



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



TCG4: SDG 4 Benchmarking: Recommendations

TCG4/34

16-18 January 2018

Dusit Thani Dubai

133, Sheikh Zayed Road, Trade Centre,

Dubai, United Arab Emirates





Contents

Considerations	3
Global Vs. Other Levels	3
Absolute Vs. Relative.....	4
External and Functional Benchmarking.....	4
Caution on Measures/Methods that Consider Only Economic Aspect.....	5
Considerations in the Linking of National and Cross-national Assessments	5
Taking a Balanced Approach	5
Using Universally Applicable and Consistent Definitions	6
Differences in Starting Points and National Capabilities.....	6
Recommendations	7

DRAFT

Considerations

Despite numerous challenges, it is worthwhile to note that the SDGs aim to be more inclusive in what qualifies as sustainable development and promoting country ownership of SDG progress. The global desire to measure learning, with the galvanizing belief that better data will lead to more students achieving, is found at the heart of many of the current initiatives studied in this report. In this context, our recommendation is that establishing national and regional benchmarking is a better option for SDG 4. However, we present the following initial pointers and recommendations relating to benchmarking efforts of SDG 4 for consideration as its discussions move forward.

Global Vs. Other Levels

Setting benchmarks for all levels of SDG 4 and the monitoring of those will certainly overwhelm UIS as well as participating countries. Discussions are still ongoing as to which level should be focused on. The 2016 GEM report recommends that rather than overhauling the ways in which data is collected in education, better coordination between agencies and more resources to implement plans would be more effective in the changes needed to monitor the 2030 education agenda (UNESCO, 2016a). Some scholars argue that the agenda itself is not actually universal because of how targets are set within the goals that direct attention to developing nations more than developed ones, such as nutrition issues in Goal 2 being dominated by malnutrition and not shared with the equally threatening problem of obesity (Vandemoortele, 2017).

Regardless, every region and nation, no matter the number of similarities, have their own interests when it comes to their specific development. It is unlikely that all of those interests will align at a large scale. For example, Fijian education stakeholders prioritized 14 indicators to measure inclusive education, only four of which were covered by current SDG targets (Sprunt et al., 2017). Therefore, one of the recommendations by the OWG that seems meaningful is each government setting its own national targets inspired by those at the global level due to country context (King, 2015). It is important that countries have ownership of the SDGs to realize the necessary changes. Nations may choose a combination of the global reporting indicators and complementary national indicators to harmonize global and national reporting. Vandemoortele (2017) suggests that global assessment needs to pay more attention to how the global targets make a difference at the national and sub-national levels.

Countries in a given region tend to have common education contexts, thus setting benchmarks at the regional level may have better applicability and political consensus among them rather than focusing on benchmarks for global goals. The OECD countries in particular have had very little difficulty in establishing benchmarks for the SDGs, thanks mostly its function as an ideational agency, separating goal setting from politics as much as possible, and the minimal competition among its member states. For regional reporting, existing mechanisms, such as the Regional Economic Commissions, should work as a foundation to foster benchmarking dialogue and knowledge-sharing among countries. Regional monitoring processes can also negotiate what is being measured at the national and global levels, especially if organizations are already subsidiaries of international organizations. Thematic reporting could be left to the coordination among specialized organizations, universities, and even businesses, which may have access to valuable SDG data.

Global targets are meant to help accelerate progress towards quality and equitable education, yet their applicability can only be judged within the country-specific context. Because national



governments are expected to integrate the global SDG 4 commitments into national education development efforts, appropriate intermediate national/local benchmarks must be established. The intermediate benchmarks for each target can serve as quantitative goalposts for review of overall progress vis-à-vis the longer-term goals (UNESCO, 2017). Combining those with regional benchmarks seems to be an effective manner with which to monitor progress towards SDG 4.

Absolute Vs. Relative

While some initiatives included in this review suggest that the SDGs require absolute benchmarks, others suggest that this alone is an ineffective practice, since absolute standards are much more difficult to achieve at the international or cross-country level. For example, the OECD examines the distance to travel in order to reach each target level that involves determining levels of achievement on each target level. The level was pre-determined in the 2030 Agenda, either as a fixed value or as a relative improvement on a country's starting position (OECD, 2017c). Likewise, even before the SDGs, global goals and targets were expressed in either absolute terms or as combined relative and absolute benchmarks. Some scholars argue that neither type of benchmark taken alone provides the full picture of a country's progress or situation (Vandermoortele & Delamonica, 2010).

A combination of relative and absolute benchmarks arguably constitutes the best guarantee against possible biases in setting global targets. Therefore, our recommendation is not to get confined with one method. It will depend on a multiple factors, such as the how the target is set, whether it is clearly quantifiable, to what extent the initial position is important, and so on. Therefore, a benchmark will have to be expressed using either or both terms, but should also have other methods considered as well if possible. Experts will need to decide on this depending on the indicators and what is currently available to measure them.

External and Functional Benchmarking

Global benchmarking is undoubtedly complicated. Because the SDGs are a set of global commitments, developing external benchmarks is a natural process. However, such top-down benchmarking may lead countries to focus on what indicators and results are measurable, rather than on "real" performance (Groenendijk, 2009). The effects of top-down benchmarking may be unavoidable with global benchmarks, but without it, measuring progress on the SDGs may be more disjointed than is preferred. The effects of external benchmarking can be reduced through the involvement of relevant stakeholders such that ownership is achieved.

Functional benchmarking sets out to analyze aspects such as functions and processes of participating entities independently of characteristics like output and sector, whereas generic benchmarking involves all aspects of involved entities. In a way, functional benchmarking can be viewed as the reduced form of generic benchmarking, by reducing larger units to smaller and more comparable entities. A smaller benchmarking scale of more comparable entities allows the easier exclusion external factors in favor of those factors that really do make a difference in education systems. Rather than creating an unwieldy set of "all-aspects" (i.e., generic) benchmarks, it will be more manageable to establish more specific, functional benchmarks.



Caution on Measures/Methods that Consider Only Economic Aspect

Another trend identified in this review is the tendency for progress to be measured in terms of economic progress. As shown by Dill and Gebhart (2016), many of the current indices used to track SDG progress inherently favor developed countries over developing countries because of their reliance on economic status. While certainly applicable, economic growth is not the only barometer against which countries and its individual citizens change and develop in education. Especially considering the rise of BRICS and MINT countries and the ways in which they resource and tackle education challenges, benchmarking countries based on old ideas of development must give way to transformational change that is structural, institutional and normative. The push for more qualitative data by countries and institutions alike is also a promising start, and strategies to incorporate them into SDG 4 reporting must consider how they can better measure components of education, such as quality.

Considerations in the Linking of National and Cross-national Assessments

Recent benchmarking efforts at both the GAML and UIS have centered on establishing a proficiency scale linking with NAs and CNAs through the process of social moderation/policy linking to measure progress on SDG 4.1.1. Despite the statistical challenge this effort presents, taking advantage of current measurement tools that are already increasingly used presents an opportunity to advance their usefulness in measuring learning. For example, countries like Mexico and Brazil have already benchmarked student performance against PISA. Developing reliable and valid items to cross-link the existing regional and international assessments is an important step in the right direction towards advancing the world's knowledge on benchmarking at a global level.

However, there are important considerations to make. In Treviño & Ordenes' (2017) four time-bound strategies for assessing SDG 4, the three mid- to long-term strategies (including the development of a worldwide assessment) all greatly reduce the external validity in representing national curricula. Not aligning metrics to national policy and curricula will reduce their use and usefulness in informing policy development and supporting classroom interventions as they diverge from countries' needs and priorities. Such effects could go in at least two ways: either the measurement of SDG 4 fails to accurately capture and influence learning, or that countries will increasingly push towards a homogenized curriculum (and with that the erasure of culture in curriculum). While some suggest that ignoring the impact of culture on learning and moving beyond static measures allows for the definition/creation of a global curriculum (and therefore global benchmarks), culture stills plays a significant part in the education system to be ignored. Establishing global content standards, even with experts meeting to determine them, elicits further questions of which standards are chosen and what/who they represent, as well as what might be ignored or forgotten in benchmark development based on which experts are present.

Taking a Balanced Approach

A set of 43 indicators, including the 11 global indicators recommended by the Inter-Agency and Expert Group on SDG indicators have been approved by the Technical Cooperation Group for SDG4-Education 2030 Indicators. However, establishing benchmarks for each of these 43 indicators is not an optimal option, particularly for many developing countries that lack the resources in monitoring



their educational goals. Therefore, we suggest using a selected number of indicators for this benchmarking process, which should be determined by existing team of experts working in this area.

Using Universally Applicable and Consistent Definitions

Characteristics such as parity and access are easier to measure in education, but what constitutes learning remains elusive. Detailing the developmental stages of knowledge and knowledge transfer, especially that of commonalities among countries/cultures, is beyond the scope of this report, but current understandings and foundations of such areas and even the “globalness” in education remain conceptually ambiguous (Sparapani et al., 2014). Given that challenge, it may be more appropriate to establish content-based standards that “are informed by and can be mapped to local curricula and relevant national and international standards” (ACER, 2017: 7).

In order to more broadly monitor progress against indicators, within and across countries, it is important to establish uniformly applicable definitions (for example, the minimum proficiency level for reading at a particular grade or age level) of all the relevant features a particular indicator. In this regard, the use of ISCED is recommended in providing a cross-nationally standardized way of referring to the measurement points in indicators (ACER, 2017). Likewise, as ACER's report points out that meaningful comparisons across education systems require consistency in the developmental context (such as the length of schooling). Such effort “poses a number of practical and political issues associated with the different structures of schooling across systems and the political imperatives that guide policy development within and across countries” (ACER, 2017, p. 4).

Differences in Starting Points and National Capabilities

Another important aspect to consider is to use measures that go beyond the assessment of status and rates of changes alone. Certain indices are biased towards developed countries, and equating economic prosperity to high educational outcomes is not always a strong relationship. Ensuring that countries like Namibia won't be both disadvantaged or misrepresented when compared to a country like France in progress or final reports on SDG attainment is fair by taking into account both status and rates of change. Setting one-size-fits-all quantitative and time-bound targets without taking account of differences in starting points and national capacities might not be realistic (Clemens, Kenny, & Moss, 2007) and could be unfair to countries that start farther from the target and face larger resource and other capacity constraints (Easterly, 2009; Fukuda-Parr, Greenstein, & Stewart, 2013). The benchmarking option should therefore consider different starting points or levels of development and available resources to avoid such risks. One such approach is frontier analysis that identifies benchmark rates using the rate of the historically best performing country among those at a similar level of coverage or attainment (see Cook et al., 2014; Luh et al., 2016). It may not be possible to capture all the data required of the SDGs with this process, but it can be a helpful start.

Recommendations

- Nations may choose a combination of the global reporting indicators and the complementary national indicators to harmonize global and national reporting. This report suggests establishing appropriate intermediate national/local benchmarks, where the benchmarks for each target can serve as quantitative goalposts for review of overall progress vis-à-vis the longer-term goals. Combining them with regional benchmarks can be an effective manner with which to monitor progress towards SDG 4.
- A combination of relative and absolute benchmarks arguably constitutes the best guarantee against possible biases in setting global targets. Therefore, benchmarks will have to be expressed using either or both terms, but should also have other methods considered as well if possible. Experts will need to decide on this depending on the indicators and what is currently available to measure them.
- The SDGs are a set of global commitments, but external benchmarking may lead countries to focus on what indicators and results are measurable, rather than on “real” performance. The effects of top-down benchmarking may be unavoidable with global benchmarks, but can be reduced through the involvement of relevant stakeholders such that ownership is achieved. Functional benchmarking reduces larger units of comparison into smaller and more comparable entities. With a smaller benchmarking scale, it is easier to exclude external factors in favor of those factors that really do make a difference in education systems.
- In the process of linking national and cross-national assessments to one another, many considerations should be noted. For example, not aligning metrics to national policy and curricula will reduce their use and usefulness in informing a country’s policy development and supporting classroom interventions as they diverge from countries’ needs and priorities.
- Establishing benchmarks for each of the 43 indicators is not an optimal option, particularly for many developing countries that lack the resources in monitoring their educational goals. Therefore, selected indicators should be used for the benchmarking process, which should be determined by existing team of experts working in this area.
- Meaningful comparisons across education systems require consistency in the developmental context (such as the length of schooling). In order to more broadly monitor progress against indicators, within and across countries, it is important to establish uniformly applicable definitions of all the relevant features of a particular indicator. In this regard, the use of the ISCED is recommended in providing a cross-nationally standardized way of referring to the measurement points in indicators.
- Measures that go beyond the assessment of status and rates of changes alone must be used. One such approach is frontier analysis that identifies benchmark rates using the rate of the historically best performing country among those at a similar level of coverage or attainment.