

Learning Assessments in the Time of SDGs: New Actors and Evolving Alliances in the Construction of a Global Field

Clara Fontdevila

1 Introduction

Large-scale learning assessments feature prominently in the indicator framework associated with the SDG 4 agenda as key data sources to monitor progress on this area, particularly given measurement challenges relative to comparability and data availability. While the assessment of student learning has been a practice on the rise since the mid-1990s (Benavot & Tanner, 2007; Heyneman & Lee, 2016; Kamens & Benavot, 2011; Pizmony-Levy, 2013), the adoption of the SDG 4/Education 2030 agenda represents a milestone in the consolidation of such a trend. Both the global and the thematic indicator frameworks established for the monitoring of Education 2030 are unambiguous on the need for countries to adopt or participate in some form of learning assessment so that student achievement can be reported on an internationally comparable scale. Up to five targets in SDG 4 include one or more learning-related indicator, and five out of 11 global indicators require reporting on student learning, skills, or knowledge. These measurement needs turn the adoption and use of large-scale assessments (LSAs) into an essential condition for tracing and tracking progress in the new education goals.

The magnitude of the changes and the new dynamics brought about by the SDG 4/Education 2030 agenda lend themselves to a productive enquiry on the relationship between the different organisations involved in the promotion, administration, or use of LSA. The shift entailed by the new education agenda is not only likely to have a direct and positive impact on the centrality and legitimacy of assessment. More importantly, this push is likely to penetrate the agendas of these organisations and to affect the interrelationships among the different actors with a role in LSA.

So far, the relationships among these organisations have received limited attention. Particularly when it comes to developing countries, the links between international large-scale assessments (ILSAs) and national large-scale assessments (NLSAs) have rarely been explored in depth. To some extent, this

could be the result of a particular division of labour within the scholarly analysis of LSAs. However, even the relationship among different ILSAs remains only partially explored, and most of the work has focussed on the role, evolution, and influence of major ‘players’, including the OECD and the International Association for the Evaluation of Educational Achievement (IEA) (e.g., see Grek [2009] and Pizmony-Levy [2013], respectively). In short, the global learning assessment landscape has been explored only to a limited extent. Changes brought about by SDG 4 constitute a useful entry-point to understand both its structure and recent changes.

2 Key Concepts

In this chapter, I inquire into the reconfiguration or reshaping of the global assessment field as entailed by the negotiation and adoption of SDG 4 in order to gain a better understanding of the composition and structure of this community of practice. To this aim, I examine changes both in the institutional agendas of the different organisations involved in the promotion and administration of LSAs, and in the relationship among these agencies.

The working hypothesis that orients this chapter is that the negotiation of SDG 4 targets and indicators has decisively contributed to the consolidation, integration, and diversification of a global field of assessment. The notion of *field* is here used in the sense advanced by Bourdieu (1983) – a structured and differentiated social space of specialised practice revolving around distinctive beliefs and institutions, and in which different actors struggle and compete for the preservation or transformation of their own relative positions.

More specifically, the degree and impact of the SDG 4 agenda are examined in relation to the key dimensions of field autonomy put forward by Buchholz (2016) in her discussion of global fields. These mechanisms include (a) the establishment of a distinctive set of institutions (i.e., the articulation of an institutional infrastructure in which the field-specific principles become objectified); (b) the existence of an distinctive ‘form of belief’, which construes a specific sphere of practice as distinctive, independent, and valuable in its own right; and (c) the emergence of autonomous and field-specific principles of hierarchisation. Later sections of the chapter, dedicated to presentation of the main findings, are structured around these three subprocesses.

Methodologically, this chapter builds on the combination of three methods. First, the research draws on the analysis of documentary data derived from the main reports, policy briefs, and presentations prepared by the main

organisations involved in the different consultation and coordination mechanisms described (see Table 12.1). The analysis aimed to identify frequent themes and foci of debate within these documents, as well as the positions, priorities, and framings mobilised by the different stakeholders.

Second, semi-structured interviews were conducted with 41 key informants, experts, and representatives of these organisations (see Table 12.1). The interviews were transcribed and correspondingly anonymised in order to guarantee the confidentiality of the information. The purpose of the interviews was to gain an understanding of the different motivations and incentives driving different organisations' engagement in the SDG 4 process, as well as to identify supporting frameworks, normative beliefs, and main areas of contention.

Finally, the research benefited also from the observation of one of the meetings convening and fostering debate and discussion among these actors, the third meeting of the Global Alliance to Monitor Learning (GAML), in Mexico City, May 2017. The main focus of observation was on the informal and formal relationships and communication patterns among different actors within

TABLE 12.1 List of interviewees

Group 1	Bilateral and multilateral development agencies and banks; multistakeholder partnerships	WB, IADB, USAID, GPE, DFID	7
Group 2	UN system agencies	UIS, UNESCO, UNICEF, UN-Women, GEMR	11
Group 3	Private sector organisations (research institutes, foundations, think tanks)	RTI, Pearson, CUE-Brookings, Hewlett Foundation, ACER	9
Group 4	NGOs and civil society organisations	Save the Children, ActionAid, GCE	7
Group 5	Regional assessment networks; agencies and organisations in charge of ILSAs, citizen-led assessments	OECD, IEA, ASER Centre, PASEC	4
Group 6	Others	University-affiliated experts, UN-related or UN-supported initiatives (e.g., Education Commission)	3

the field, as well as the procedures and mechanisms of decision-making and consensus-building.

3 The Institutionalisation of a Global Assessment Infrastructure

In this section, I focus on the first of the mechanisms described by Buchholz in the construction of a field, that is, is the ‘establishment of an institutional infrastructure for the worldwide circulation of ideas, persons, and goods across borders’ (2016, p. 14). The establishment of forums and spaces that enable regular exchange (and competition), and help connect the different players in the field, is instrumental in ensuring the global outlook of the different stakeholders. These organisations and events play an integrative and unifying role, crucial for the construction of a field as a space of specialised practice, and an arena in which organisations position themselves vis-à-vis other stakeholders involved in a given sphere of activity.

From the outset, the many national, regional, and international assessments established in the mid-1990s have been connected, especially through their relationships with international agencies, nongovernmental organisations, regional organisations, and professional networks. In fact, some scholars have identified international organisations, nongovernmental organisations (NGOs), and regional associations as the main carriers and agents of diffusion of both national and international assessment (Chabbott, 2003; Kamens & Benavot, 2011; Kamens & McNeely, 2010; Lockheed, 2013). However, these promotion and diffusion efforts seem to occur in a disorganised or highly decentralised way. Hence, the general interconnectedness has not translated into the articulation of an institutional infrastructure in which the field-specific principles become objectified. The following section will describe some of the mechanisms and spaces that have contributed to some degree of integration and institutionalisation.

3.1 *The Learning Metrics Task Force Initiative: Laying the Foundations for a Global Debate*

Early stages of the run-up to the articulation of SDG 4 prompted some key changes in a scarcely institutionalised community of practice. As early as 2012, the establishment of the Learning Metrics Task Force (LMTF)¹ laid the foundations for the construction of a global infrastructure able to foster a minimal integration of the different efforts in place. More specifically, the socialisation and familiarisation effect brought about by LMTF meetings played an important role in constructing new and shared meanings and the legitimation of the assessment programme.

The LMTF was envisaged as a multistakeholder partnership co-convened by the Center for Universal Education (CUE) at the Brookings Institution and the UNESCO Institute for Statistics (UIS). The taskforce kicked off in the early years of the post-2015 global education debate, and in its first phase, it explicitly focussed on ‘catalyzing global dialogue and developing a series of recommendations on learning assessments’ (Anderson & Ditmore, 2016, p. 4). To this end, the taskforce organised three open consultations and launched three thematic reports and a summary report. Beyond the specific recommendations that resulted from the discussion (a question to which I will return later in this chapter), the impact of the LMTF was central in creating a sense of community and common purpose among different agents involved with assessment and monitoring activities. A UN staff member (Group 2) interviewed for this study stated:

If you look at the work of the LMTF you will see that the discussion of indicators was already there and was quite influential. Of course, the LMTF was launched in 2012, and the agenda evolved. ... But the technical work involved the same people, or more, that were involved later in the GAML. There was a natural link between the two ... so you can see a political agenda on learning ... on the technical seminal work of the LMTF ... and then of course the SDG 4 itself. So, I do really think the LMTF was quite a rich experience because of the diversity of people involved.

It is also important to take into account that, while the LMTF was formally a collaboration between CUE and the UIS, different interviewees suggest that CUE was always far more in control of the agenda than the UIS. The fact that the UIS joined these efforts only at a subsequent stage (and, apparently, in a rather accidental way) suggests that, during this early phase, the Institute had a rather limited political clout. A private sector representative (Group 3) commented:

It’s interesting because the LMTF was co-led by the UIS and the Brookings Institution ... but all the networking, consultation ... was done by the Brookings Institution. So the UIS was doing things related to mandate ... but really, the Brookings Institution was instrumental.

3.2 *From the LMTF to the LMP and GAML: The UIS in the Driving Seat?*

The constitution of the Learning Metrics Partnership (LMP) and, later on, of the GAML, emerged to some extent as a continuation of the LMTF effort. Most

of the interviewees for this study regard them as a sort of prolongation of the LMTF. However, these new spaces proved a key opportunity for UIS to regain a position of authority. Particularly in the GAML, the guidance and ascendancy exerted by the Institute have been clearer and stronger than in the LMTF. According to different interviewees, this was a deliberate move prompted to a large extent by a change in the leadership of the Institute. According to an interviewee from a development agency (Group 1):

I think the first thing Silvia Montoya [the current director of UIS] did when she came, she stopped the partnership. Because she wanted to understand all the moves, and why we were engaging and why they were driving this. Because Brookings was driving the LMTF, and in the work that was done, UIS was not contributing much. We convened meetings and contributed to the publications, but we didn't really contribute a lot to this. And it was not participatory, in essence. And UIS wanted something participatory. That's why GAML was created, it's more participatory.

There is a certain consensus among the interviewees that one of the key factors explaining the repositioning of UIS in the debate and the progress in the 'global assessment conversation' was the appointment of a new Director of UIS whose motivation and leadership capacity would differ notably from her predecessor's. However, the leading position enjoyed by UIS from mid-2015 also owes much to a series of external changes not directly connected to the agency exerted by UIS. In particular, it is largely the result of the formal recognition of the Institute as the custodian agency for the reporting of global indicators (UIS, 2016b), and as the main source of cross-comparable data on education (WEF, 2015).

It is however unclear to what extent UIS was prepared or willing to assume this leading position and whether other agencies in the development field regarded UIS in these terms. In this sense, it is important to take into account that UNESCO-led production of statistics was in fact subjected to heavy criticism during the 1990s, a situation that the creation of UIS in 1999 could only partially reverse. The legitimacy and authority enjoyed by UIS have thus tended to be limited and subject to strong competition from other international organisations, including the World Bank and the OECD (Heyneman, 1999; Cussó & D'Amico, 2007).

The LMP was conceived as a joint initiative of the UIS and the Centre for Global Education Monitoring of the Australian Council for Educational Research (ACER-GEM),² in partnership with the Australian Government's

Department of Foreign Affairs and Trade (DFAT). The LMP was oriented to 'develop a set of nationally and internationally-comparable learning metrics in mathematics and reading, and to facilitate and support their use for monitoring purposes in partnership with interested countries' (UIS-ACER, 2014, p. 1). The flagship products of the initiative included advances in the UIS Catalogue and Database of Learning Assessments, and the Learning Assessment Capacity Index database. These two products, in fact, were to some extent the embryonic version of the work that would be further developed by the GAML. In any case, and beyond the more tangible outputs of that effort, the LMP contributed to the development of an institutional structure enabling and furthering the circulation of ideas and people within the field of assessment.

The GAML, in turn, was originally defined as an 'umbrella initiative to monitor and track progress towards all learning-related Education 2030 targets' (UIS, 2016b) and has been characterised by an evolving structure as well as by its changing composition. The singularity of the GAML lies in the fact that it constitutes a space of debate separate from the Technical Cooperation Group (TCG), a platform convened by both UIS and the UNESCO Education Sector's Division of Education 2030 Support and Coordination. While the TCG has a political mandate to develop and debate the SDG 4 thematic indicators, the UIS decided to keep the debate on learning outcomes within GAML, an ad hoc platform.

The separate existence of the GAML has major implications, particularly given the initially limited presence of country representatives, compared to TCG. According to the last note on its governance structure (UIS, 2017d), its membership is open to any individual willing to contribute to the work of GAML, with members typically falling under nine different categories: international organisations, development partners, regional organisations, regional development banks, civil society organisations, UIS technical partners, assessment organisations, scholars and academics, and representatives of UN member states (who remain however a very limited fraction of the total).

In terms of governance, the management of the platform (defined as coordination, support, and logistic functions) is handled by a Secretariat hosted by UIS, while its general oversight and guidance on priorities is the responsibility of a Strategic Planning Committee. The latter is composed of the UIS Director, a chair, and several vice co-chairs, including representatives of international organisations, civil society, teachers' unions, regional assessments, global assessments, and country representatives. Decisions are to be endorsed by the GAML plenary during in-person meetings, which it assumed all the members will attend. In this sense, GAML is by definition an inclusive and accountable space. The decision-making procedures, however, remain relatively

underdeveloped, unspecified, or unclear to most of the interviewees for this study. The protocol and procedures for invitation, as well as the channels used to encourage participation, also remain rather unclear at the time of writing.

3.3 *A First-of-Its-Kind Initiative?*

As noted above, the GAML has not by any means been the first global space to serve socialisation purposes within the learning assessment community. For example, the unifying or brokering role played by the PISA for Development (PISA-D), which parallels the efforts described in this section, should not be underestimated. The PISA-D project was devised as an extension of PISA to lower- and middle-income countries (Addey, 2017) and relies to an important extent on the funding provided by bilateral and multilateral donors who have also been involved in the GAML efforts.

In this sense, the assessment field has never been a fully 'balkanised' or fragmented realm. On the contrary, the different professionals and organisations concerned with assessment (as promoters, designers, funders, and so on) have nurtured informal ties and formalised relations of cooperation for years. However, the specificity and novelty of the efforts described previously related to the *global* perspective they foster – the fact that they are not organisation- or region-specific. In addition, the coordination efforts under way are likely to ensure a greater legitimacy of the global assessment field within its broader environment, the global education policy field. Because democratisation is increasingly regarded as a key element when it comes to securing legitimacy in the context of globalisation (Buchanan & Keohane, 2006), the inclusive and accountable nature of the UIS-promoted platforms is likely to ensure high levels of social acceptance.

However, the distribution of roles and power is far from settled; it is in fact continuously negotiated and built by a variety of actors and forces. On the one hand, the GAML has contributed to ensure a much more central and leading role for UIS (especially in comparison to the LMTF initiative). However, the limited authority and normative capacity enjoyed by UIS make such a position rather vulnerable. Also, UIS is considered a latecomer in the learning assessment landscape and has long focussed on the adult literacy field through the LAMP programme (Literacy Assessment and Monitoring Programme), which has been affected by resource and prestige challenges since its inception (Addey et al., 2017). Some of the interviewees for this study referred to the limited financial capacity of UIS, noting that it could put in jeopardy both the success of the global reporting effort and UIS's leadership or steering capacity. A staff member of a development agency (Group 1) expressed this view:

I think there's a problem of strategic orientation of UNESCO. UNESCO is simply incapable of prioritising. This is a massive area of public good where UNESCO has a comparative advantage from its position to provide. ... The financial situation of UIS is ... a sign ... that UNESCO actually does not understand that. ... And that's sad, because it leaves the door open to another organisation that may be less well-placed to guarantee minimum standards for such a process to be beneficial for the world. Of course, UIS is trying and they will get some funding for that ... but it's not the way it should be.

In fact, much of the work of the GAML appears to be highly dependent on the funding of UIS and a very limited group of agencies and organisations – the UK Department for International Development (DFID), the Australian Department of Foreign Affairs and Trade (DFAT), and the Hewlett Foundation (UIS, 2017d). While these financial contributions do not necessarily translate into higher levels of influence, the overreliance on such a limited number of partners could have an impact on UIS's image of neutrality and impartiality, especially given the limited formalisation regarding decision-making and invitation procedures.

The relatively limited technical expertise on LSAs currently available within UIS tends to perpetuate a certain relation of dependence on external partners such as consultancy firms, university-affiliated scholars, or independent research organisations. While such collaboration ensures a certain degree of sophistication in the execution of the objectives, it may also pose significant risks in terms of sustainability, and even of legitimacy if not accompanied by the necessary levels of public scrutiny and institutional capacity-building.

4 A New Vision for a New Agenda: The Imperative of Assessment

This section considers a second key mechanism in the articulation of a social field, that is, the existence of a distinctive 'form of belief' that construes a specific sphere of practice as distinctive, independent, and valuable in its own right (Buchholz, 2016). The formation of a field does not arise out of the establishment of institutionalised spaces in a *mechanical* way. Another crucial dimension of field autonomy is thus the articulation of an autonomous ideology or vision which, crucially, defines and legitimatises a sphere of practice as singular and valuable in its own right, superior in some way to other practices (Bourdieu, 1983; Buchholz, 2016; Gorski, 2013).

The emergence of a relatively integrated global assessment field has revolved around the identification of assessment as the policy *solution* to an

institutionalised *problem* (i.e., one that has entered institutional agendas). The problem identified is the learning crisis. Such a coupling has been very much enabled and fostered by the particular framing of the post-2015 debate.

As early as in 2006, three economists connected to the World Bank and the Center for Global Development proposed to replace the education-related Millennium Development Goal (MDG) with a Millennium Learning Goal so that education systems would be judged and held accountable by their learning outcomes (Filmer, Hasan, & Pritchett, 2006). These scholars were vocal in their criticism of an inputs-based approach to school quality, and portrayed learning outcomes (as measured by regional or international assessments) as a proxy for education quality (Filmer et al., 2006; see also Barrett, 2011). A similar focus on student achievement was also embraced in the World Bank Education Strategy 2020 (World Bank, 2011).

Such an approach raised a variety of concerns, particularly regarding the possibility of unintended consequences, such as standardisation of the curriculum, diversion of attention from other less easily measurable purposes of education, or lack of attention to the quality of the process (Barrett, 2011; Bonal, Verger, & Zancajo, 2015; Klees, 2013; Rose, 2015). Eventually, the final wording and formulation of the goals and targets avoided such pitfalls by including learning targets with other quality-related targets related to inputs and processes (Bonal et al., 2015; Rose, 2014).

However, most of the debate in the run-up to the formulation of SDG 4 continued to revolve around the so-called 'learning agenda'. The increasing availability of data evidencing low levels of learning despite global progress in enrolment, contributed significantly to the growing visibility and centrality of the learning/quality binary. The 'global learning crisis' spotlighted by the EFA Global Monitoring Report (UNESCO, 2012b, 2014f) and its equity-oriented framework were instrumental in giving currency to the issue and fostered an alignment of a variety of stakeholders around the need to pay greater attention to quality and/or learning outcomes. The negotiation and adoption of SDG 4, as well as the development of monitoring mechanisms, contributed greatly to secure and disseminate the 'quality turn' within the global discourse on education, understood as an effort to transcend a focus on schooling and enrolment figures as key indicators of progress. (For further discussion of this transformation, see Chapter 9 by Yusuf Sayed and Kate Moriarty; Sachs-Israel, 2017; Sayed, Ahmed, & Mogliacci, 2018.)

The coupling of the learning/quality problem with the 'assessment solution' and their rise in the global agenda are the consequence of a wide range of predisposing and precipitating factors (whose complexity is beyond the scope of this chapter; for more detail, see Chapter 11 by Aaron Benavot and William C. Smith). The run-up to SDG 4 contributed decisively to securing this

connection and making it visible. Such intertwining is particularly clear in the first works produced by the LMTF.³ Indeed, most of its activity revolved in fact around measurement-related interventions. According to three promoters of the LMTF initiative, ‘The real debates now center on how to conceptualise and assess learning within a global framework’ (Winthrop, Anderson, & Cruzalegui, 2015, p. 298). The different reports produced and resulting from the LMTF consultations tended to emphasise measurement as an essential and necessary (although not sufficient) part of the policy solution to the learning problem.

The importance attributed to measurement as a key part of the equation is explicit in Winthrop et al. (2015, p. 300):

[MDG] goals and indicators were chosen over other EFA goals because they were easier to measure at the global level (Winthrop and Anderson, 2013). Robust data were available in a majority of countries and comparability across national contexts was possible. ... Availability of metrics has especially driven funding from donors who choose funding priorities based on areas where they perceive their external support can have a measurable impact.

This interpretation of the comparatively poor traction of the EFA agenda and the limited progress of the education-related MDGs (as opposed to other areas, including health) appears to have inspired to a large extent the quest for globally comparable learning and its framing as a key part of the solution. Given the high levels of visibility of the LMTF, one of the unintended effects of the initiative could be what has been described as a ‘conflation’ between quality and benchmarking (Soudien, 2013). Assessment thus became a key policy route to address the quality and equity imperatives, making it a priority area for most bilateral and multilateral aid agencies or lending programmes.

5 Principles of Hierarchisation: An Improbable Agreement?

This section examines the third mechanism described by Buchholz (2016) as key in the articulation of a field, that is, emergence of autonomous and field-specific principles of hierarchisation. In fact, the emphasis on the existence of a *common vision* of the prior section risks obscuring the existence of contending approaches and visions within a field. Similarly, theories of the global diffusion of assessment risk eliding the existence of competition dynamics (or power relations) within a given sphere of practice.

The assessment field is far from a unified bloc. Rather, like most social fields, it is an arena of struggle about the kind of knowledge that is valued and of competition for dominant positions (Go & Krause, 2016). Far from being equal partners in a flat world, the different agencies and organisations involved in the debate strive and compete for global legitimacy. In that sense, the nascent global field of assessment is far from being settled, that is to say, one that ‘has reached a higher degree of consolidation, being characterised by a “robust social order” and established “rules of the game”’ (Fligstein & McAdam, 2012, p. 92).

5.1 *Divisive Issues and Classification Struggles*

Most areas of disagreement were identified at an early stage by some of the leaders within the LMTF. Taking stock of the different debates fostered by the LMTF activity, Winthrop et al. (2015) summarised the areas with a lack of consensus as follows: a narrow versus broad scope of learning measurement, globally comparable versus nationally defined goals and targets, universal versus country-determined benchmarks, measuring learning for all versus only those who are in school, and top-down versus bottom-up implementation.

While not necessarily constituting fault lines, these different positions emphasise and value different aspects of assessment and suggest two more general positions, confirmed in interviews for this study. In general terms, an opposition or divisive line emerges between those who value scientific sophistication and a particular variety of expertise cultivated in highly specialised agencies with a proven record, and those who prioritise the value of local relevance, context-sensitivity, or country ownership. To some degree, this struggle intersects with (but does not completely equate with) the different value attributed to cross-country comparability and, more generally, to the ultimate purpose given to assessment. While some agents expect that assessment will trigger change by providing domestic governments with better or more reliable information necessary to improve policy formulation and planning, others emphasise the value of cross-comparability or the ‘disciplining’ effect of national assessments within the global arena, that is, cross-accountability pressures resulting from global reporting. Hence, although these theories are not incompatible, they necessarily end up placing a different value on comparability. A UN staff member (Group 2) interviewed for this study noted:

The two logics are the two ends of the spectrum. If you are sitting at the international global level, the internationally comparable – that’s how you see things. And if you are at the other end of spectrum, you’re looking at ... you want some kind of assessment that is easy to design, which

doesn't take long, which is cheap. ... You need, in your context, to improve your interventions, to better inform yourself, to better design or to read-just interventions.

These 'classification struggles' are more likely to become visible as some organisations face the need to privilege a particular approach by means of providing financial aid, policy recommendations, technical support, etc. This is clearly the case of the GAML, whose participants are expected to reach consensus on the most appropriate tools for the reporting of global indicators, and to provide countries with guidance to improve learning assessments. The GAML and other SDG 4 fora are not exactly interest-free realms – as different organisations are likely to use them to promote and disseminate their own vision or products. Some of the representatives of the most reputable and/or long-standing assessment programmes have indeed been particularly vocal in asserting the superiority of cross-comparable assessments. Remarkably, these are strategically framed as appropriate not only given their readiness for global and thematic reporting but also as a fast track to build capacity at a national level. Representatives from the OECD and the IEA not only emphasise the technical superiority and cost-efficiency of their flagship programmes but also portray them as learning and training opportunities for participant countries.

5.2 *A Balancing Act*

At the time of writing, it is still unclear which (if any) approach will be privileged. On the one side, and at least in relation to Target 4.1,⁴ the GAML has unequivocally encouraged the expansion and strengthening of national assessments, emphasising ownership as well as the logics of assessment as a public good. These are to be plotted or anchored against a global reporting scale, after an assessment of its quality. While in the short term, global monitoring will be based on these cross-national assessments, this is not the approach privileged in the long term (Montoya, 2017; UIS, 2017e, 2017g). A development agency staff member (Group 1) argued:

You cannot make progress in this work without involving organisations with high capacity ... but then the question is how do you make sure that the outputs of that do not privilege a particular organisation. ... It's a really delicate balancing exercise. ... But they also need to satisfy certain standards in terms of how they collaborate, and what they make public, and what their agenda is ... and [it] is not that easy ... but from that point of view, I think the GAML is trying to accommodate as many players as possible.

Much of the GAML effort has been recently directed at the construction of reporting scales expected to enable a solid and rigorous linking of both national and cross-national assessments to a common scale of performance levels (UIS, 2017). Those efforts seem to avoid a zero-sum approach and to accommodate the use of different assessments.

The possibility of using national assessments for global reporting purposes represents a significant shift in learning assessment practices. While cross-national assessments had been long considered useful for international comparisons and as a means to inform national policy, the same was not true for national assessments, in that NLSAs were deemed far less appropriate for global reporting. As noted by Benavot and Koseleci (2015, p. 19), 'National learning assessments are not designed for comparing learning outcomes across education systems'⁵. Scholars had given some thought to this possibility and worked on theorising its requirements (e.g., Lockheed, 2016). This seemed more of a technical feasibility than a ready-to-be-implemented approach. To some extent, the conversation fostered by the LMTF opened the door or created the conditions for the coordination efforts that such a goal would require.

Such a shift represents a change in the relationship between dominant players, who historically monopolised visibility in relation to large-scale assessments as monitoring tools, and 'pretenders', who may be seeking a more visible and central position as new market niches unfold. The disruptive potential of such an approach lies in the fact that ILSAs have precisely constructed their authority by way of emphasising their potential for cross-comparability purposes. According to Martens (2007), the 'comparative turn' (or 'governance by comparison') was one the main drivers of OECD success. Similarly, as Grek (2009, p. 25) noted, 'the OECD has created a niche as a technically highly competent agency for the development of educational indicators and comparative educational performance measures'. The construction of a universal scale, against which any national assessment can be anchored or plotted, puts into question the comparative advantage of the IEA or OECD in that regard.

Ultimately, the privileging of any particular approach (cross-national vs. national; open-source vs. licensed models of assessment, etc.) does not depend solely on GAML, let alone UIS, guidance. The advice or support of aid and lending agencies is going to be a determinant in consolidation or spread. Through financial and technical assistance, initiatives like the Global Partnership for Education, as well as bilateral aid agencies and multilateral development banks, are likely to have a crucial impact in fostering specific models of their preference. Most interviewees for this study noted the ambivalent, divided,

or evolving attitude of most of these organisations, or do not agree about the direction of their preferences. Existing assessment programmes are free to continue advancing their own agenda, regardless of the direction that the GAML is taking. Efforts from OECD to advance the PISA-D are likely to proceed even if they are not necessarily the sole or priority approach favoured by GAML discussions.

6 Final Remarks

Most of the processes described previously are still in progress and so is the scholarship exploring them. The empirical basis informing and supporting this chapter is limited. The chapter aims only to propose some tentative explanations that will require further elaboration. Nonetheless, some preliminary conclusions can be drawn at this point.

First, evidence suggests that there is a global field of learning assessment in the making, although this is very much in a nascent stage and with little integration. The establishment of an incipient infrastructure and the development of a shared language is partly due to the growing interdependence of different types of organisations involved in assessment-related work. The existence of different and competing criteria for the categorisation of LSAs suggests that the field is in an emerging and evolving stage with its boundaries and organising principles open to (re)definition.

In fact, the articulation and unfolding of a field should not be equated to the emergence of a complete consensus among the multiple actors populating this social space. Different actors in the assessment field tend to emphasise different purposes of NLSAs and, consequently, place a different value on cross-comparability, efforts to develop domestic capacity, etc. The fact that no assessment programme enjoys a hegemonic position at the moment leaves the space open to competitive dynamics among concerned organisations. The same applies to the international organisations in charge of collecting and harmonising this data. While UIS attempts to regain a central role, the lead is likely to be disputed by other organisations, which are better resourced and enjoy even higher reputations.

Second, the field seems to be increasingly diverse in its composition, and the production of metrics and harmonisation of data is not by any means the remit of international organisations. Paradoxically, the growing integration and consolidation of the assessment field have been accompanied by its opening to a wider range of stakeholders. The negotiation and early implementation of SDG 4 have increased the number further. Certain private actors, including

think tanks and research institutes, seem to have deployed considerable influence in the configuration of the field.

The self-ruling nature of these organisations raises some concerns regarding accountability and transparency. While this 'private' status does not preclude the possibility of productive exchanges, it is very likely to generate conflicts of interests in the medium- and long-term, or to make public scrutiny increasingly difficult. It is thus important to develop mechanisms to hold these actors accountable, as well as to ensure that their contributions are guided by democratic and transparency principles. The emerging institutional architecture of the field should be equipped with a clear and well-defined governance structure. While the GAML has the potential to fulfil the role, its convening capacity and its normative and scientific authority are far from consolidated. Uncoordinated efforts on the part of lending agencies or assessment programmes may reinforce the centrifugal dynamics and fragmentation referred to previously. At this stage, it is unclear if the monitoring structures implemented as a result of the SDG 4 agenda will be able to counter these dynamics.

Third, the emergent nature of the field risks having a diverting effect in relation to other areas that also require improved measurement, especially in terms of political and technical attention. The assessment needs associated with the new agenda could create a perverse incentive for organisations involved in the collection of education indicators and even for other organisations in the development field, not traditionally engaged in data collection. As a global 'assessment market' unfolds, the prestige and visibility gains associated with its central positions may motivate some organisations to put additional effort into this area. As a consequence, other education dimensions that are indeed central to the SDG 4 agenda may remain underdeveloped or underscrutinised in practice.

Other challenges created by the push for LSA concern the ultimate potential of assessment and monitoring as levers for change. However, more empirical work is needed to better understand whether LSAs can live up to their promise, and, especially, under which circumstances. It is not clear, for instance, how to ensure that countries' *participation* in LSAs translates into greater capacity to make use of data or to eventually develop their own assessment capacity. Similarly, risks associated with the narrowing of curriculum, conflicts of interest among providers, or countries' dependence on external support should not be underestimated. While this is not necessarily the case, and while capacity-building and technical programmes projected by the SDG agenda could play a central role as enablers of an effective and balanced use of such tools, those risks constitute an empirical question that can only be addressed through research and accurate monitoring.

Finally, and in spite of the abovementioned limitations, the preliminary results of this research suggest some possible future research lines.

First, while this research focussed on assessment programmes for basic education (primary and lower secondary), learning assessments cover a range of educational sectors (early years, adult education, higher education, etc.). Each of these engages a different combination of interest groups and presents different trajectories that could be tracked empirically.

Second, the past and future development of the assessment field could be explored with a clearer emphasis on its relationship with its broadest environment, that is, the global education policy field (Jakobi, 2009) as well as extra-educational structures, events, and processes (Dale, 2005). The links between LSAs and SDG 4 and wider SDG processes are obvious starting points. The relative autonomy of this global field in relation to national assessment fields should be also examined in more depth in order to understand how different national assessment cultures (or education policy dynamics) are reflected in the global context.

Finally, it would be worth exploring the impact and recontextualisation of this global assessment agenda at national or subnational levels in order to understand which local processes are advanced or affected by the evolution of the global field.

Notes

- 1 Sometimes referred to as LMTF 1.0, to contrast it with a second phase, which I discuss below.
- 2 ACER is an Australian-based, not-for-profit, research-oriented organisation with a focus on education. ACER depends financially on research and consultancy contracts commissioned by education administrators as well as private, non-governmental, and international organisations. Historically, ACER has played a key role in the implementation and administration of cross-country, large-scale assessments including Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literary Study (PIRLS).
- 3 This is not to suggest such coupling was the direct result of LMTF. Before its creation, the assessment programme had already entered the political agenda of different organisations in the development field, after having been gathering momentum for a while.

- 4 Some interviewees suggest that efforts in relation to Target 4.2 (related to early childhood development, care, and preprimary education) could be headed in the opposite direction. However, this debate appears to still be developing and falls beyond the scope of this specific section of the chapter.
- 5 Similarly, the emergence of hybrid assessments combining elements from LSAs with household-based educational surveys (Wagner, 2011) would have had comparable 'diluting' effects.