must be acted upon.
- Public controversy must not be shunned if findings indicate unfavourable results, because it stimulates positive progress.

Source: Australian Association of Research in Education, 1995: Code of Ethics for Research in Education, Victoria, Cold Stream.

4 Communicating the Data Analysis Results

All the effort of data collection and production is to support management and leadership to make good decisions on education planning, policymaking and resource allocation. Openly structured and communicated data makes authorities and community stakeholders accountable for effective management of the education sector[18]. It is important that the data produced is understood by different audiences at different levels. Therefore, data communication is always a very important aspect succeeding the data analysis process.

In the previous sections, we have already seen many examples of data presentations. This section will delve into the theoretical elements of data presentation to understand why we choose certain presentation methods and not others.

4.1 Presenting data to an audience

When preparing a presentation of data and indicators, we should ask ourselves the following questions:

- What am I trying to communicate?
- Who are my audiences?
- What kind of presentation will be most effective?
- What might prevent my audience from understanding the data in the presentation?

Every presentation technique has its own characteristics, advantages and disadvantages; but the choice will depend upon the context within which the presentation is delivered and the message to be conveyed. For example, a chart may highlight distributions and trends visually with immediacy, that is easy to grasp, but it may fail to explain underlying patterns, causes and effects and these can be easily manipulated. Descriptive text can draw attention to salient findings and possible causes and implications. In visual representation of analyses, it is therefore crucial to apply consistency of style and logic to minimize misinterpretation.

The goal of a visual presentation is to make the reader want to interpret it and understand the messages with minimum effort. As a rule of thumb, when creating a

[18] UNESCO-IIEP, 2018: International Policy Forum in Manila, Philippines, Recommendations to promote transparency and accountability, access: http://www.iiep.unesco.org/en/10-ways-promote-transparency-and-accountability-education-4307

visual representation of data, it should follow the following style rules:

- **INFORMATIVE;**
- **SELF-EXPLANATORY;**
- **PLEASANT IN APPEARANCE;**
- **EASY TO UNDERSTAND.**

Always try to communicate with your audience in the most simple, direct and efficient manner possible. Do not overload presentations with too many tables, charts, numbers and text. Identify the findings and ideas and then present them in the simplest possible way. The details are finally for the audience to be accessed in the background report from which the visual presentation has been drawn.

4.2 Tabular presentation of data

Tables are a systematic arrangement of words, numbers or signs, in parallel columns, to depict and summarize statistical data for eventual relationships. They are good for presenting large amounts of numerical information that would otherwise be confusing to describe in text.

A table should facilitate subsequent interpretation of the presented information, especially when various parameters in two or more groups are to be compared. The following section gives a guideline to constructing a table with the basic components so the information can be presented in a manner that is concise, clear, direct and effective.

TITLE: The title is the main description of the table. It should be concise and for the sake of interpretation and record keeping, informative and meaningful. The title should include a date or year of reference regarding the data; the place such as a country, regions, provinces, districts, villages or schools; and any other attribute that is common to all the data entities in the table (see 'Title' in the example below). Note the title can be displayed embedded in the chart, or as the caption.

ROWS AND COLUMNS HEADINGS: Each row and column need a heading that describes the data in that row or column, using labels such as 'number of female students,' or 'percentage of youth and adults in TVET'. Headings usually vary from row to row and from column to column so as to distinguish between them. They may be re-grouped under a major common heading (see below Example Tabular Presentation of Data).

UNIT OF MEASUREMENT: A unit of measurement should be specified for every entry in the table, such as percentages, numbers in millions, or currencies. The unit of measurement is usually given in the title (if the same unit of measurement applies to the whole table), or in the column or row heading (if rows and headings use different units of measurement). To facilitate understanding and interpretation, it is advized to establish a convention used within the country, at best in line with established SDG 4 indicators and their specific unit of measurement.

DATA FIELDS: Actual numbers are included in the data field. Such data should correspond with the rows and columns. No data field should be left vacant. If there is

no data, codes are used to provide explanation for not having data, e.g., missing data with 'm'; not applicable with 'n/a', etc. While recording data, sometimes it is difficult to record the precise value. In that cases, numbers may be rounded to a nearby value using a degree of accuracy that is appropriate to achieve the goal of the presentation. When rounding numbers, unnecessary trailing zeros should be removed and the units of measure altered accordingly.

FOOTNOTES: Footnotes provide important information that help us understand the data in the table. The following information should be provided in footnotes:

- Source of the data which enables interested readers to pursue the underlying information and estimate the quality of the data;
- Conventions used and further explanations of terms found in the table where necessary;
- Complete annotation of row, or column headings, or the title if the labels are too long to insert in full in the table cells;
- Differences in status of some entries in the table;
- Data limitations;
- Any other exceptions/deviations from the stated norm (see 'Footnote' in Example below Tabular Presentation of Data).

Box 5: Tips for presenting data

- **Font style:** Different font styles may be used to highlight specific items of a table that require special attention.
- **The ordering of rows and columns:** This is critical for clarity. As a general rule, rows and columns should be arranged following a natural or logical order, ranked, for example, by alphabetical order, geographical location, year or magnitude. Alphabetical and geographical ordering are both useful for reference, whereas ordering by magnitude makes the ranking of the different entities immediately obvious.
- **Numbers:** Numbers are easier to compare when the table has a vertical orientation. The human eye can make comparisons more easily when reading down a column of data than across several columns.
- **Consistent appearance:** The appearance of the table should be kept consistent throughout the report. For example, conventions for labelling and ordering rows and columns should be kept consistent as much as possible. A common mistake is to switch, or mix up the rows and columns and labelling across tables within a report.
- **Number the table:** Tables should be properly numbered for ease of reference.
- **Unnecessary distraction:** Avoid designing complex tables with many layers of headings for row and columns. Break them into smaller tables. Many tables that should logically appear together should be placed in the appendices.

Figure 18: Example of tabular presentation of data

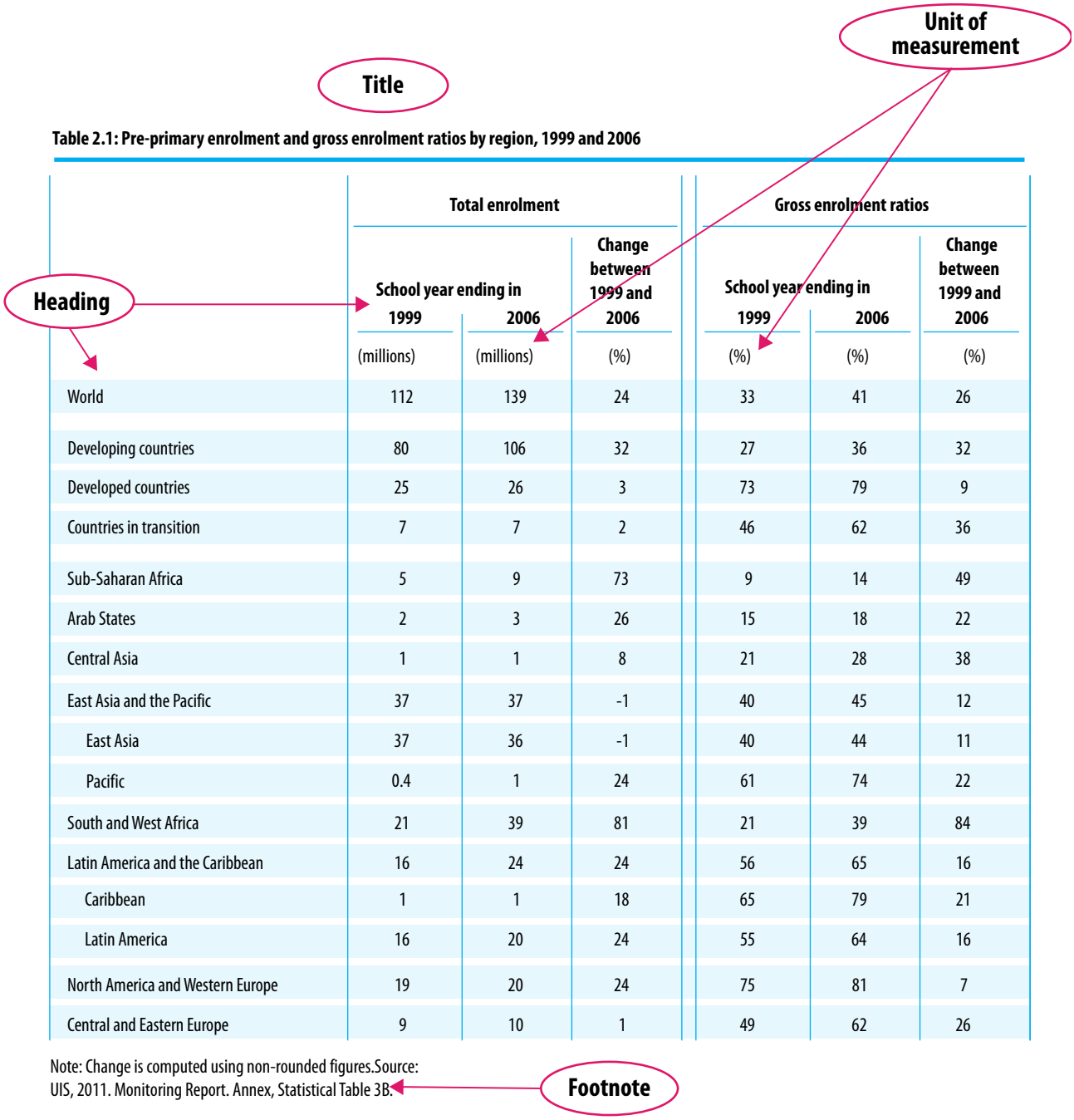


Table 2.1: Pre-primary enrolment and gross enrolment ratios by region, 1999 and 2006

	Total enrolment			Gross enrolment ratios		
	School year ending in		Change between 1999 and 2006	School year ending in		Change between 1999 and 2006
	1999	2006		1999	2006	
	(millions)	(millions)	(%)	(%)	(%)	(%)
World	112	139	24	33	41	26
Developing countries	80	106	32	27	36	32
Developed countries	25	26	3	73	79	9
Countries in transition	7	7	2	46	62	36
Sub-Saharan Africa	5	9	73	9	14	49
Arab States	2	3	26	15	18	22
Central Asia	1	1	8	21	28	38
East Asia and the Pacific	37	37	-1	40	45	12
East Asia	37	36	-1	40	44	11
Pacific	0.4	1	24	61	74	22
South and West Africa	21	39	81	21	39	84
Latin America and the Caribbean	16	24	24	56	65	16
Caribbean	1	1	18	65	79	21
Latin America	16	20	24	55	64	16
North America and Western Europe	19	20	24	75	81	7
Central and Eastern Europe	9	10	1	49	62	26

Note: Change is computed using non-rounded figures.Source:
UIS, 2011. Monitoring Report. Annex, Statistical Table 3B.

Source: UNESCO, 2011: Systemic Monitoring of Education for All. Training Modules for Asia-Pacific. Bangkok, UNESCO, **access:** http://uis.unesco.org/sites/default/files/documents/systematic-monitoring-of-education-for-all-training-modules-for-asia-pacific-en.pdf

4.3 Visual presentation of data

Charts with graphs can be easier to understand than tables and are often more effective for highlighting important information quickly for readers who are less apt with numbers. Charts are especially useful for:

- Faster understanding of numbers;
- Recognizing distributions in data, showing patterns and comparing trends;
- Easing comparing numerical information;
- Allowing information to be presented in various ways.

There are numerous ways of visually presenting statistical information. In short, the main purpose of charts is to visually impart information that cannot be easily read and interpreted from a table of data. In other words, the advantage of charts is that they are visually more attractive than tables, and can ease presentations. With the help of computer software packages, graphical visualization of data can be made in a variety of ways.

As mentioned, charts are not suitable for communicating detailed and precise information, and can be time-consuming and expensive to design.

The following section will explain a few charts and provide a structured overview (Figures 19 and 20) to help decide on which chart is appropriate to pick for which purpose.

4.3.1 Explaining when to use certain charts

LINE CHART: These are used to track changes over short and long periods of time. When smaller changes exist, line graphs are better to use than bar graphs. Line graphs can also be used to compare changes over the same period of time for more than one group.

BAR CHART: These are used for categorical data or metric data that are transformed into categorical data and are used to compare things between different groups, or to track changes over time Categories are shown on the horizontal axis. Frequency, percentage, or proportion is shown on the vertical axis. Bars are separated from each other to emphasize the distinctness of the categories. The bars must be of the same width. The length of each bar is proportional to the frequency, percentage, or proportion in the category. Levels ought to be provided on both axes. However, when trying to measure change over time, bar graphs are best when the changes are larger.

PIE CHART: These are best to use when trying to compare parts of a whole. They do not show changes over time. Like bar charts, pie charts are also used for categorical data. A circle is divided into segments, the areas of which are proportional to the values in the question. But the areas are proportional to the angles the corresponding segments make at the centre of the circle. Thus, segments of the circle are cut in such a way that their values are proportional to the angles.

AREA CHART: These are very similar to line graphs. They can be used to track changes over time for one or more groups. Area graphs are good to use when tracking the changes

in two or more related groups that make up one whole category (for example male and female students).

SCATTER PLOTS: These are used to determine relationships between the two different things. We have already seen this type in regression analysis. The x-axis is used to measure one event (or variable) and the y-axis is used to measure the other. If both variables increase at the same time, they have a positive relationship. If one variable decreases while the other increases, they have a negative relationship. Sometimes the variables don't follow any pattern and have no relationship.

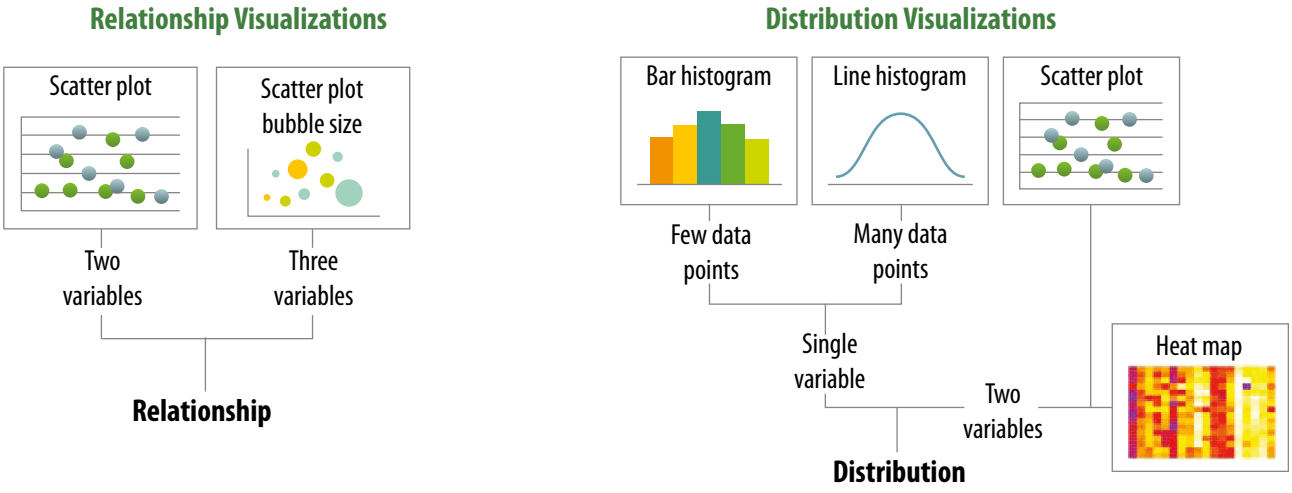
DOT PLOTS: These allow for comparison of one, or several groups by different characteristics. We have seen an example in the gap analysis. This representation allows for an easy comparison across selected groups and their background characteristics.

DISPARITY GRAPHS (OR THREE GRAPHS): This type of graph is used to present disparities of performance across different population groups for one particular indicator. In the following example, the visual representation allows for comparison between the overall national average, and the most disadvantaged subgroup considered (girls from poorest 20 per cent) as well as the most privileged group (boys from the richest 20 per cent).

GEOGRAPHICAL MAPS: Maps are a great way to represent data/indicators which vary across different geographical areas, in particular to advocate for more equitable distribution of resources towards the most disadvantaged, or to target support where it is the most needed. They can be a powerful advocacy tool to present information to decision makers for geographic targeting. Maps can be created at different administrative levels, and therefore can serve the same purpose at various levels of governments.

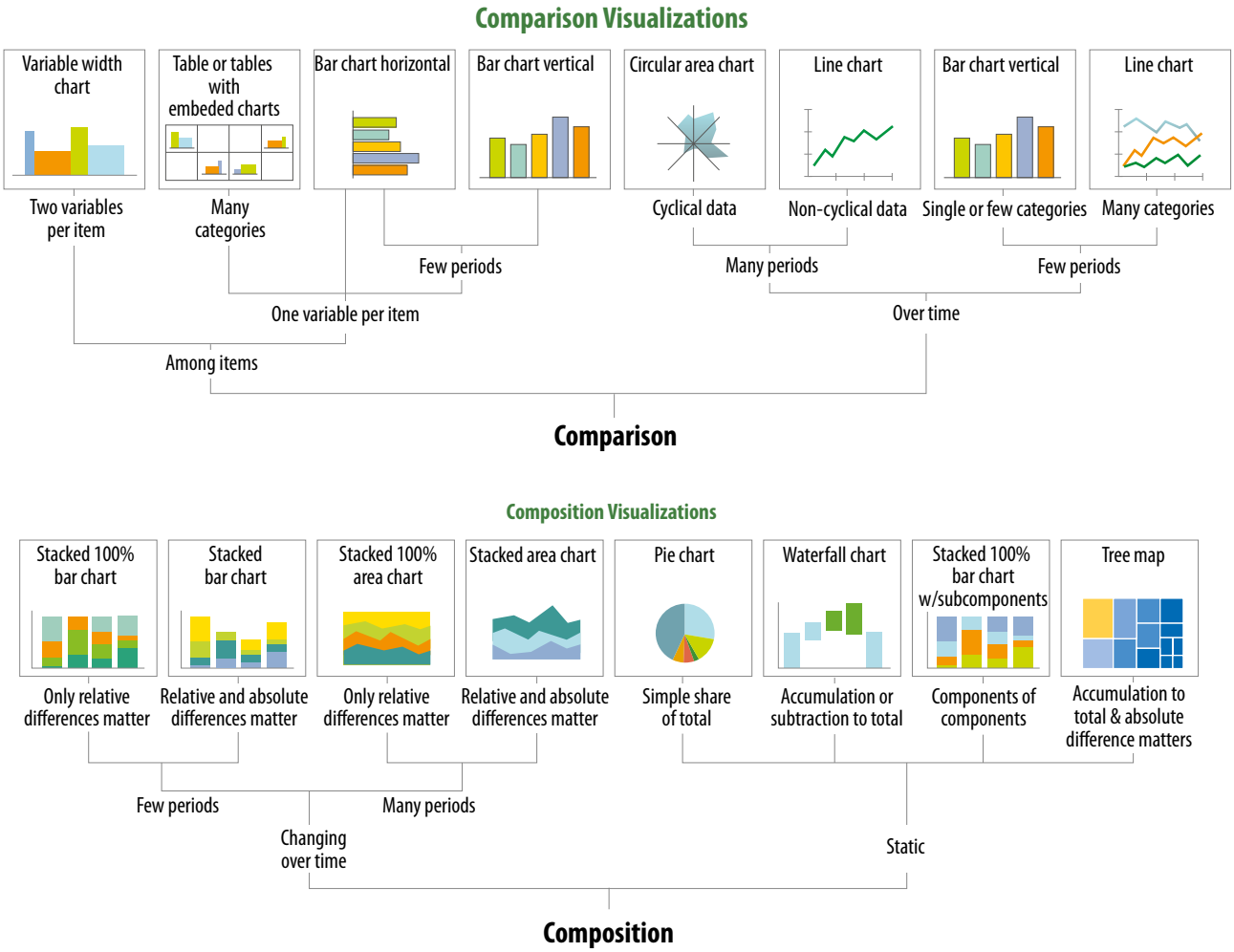
HEAT MAPS: The term 'heat map' refers to displaying colour scaled representation of values on a map, similar to the above geographical map. The colours here are, however, not primarily used to distinguish between districts but to indicate a degree of performance, or shortcomings.

Figure 19: Chart examples appropriate for specific purposes



Source: Adapted from Big Data, 2016: The Four Pillars of Visual Analytics, **access:** http://bigdata.black/featured/the-four-pillars-visual-analytics/

Figure 20: Chart examples appropriate for specific purposes



Source: Big Data, 2016: The Four Pillars of Visual Analytics, **access:** http://bigdata.black/featured/the-four-pillars-visual-analytics/black/featured/the-four-pillars-visual-analytics/

4.3.2 Basic components of a chart

Similar to tables, charts should have a title, axis labels (including the units of measurement), tick marks on the axes (with labels for some tick marks and sub-groups) and footnotes and references to the source data.

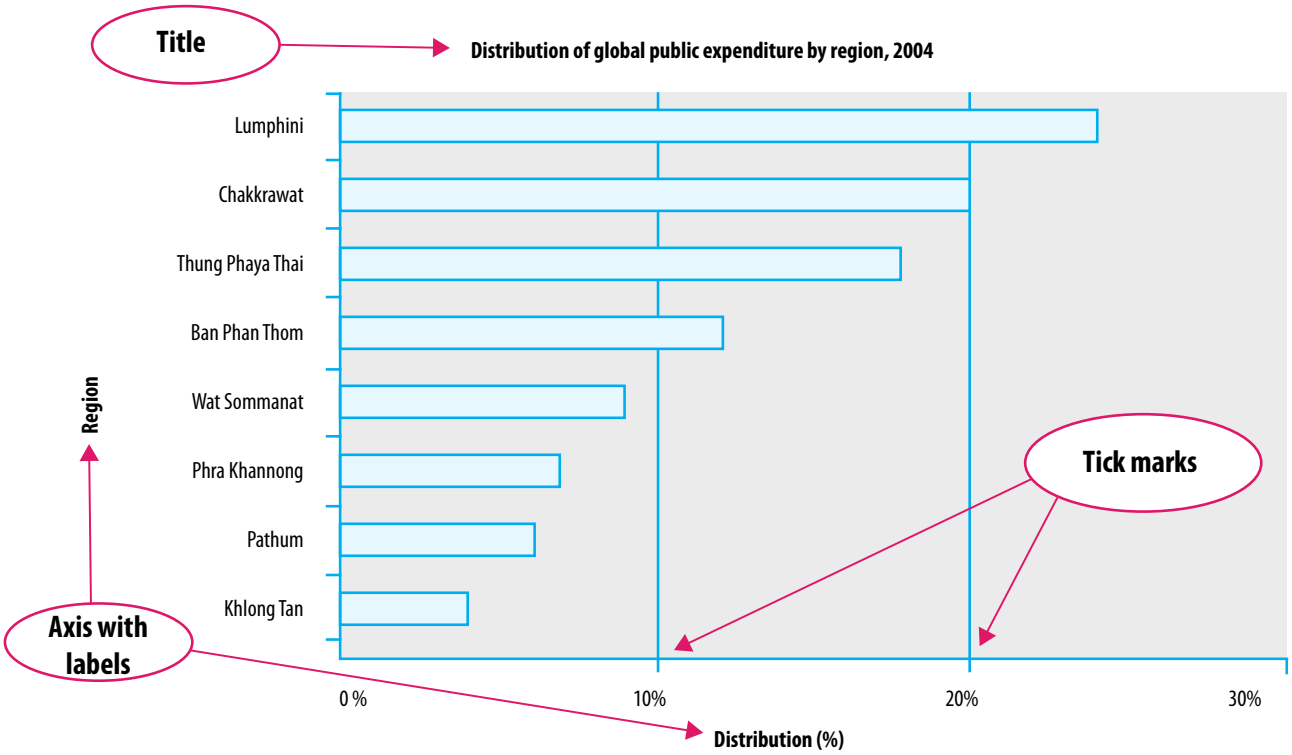
An explanatory title, labels and footnotes are essential for understanding and interpreting a chart. One must nevertheless bear in mind that too many text details and labels can distract the reader from the main message.

An effective chart also has the following characteristics:

- Clear objective and messages to be presented;

- Good choice of graph type for the information that is to be presented;
- Appropriate level of simplicity or complexity, depending on the readers' abilities to analyse, interpret and understand.

Figure 21: Example of a Graphical Presentation of Data



Note: Distribution is calculated using PPP US$.

Source: UNESCO, 2011: Systemic Monitoring of Education for All. Training Modules for Asia-Pacific. Bangkok, UNESCO, **access:** http://uis.unesco.org/sites/default/files/documents/systematic-monitoring-of-education-for-all-training-modules-for-asia-pacific-en.pdf

4.4 Strategic Dissemination

The dissemination of education data is as important as the collection of data. While our data collection and curation serve as the foundation on which to base a decision for policy makers, planners and programme implementers, the resultant information is also relevant for other stakeholders, such as educational institutions, the press, research scholars, research institutions, academia, relevant NGOs, community organizations and citizens in general.

Only when all stakeholders are informed properly can they make adequate decisions that help improve a situation. Ill-fed citizens are unlikely to engage in improving, for example, school conditions by reporting ineffective use of resources. The following table will

provide you with an overview of the possible communication channels to make use of to reach specific audiences.

Table 4: Suggested communication channels for specific audiences

Government officials and policy makers	General public
• Dashboards; • Face-to-face meetings; • Policy briefs, brochures and executive summaries; • Websites	• Magazines; • News media; • Radio and television; • Web-based media (e.g. Social Media); • Websites; • School report cards.
Programme managers and implementers	**Technical and development agencies**
• Monthly/quarterly reports; • Executive summaries; • Audio-visual presentations; • Infographics; • Public websites.	• Full annual reports; • Audio-visual presentations; • Public websites; • Brochures.
Civil society members and organizations	**Academic researchers**
• Fact sheets; • Infographics; • Brochures; • Websites.	• Technical reports; • Special topic articles; • Research databases; • Websites.

Source: Adapted from UN ESCAP Statistics Division, 2019: Stats Brief. Dissemination and Communication of Vital Statistics. Issue No. 17. Bangkok, UN ESCAP.

4.4.1 School report cards – feeding back to the community

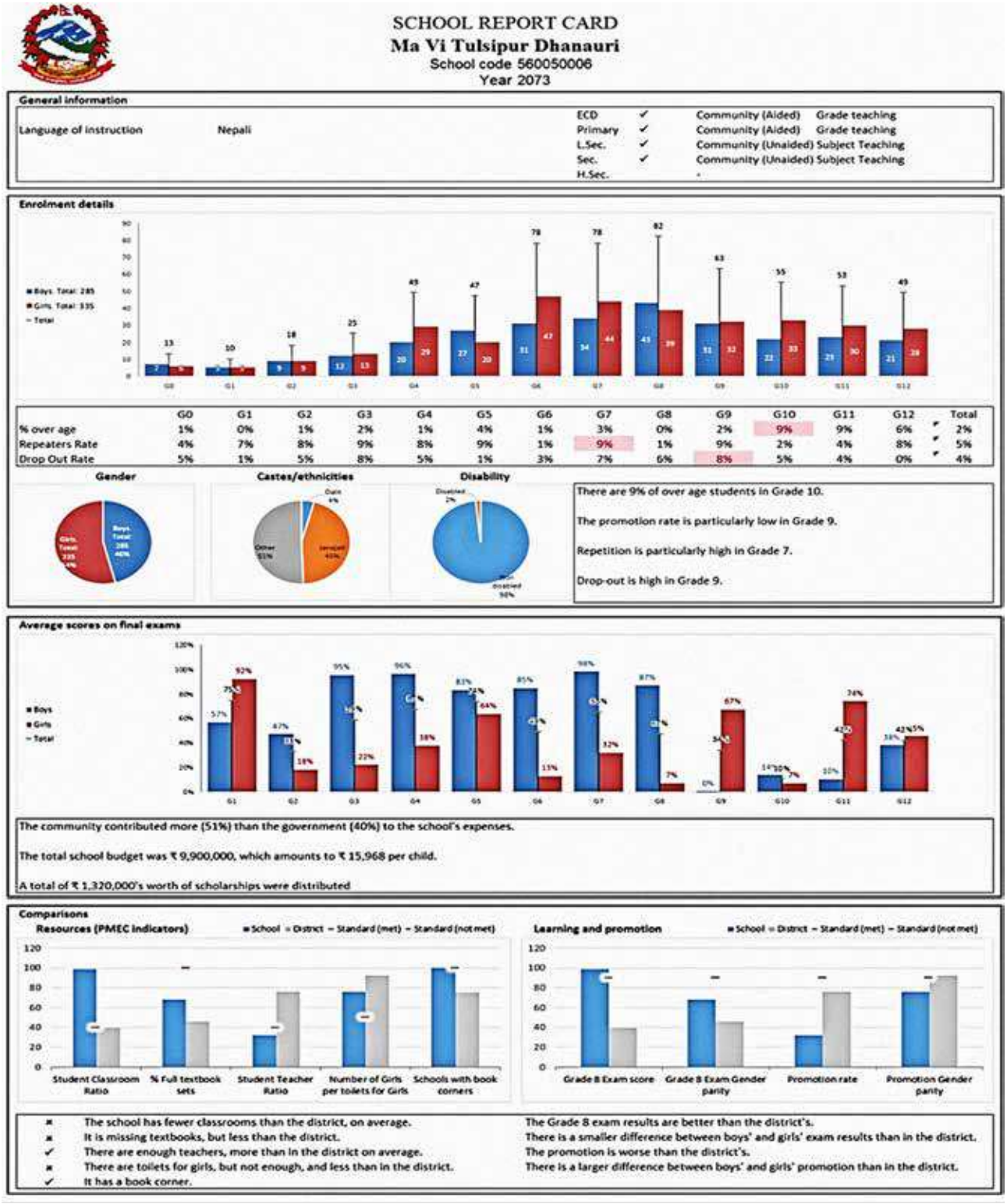
It is worth noting at this point that data should be disseminated at levels as close to their collection as possible. This is the reason for involving the community level, particularly the community around the points of data collection, such as schools, to gain and maintain support in monitoring education reforms. In the end, when we submit data about our circumstances, we would want to see the data being used and having an effect in some form. School report cards provide several functions:

- Support school management function by providing regularly, automatically updated data snapshots about enrolment, participation, learning, teachers, school improvements etc. for each school.
- Improve the feedback loop from the national EMIS database and the schools.

- Improve communication about school-level performance, to increase community awareness as well as hold the schools accountable for improving results as per plan.
- Increase community participation and develop a sense of 'ownership' among community members.

As EMIS collects a wealth of data from schools, it has the capacity to create school report cards automatically and thus, facilitate the dissemination of information back to the school and community. The following example allows for a compact overview of most, if not all, important characteristics.

Figure 23: Example - school report card from Nepal

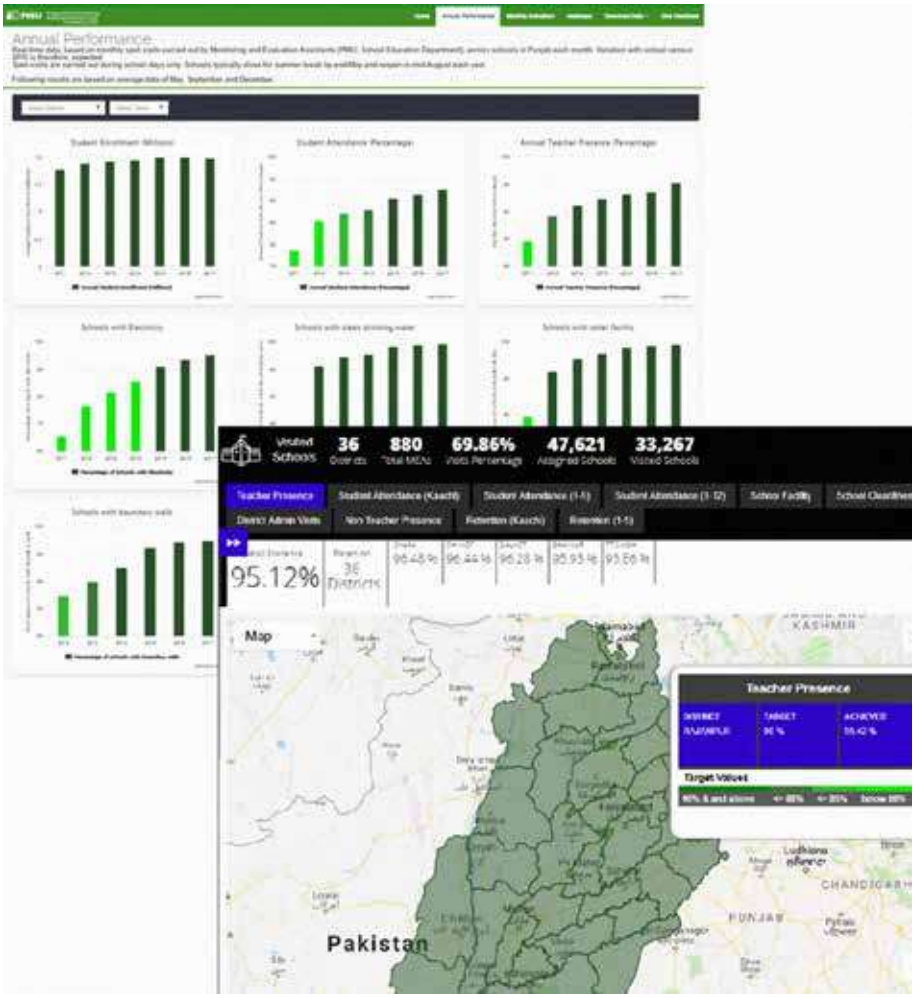


Source: Presentation by the Ministry of Education in Nepal on the lessons learned and outcomes of the collaboration under the Data Must Speak initiative. Prepared for the GPE Webinar, 27 July 2017, Kathmandu, Nepal, **access:** https://www.unicef.org/sites/default/files/2018-08/General%20Presentation%20Nepal.pdf

4.4.2 Supplementing with digital information

As we more and more move towards distributing information through online and mobile platforms to reach families and communities, locally and internationally, at this point we may want to consider if we can supplement or strengthen paper-based school report cards with digital school report cards that can be accessed by all education stakeholders with mobile and computer devices if and when needed.

Figure 24: Example for free, open and real-time digital reporting of school information, Punjab, Pakistan



Source: Programme Monitoring and Implementation Unit, **access:** https://open.punjab.gov.pk/schools/

For a planning guide on online school report cards and the steps to keep in mind, see:

Building Online School Report Cards[19]

Case Study: Unified District Information System for Education and Student Database Management Information System - India[20]

As schools in India suffer from accountability deficits, teacher absenteeism, school dropouts and poor learning outcomes, the government developed the Unified District Information System for Education (U-DISE) as one response to the deficits by means of proliferating data on school systems. The U-DISE is the single largest EMIS database on currently available information on schools and education in this vast country. In 2010, school report cards under U-DISE won a number of awards for bringing higher accountability and capturing education data. U-DISE support is available online, including downloadable software for database organization and instruction manuals, among others. A report module generator enables the selection of specific indicators to generate tailored reports on:

- School profile: management, sources of funding, school type, language of instruction, etc;
- Enrolment and repeater information: by age, sex, social class, etc;
- Teacher provisioning: availability, qualifications, teacher training;
- Infrastructure and learning facilities: physical conditions, water, electricity, etc.;
- Examination results;
- Receipts of school grants;
- Compliance information on the constitution and functionality of school management committees;
- Time of operation: number of instructional days in the school, operating hours of schools;
- Number of remedial classes provided to students;
- Availability of educational resources: textbooks, uniforms, and other provisions guaranteed under the Right to Education Act.

[19] Building Online School Report Cards, access: https://www.excelined.org/wp-content/uploads/2018/08/ExcelinEd.AStateGuideToBuildingOnlineReportCards.May2017.pdf

[20] All U-DISE related data/information, activities, revisions, and events are available on a NIEPA-designed portal, access: http://udise.in/index.html. The accompanying school report card website, access: http://schoolreportcards.in/SRC-New/

To ensure access for those with low literacy skills, instructions are sent to each state that the U-DISE school report cards are to be read out to the local community during teacher/parent meetings for transparency, engagement and accountability purposes.

Expansion of U-DISE on academic performance:

In a bid to keep an annual record of academic performances and basic individual details of all the students across their school education years in the country, the government developed a Student Database Management Information System (SDMIS) and this was launched in 2016.

SDMIS collects additional information on:

- Student entitlements: textbooks, uniform, transport, scholarship, etc;
- Student achievements: attendance, examination-related information, etc;
- Student profile: gender, religion, disability, mother tongue, health, below-poverty-line status and financial status of the child, among other variables.

SDMIS can automatically calculate enrolment ratios (gross enrolment ratio and NER), and flow rates (dropout, repetition, promotion, retention, transition, survival rates, etc).

For more information, see: Bordoloi, Mridusmita and Varun Kapoor, 2018: Mridusmita Bordoloi and Varun Kapoor. Paris, UNESCO-IIEP. http://unesdoc.unesco.org/images/0026/002659/265933E.pdf

Lessons to learn

As no system is ever perfect, even a comprehensive system like U-DISE in combination with SDMIS from the above case study requires adjusting collection, processing and dissemination with all stakeholders. Intergovernmental coordination and interoperability across databases are key to effective and efficient data collection, processing and dissemination. Common problems have been:

- When data is collected annually on a particular date of the year, sub-national agencies may create parallel real-time management information systems with different databases, independent of one another.
- Data on different education aspects may end up being collected by a number of different departments, which in turn results in data silos.
- The subsequent detailed verification and validation processes of the data from different departments and the scale of the process and strict timelines results in data quality issues, either due to schools submitting erroneous data, or the processing being subject to errors.
- Furthermore, large-scale data processing results in time gaps of up to one year from the moment of schools submitting the data and school report cards being available to the public.
- When subnational agencies, including schools, are tasked to submit budget

requests in line with their reported school performances, the central government must pay attention to the budget requirements made by the agencies and schools. Otherwise, the purpose of collecting the data and cross-verifying performance and requirements are obsolete.

Case Study: The benefit of feeding back data analyses to the community - Bangladesh report cards

As a citizen-led initiative, Transparency International Bangladesh (TIB) strives to involve parents in education through useful school data published in leaflets, information boards and desks, interactive discussions at mothers' gatherings and meetings with authorities. The TIB initiative discloses crucial information pertinent to SDG 4 monitoring, as for Target 4.1 on learning outcomes and a free-education proxy, and for Target 4.a on safe learning environments as well as for Target 4.c on teachers.

TIB's survey generates data on school funding, the condition of school facilities, teachers' qualifications, teachers' behaviour, school management, students' learning outcomes, undue payment of extra fees, corporal punishment, among others.

TIB creates baseline survey data, develops reports, leaflets and other information material purposefully made accessible to the local community through e.g. parental school meetings. This information is also shared with the local and national education authorities.

The TIB initiative collects data directly from parents, teachers, the school management committees and education officers. TIB also generates information from the websites of the Ministry of Primary and Mass Education, DPE, and the Bangladesh Bureau of Educational Information and Statistics, besides from newspapers, policies and reports. TIB staff at the national level are responsible for crosschecking and guidance to ensure data accuracy.

The approach that TIB has taken has led to side benefits that have increased the usability of the collected data. Feeding the collected information back to the community, especially parents, makes them participate in the data generation process more actively.

Moreover, due to the initiative making the processed data available through school meetings and by means of accessible information products, the local community (parents) has been able to understand the issues of improving school and learning conditions; and in this process serve as a self-regulatory inspection mechanism on, for example, the quality of teaching, the attendance of teachers, the maintenance of facilities, accurate student number reporting and of course keeping children in school. **For the education authorities, this intervention functions as an additional monitoring mechanism that increases accountability and transparency in education.**

Parents of non-intervention schools, on the other hand, were unable to understand the value of open data and thus failed to raise questions over quality education, or school management issues related to open school data with the eventual consequences for low quality education.

Source: Roy, Dipu and Abu Said Md. Juel Miah, 2018: Bangladesh: using open school data to improve transparency and accountability, Paris, UNESCO-IIEP. Access: http://unesdoc.unesco.org/images/0026/002659/265930e.pdf

4.5 Audio-visual media tools: infographics, animations and documentaries

Various representations can be applied, i.e. concept maps, graphic organizers, flow diagrams, simulations, pictograms to visualize information so that it can be relayed to the audience in a more effective and efficient manner. Borkin et al. (2013) studied visualization types and relationships with memory level. They claimed that visuals containing pictograms may be remembered better than other visual means and moreover infographics are the highest-level recalling type of visual means[21].

Infographics are an essential part of the survey analysts' tool box because they convey complex data in an easy to follow and visually appealing format. From blog posts and web articles to glossy brochures and of course, data analysis presentation, infographics are a ubiquitous part of the information landscape.

Another step up from creating infographics are animations in form of videos. Short, animated videos can be a powerful communications and advocacy tool that captivates the mind of the viewer with moving visuals and easy to absorb narrative. Easily shareable through social media, this can really help to spread a message, raise awareness of issues, or efforts undertaken to address them.

Another valuable tool are videos in a documentary style. We all have seen documentaries, be they on wildlife, or manufacturing processes. Documentary style videos have also been applied in education advocacy, especially to communicate the importance of education.

Combined with statistical information, documentaries tell a real story that can inform crucial information with real life examples and these are reusable whenever needed, while maintaining a consistent message. The overarching goal for these media formats is to create simple public information that is intelligible for the masses.

4.5.1 Characteristics of audio-visual media

Whether infographics or videos, they are typically designed by graphic designers, animators and video editors in collaboration with a communications expert of an institution, as well as a thematic sector expert to create specific visualizations to be used either online, or for posters, banners, public screenings or any other type of communication purpose.

They often involve icons, characters or symbols which are very visual and aim at reducing the amount of text to be read. Here are three reason why audio-visually appealing media should find consideration for statistics dissemination:

- Audio-visual media is easier than ever to create in today's digital age.

[21] Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H., 2013: What makes a visualization memorable?. In: Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2306-2315.

- Decision makers have faster access to more data than ever before.
- It is easier to look at visual media than reading long text.

4.5.2 The basic steps of creating audio-visual material

Like every initiative, creating an infographic, video, or any other kind of presentation material follows phases of planning and execution. As this is a highly creative process, it will require additional expertise.

This will also require additional cost, unless you are talented in media creation yourself and have the necessary tools at your disposal. Otherwise, perhaps someone in your department has a background in traditional, or social media, graphic design, video editing, photography, or animation who may help you in this process.

In any case, it is advisable to consult a professional in the aforementioned areas as they not only have the required creative mind-set but also the required media creation tools. Creating media for the public is always a collaborative effort of experts of the relevant disciplines.

Creating such media is not as simple as it sometimes looks. The following section includes an insight into the general creation process, with information on how you can make the necessary preparations for creating audio-visual materials.

1. **CONCEPTION:** Outline the objectives for your material. What is your theme and what is the central message you want to convey?
2. **PLANNING:** Determine what statistics and identified issues and recommendations you have at your disposal. If necessary, process raw data to obtain the needed statistics. Also, create a brief narrative that helps you order the statistics and the key points to raise that serve to explain your central message.

   For example, outline on paper in writing and in sketched pictures a sequence. For an image like an infographic, draw where and in what order your information should be displayed.

   For a video, create a sequence-by-sequence description (this may include visual sketches of yours) which includes the data and key points to talk about. This is called a 'storyboard'.
3. **EXECUTION:** Contact an expert who is proficient in creating the desired material for your desired product. It is best to request some sample ideas to help you decide what tone your material should take (colourful vs monotone, happy vs serious, elegant vs playful, etc.).

   Depending on your desire, you will receive a set of examples to choose from comprised of images, font styles, sounds (if necessary), layouts (where applicable) and other details. Ensure communication about all the steps involved in the creation process to avoid ending up with a product not of your liking, or displaying the wrong message.

4. **PROOFING:** Proofread the final material. Make sure everything is not only spelled correctly, but ensure the messages are exactly as you want them to be without any creative changes (as the designer/creator may not be an expert in your field, or he or she may likely misinterpret certain messages).

   Also ensure that any external data, or information is appropriately referenced somewhere in your material to ensure transparency and accountability of all information.

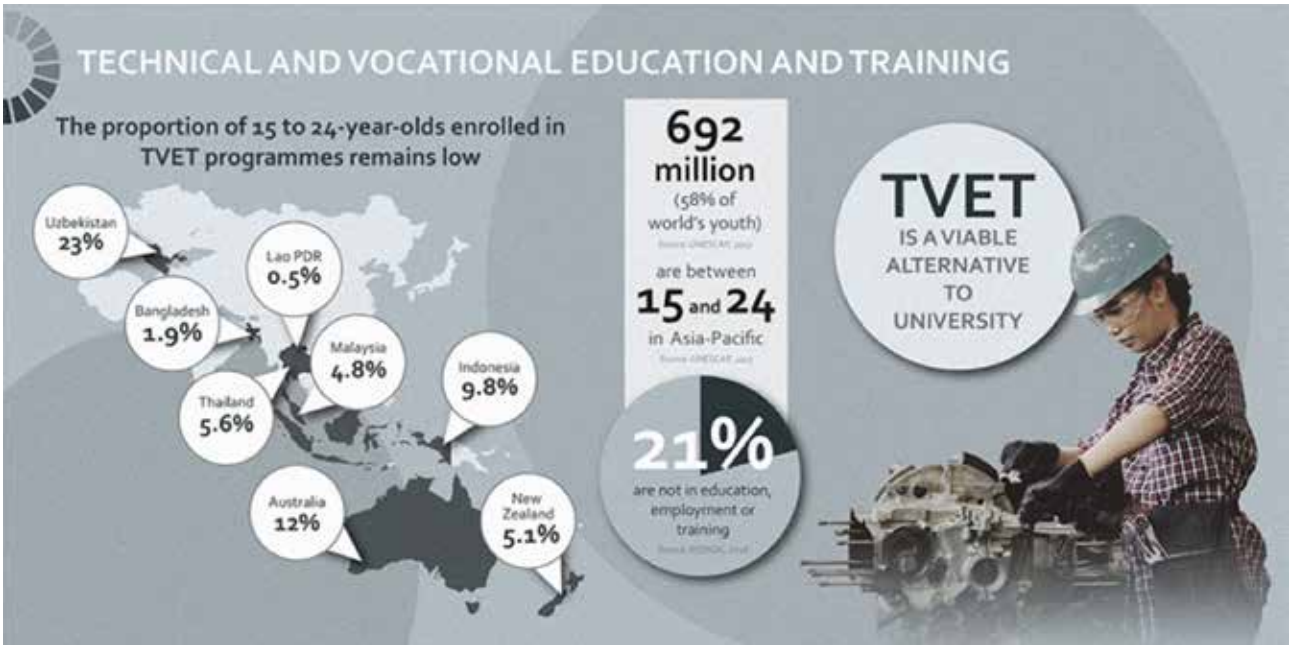
   Likewise, promote your own data source if the data has been collected and curated by yourself, or your department.

Irrelevant of the medium, a central part is to have a good story around the data presented with a focus on who, what, where, when and how in a language the target audience will understand.

4.5.3 Examples of effective audio-visual material for public dissemination

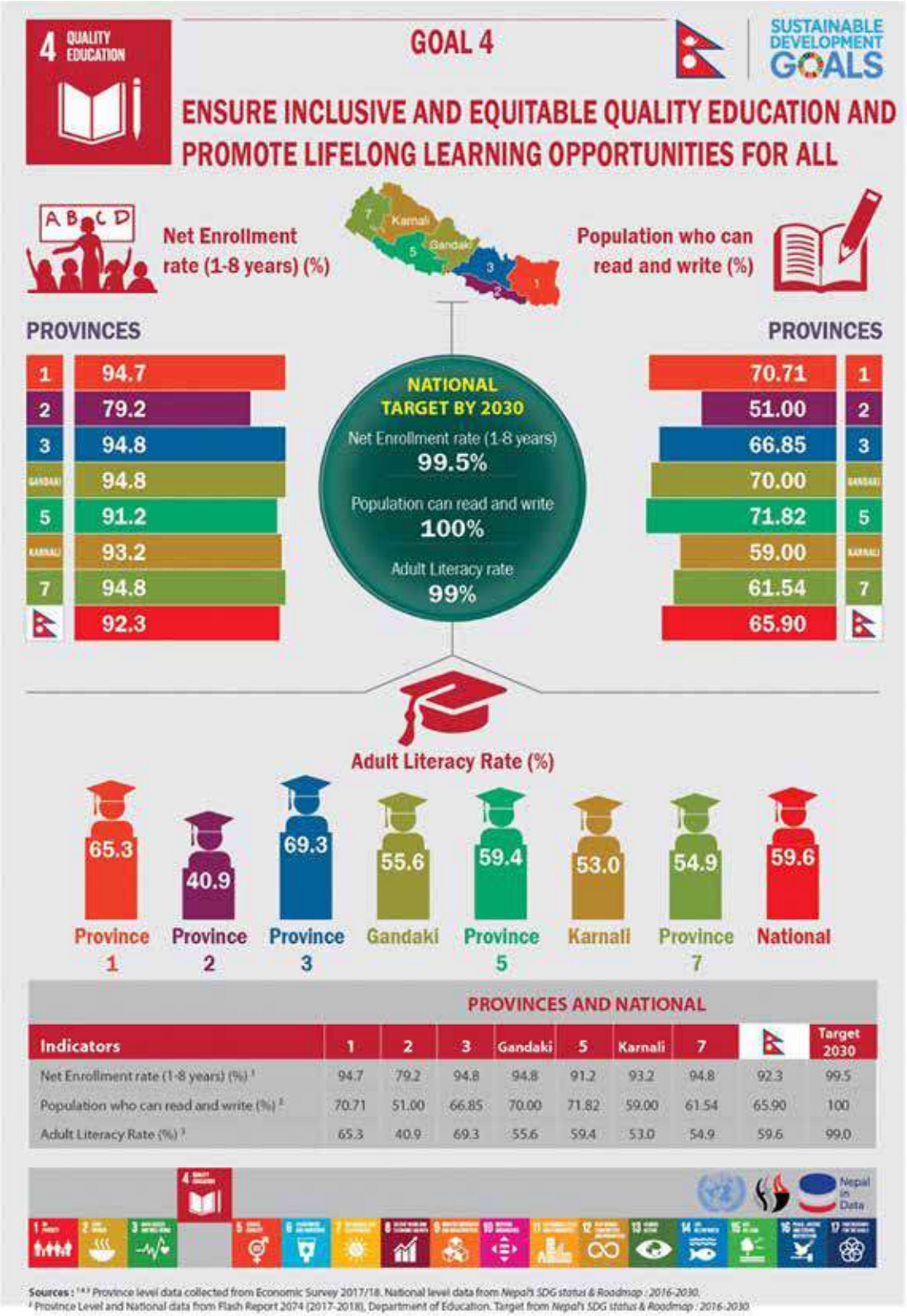
Finally, we will present a few examples for inspiration.

Figure 25: Infographic - Proportion of 15-to-24-yer-olds enrolled in TVET programmes



Source: UNESCO, 2018: Technical and Vocational Education and Training: UNESCO Asia-Pacific In Graphic Detail number 4, **access:** https://drupal.unescobkk.org/content/technical-and-vocational-education-and-training-unesco-asia-pacific-graphic-detail-4

Figure 26: Infographic – SDG 4 in Nepal



Source: UNDP Nepal, 2018: SDGs in Nepal: Infographics, **access:** http://www.np.undp.org/content/nepal/en/home/ library/sustainable-development-goals-national-report---nepal/sdgs-in-nepal-infographics.html

Figure 27: Examples on animation and documentary style videos

Animation style videos

To watch the animation, access:
https://www.youtube.com/watch?v=_tnY8PkJVMY

To watch the animation, access:
https://www.youtube.com/watch?v=JMEueYuMadU

Documentary style videos

To watch the animation, access:
https://www.youtube.com/watch?v=3llt7wqQGJ8

To watch the animation, access:
https://www.youtube.com/watch?v=U3BsPo3NFqE

Source: UNESCO, 2018: Technical and Vocational Education and Training: UNESCO Asia-Pacific In Graphic Detail number 4, **access:** https://drupal.unescobkk.org/content/technical-and-vocational-education-and-training-unesco-asia-pacific-graphic-detail-4

Technical note

Substituting missing values

Imputation is a technique to replace missing values with substitute values. There are different types of imputation techniques.

i. *Average Imputation*

This technique calculates the average value for the variable and then replaces the missing value with the calculated average value. This choice is not always recommended because it can artificially reduce the variability of your data, but in some cases it makes sense. To give an example, where no information is available for a country, the unweighted regional group mean of the given indicator can be used as the imputed value. Because this method is sensitive to the weight of countries in the region, the rule is not applied to countries with substantial relative weights with respect to their region (for example, China in East Asia and the Pacific). In such cases, manual imputation is required even if it results in a non-publishable estimate. Currently such estimates are made for about a dozen countries.

ii. *Common-Point Imputation*

In this technique, the median or mode is calculated to replace the missing average value. This technique is more structured than plain guessing. Yet, it is still a more imprecise and for planning it is a riskier option. Use caution unless you have good reason and data to support using the substitute value.

iii. *Regression imputation*

In regression imputations, the imputed value is predicted from a regression equation. For this method, the information from complete observations is used to predict the values for the missing observations. Regression assumes that the imputed values fall directly on a regression line with a non-zero slope, so it implies a correlation of one between the predictors and the missing outcome variable. The regression imputation will overestimate the correlations, while the variances and covariance are underestimated.

iv. *Hot-Deck Imputation*

UIS uses an automated single imputation method, also referred to as 'hot-deck' that creates a single estimate to replace the missing value in a dataset. Missing values are imputed by substituting these with available data from the next/previous year(s) for the same indicator in question. This method assumes that if a measurement is missing, the next best guess is that the value has not changed in comparison to the next/previously available.

Another technique is to use statistically correlated indicators to impute the missing value of the given indicator (even across time). For example, if the pupil-teacher ratio for primary education for both the public and private sector is missing in a given year, but data on the public sector is available for another year, the rate of change of the public-sector ratio between the two years can be applied to derive the pupil-teacher ratio for the public and private sector for the missing year.

This approach assumes that the value for the required indicator changes in the same way as the value from the substitute indicator. If values of the indicator are only available for years prior to the year of the missing value, the most recent year's value is used as the imputed value for the missing year. If values of the indicator are only available for years after the year of the missing value, the earliest year's value is used as the imputed value for the missing year. If values of the indicator are available for both years before and after, data is imputed using linear interpolation between the two years that are closest to the year of the missing data.