Edinburgh Research Explorer

The education Sustainable Development Goal and the generative power of failing metrics

The education Sustainable Development Goal and the generative power of failing metrics

Sotiria Grek

University of Edinburgh, Edinburgh EH8 9YL, Scotland
Corresponding author: S. Grek, Social Policy, School of Social and Political Science, 22 George Square, Edinburgh EH8 9LD,
Scotland. Email: Sotiria.Grek@ed.ac.uk

Abstract

The article traces the development of the epistemic infrastructure of the education sustainable development goal (SDG) in order to examine the ways that the incremental buildup of the discourse, technical expertise, and necessary—although always fragile—alliances facilitated a paradigmatic policy shift in the field of education: This is the move from the measurement of schooling to the measurement of learning. Through an analytical lens that examines the entanglement of the material, semiotic, and political and temporal/spatial elements of the infrastructure, the article shows how the sustainable development goal 4 (SDG4) as an epistemic infrastructure enabled a fundamental reorientation in the field of global education governance. The article discusses the ways that quantification, despite—and often thanks to—its failings, folded contested discourses, decision-making, politics, and ideas into its processes. Thus, the paper argues that the making of the SDG4 represents a paradigmatic policy shift; one that is not only to be traced in the move from schooling to the policy prioritization of learning outcomes but also in the very production of global public policy through the work of the SDGs as epistemic infrastructures.

Keywords: SDGs, education, quantification, infrastructures, global public policy

In May 2015, the World Education Forum (WEF) was celebrated in Incheon, the Republic of Korea, with the participation of over 1,500 people, including 120 Ministers of Education and representatives from a wide range of international governmental and nongovernmental organizations. The event at Incheon represented a milestone in the history of the UNESCO summitry, a long trajectory of large education conferences that demanded fair, free, and quality education for all. Similar to others prior to it, the main product of WEF 2015 was the so-called Incheon Declaration, along with the Framework for Action adopted by UNESCO member states few months later, in November 2015. In conjunction, both documents established an ambitious and highly aspirational education agenda for the period 2015–2030, condensed in the overarching goal to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" and a number of associated targets; this is the sustainable development goal 4 (SDG4; UNESCO, 2016).

The SDG4 is one of the 17 sustainable development goals (SDGs) that are "integrated and indivisible and balance the three dimensions of sustainable development: the economic, social and environmental" (UN, 2015, p. 5). According to the United Nations (UN), "they result from what is arguably the

most inclusive process of consultation in the history of the United Nations, reflective of substantive input from all sectors of society, all actors of the international community and all parts of the world" (UNESCO, 2017, p. 4). Indeed, as will be shown here, it is precisely this inclusive and participatory governance model that became key in the formation of many aspects of the SDG4 agenda and its implementation. As a programmatic document oriented at nurturing and securing a form of collective commitment toward a shared set of aspirations, the new agenda builds on a well-established tradition of consultation and collaboration that has come to be recognized as a characteristic of the UN system.

The SDG4 represented simultaneously a form of continuity and a departure from previous instances of goal setting such as the Education For All (EFA)[1] and the millennium development goals (MDGs). First, the new set of goals is characterized by an unprecedented degree of ambition, shifting away from the mere focus on primary education and gender equality in the Global South that characterized the MDG era. It establishes a *universal* agenda[2] that contrasts with the prior focus on developing countries (King, 2017; Unterhalter, 2019). Secondly, the very making of Education 2030 (and of the SDGs in general) represents a path-breaking development in the long history of goal-setting practices and UN summitry. The open, inclusive, and participatory nature of the consultative process around the SDG4 was in many ways unprecedented, and the openly-negotiated and improvisatory character of the SDG debate contrasted with the technocratic origins of the MDGs (cf. Fukuda-Parr & McNeill, 2019).

Thus, this article traces the development of the epistemic infrastructure of the SDG4 in order to show the ways that the incremental buildup of the discourse, technical expertise, and, given this apparent universality of the SDG agenda, the fragile but necessary actor alliances facilitated a paradigmatic policy shift in the field of education: This is the move from the measurement of schooling (Barro & Lee, 1996) to the measurement of learning. The shift entailed the prioritization of an emphasis on learning outcomes, skills, and competencies, measured through what children "can do" with the knowledge they acquire at school. In other words, instead of the traditional education statistics that measured inputs such as education expenditure, teacher salaries, or length of the school year, the pendulum shifted to a greater interest in decontextualized, applied knowledge, measured in real-life contexts. Although the work around the construction of the SDG4 (both prior to and after 2015) is not the only process that facilitated this shift (indeed its origins lie in New Public Management and the economization of education discourse in the 1980s and early 1990s—see Gunter et al., 2016; Ozga et al., 2009), the global nature of the SDG4 process and the active involvement of most key education actors in its production led to a concerted effort to devise *global* learning metrics (Crouch & Montoya, 2019). Thus, alongside other key venues (one of them being the Organisation of Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA), as will be discussed further on), the SDG4 became a prime site of the production of this radical reconceptualization of educational measurement and policy with implications across the world.

Indeed, the complexity and length of the SDG4 process render the painting of a comprehensive picture of all related events and actors as a futile endeavor; in a sense, infrastructures are never finished. They are ongoing projects, always developing in new forms and directions. A focused analysis of the production of the SDG4, viewed through the lens of the notion of "epistemic infrastructures," allows for a close-up on the interdependency of materialities, technologies, individual actors, and organizations that participated in its making. Indeed, the paper adopts the definition of an epistemic infrastructure as the "complex interplay of material, techno-political and organisational structures within which (statistical) knowledge is produced, disseminated and translated into global public policy" (Tichenor et al., 2022, this volume).

Earlier literature on infrastructure studies (Star, 1999; Winner, 1986) highlighted their invisibility; infrastructures were seen as comprised of social, material, and technological elements that are interdependent and flow seamlessly into one another, facilitating the unobstructed move of numbers, people, goods, and ideas in the production of new ways of measuring, viewing, and living in this world. However, in contrast to the neat accounts of global education reforms flowing top-down, the SDG4 has never been the perfect invisible infrastructure, moving ideas and practices from some imaginary "centre of

---

[1]   The EFA was a long, multilayered, and multisited negotiation process that involved numerous meetings and consultations, largely led by UNESCO. At the same time, the EFA-related efforts were being paralleled by the negotiation of the SDGs. As it happened, the two processes—i.e., the EFA-led process and the debates facilitated by the UN Open Working Group on the SDGs—reinforced and informed one another through an intricate, often conflictual, political process.

[2]   This can be understood as a universality of principles (human rights), universality of reach (focus on equity and inclusion), and universality of country coverage.

calculation" (B. Latour, 1987) to the periphery. Instead, long before its inception, it has been a site of conflict and contestation, a space where relationships break down and—more often than not—metrics fail. Since the idea of metric "failure" might have normative connotations, it needs to be clarified that the article sees "failing metrics" as those that lose their policy momentum, by increasingly being perceived by the policy, expert, and professional communities as irrelevant or even misleading; ultimately, their continued measurement is seen as having detrimental, rather than positive effects on the policy arenas they are meant to contribute. Such failings can be either real or manufactured, yet the outcome is the same: The failure of achieving global goals (irrespective of whether they are misplaced or, in fact, unattainable in the first place) sparks quests for improved metrics that will excite, persuade, and "stick" anew (Bandola-Gill, 2021). Yet, despite such perceived failures, it is the infrastructure's breakdown that fuels its growth and expansion. As this paper will show, the paradigmatic shift from the policy focus on schooling to learning happened not despite but *because of* the failures, contestations, and breakdowns in the process of the production of the indicators.

The policy prioritization of learning and its associated outcomes is not a novel topic in education research. Although there has been scholarship on the discursive expansion of the language of learning outcomes and skills (Klees et al., 2019), as well as some critical literature on the validity and robustness of the new learning metrics (Benavot & Smith, 2020), and on their effects on global education policy reforms (Mundy et al., 2016), the paper's contribution lies in its analysis of the *entanglement* of materialities, discourses, ideas, and practices into the building of a new epistemic infrastructure and the new "policy work" (Colebatch, 2007) that these entanglements have brought about.

Indeed, these entanglements have allowed a plethora of contestations to unfold: One of the most prominent ones is the large emphasis on some indicators versus others, as well as the issue of the democratic decision-making process. After a brief overview of the intellectual terrain on infrastructures and some methodological considerations (*The rise of the study of infrastructures: vogue, vague, or "really useful knowledge"?* section), the following section (*From schooling to learning: the incremental building of an infrastructural base, 2006–2013* section) will discuss the history of the shift of education discourse from the measurement of inputs to skills and outcomes. In particular, I will discuss the ways in which some powerful actors prepared the ground for a move away from the measurement of schooling (through measuring access and completion) to learning (through the measurement of literacy skills). The primary means of facilitating this change was through presenting the MDG education targets as misleading and thus as "failing" metrics; the mobilization of new evidence and a "killer" number (Stevens, 2011) was used in order to create the space for contestation and change. The building of this discourse (the discursive element), together with the production of new metrics to replace the old ones (the material one), as well as the role of expert brokerage (the role of actors and interdependencies), all became a vital mix and thus the building block in the construction of the infrastructure of the SDG4. In addition, I will highlight the importance of the *temporal* dimension in the building of epistemic infrastructures, in terms of first, their temporal discursive framing of "past failures/current crisis/future projections," as well as in relation to the slow, step-by-step buildup of the measurement infrastructure in order to gather steam, create the evidence, build a support base and thus have greater policy influence.

Section 4 will then move on to the analysis of the workings of the Technical Advisory (later Cooperation) Group, in charge of the development of some of the SDG4 indicators. The section will show how the Technical Advisory Group/Technical Cooperation Group (TAG/TCG) began its work in 2014 primarily as a group of expert IO statisticians and later expanded into a much larger—and with a different function—grouping that included country and civil society representatives, all in the name of democratizing the measurement agenda and process. Thus, beginning with the small, highly technical, and elitist group in 2014, we observe how the slow building of a much larger infrastructure of actors and materialities came together to support, prop up, and legitimize the work of the production of numbers. Thus, this section will focus more on the *spatial* features of the infrastructure, as it expanded across contexts and fields of practice, to include a much wider actor membership and achieve greater coordination across the local, national, and global levels.

The key role of the meeting in the practice of "infrastructuring" section will discuss the infrastructural qualities of meetings of the SDG4, by showing how, instead of a seamless flow of coordination and cooperation, it was failing metrics and the continued breakdown of the proceedings that both acted as generative forces that ensured its continuity and growth. Finally, the concluding discussion will bring together the mix of new ideas, materials, and actors, coupled with the temporal and spatial aspects

of the infrastructure, in order to show the ways that a new paradigmatic shift in the field of education took place, with implications for policy work and reform globally.

The rise of the study of infrastructures: vogue, vague, or 'really useful knowledge'[3]?

"Infrastructures are conceptually unruly" (Larkin, 2013, p. 329), Brian Larkin wrote, and there could not have been a more accurate description for the varied application of the term. In fact, it is precisely the conceptual plasticity *and* the focus on materiality that has made infrastructures such a popular concept in social theory. Nonetheless, they have not always been as vogue as they are today: In fact, it was only the mid-1990s when Bowker (1995) first pointed toward the materiality of infrastructures as a way of understanding their function and effects. Bowker saw infrastructures as largely invisible backdrops to social action and thus analytically not penetrable; he therefore proposed the notion of "infrastructural inversion," as a way of breaking the invisibility and flow of the infrastructure. Infrastructural inversion (Bowker, 1995) was about making the invisible visible, through a focus on material relations and the ways they reconfigure how we know and live in the world.

Similarly, in 1996, Susan Leigh Star and Karen Ruhleder saw invisibility as a key quality of infrastructural systems. Nonetheless, they also identified the seamless flow of the infrastructure as a fragile achievement that was prone to breakdown and failure (Star & Ruhleder, 1996). The invisibility/visibility conundrum was further discussed by Larkin (2013), who suggested that infrastructures can be invisible but can also become a spectacle, and thus depend on their visibility for their success. However, following Larkin, even when an infrastructure is open, visible, and ready to be experienced, what is there to see? According to Harvey et al.,

> Provisionally, and minimally, we might say that we are dealing with technologically mediated, dynamic forms that continuously produce and transform sociotechnical relations. That is, infrastructures are extended material assemblages that generate effects and structure social relations, either through engineered (i.e. planned and purposefully crafted) or non-engineered (i.e. unplanned and emergent) activities. (Harvey et al., 2017, p. 5)

This article aims to contribute to the literature on infrastructures, by showing the particularities of the mix of materials, practices, and meanings in the making of measurement agendas, such as the SDGs. Given the centrality of knowledge and data production in global governance, the study finds Tichenor et al., (2022) concept of "epistemic infrastructures" as particularly apt for the analysis, since it advances the analytical purchase of the—Science and Technology Studies-primarily informed—concept to bring it much closer to policy theory and practice. In particular, as the paper shows, both the flow and the failures, the unlikely alliances and the clashes, did not only facilitate the production of a system of measurement and a particular way of naming and understanding educational realities in the 21st century. They also brought about a much more fundamental and—as it appears—permanent policy shift: This was the change of the paradigm in educational measurement, practice, and values. The move away from the measurement and thus prioritization of educational inputs (numbers of teachers, school facilities, financial support, and others) toward outputs (learning outcomes, test results, skills, and competencies) has not merely taken place at the discursive level, or the measurement one. Neither has it only been circulated nor promoted among organizations and actors, experts and professionals, that work in the field of education. Rather, it has produced a paradigmatic policy shift that has had dramatic consequences on the way education policies at the country level are made (Verger et al., 2019). The intention of the article is not to criticize this shift or to evaluate it; instead, the aim is to utilize the three orders of the epistemic infrastructure (the materialities, the interdependencies, and the paradigmatic shifts) in order to place emphasis on the role of the incremental, the informal, sometimes chaotic, and almost certainly confrontational nature of producing knowledge for policy. The paper shows that rather than objective, seamless, and invisible, the traveling of numbers and ideas in the SDG4 was an adventurous journey, with some loud passengers, several breakdowns and accidents, and a destination unknown.

---

[3] The term "really useful knowledge" is derived from radical education thought of the 19th century; it was supporting a critical understanding of self and society; and it was knowledge meant "to set you free."

Finally, in terms of methods, the study focused on three main sources of data: First, the discourse analysis of documents relating to the production of the SDG4, as well as materials that predated it. Critical Discourse Analysis is a particularly apt method for the analysis of the making of infrastructures because it sees the text as a key aspect of how certain understandings of the world are shaped and perpetuated by practices of infrastructuring (Fairclough, 1995; Wodak & Meyer, 2001). Hence, the analysis of these documents is useful for, on the one hand, showing what is technically possible, while, on the other hand, explaining what the principles and perspectives of those participating in the production of the infrastructure are.

Second, the study is informed by 20 in-depth interviews with key actors of international organizations (IOs) and the civil society. Research participants—IO actors in their majority—discussed the importance of meetings in the work they do. They talked about their significance in "winning hearts and minds" and discussed their role in taking decisions in regard to the choice of metrics and indicators. Finally, the social network analysis component focused on an exploration of the role of the SDG4 meetings and the alliances and connections they generated. The combination of these methods allowed for a study of the discursive meaning produced by relevant IO and research reports. Interviews gave me an insight into the experience, views, positionings, and choices of the key actors that participated in the infrastructure. Lastly, a social network analysis, focusing on the two main indicator technical groups, explored their meetings as the key stabilizing moment when negotiations achieved the desired *pax romana* before disagreement and conflict unraveled again. Thus, the research design offered the capacity to study different elements of the infrastructure as well as their entanglements and effects.

From schooling to learning: the incremental building of an infrastructural base, 2006–2013

The discursive and logical shift that moved the measurement agenda from a focus on schooling to learning began as early as the 2000s. On the one hand, the OECD PISA, although measuring the skills and competencies of 15-year-olds in the global North (at least in the first rounds of the learning assessment and before its expansion in 2012 and 2015), received unprecedented media and policy attention worldwide; this was due to PISA's ranking of countries according to their education performance. PISA and subsequently the OECD prided itself in decontextualizing education by focusing on global, comparative testing not on the knowledge that students acquire at school (thus moving away from traditional ways of approaching schooling and curricula) but on what students can *do* with this knowledge. The OECD made direct links between countries' future competitiveness to how well schools prepare students to enter the labor market. PISA results were announced at the end of each testing cycle (every 3 years) and caused "shock and awe" to many European countries in particular (and increasingly globally) including the "education catastrophe" that hit Germany, or the "education miracle" that turned Finland into an education tourist hotspot for education ministers and experts from around the world (author). In many senses, OECD PISA became the flagship international comparative test that shifted the focus of education policy makers to outputs, rather than inputs, and to learning rather than schooling. The significance of PISA data is undisputable, given that European education governance became dependent on it, in order to—for the first time ever—create indicators and benchmarks to measure education performance in EU member states—what was called the Lisbon agenda (author).

Nonetheless, perhaps more so than the OECD, it was the work of the World Bank that shifted the education debate, given the Bank's influence in the Global South (Prada Uribe, 2012). The World Bank opposed the MDG emphasis on access to education, suggesting that lack of education had never been only a matter of whether children are in school or not; instead, it was suggested that the focus should be on what children achieve at school when they are there. The work was undertaken by senior economists at the World Bank and the links to improved national economic growth were explicit from the start: In two seminal research reports (Glewwe, 2002; Hanushek & Kimko, 2000), it was suggested that individual mobility and better economic outcomes were achieved in countries that focused on knowledge and skills acquired in primary schools, rather than those systems that merely aimed to increase access. In 2006, another World Bank report became a milestone moment for education measurement, as it shifted the debate not only in education policy circles but also in development ones. The report, provocatively entitled "From Schooling to Learning" (IEG-WB, 2006), was written by the Independent Evaluation

Group and created a polemical discourse against the MDGs' focus on access and completion: It suggested that the current emphasis was misplaced and that much more attention should be given to the improvement of skills and competencies, as it is the latter that leads to economic prosperity and better outcomes. As a consequence, the Center for Global Development appointed three World Bank economists to further explore the issue; their report, *A Millennium Learning Goal: Measuring Real Progress in Education* (Filmer et al., 2006), unequivocally suggested that there was no evidence that showed that completion of primary school guaranteed the achievement of minimal levels of literacy and numeracy and that a rethink was long overdue. The example of the failed MDGs is an excellent illustration of the core argument of this paper in regard to the power of metrics not only to influence policy direction but in fact to be the space where policy work is done: It was the production of new metrics that pushed for the idea that previous metrics had failed. And it is precisely the perceived failure of the MDGs that created the new space for contestation around which new metrics (and thus policy priorities) should replace them. The materiality of data, reports, and meetings intersected with the work of specific expert organizations and actors and led to a substantial policy shift, which was first taken up by specific governments.

Indeed, the arguments developed by the OECD and the World Bank had far more purchase in the development community groups, rather than in education (at least at the start). Both DFID (the UK's former Department for International Development) and USAID (the United States Agency for International Development) produced new strategies in the period of 2010–2015 that identified the measurement of learning outcomes as an institutional priority and consequently channeled their education investments accordingly. Although there were a number of voices from academia that suggested that a singular focus on learning outcomes would take the attention away from other important pedagogical aspects (Barrett, 2011; Tikly, 2015), their commentary remained "academic"; they had little policy influence and impact. Yet, there were still quite a few voices in education, especially those from UNESCO and the civil society, which were worried about the new trend and the misplacement, as they saw it, of education and schooling measures with those of outputs. Once again, the two functions of education, the humanistic and the economic one, were pitted against one another. The result was the slow emergence of "a divide between those emphasizing quality and those primarily concerned with learning outcomes…Even if the differences between the two approaches were originally a matter of nuance or emphasis, they ended up forming two distinct communities of understanding, informed by different sets of ideas" (Fontdevila, 2021, p. 177).

Indeed, as the decade progressed and the end of the MDG timeframe was drawing to a close, we can observe a much more concerted effort to change not only the discourse (that had already been achieved) but to start building an infrastructure for the establishment of a new measurement agenda, one in which learning, skills, and competencies would be center-stage and would replace the previous failing targets. The key protagonist in this new era was not the World Bank (although it was always supporting at the background) but a new initiative, the Global Compact for Learning (GCL), which was launched in 2011 by the Brookings Institute Center for Universal Education. GCL quickly became an advocacy tool; through its reports, it created a sense of urgency, putting forward the idea that there was a learning crisis that was "hitting the poorest, most marginalized and the youth particularly hard" (CUE, 2011). Just a year later, UNESCO in conjunction with the Global Education Monitoring Report (GEMR, 2012) published an estimate of the number of children not achieving basic literacy skills as reaching 250 million. The shocking figure became further ammunition not only for those that were pushing for the learning turn but also for those who were suggesting the benefits of international learning assessments; without them, there would have been no evidence of this crisis. Thus, the crisis discourse had created a sense of urgency and would quickly turn into the need for *action*. Not only was it obvious that the MDG targets, set in 2000, were not going to be met but also it had become evident—to some, at least—that these targets were ill-defined and misplaced and thus were failing millions of children around the world.

Crucially, GCL prepared the ground for the launch of another key initiative: The Learning Metrics Task Force (LMTF) was established in 2013 with the aim to "catalyze a shift in the global conversation on education from a focus on access to access *plus* learning" (UIS/CUE, 2013; emphasis mine). This was a subtle, yet fundamental change and an open invitation to the two measurement camps to come together in search of the post-2015 agenda. Brookings invited the UNESCO Institute of Statistics (UIS) to head the task force, an important gesture toward an actor that appeared more trustworthy (to teacher

organizations and civil society, at least) than the World Bank. More crucially, this was not an elite exercise; rather, LMTF was a very diverse organization that included a wide range of actors not only from the IOs' expert world but also from regional organizations, donors, governments, statistical agencies, and civil society. The pluralistic nature of the membership, coupled with its UIS leadership and the timing (the preparations for the post-2015 agenda had already begun) made the LMTF the perfect opportunity to build the measurement infrastructure not only up but wide. This was the moment when the buildup of the new measurement agenda was to stretch across contexts and organizations to expand spatially, too. Essentially, the establishment of the LMTF became the foundation for building—what would later be called—the SDG4.

From IOs' advisory to cooperative role: brokerage and the spatial infrastructuring of numbers

LMTF brought together a vast array of actors and organizations in its efforts to offer legitimacy to the task of shifting the debate and subsequently the post-2015 goals for education. As the previous section showed, it approached the contentious topic of the prioritization of metrics and goals diplomatically, suggesting that they were interested in exploring "access *plus* learning" metrics. Thus, an olive branch was extended to academics, the civil society, and professional organizations that perceived the learning focus as reductionist and as reflecting merely the economistic lens of the Bank's ideological positioning. Additionally, UIS' leadership (and not the World Bank's, for example) gave the project not only credibility but also a ticket to move away from merely debating over priorities (the 250 million failing children was an alarm that kept on ringing) toward trying to find practical measurement solutions for their aims—in light of PISA and other regional, cross-national tests, the attention turned to the production of learning assessments, which, as it happened, have become the key data production machines for the SDG4 agenda (Fontdevila, 2021).

Despite the seemingly celebratory and ambitious language, the work of the LMTF was challenging, given that consensus had to be found not only on the aims themselves but also in relation to how these aims would translate into measurable indicators, as well as which spaces of deliberation would constitute the legitimate decision-making venues for making these choices. This is due to the fact that the efforts to devise the SDG4 indicator framework did not start by the UN Statistical Commission, but dated back to the establishment of an interagency, ad-hoc platform known as the TAG. Originally, the TAG was established by UNESCO in 2014 and recruited experts from UNESCO itself but also from the GMR, the OECD, UNICEF, and the World Bank. In many senses, while after 2014 LMTF 2.0—as the version came to be called—continued the debate at the country level (Anderson, 2014), TAG adopted the work of the original LMTF with its focus on "seven learning domains, and recommendations for global measurement areas" (Anderson, 2014). Chaired by the UIS, TAG was a much smaller grouping, with its membership limited to IO experts, and with the task to devise the "post-2015" indicator agenda.

From March 2014 to May 2015, the TAG embarked on the process of mapping existing and potential education indicators, taking into consideration both their alignment with the (anticipated) targets and questions of data availability. Importantly, the work of the TAG benefitted from the input of a global consultation process, running from November 2014 to January 2015. In May 2015 the group's proposal was incorporated to the Framework for Action at the WEF in Incheon. That was a pivotal moment for the group's continuity, since the WEF recommended that the TAG is expanded, in order to include civil society and UNESCO member states organizations' representatives. It was partly the distrust toward the IOs leading the measurement agenda by the EFA actors, and partly the universalistic and participatory agenda of the SDGs that had brought this significant change, which also led to the renaming of TAG as the "Extended TAG." Subsequently, the Extended TAG conducted ongoing open consultations led by regional leaders. Very quickly, what was a small, rather swift, and efficient technical team of IO experts and representatives (with their own of course internal conflicts and competitions) had suddenly opened up to a much larger governing structure that required coordination, continuity, funding, support, meaning, and a sense of purpose and unity: in other words, it became a complex infrastructure, ever expanding and changing, but always propping up and pushing the work of numbers.

Areas of concern for ETAG related to the issue of whether "temporary placeholder" indicators should be devised, especially in relation to the lack of a universally comparable metric for learning outcomes.

Above all, a major qualitative difference had already taken place in comparison to the previous education MDGs: Five of the seven SDG4 targets now focused on learning outcomes and skills, a major departure from previous targets, which focused on access and completion. In 2016, with the new SDG4 agenda formally adopted, the ETAG shifted again, giving rise to the TCG, with the same broad membership (UIS, 2017) and remaining operative to date.

Additionally, in parallel to the TCG, another group came into existence, following on the footsteps from the LMTF: This was the "Global Alliance for Monitoring Learning" (or GAML in short), the successor of the LMTF. Also created in 2016, GAML was originally defined as an "umbrella initiative to monitor and track progress towards all learning-related Education 2030 targets" (UIS, 2016, p. 49), and was tasked with the development of tools, methodologiess and shared standards to measure learning outcomes in the context of SDG4. Following the TCG, its membership is open to any individual or organization willing to contribute to the work of GAML and includes IOs, civil society organizations, a variety of technical partners and assessment organizations, and representatives of UN member states.

Therefore, the political game of numbers became too high-stakes to leave it to the technical experts only. Wider legitimacy was sought and gained through the expansion of the measurement infrastructure into an epistemic one: one that became legitimate and dominant through its active involvement of actors from across sectors and countries. Even though the involvement of the majority of these actors, as the next section will show, was generally passive, the language of the new indicators became the new *episteme*: That is, a way of knowing, describing, and communicating about the world that was not merely about the craft of numbers but involved the production of a new governing paradigm.

The key role of the meeting in the practice of "infrastructuring"

The transitions from the ETAG to the TCG and from the LMTF to GAML were not without problems. Some original members of the TAG saw the TCG as a marker of the increasingly politicized nature of the indicators debate. At the same time, certain countries represented in the TCG and in GAML perceived that their input had not been sufficiently taken into consideration but simply used for rubber-stamping purposes. Others saw their role as primarily watchdogs, rather than full participants in the process. A civil society representative—involved in the TCG over a long period of time—elaborated on such tensions in the excerpt below:

> We were of course invited to be part of this, which was a clever move because we had probably been, if not the, at least one of the most critical voices in the room. So we had a dilemma and ended up actually agreeing to be part of this committee … I think what we struggle with is the fact that we know that just by being in the room, we are giving an indirect blessing of what the […] is doing. And at the same time, if we are not in the room, then we have no access to the conversations. We don't know what's going on (Civil society 1).

Thus, in this last empirical section, the article offers some observations on the process and practice of these groups' gatherings as the site where social, technological, and material elements come together and stabilize an otherwise fluid and contested field. Anthropologist Clifford Geertz's idea of the "poetics of power" (Geertz, 1980) is useful for unraveling the thick layer of dramaturgy coating this apparently technocratic regime. Several of the study's interviewees suggested that most meetings are performative events, which follow a certain ritual, allowing enough free space to conclude with some loose decisions that determine the agenda for the follow-up meeting. The predominance of interviewees suggested that there is a clear-cut distinction between participants from the Global North, whose presence and contributions dominate the meetings, while representatives from countries of the Global South most of the time have a very passive presence, if any at all. This of course does not negate the agency and power of participants from the Global South, especially in relation to exploiting their own perceived weak positioning in order to accomplish specific goals.

Further, the ambiguity and informality of the process, despite being an issue for some in the room, becomes a valuable, malleable tool in ensuring participation while at the same time also pushing on with a specific agenda. Interestingly, however, frustration and discord about the lack of transparency are not sufficient reasons to disassociate oneself from these alliances; being present at the discussions even when one is at the receiving end is still considered more valuable than not participating in such meetings:

Are we working on consensus basis? How do we deal with the fact that so many people have a conflict of interest? Who will draw conclusions? If there's voting, with what numbers would something have to be supported for it to be carried? And this was a frustration that grew as every session basically just ended with a broad sweeping, this was a very good discussion, thanks guys. And it was never really clear what anything would result in. (Civil society 2)

Meetings are therefore key sites, where multiple elements of the infrastructure mix: material inscriptions in the form of data, PowerPoints, documents distributed prior and during the meeting, as well as the production and pursuit of common meaning and aims. A plethora of actors come together, irrespective of their own interests and ideas, in order to achieve a compromise, specific enough to keep the infrastructure going but also flexible enough so that can be adapted and translated in their own contexts. Not everyone's participation has the same centrality and weight in these proceedings; nonetheless, the expansion and approval of actors are necessary in order to create some, even partial agreement and continuity (author). As Luhmann suggested (Luhmann, 1969), this is legitimacy achieved via procedure: However, in the case of an epistemic infrastructure, its technoscientific element, in the form of data, standards, and protocols is as key as the political agreement, sought after by all participants in the room who have a stake—irrespective on how little or large—in keeping the game going. Meeting by meeting, compromise by compromise, the epistemic infrastructure achieves more than just the production of knowledge for policy: Instead, it has slowly shifted the needle toward a new, paradigmatic policy change. In the case of the SDG4, as we have seen, it was the fundamental policy shift from the measurement of access and enrolment data to the data and policy emphasis on learning outcomes.

The visual depiction of how actors are connected through meetings is helpful here (as is shown in Figure 1) because it allows us to explore the ways that some actors are central (measured through their participation in most meetings) versus those that are more peripheral and those that might have attended only one or two meetings. Frequency of attendance denotes a more key positioning within the infrastructure, whereas less active participants are no less important; their inclusion and participation at least in some of the meetings adds legitimacy to the project and strengthens the infrastructure as a politically sanctioned operation. This is not an invisible infrastructure: In fact, its visibility is heightened the wider its participant group is. On the other hand, the more some parts of it become visible (official meetings, for example), the more opaque and hidden other parts of it are (informal communications and negotiations). Finally, although the limitations of the print medium do not allow us to interactively see the evolution of meetings over time, it is important to understand these networks of meetings and actors as temporal and spatial structures, showing the accumulation and buildup of the infrastructure over time.

Discussion: the work of infrastructuring in the case of the education SDG

This article focused on an analysis of the development of the SDG4 as an epistemic infrastructure. The infrastructure is propped up and legitimized through the centrifugal forces of technocracy versus the perceived need for SDG4's inclusivity and political acceptance. The case of failing metrics and conflicting ideas and interests reveal how epistemic infrastructures, rather than being monolithic blocks, remain fragile and, despite their claims to data and objectivity, it is precisely in the failure of (some) of their evidence-production work that they capitalize on, in order to shift the agenda and move it along: The article showed how the perceived failing of the education MDGs (with the use of flagship numbers of emergency, such as the 250 million children not having basic literacy skills) was used as a vehicle to slowly build an agenda that, although having made plenty of "concessions," is now perceived as the dominant policy regime in global education reforms. There were plentiful of circumstances that the disagreement was such that a possible breakdown seemed almost unavoidable: For example, the reason of the compromise in the drawing of the main parameters of the SDG4 was the real possibility of the exclusion of an education-focused goal, due to the polarization of the two "camps." Yet, it is precisely the diversity and entanglement of the infrastructure's social, technical, and political elements that sustained and even strengthened it, almost against the odds.

It is this incompleteness and fragility of the infrastructure that the paper focuses upon, alongside the generative power of failing metrics to provide fertile ground for more—and allegedly more precise
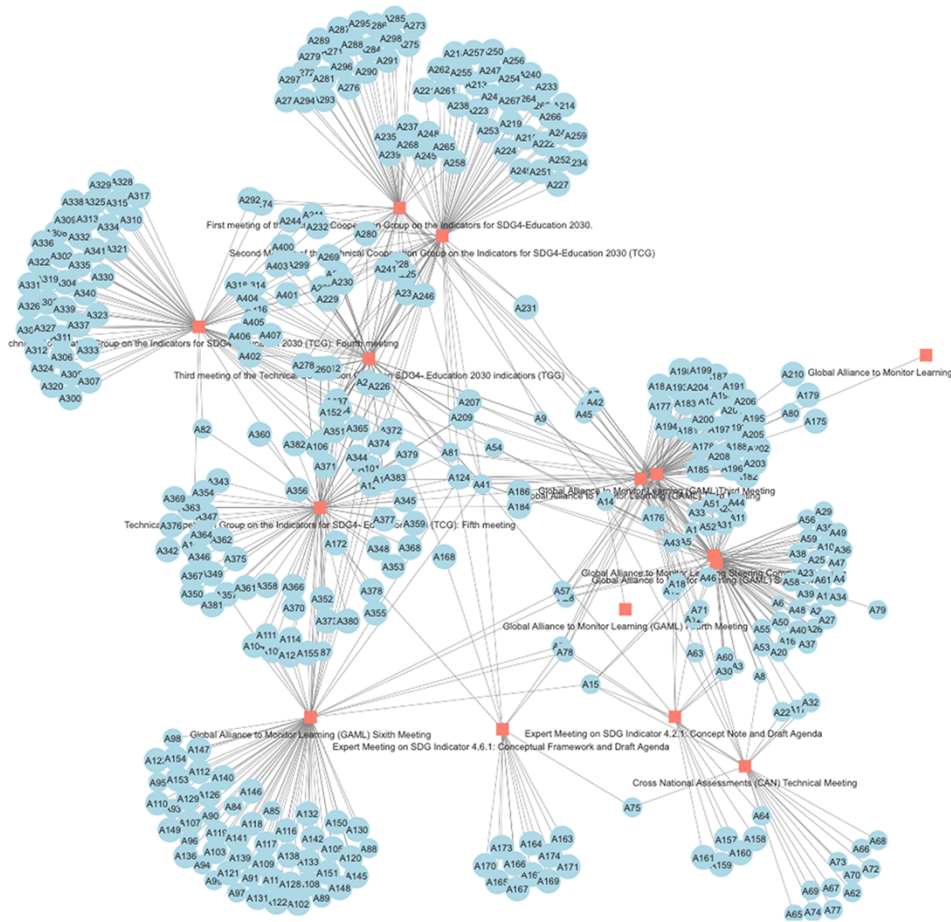
Figure 1. SDG4 Expert group networks: meetings (in squares) and actors (in circles).

and truthful—evidence production. Here, the article's focus aligns with Calkins and Rottenburg (2017) in their engagement with "infrastructuring as a material-semiotic practice": Although the stable materiality and the technoscientific dimensions of infrastructural work remain in place, the term is meant to denote the ongoing, continuous nature of infrastructuring as *practice* rather than as a solid, stable space of production. Quantification in epistemic infrastructures becomes the fuel and language of practice, as it brings together ideas and objectivity in one entangled mix. The story of the SDG4 is not one of a dominant IO versus weaker ones; instead, quantification is the hegemonic force enabling not only the technical aspects of the agenda but crucially the political ones too.

Indeed, one of the main findings of the METRO project (within which the case of the SDG4 was studied) is the changing role of IOs; that is, rather than assuming the expert role of the data producers (therefore asserting their credibility through the production of scientific truth), they have taken a new, brokerage role (author; Bandola-Gill, 2021), working through creating zones of visibility and intervention while simultaneously producing areas of opaqueness and invisibility. As such, they appear to capitalize on the complexity of a governing and measurement architecture that does not allow comprehensive knowledge and understanding of all its parts; in fact, not only IOs but all participant actors appear content to offer minimum agreement on the basis that the fluidity and the gaps in the process allow them to manipulate the infrastructure in various ways, without needing to have any more than localized knowledge and influence.

The article pointed to two further aspects of the work of infrastructuring that we need to take into account: that is, their temporal and spatial elements. First, starting with the concept of time, any

infrastructural investment has a temporal element that is not only evident in the passage of chronological time but is also palpable in the transformational intent and the promise of improvement and of a utopian perfectibility, accentuated by the power of numbers and their perceived infallibility. This promise of an anticipatory better future is central in the work of the SDGs—when it comes to SDG4, it has almost taken a moral dimension and sense of urgency (author), capitalized to either speed up or slow down the process depending on context. The SDG4 discursive analysis of reports and declarations (author) shows infrastructural meaning to be produced through gathering past failures and future ideals into an unfolding anticipation in the present. The case showed that apart from the anticipatory talk, a certain slowness of time was important in laying down the foundations of the new agenda, avoiding shocks and too sudden changes. Once the groundwork was done, after 2015, we see the process speeding up, coupled with an emphasis on expanding the infrastructure spatially and including a great variety of actors, both geographically, in terms of sectors as well as the ideas and interests that contributed to its production.

Indeed, the analysis of the SDG4 is a telling example of how problematic the conceptualization of global governance as a top-down agenda, from some imagined center to the peripheries, is. Although power asymmetries are of course ever present, the continuous expansion and complexity create new gaps and zones of opacity. The work of infrastructuring is spatial: It is about creating multiple centers and peripheries: "rather than a margin responding to the center, we are confronted with a fractal image, in which a multiplication of margins and centers proliferate alongside infrastructural developments" (Harvey et al., 2017, p. 17). As a result, failures and dead-ends are frequent. Uncertainty and incomprehensibility extend beyond end-users to also encompass designers and experts. Different hierarchies of expert work emerge, since expert knowledge is necessary not only to negotiate the goals but also to decide which expert functions have to "sink" into the purely technical, statistical work, and which ones need to be assigned to the "higher" knowledge producer, i.e., more experienced or better-networked experts.

Finally, the article used the practice of the meeting, as the site where multiple infrastructural elements meet: the technical, the political, as well as the material and the social. The paper showed how failing metrics (i.e., disputable evidence) become significant capital as they help repair gaps in the infrastructure by calling for the production of new, better data; thus mishaps and inconsistencies acquire generative power that leads to the production of new discourses and new evidence which, in turn, create further conflicts and alliances. Ultimately, the article showed the ways that quantification enabled a paradigmatic policy shift, through the imbrication of measurement and politics into one entangled mix. It showed the ways that the work of numbers—in fact, the work of *failing* numbers—folded all the contested discourses, decision-making, politics, and ideas into its processes. The actors the paper examined had plenty of technical disagreements, ideological clashes, and conflicting interests; however, the one perspective they all shared was that quantification was the only way to engage with setting the policy agenda and creating policy momentum. Thus, the making of the SDG4 represents a paradigmatic policy shift; one that is not only to be traced in the move from schooling to the policy prioritization of learning outcomes but also in the very production of global public policy as an epistemic infrastructure.

Funding

Conflict of interest

None declared.

References

Anderson, K. (2014, September 1). The Learning Metrics Task Force 2.0: Taking the Global Dialogues on Measuring Learning to the Country Level. https://neqmap.bangkok.unesco.org/wp-content/uploads/2019/06/The-Learning-Metrics-Task-Force-2.0.pdf.

Bandola-Gill, J. (2021). The legitimacy of experts in policy: Navigating technocratic and political accountability in the case of global poverty governance. *Evidence and Policy: A Journal of Research, Debate and Practice*, *17*(04), 615–633(19). https://doi.org/10.1332/174426420X16000980489195.

Barrett, A. M. (2011). A Millennium Learning Goal for education post-2015: A question of outcomes or processes. *Comparative Education*, *47*(01), 119–133. https://doi.org/10.1080/03050068.2011.541682.

Barro, R. J., & Lee, J. W. (1996). International measures of schooling years and schooling quality. *The American Economic Review*, *86*(02), 218–223.

Benavot, A., & Smith, W. (2020). Global Learning Metrics as a ready-made solution to a manufactured crisis. In A. Wulff (Ed.), *Grading Goal Four: Tensions, threats and opportunities in the SDG on quality education* (pp. 238–261). Brill.

Bowker, G. (1995). Second nature once removed: Time, space and representations. *Time and Society*, *4*(01), 47–66.

Calkins, S., & Rottenburg, R. (2017). Evidence, infrastructure and worth. In P. Harvey, C. B. Jensen & A. Morita (Eds.), *Infrastructures and social complexity* (pp. 253–265). Routledge.

Colebatch, H. K. (2007). What work makes policy? *Policy Sciences*, *39*(04), 309–321.

Crouch, L., & Montoya, S. (2019). Measuring and Reporting on SDG 4 and Beyond: Lessons from the Innovations in Global Learning Metrics Symposium'. In UIS (blog) http://uis.unesco.org/en/blog/measuring-and-reporting-sdg-4-and-beyond-lessons-innovations-global-learning-metrics-symposium.

CUE. (2011). *A global compact on learning. Taking action on education in developing countries*. Center for Universal Education at Brookings.

Fairclough, N. (1995). *Critical Discourse Analysis*. Longman.

Filmer, D., Hasan, A., & Pritchett, L. (2006). A millennium learning goal: Measuring real progress in education (CGD Working Paper No. 97). Washington, DC: Center for Global Development.

Fontdevila, C. (2021). *Global governance as promise-making. Negotiating and monitoring learning goals in the time of the SDGs* [PhD thesis]. Facultat de Ciènces Polítiques I Sociologia, Universitat Autònoma de Barcelona.

Fukuda-Parr, S., & McNeill, D. (2019). Knowledge and politics in setting and measuring the SDGs: Introduction to special issue. *Global Policy*, *10*(0Suppl. 1), 5–15.

Geertz, C. (1980). *Negara: The theatre state in nineteenth-century Bali*. Princeton University Press.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature*, *40*(02), 436–482.

Global Education Monitoring Work. (2012). *Youth and skills: Putting education to work*. UNESCO.

Gunter, H., Grimaldi, E., Hall, D., & Serpieri, R. (2016). *New public management and the reform of education: European lessons for policy and practice*. Routledge.

Hanushek, E., & Kimko, D. (2000). Schooling, labor force quality and the growth of nations. *American Economic Review*, *95*(05), 1184–1208.

Harvey, P., Jensen, C. B., & Morita, A. (2017). *Infrastructures and social complexity*. Routledge.

IEG-WB. (2006). *From schooling access to learning outcomes. An unfinished agenda. An evaluation of World Bank support to primary education*. The World Bank.

King, K. (2017). Lost in translation? The challenge of translating the global education goal and targets into global indicators. *Compare: A Journal of Comparative and International Education*, *47*(06), 801–817.

Klees, S. J., Stromquist, N., Samoff, J., & Vally, S. (2019). The world development report on education: A critical analysis. *Development and Change*, *50*(2), 1–18.

Larkin, B. (2013). The poetics and politics of infrastructure. *Annual Review of Anthropology*, *42*(01), 327–343.

Latour, B. (1987). *Science in action*. Harvard University Press.

Luhmann, N. (1969). *Legitimation durch Verfahren*. Suhrkamp.

Mundy, K., Green, A., Lingard, B., & Verger, A. (2016). *Handbook of global education policy*. Wiley-Blackwell.

Ozga, J., Grek, S., & Lawn, M. (2009). The new production of governing knowledge: Education research in England. *Soziale Welt*, *60*(04), 353–369.

Prada Uribe, M. A. (2012). Development through data? A case study on the World Bank's performance indicators and their impact on development in the Global South. IRPA Working Paper GAL Series n. 5/2012 Society. Cambridge, MA: Harvard University Press.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, *43*(03), 377–391.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(01), 111–134.

Stevens, A. (2011). Telling policy stories: An ethnographic study of the use of evidence in policy-making in the UK. *Journal of Social Policy*, 40(02), 237–255.

Tichenor, M., Merry, S., Grek, S., & Bandola-Gill, J. (2022). Global public policy in a quantified world: Sustainable Development Goals as epistemic infrastructures. *Policy & Society*, 41(4), 431–444.

Tikly, L. (2015). What works, for whom, and in what circumstances? Towards a critical realist understanding of learning in international and comparative education. *International Journal of Educational Development*, 40, 237–249. https://doi.org/10.1016/j.ijedudev.2014.11.008.

UIS. (2016). First meeting of the Global Alliance to Monitor Learning (GAML). UIS. http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/08/GAML-first-meeting-concept-note.pdf.

UIS. (2017). Technical Cooperation Group on the SDG 4 - Education 2030 indicators (TCG). Terms of Reference. http://tcg.uis.unesco.org/wp-content/uploads/sites/4/2018/08/TCG_ToRdraft_20170929.pdf.

UIS/CUE. (2013). *Toward universal learning. What every child should learn*. UIS/Center for Universal Education at Brookings.

UNESCO. (2016). *First meeting of the SDG-education 2030 steering committee*, 25–26 May 2016 (ED-2016/ED2030/ME/1).

UNESCO. (2017). *Unpacking sustainable development goal 4/education 2030*.

United Nations. (2015). *Transforming our world: The 2030 Agenda or Sustainable Development*. UN.

Unterhalter, E. (2019). The many meanings of quality education: Politics of targets and indicators in SDG 4. *Global Policy*, 10(0S1), 1.

Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(01), 5–30.

Winner, L. (1986). Do statistics have politics? In D. Mackenzie & J. Wajeman (Eds.), *The social shaping of technology* (pp. 28–40). Open University Press.

Wodak, R., & Meyer, M. (2001). *Methods of Critical Discourse Analysis*. SAGE.