Digital Development

Working Paper Series

The Digital Development (formerly Development Informatics) working paper series discusses the broad issues surrounding digital data, information, knowledge, information systems, and information and communication technologies in the process of socio-economic development

Paper No. 89

Strengthening the Skills Pipeline for Statistical Capacity Development to Meet the Demands of Sustainable Development: Implementing a Data Fellowship Model in Colombia

PETE JONES, JACKIE CARTER, JACO RENKEN & MAGDALENA ARBELÁEZ TOBÓN

2021

Publisher: ***Centre for Digital Development***
Global Development Institute, SEED
University of Manchester, Arthur Lewis Building, Manchester, M13 9PL, UK
Email: cdd@manchester.ac.uk Web: https://www.cdd.manchester.ac.uk/

View/Download from:
http://www.gdi.manchester.ac.uk/research/publications/di/

Table of Contents

Strengthening the Skills Pipeline for Statistical Capacity Development to Meet the Demands of Sustainable Development: Implementing a Data Fellowship Model in Colombia

Pete Jones[1], Jackie Carter[1], Jaco Renken[2] & Magdalena Arbeláez Tobón[3]
[1]Department of Social Statistics, University of Manchester, UK
[2]Centre for Digital Development, University of Manchester, UK
[3]Data-Pop Alliance

2021

Abstract

The UN Sustainable Development Goals (SDGs) create a global need to deliver on an agenda which relies heavily on quantitative data. This creates a demand for human capital with the requisite statistical and data skills to work on the challenges represented by the SDGs. Increasingly, long-term domestic solutions to the building of statistical capacities are being sought, in order to decrease the dependency on external support and ensure a sustainable flow of qualified professionals.

In this paper we argue for the value of thinking of statistical capacity as a pipeline which needs to begin in the education system, and illustrate how investment in this end of the pipeline can deliver a more sustainable and long-term pathway to building up the holistic skills-base to enable the challenges of the SDGs to be tackled. We describe an existing tried-and-tested experiential learning model called Q-Step based on developing skills in the workplace, and propose that the successes of this partnership-driven model illustrate how 'data fellowships' can fulfil some of the unmet capacity needs of the data revolution for sustainable development.

We illustrate our argument through a practical exploration of the development of such a learning model in Colombia. Although there are challenges in ensuring that educational access is equal, we demonstrate that there are significant opportunities and a data fellowships model opens doors for addressing local skills gaps to help deliver the SDGs.

A. Introduction

The ability to collect, handle and analyse quantitative data has never been more central to civic participation and social progress. The United Nations' 2030 agenda for sustainable development has established the Sustainable Development Goals (SDGs) as the primary framework through which societal progress is defined globally (UN, 2015). The associated Global Indicator Framework, largely the work of statisticians, seeks to measure these goals through a host of quantitative indicators, a strategy which places data and statistics at the heart of the international agenda for social, economic and environmental development (UN, 2020). There is a global ambition to deliver on these goals, yet the majority of countries continue to face challenges in their access to a workforce with the requisite statistical and data literacies to meet these demands. In this paper we aim to address the question of where these literacies are going to come from by highlighting the role university-based education interventions can play.

Our goal is to provide a critical review of the relevant literature for thinking about how statistical education initiatives can contribute to the delivery of the data-oriented agenda of the SDGs. We argue that the existing capacity development literature has primarily considered statistical capacity at the level of providing training and support to those working in national statistical offices (NSOs). However, we suggest that there are unmet needs in this area, and that there is much to be gained from thinking more about capacity development in terms of strengthening the pipeline for statistical and data literacies and skills. We propose that leveraging the SDGs as an educational tool and building partnerships with organisations working on SDG-related social challenges can help to build engaging, incentivised pathways for the development of home-grown statistical literacy. Focusing on Colombia, we offer a practical example to illustrate this discussion by exploring the opportunities and challenges of adapting an established internship-based experiential learning model for the development of statistical and data literacy. This focus on Colombia is the result of a partnership between researchers at the University of Manchester and Del Rosario University in Bogotá. Ultimately, we argue that how the SDGs and the global need for quantitative skills can be used to mobilise statistical education is an open question, but that there are a number of open doors and opportunities for strengthening the statistical capacity pipeline in Colombia and beyond.

The rest of the paper is structured as follows. In Section B, we discuss the connection between the SDGs and statistical education programmes in more detail. First, we provide some context on the SDGs, their importance, and their statistical implications. We then review the literature on statistical capacity development in order to highlight the growing recognition that more human-centred understandings of statistical capacity are needed to address the skills shortages many countries face. In Section C we explore the value of a university-based research internship or 'data fellowship' model for strengthening the statistical and data literacy pipeline in countries. We reflect on the statistical and data literacy literature, and the kinds of skills and competencies most needed and coveted for working on data-driven approaches to addressing social challenges. We describe an effective and proven internship-based model for experiential learning in the UK, and highlight how this model taps into many of the currents and unmet needs outlined in Section B. We then open up the discussion to consider how this model can be expanded

outside of the UK, arguing for the need for country-led implementations tailored to the particular national and regional needs therein. In Section D we explore the challenges and opportunities of developing this model outside of the UK more directly through a focus on Colombia. We briefly consider the baseline regional capacity, before zooming in on Colombia as a case study for evaluating how this kind of statistical education programme can achieve the highest impact in that country.

B. Statistical Capacity Building and the SDGs

B1. Background: The SDGs and their Statistical Implications

In 2015, the United Nations launched its 2030 Agenda for Sustainable Development (UN, 2015). The Agenda constitutes a plan for sustainable development which centres on 17 overarching goals, comprising 169 associated targets and 231 indicators with which to measure progress. These goals build on the earlier Millennium Development Goals (MDGs), widely seen to have been well-intentioned but hindered by data gaps, measurability issues, and an inattentiveness to the needs and perspectives of global South governments (Attaran, 2005; Fukuda-Parr and McNeill, 2019; Merry, 2019; OECD, 2015; Yayboke et al, 2017). The SDGs eclipse the MDGs in quantity (compare the SDG indicator framework's 231 indicators with the MDGs' 60), breadth and ambition, taking a much more holistic approach to sustainable development which applies to all countries and not just those considered low-income, while acknowledging that the particular path to realising them will vary country to country (UN, 2015). As such, there is a truly global need and desire to deliver on the SDGs.

This brings significant implications for the statistical education research community, as statistics and data skills and competencies are at the heart of the 2030 agenda. In the most direct sense, an unprecedented amount of data is required to populate the SDG indicators, and this creates new demand for access to individuals with the capacity to collect, handle and analyse such data. However, there are more indirect ways in which countries need to draw on a statistically skilled talent pool in order to deliver on the SDGs. A significant number of indicators are classified as "Tier 3", which means there is no established international standards or methodology for measuring them. This only increases the burden at a national level to identify appropriate data sources and proxies, a task which is complicated by the increasing consideration being given to "unofficial" statistics and data, either for SDG monitoring specifically or for evidence-based policy and social innovation more generally (eg see Lokanathan et al, 2017; MacFeely, 2019; Stuart et al, 2015). These new forms of data open up questions of data justice, ownership and privacy (Heeks and Renken, 2018). Thus, in addition to statistical analysis skills, being able to critically evaluate the value of quantitative data, as well as understanding the legal and ethical issues which such data can present, are skills which are much needed.

Moreover, the effect of the SDGs on the statistical landscape creates the need not only for more skills, but also for new skills. The next generation of statisticians and data professionals working on SDG problems need not only to possess critical data skills, but also other competencies required to navigate the new landscape. These competencies include domain knowledge about sustainability, including understanding of the complex,

interconnected nature of the goals, targets and indicators in the SDG framework (Barbier and Burgess, 2019; Dang and Serajuddin, 2020). In addition, working on the SDGs entails multiple levels of analysis (organisational, sectoral, national, regional, global) which introduces further analytical complexity and practical challenges. Moreover, efforts at addressing the SDGs involve a diversity of stakeholders, including private organisations, governments, civil society, NGOs and multilateral development bodies. A combination of analytical, technical, professional and communicative skills are therefore required in order to navigate the multi-disciplinary configurations which often mark work related to delivering the SDGs. This combination of factors implies a workforce which is equipped not only with a strong textbook understanding of statistics, but also the ability to engage with the fluidity and complexity of real-world research contexts by drawing on a hybrid skillset of competencies.

In sum, the SDGs require a steady supply of statistically and data literate social researchers. In reality, these skills are in short supply in many countries. As such, the UN Independent Expert Advisory Group on a Data Revolution for Sustainable Development has acknowledged that delivering the current agenda requires "a huge increase in the capacity of many governments, institutions and individuals" (UN IEAG, 2014: 4), given that national statistical systems around the world face significant challenges in the form of limited budgets, resources, and access to staff with the required statistical and data literacies to gather and analyse the relevant data (Fisher and Fukuda-Parr, 2019). What role can statistics education programmes play in achieving this necessary increase in capacity?

B2. Statistical Capacity Development

In order to explore how educational interventions can contribute in the space of statistical capacity development, we first need to establish that the term statistical capacity refers to "the ability of a country's national statistical system, its organisations and individuals to collect, produce, analyse and disseminate high quality and reliable statistics and data to meet users' needs" (PARIS21, 2019: 13). As we argue above, the implications of the SDGs mean that this capacity is marked by a hybrid skillset incorporating analytical, technical and professional competencies. Efforts at evaluating and developing statistical capacity are primarily coordinated by multilateral organisations such as the World Bank, the UN Statistical Division, and the Partnership in Statistics for Development in the 21st Century (PARIS21), as well as national governmental bodies such as the UK's Office for National Statistics, Statistics Canada and the French INSEE.

There is an increasing recognition in the capacity development literature that efforts at building statistical capacity need to be focused not on importing capacities from outside, but empowering those in-country in order "to provide the long-term foundation for transformation" (Stafford-Smith et al, 2017: 914). As MacFeely and Barnat put it, "capacity building is most effective when it is home-grown, long-term in perspective and managed collectively by those who stand to benefit" (MacFeely and Barnat, 2017: 897). Ensuring that there is a supply of statistical and data skills sufficient to meet the demand of the SDGs is a critically important pillar in this regard. In order to have home-grown capacities in place for statistical measurement and reporting, more attention needs to be placed on connecting

the higher-order development outcomes with the earlier stages of the pipeline of a statistically and data literate workforce capable of addressing the challenges of the SDGs.

However, statistical capacity is generally evaluated at the level of the national statistical system (NSS), without regard for the flow of sufficiently skilled talent into the system. For example, the World Bank's Statistical Capacity Indicator assessment system evaluates nations' capacity only in terms of methodology, data sources and periodicity (World Bank, n.d.). This effectively frames statistical capacity as the ability to provide data that can fulfil international monitoring standards, detaching data from the individuals that use (and stand to benefit from) the data within countries. We see two shortcomings of this perspective on statistical capacity which can be addressed through an increased role for higher education. Firstly, capacity development is often thought of as an interaction between national statistical systems and other actors such as NGOs and foreign aid and multilateral organisations. Focusing on the NSS without regard for the statistical and data literacy pipeline that feeds into it seems to either implicitly assume that the relevant capacities will be built from outside the system (e.g. by international actors providing training for NSS staff), or fail to account for the realities of statistical education in many countries, where general statistical competency rates may be low and very unequally distributed through the population (North et al, 2014; Stuart et al, 2015). Such a perspective therefore overlooks the role that investment in statistical education interventions can play in strengthening the in-country pipeline and reducing dependency on external support.

Secondly, data alone will not be sufficient for meeting the challenges of the SDGs, and countries require investment in data literacy and skills in order to sustainably improve the capacity for effective use and understanding of data (PARIS21, 2015; Stuart et al, 2015). While the benefits of investing in data production, infrastructure and technology are well-known and understood, there is still comparatively little energy directed towards the need to invest in the individual capabilities necessary to make effective use of these data and technologies. In a recent survey of national statistical offices (NSOs) around the world on statistical capacity assessment, PARIS21 found that "while only 2% of assessed capabilities target the individual, 32% of countries expressed that individual capabilities need to be improved to rise to new data ecosystem challenges" (PARIS21, 2019: 16). Moreover, Keijzer and Klingebiel caution against "the production of data and information at the expense of statistical capacity" (2017: 10), suggesting that care must be taken to ensure that the short-term desire to populate the SDG indicators does not divert resources away from the longer-term development of the overall system capacity (MacFeely and Barnat, 2017).

Recent advances in the capacity development literature begin to address these shortcomings by acknowledging that a more human-centred vision of statistical capacity is needed to respond to the current global demands. This is reflected in the "Capacity Development 4.0" (CD4.0) framework recently developed by PARIS21, which aims to "go beyond the traditional production-side interventions to also include the strengthening of data use, literacy and results" (PARIS21, n.d.). CD4.0 offers a conceptual refinement of statistical capacity in light of the 2030 agenda, which recognises that data alone are insufficient to generate the insight and knowledge required to form effective evidence-based policy, and that more work needs to be done to improve the skills base on which the delivery of the SDGs relies.

Thus, raising the level of individual skills is now being made a priority in the statistical capacity discourse, with a mixture of professional, technical and analytical skills being promoted. This addresses prior concerns that one of the barriers to the data revolution is that there is "too little investment in people and skills" (PARIS21, 2015: 19). By focusing more on developing the skills and resources necessary for the effective use of data, the CD4.0 approach acknowledges that both the short-term monitoring requirements and the longer-term development goals can be met by promoting in-country skills development. On a practical level, CD4.0 remains focused on the national statistical system and, as such, does not explicitly consider the role of educational interventions in helping to build the statistical capacities it calls for. However, it usefully emblematises the appetite for new ways of thinking about statistical capacity, as well as the clear need for the relevant skills and competencies to be built. We argue that connecting the desired outcomes of this capacity building perspective with the education system through novel approaches and partnerships can help build a stronger, more direct pipeline for a workforce with the required skills and competencies to tackle the SDGs.

B3. Partnerships for Building Statistical Capacities

While much of the focus of statistical capacity building has been on NSOs and the broader NSSs, there is growing recognition that there is a lot of valuable expertise and data which lies outside of these systems. As such, statistical capacity building is increasingly focused on the broader "data ecosystem" (PARIS21, 2020), which includes the NSO, the wider NSS, the private sector, academia, civil society and individual citizens. This means not only considering non-traditional data sources, but also the vital role of partnerships between different sectors to capitalise on the skills and experience that already exist in countries looking to develop their statistical capacity (Oburu and Yoshikawa, 2018). The SDGs often cut across traditional realms of expertise and require co-ordination between multiple actors within the data ecosystem. As MacFeely and Barnat argue, "availing of new opportunities arising from new data sources or meeting the challenges of greater demands for climate related data, geo-spatial information, better visualisation or improved explanations will require new strategic partnerships, both nationally and internationally in order to get the mix of skills and experience necessary" (MacFeely and Barnat, 2017: 898).

In fact, partnerships are represented explicitly by Goal 17 in the SDGs, a broad goal targeted at improving the means of implementation of the SDG framework, which includes capacity building and engaging "effective public, public-private and civil society partnerships, building on the experience and resourcing strategies of partnerships" (UN, 2020). The UN's *A World That Counts* report frames such partnerships as a key aspect of how the data revolution can be achieved, listing among its hopes for the future that "governments, civil society, academia and the philanthropic sector work together to raise awareness of publicly available data, to strengthen the data and statistical literacy ("numeracy") of citizens" (UN, 2014: 18).

This follows the recognition in the statistical capacity literature that the SDGs require building capacities at deeper, more foundational levels in society in order to bring about the data revolution for sustainable development (Jütting, 2016; Sachs, 2012). Thus, one of the

ways in which the current development agenda can be more sustainable is by increasing the role of partnerships with education providers to create stronger pipelines for the skills needed to work on the SDGs. As Akinsola and Sharma argue, "as we move towards 2030, what is becoming increasingly clear is that traditional approaches to promoting partnerships might not be enough to meet the objectives of SDGs. Bilateral partnerships that focus on increasing capacity at the grassroots and decreasing dependency on external support need to be promoted" (Akinsola and Sharma, 2018: 416). We suggest that, in this way, partnerships between higher education institutions and organisations producing or working with data related to the SDGs can be leveraged as an effective means for sustainably developing statistical capacity. In Section C, we discuss how this can be achieved in more detail.

B4. Summary

The SDGs create a global need to deliver on an agenda which relies heavily on quantitative data. This creates a demand for human capital with the requisite statistical and data skills to work on the challenges represented by the SDGs. However, the dominant paradigm of statistical capacity development is focused on the national statistical office, including the training of the staff therein by external actors. Increasingly, more home-grown and long-term solutions to the building of statistical capacities are being sought, in order to decrease the dependency on external support and ensure a sustainable flow of qualified labour from the grassroots level. This shift to a more human-centred and holistic vision of statistical capacity, along with an increased recognition of the value of cross-sector partnerships for building statistical capacity, opens doors for novel statistical education initiatives to play a role in strengthening the pipeline for statistical capacities. In the next section, we suggest and discuss an experiential learning-based intervention which offers a promising model for achieving impact in this area.

C. A Data Fellowship Model for Building Statistical and Data Literacies

C1. Statistical Literacy and Data Literacy

Before proceeding further, it is worth clarifying how we are using the terms "statistical literacy" and "data literacy" here, in order to establish the kinds of skills that are most important for an educational intervention that can address the unmet needs discussed in Section B. In the education community, questions of data provenance and handling have typically been considered components of "data literacy", while the more analytical and conceptual questions of quantitative analysis of data have been considered under "statistical literacy" (e.g. Schield, 2004). From this perspective, we can understand *statistical literacy* as a combination of needs and competencies comprising "the ability to produce, analyse and summarise detailed statistics in surveys and studies" and "the ability to read and interpret summary statistics in the everyday media: in graphs, tables, statements and essays" (Schield, 2010: 135), and *data literacy* as a broader "ability to collect, manage, evaluate, and apply data, in a critical manner" (Ridsdale et al, 2015). Separating out these literacies assists us in considerations of the way in which these could be addressed in the

curriculum, either in the school or university classroom, or through training in the workplace.

However, the demands of the data revolution and the rise of more holistic data science approaches to statistical research have begun to erode the boundaries between the two, along with distinctions between the needs of producers and users of data. Statistical literacy scholarship increasingly recognises that "as forms of communication and ways of presenting and receiving data change so too does what it takes for someone to be statistically literate" (Wild, 2017: 31). The critical data skills that are typically thought of as comprising data literacy are now being argued to be necessary parts of the statistical curriculum (Gould, 2017). The current age of big data, open data and new forms of data requires a more holistic skill set which includes and goes beyond understanding the basic formulas, principles and analytical aspects of statistical reasoning. In order to both leverage the power and understand the limits of the varied data sources which now permeate our world, a kind of literacy is needed which includes “a critical appreciation of data provenance and quality” (Ridgway et al, 2013), “understanding issues of data privacy and ownership” (Gould, 2017), and “understanding how data are stored” (ibid.). These skills are in addition to, rather than instead of conventional statistical literacy skills.

The conceptual, mechanical and critical aspects of doing statistics with real data are all important elements of the skillset required to work on understanding and addressing social challenges such as those represented by the SDGs (Engel, 2017; Gould, 2017; Prodromou and Dunne, 2017). Critical data skills are important as, unlike textbook data, “real data about society are often more complex and messy” (Engel, 2017: 46). If learners are to use their statistical skills to work on SDG problems, they need to be prepared to work with such real-world data and develop the various data literacy requirements that this entails. This is particularly important for ensuring that data sources which fall outside the traditional sphere of official statistics are adequately understood when these are utilised towards the SDGs. Likewise, the ability to draw insights from such data also requires statistical literacy: the social problems represented by the SDGs are highly complex, and understanding them therefore “requires the ability to explore, understand, and reason about complex multivariate data, because social phenomena do not happen in a vacuum, and their understanding requires awareness of how variables co-vary, or affect each other, or are situated in a network of causal factors that may change over time in manifold ways” (Engel, 2017: 45).

The recognition of the need for both sets of competencies does not mean that the distinction should be dispensed with altogether, or that it is appropriate to assume that the terms may be used interchangeably. By referring only to “data literacy”, one may implicitly centre the tools and perspectives of data science at the expense of the important principles of statistical reasoning that remain central in statistics education. Similarly, referring only to “statistical literacy” may privilege the analytical dimension of working with data at the expense of the critical and practical aspects which frequently define such work outside of the classroom. We do not aim (nor claim) to reconcile these definitional issues in this paper. For these reasons, we generally refer to “statistical and data literacies”, as this collapsing of the two effectively captures that *hybridity* is at the core of how quantitative education curricula can contribute to the unmet needs of the data revolution. Moreover, as our

proposed model is premised on the application of statistics in the workplace, we regard the need to have the full set of data and statistical literacies to provide an inclusive set of skills which will have most benefit in strengthening the skills pipeline nationally.

In the following section, we set out how an experiential learning model can help to capture the value of a hybrid approach by giving learners hands-on experience with real data in addition to classroom-based learning.

C2. Data Fellowships and the Value of Experiential Learning

The model we propose is designed to build on the success of an established model from the UK called Q-Step (see Carter et al, 2017). Q-Step was a national educational step-change programme developed as a strategic response to the shortage of quantitatively-skilled social science graduates in the United Kingdom (Nuffield, 2014). The programme funded 15 Q-Step Centres within higher education institutions. As an experimental programme, centres were encouraged to trial and evaluate different approaches to training undergraduates in statistics and data analysis. The University of Manchester Q-Step Centre, with many successful partnerships already established with industry, developed a paid internship programme which placed 250 undergraduates since 2013 in organisations including local and national government, charities, statistical institutes and research consultancies. Students spent two months working on real quantitative data-driven projects that could help them to practice and develop their professional and analytical skills. The data-driven research projects, which were co-designed by the host organisation, had to matter and have relevance to the host. Although not formally assessed as part of the student's undergraduate degree, the interns were required to develop a poster at the end of their placement. A half-day conference was then held to provide an opportunity to celebrate the placement learning, with prizes being awarded for talks and posters, the external judges panel comprising practitioners, academics and former interns. The intention was to share the value of the experiential learning opportunities, and to make visible the analytical and professional learning that took place. Many of these internships resulted in graduate job offers, and even more led to valuable connections between the partner organisations and the university.

The value of the Q-Step paid internship model is that it exposes students not only to classroom-based instruction in the basic statistical principles and analytical techniques which underpin statistical literacy, but also hands-on experience working with real data. Students participating in the programme first receive intensive training in the classroom to develop their statistical and data literacy skills, before moving to the workplace where they undertake real quantitative data-driven projects. In these research internships, students are able to practice and develop the skills learned in the classroom while building the accompanying professional and analytical skills that are necessary for working with and communicating about data. These placements provide a win-win scenario, where partner organisations receive an extra pair of hands to work on a short-term project of value to them, while the student gets to hone and develop their skills doing research that matters. Whilst we named these 'internships', they form the basis of the data fellows programme that we propose here.

The data fellowship model is an example of experiential learning, a pedagogical approach rooted in the theory of Kolb (1984), of which "a key part is personal hands-on practical experience in the learning process" (Purdam, 2016: 258). Experiential modes of learning such as internships, study away programmes and project-based learning are "proving to be critical and powerful in the overall quality and meaning of the undergraduate experience" (Bass, 2012: 24), with growing evidence that there are material benefits to both learners and educators (Roberts, 2018). This is particularly important for the teaching of quantitative skills to social science students, where there is a growing recognition that "the learning of social science research skills should be closer to the reality of social science research in practice" (Purdam, 2016: 266). This means getting students to practice and develop the hybrid set of data and statistical literacies required through hands-on experience with real data and real research problems.

This experiential learning can help to bridge the gap between the formulae learned and the conceptual understandings underpinning them by putting statistics into the context of front-line social research. A key challenge for statistics education is to ensure that students develop more than just the ability to recall and apply formulae and calculations from textbooks, but also a higher-level understanding of the principles of making sense of data. As Gail Burrill observes, "you could go to a class you have of undergraduates and ask them what the mean is, and you're probably going to get an answer like 'you add them up and divide'. So, there's really not a conceptual foundation. It's a recipe" (Burrill, 2020).
This is illustrated by Hobden (2014), who reports on a teacher re-skilling programme in South Africa which aimed to provide educators with the quantitative foundations to teach a course in Mathematical Literacy. When asked to provide an interpretive explanation of a sample sentence reporting the median survival times for men and women once infected with HIV, the trainees displayed generally low levels of conceptual understanding which led them to misunderstand the underlying health information being communicated in the sentence. Hobden notes that by focusing on a free-text, explanatory task rather than a mechanical application of a formula or calculation, the study draws "attention to the chasm that exists between the procedural computation of basic statistical measures and the conceptual understanding that enables and informs important opinions" (Hobden, 2014: 80). This chimes with Chris Wild's assessment that capacity building in statistical literacy ought to prioritise "conceptual understandings that enhance *human understanding* of data and what data can and cannot do, elements that are *empowering* and of *lasting value* to citizens" (Wild, 2017: 36, emphasis in original). Crucially, the hybrid skillset needed to close this gap between procedural knowledge and effective understanding for everyday use is "unlikely to develop without opportunity to learn through exposure to practice, and this seems to be the key" (Hobden, 2014: 81).

We propose that the Q-Step model developed at the University of Manchester can provide a valuable intervention in light of the challenges identified in Section B by helping to develop the unmet skills needs and thus contribute to the development of the next generation of quantitative social researchers. The approach achieves this in a number of ways. First, the paid internship model incentivises students to take an interest in statistical and data-driven approaches to solving social challenges, providing motivation for more students to embark on pathways into quantitative careers. Moreover, as the learners are substantively interested in complex social phenomena, as social science students, they are able to bring

their new knowledge and understanding back to their studies and use this to direct their interests and future career choices. Second, the two-way feedback between the partner organisations and the educators helps build understanding of which skills are most valued and expected by prospective employers working on data-oriented projects. Importantly, this includes those professional and communicative skills which we have argued are particularly needed for working on the SDGs. Third, by working with partners in relevant sectors, the model capitalises on the already existing skills and data available in-country. This not only reduces the dependency on external actors for capacity building support, but also ensures that learners will be honing their data skills on problems that are relevant to their society. With the potential value of this model in mind, we consider in the next section how to apply it practically in Colombia.

D. Practical Steps for Implementing Data Fellowships for the SDGs

D1. The Need for Country-Led Interventions

The Q-Step programme offers an encouraging example of how partnerships and workplace learning-based interventions can build a pipeline of statistical capacity for quantitative social research. In the remainder of this paper we wish to explore in more practical terms how this model can be developed in other countries to address the unmet needs identified in Section B. In doing so, we must first acknowledge that what works in the UK will not necessarily work elsewhere, and it is critically important to consider how the approach might serve the needs of a particular country. Both the UN (2015) and PARIS21 (2015) acknowledge that the particular social, economic and political environments in each country mean that progress towards the SDGs and efforts at statistical capacity development must be country-led processes. Language, culture, political will, economic factors, and the structure of the education system all contribute to determining what kind of implementation is practical and likely to have the desired impact in a given context.

We have argued the importance of partnerships in the process of building the in-country capacity pipeline, specifically aimed at supporting the SDG framework. However, partnerships is not an end in itself – the development sector is very familiar with tokenistic approaches to partnerships and participation potentially leading to the "the unjust and illegitimate exercise of power" (Cooke and Kothari, 2001: 1). Partnerships that lead to transformative outcomes in the capacity pipeline require in-country ownership and leadership, supplemented by international actors. The role of those trans-national actors should revolve around offering technical competence and service, whereas local actors take the lead thereby ensuring that the local context remains central to capacity development initiatives.

As such, our argument is not that the Q-Step model can be imported into other countries prefabricated. What we are arguing for is the value of countries building on the successes and strengths of this model to tailor an intervention to the particular unmet needs of that country. In practice, this means identifying where opportunities exist, such as determining where there is political will to address skills gaps in certain sectors, and working with

universities with expertise in statistical education to develop relevant learning materials. However, it also means acknowledging the challenges that may exist in a particular country, for example where access to higher education is particularly skewed, or where the baseline of statistical expertise, capacity and infrastructure required for such a programme to be impactful is not already in place.

D2. Future In-Country Research in Colombia

To this point we have argued that there are a number of unmet needs in the current statistical capacity development landscape, particularly in relation to the SDGs, and that a country-led, experiential learning-based data fellowship model can help to address these needs. In this section we wish to illustrate this argument in more practical terms through a discussion of the development of a data fellows programme in Colombia. We focus on Colombia as our research on this topic is the result of a partnership between the University of Manchester and Data-Pop Alliance, a research laboratory led by Harvard, MIT and the Overseas Development Institute that has worked extensively on data literacy in Latin America and maintains a strong network and knowledge base in the region, including in Colombia. This facilitates identifying the strengths and weaknesses of the Colombian statistical system, as well as the practical opportunities and challenges of implementing the kind of model described in Section C. Our analysis in this section is also informed and supplemented by pilot research we conducted in several Latin American countries, including an initial workshop in São Paulo in May 2019, followed by interviews with a number of key actors in the statistical community of three countries (Mexico, Colombia and Brazil), drawn from academia, civil society and the public sector (for more details see Higgins et al, 2019). This pilot research led to a partnership with colleagues at Universidad Del Rosario in Bogotá, who will lead the development of a data fellowships programme in Colombia. Here, we report some of the emerging challenges and opportunities that will shape this development. First, we focus on Colombia's statistical and data capacity for the SDGs, before turning to the Colombian pipeline for developing statistical and data literacies.

Statistical and Data Capacity for the SDGs

One of the strengths of the Colombian statistical system for developing a data fellowships programme is that Colombia is one of several countries in the region that already have well-established and resourced national statistical and academic systems, which allows more runway for getting a programme of the nature we describe up and running. The national statistical office is relatively well-resourced and developed by international standards. For example, Colombia scores well on the World Bank's Statistical Capacity Index, remaining well above the regional average as well as the average for middle-income countries over the last 15 years (World Bank, 2020). Moreover, with specific regard to the SDGs, interviewees described Colombia as an example in the region for taking the lead on the development of country indicators for measuring progress against Tier 2 and Tier 3 indicators (those where established methodologies and international standards either do not exist or are not widely used) (Higgins et al, 2019).

However, while statistical and data capacity is relatively high by international data monitoring standards, Colombia's capacity is more limited when considered from a more

human-centred perspective. Statistical and data literacy levels remain relatively low in Colombia. Colombia has been improving its national results in standardised tests such as PISA in recent years, as well as results in the national test *Pruebas Saber*. However, PISA test results are still lower than the average for OECD countries for mathematics (391 for Colombia vs 489 for OECD) and sciences (420 vs 488 for OECD). Interviewees described a need for more data literacy training across academia, particularly at the undergraduate level, as well as in the public sector and civil society. Previous research illustrates how this lack of statistical and data literacy training manifests in skills gaps in Colombia. In a survey led by the National Planning Department in 2017 concerning the use of data science in public entities, the results show low levels of exploitation of data for the generation of value in public administration. National entities state that lack of training in data analysis and applications is the main barrier (63%) to data exploitation (DNP, 2018a). This relatively low level of human capital in statistical and data skills is part of the wider issue of Colombia having an underdeveloped data culture. In many sectors, data skills are low and their value poorly understood (Higgins et al, 2019). As such, there is not yet a strong culture of using data to find innovative solutions to social problems, restricting the development of expertise in the field. In response to this shortage of qualified data professionals, the government has explicitly called for action to generate mechanisms to materialise the value of data in the public sector and to institutionalise the exploitation of data in public decision making (DNP, 2018a).

Indeed, understanding the political will and appetite for these skills and competencies in Colombia helps highlight the opportunities for Colombian higher education partners to develop a programme which is able to achieve the most impact by addressing the pressing sectoral needs and skills gaps. In this sense, the National Development Plan of the Colombian government presents a number of open doors. In particular, the Ministry of National Education, together with the National System of Science, Technology and Innovation, is committed to supporting the development of high-level human capital as a matter of objective priority (DNP, 2018b). There is strong will among policymakers to foster high-level professionals able to contribute in the 'knowledge society', and to support young people and professionals into education pathways that can contribute in this arena. As we have argued, the data fellowship model promotes a holistic skillset which helps develop well-rounded statisticians and data professionals, and there is clear demand for these skills from the Colombian government. Partnering with key organisations and public institutions in which shortages of statistical and data literacy have been identified helps to ensure that the programme builds talent pipelines directly into these gaps.

Statistical and Data Capacity Development Pipeline

A key strength of the capacity development pipeline in Colombia is that the country has a relatively strong academic sector from which expertise in teaching can be leveraged. However, there is a somewhat limited offer of courses in relevant subjects: in 2017 there were only 75 formal education programs related to data science and analytics in the country, with only five of these at postgraduate level and just one PhD programme, while only 2.5% of the scholarships offered by the Colombian government were in data skills-related subjects (Higgins et al, 2019: 19). In particular, teaching expertise in statistical and data skills tends to fall more under the umbrella of mathematical education subjects which

do not always develop the hybrid set of social research skills which the SDGs demand; social science subjects, on the other hand, tend to be more oriented towards theory than data analysis (Higgins et al, 2019). Interviewees from the statistical community thus emphasised "the need to build data literacy capacities towards fostering 'hybrid' professionals that can understand, use and analyze data for 'social sciences' purposes, i.e. evidence-informed policy, journalism, activism etc." (Higgins et al, 2019: 32). Through its focus on developing hybrid skills for working with data on complex social problems, the proposed model can help fill a clear skills gap in Colombia with regard to the supply of professionals capable of using data to find innovative solutions to social problems.

The government's SDG strategy recognises that academia, through both research and teaching, has a significant role to play in the field of innovation and training for the sustainability of future generations (DNP, 2018c). This presents an opportunity for educators to highlight how the data fellowship model fits within the current national development priorities and can contribute in the space of capacity development for the SDGs. The national SDG strategy also emphasises the need to establish mechanisms for dialogue with non-governmental actors, and calls for partnerships for knowledge, financing and mobilisation towards the SDGs. Moreover, the government has highlighted stimulus for the interaction between university and business as a strategic objective in its Pact for Sciences, Innovation and Technology. The data fellowship model can thus provide a platform for strengthening university-industry partnerships and bolstering the country's human capital in data-oriented fields and public innovation, both of which are development priorities in Colombia. These partnerships are at the heart of how data fellowships can achieve impact, as they involve a range of stakeholders who understand the particular needs and realities that can vary country to country. This places Colombian universities in a strong position to play a key role in supplying statistically and data literate talent for the delivery of the SDGs.

However, one of the key challenges for Colombia is that there are pre-existing inequalities in education access which structure opportunities to develop statistical and data literacies and skills. There are generally high levels of dropout and low graduation rates. In 2016, the cohort dropout indicator stood at 45.1% for university students and 53.2% for technicians and technologists, while the graduation rate was 37.4% and 27.1%, respectively (DNP, 2018b). In 2015, only 62% of young adults had finished high school, and only 38% of high school graduates progressed into higher education (ibid.). This gap in access to education is driven by structural inequalities. In Colombia, 27% of the population is low-income and 7.8% is in extreme poverty (DANE, 2018). By 2017, young people in the top quintile of the income distribution had almost four times greater access to higher education than young people from the bottom quintile. This is not only a question of high- and low-income, but also urban and rural. The population in rural areas of Colombia accumulate an average of 6 years of education, whereas in urban areas this average goes up to 9.7 years (DNP, 2018b). The gap in access to primary education is around 10% and increases to 16% for tertiary education. These gaps mean that developing the data fellowship model through partnerships between higher education institutions (particularly private universities in urban areas) and industry, civil society or government are at risk of "empowering the empowered" (Gurstein, 2011) and leaving behind the rural population when it comes to statistical and data literacy. To combat this, there must be active efforts targeting the involvement and

recruitment of marginalised populations into a data fellowship programme. On a practical level, this could involve partnerships between the large private universities and regional universities with a student pool drawn from less privileged backgrounds. Another strategy is to aim to partner with organisations that work in rural areas, in order to better target the needs of those communities when designing projects for data fellows. The use of digital course delivery methods such as MOOCs could also help to decentralise the educational components of the programme and improve rural access. Ultimately, evaluation mechanisms must be built into the intervention to ensure that partners are able to identify what works and what needs improvement to make the data fellows model work in Colombia.

E. Conclusions

The ever-expanding reach of statistics and data in the digital age means that there is a global need for capacity in the critical quantitative skills required for understanding and working with data. The Sustainable Development Goals, the most ambitious exercise in the quantification of social progress we have seen, are a clear example of where demand for such skills is outstripping supply. This demand can be a galvanising force for the development of statistics education programmes aimed at achieving impact where these skills are most needed. However, it is important that progress in this arena is not contained to the high-income countries of the Global North. Partnerships can be a key tool in building bridges between sectors and countries, to ensure that the data and expertise that already exist can be utilised in developing and motivating the next generation of quantitative talent. In this paper we have argued for the value of thinking of statistical capacity as a pipeline which begins in the education system, and illustrated why investment in this end of the pipeline can be a more sustainable and long-term pathway to building up the holistic skill-base needed to tackle the challenges of the SDGs. Through our review of the practical conditions underlying the development of such a programme in Colombia, we have illustrated that, although there are challenges in ensuring that educational access is equal, there are significant opportunities and open doors for addressing local skills gaps through data fellowships. The next step of this research is thus to begin placing students within partner organisations in Colombia and learning what works well, what needs to be refined, and what may be scalable to other countries beyond Colombia.

References

Akinsola, E. F. and Sharma, D. (2018). "Capacity Building of Developmental Scientists for Realization of the Sustainable Development Goals". In *Developmental Science and Sustainable Development Goals for Children and Youth* (eds. S. Verma, A. C. Petersen). Springer.

Attaran, A. (2005). "An Immeasurable Crisis? A Criticism of the Millennium Development Goals and Why They Cannot Be Measured". *PLoS Medicine* 2 (10), DOI: 10.1371/journal.pmed.0020318 .

Barbier, E. B. and Burgess, J. C. (2019). "Sustainable development goal indicators: Analyzing trade-offs and complementarities". *World Development* 122, 295-305.

Bass, R. (2012). "Disrupting Ourselves: The Problem of Learning in Higher Education". *EDUCAUSE Review* 47 (2), 23-33.

Burrill, G. (Interviewee). (2020, March 19). *Stats + Stories, Episode 131* [Audio podcast]. [Online: https://statsandstories.net/education1/teaching-statistics].

Carter, J., Brown, M., and Simpson, K. (2017). "From the Classroom to the Workplace: How Social Science Students Are Learning to Do Data Analysis for Real". *Statistics Education Research Journal* 16 (1), 80-101.

Cooke, B., & Kothari, U. (Eds.). (2001). *Participation: The new tyranny?*. Zed books.

DANE (2018). Departamento Administrativo Nacional de Estadística, Government of Colombia: "Boletín Técnico, Pobreza monetaria en Colombia". Retrieved from https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2018/bt_pobreza_monetaria_18.pdf (accessed 19/1/2021).

Dang, H. H. and Serajuddin, U. (2020) *World Development* 127, DOI: 10.1016/j.worlddev.2019.05.024.

DNP (2018a). Departamento Nacional de Planeación, Government of Colombia: "Documento CONPES 3920: Política Nacional de Explotación de Datos (Big Data)". Retrieved from https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%c3%b3micos/3920.pdf (accessed 19/1/2021).

DNP (2018b). Departamento Nacional de Planeación, Government of Colombia: "Plan Nacional de Desarrollo 2018 – 2022, Pacto por Colombia, pacto por la equidad". Retrieved from https://colaboracion.dnp.gov.co/CDT/Prensa/BasesPND2018-2022n.pdf (accessed 19/1/2021).

DNP (2018c). Departamento Nacional de Planeación, Government of Colombia: "Documento CONPES 3918: Estrategia para la Implementación de los Objetivos de Desarrollo Sostenible (ODS) en Colombia". Retrieved from https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3918.pdf (accessed 19/1/2021).

Engel, J. (2017). "Statistical Literacy for Active Citizenship: A Call for Data Science Education". *Statistics Education Research Journal* 16 (1), 44-49.

Fisher, A. and Fukuda-Parr, S. (2019). "Introduction – Data, Knowledge, Politics and Localizing the SDGs". *Journal of Human Development and Capabilities* 20 (4), 375-385.

Fukuda-Parr, S. and McNeill, D. (2019). "Knowledge and Politics in Setting and Measuring the SDGs: Introduction to Special Issue". *Global Policy* 10 (Supplement 1), 5-15.

Gould, R. (2017). "Data Literacy Is Statistical Literacy". *Statistics Education Research Journal* 16 (1), 22-25.

Gurstein, M. B. (2011). "Open data: Empowering the empowered or effective data use for everyone?". *First Monday* 16 (2), DOI: 10.5210/fm.v16i2.3316.

Heeks, R., and Renken, J. (2018). "Data justice for development: What would it mean?". *Information Development* 34 (1), 90-102.

Higgins, V., Casasbuenas, V., Ricard, J. and Carter, J. (2019). "EmpoderaData: Data Literacy Assessment and Sustainable Development Goals Data Gaps". Report – The University of Manchester and Data-Pop Alliance. Retrieved from https://datapopalliance.org/empoderadata-project/ (accessed 17/12/2020).

Hobden, S. (2014). "When Statistical Literacy Really Matters: Understanding published information about the HIV/AIDS epidemic in South Africa". *Statistics Education Research Journal* 13 (2), 72-82.

Jütting, J. (2016). "Capacity Building, Yes – But How to Do It?" United Nations World Data Forum. Retrieved from https://unstats.un.org/unsd/undataforum/capacity-building-yes-but-how-to-do-it/index.html (accessed 9/6/2020).

Keijzer, N. and Klingebiel, S. (2017). "Realising the Data Revolution for Sustainable Development: Towards Capacity Development 4.0." PARIS21 Discussion Paper No. 9. Retrieved from https://ssrn.com/abstract=2943055 (accessed 9/6/2020).

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. London: Prentice Hall.

Lokanathan, S., Perera-Gomez, T. and Zuhyle, S. (2017). "Mapping Big Data Solutions for the Sustinable Development Goals (draft)". LIRNEasia. Retrieved from http://hdl.handle.net/10625/56906 (accessed 17/12/2020).

MacFeely, S. (2019). "The Big (data) Bang: Opportunities and Challenges for Compiling the SDG Indicators". *Global Policy* 10 (Supplement 1), 121-133.

MacFeely, S. and Barnat, N. (2017). "Statistical capacity building for sustainable development: Developing the fundamental pillars necessary for modern national statistical systems". *Statistical Journal of the IAOS* 33, 895-909.

Merry, S. E. (2019). "The Sustainable Development Goals Confront the Infrastructure of Measurement". *Global Policy* 10 (Supplement 1), 146-148.

North, D., Gal, I. and Zewotir, T. (2014). "Building Capacity for Developing Statistical Literacy in a Developing Country: Lessons Learned from an Intervention". *Statistics Education Research Journal* 13 (2), 15-27.

Nuffield (2014). Aims and activities of the Q-Step centres. [Online: https://www.nuffieldfoundation.org/sites/default/files/files/Aims%20and%20Activities%20of%20the%20Q-Step%20Centres(1).pdf]

Oburu, P. O. and Yoshikawa, H. (2018). "Roles of Multiple Stakeholder Partnerships in Addressing Developmental and Implementation Challenges of Sustainable Development Goals". In *Developmental Science and Sustainable Development Goals for Children and Youth* (eds. S. Verma, A. C. Petersen). Springer.

OECD (2015). "Strengthening National Statistical Systems to Monitor Global Goals". The OECD Post-2015 Reflection series Element 5, Paper 1.

PARIS21 (2015). A Road Map for a Country-Led Data Revolution. [Online: https://datarevolution.paris21.org/sites/default/files/Road_map_for_a_Country_led_Data_Revolution_web.pdf]

PARIS21 (2019). Statistical Capacity Development – Outlook 2019.

PARIS21 (n.d.). Capacity Development 4.0. Webpage. [Online: https://paris21.org/capacity-development-40].

Prodromou, T., and Dunne, T. (2017). "Statistical Literacy in Data Revolution Era: Building Blocks and Instructional Dilemmas". *Statistics Education Research Journal* 16 (1), 38-43.

Purdam, K. (2016). "Task-based learning approaches for supporting the development of social science researchers' critical data skills". *International Journal of Social Research Methodology* 19 (2), 257-267.

Ridgway, J., Nicholson, J., and McCusker, S. (2013). "'Open Data' and the Semantic Web Require a Rethink on Statistics Teaching". *Technology Innovations in Statistics Education* 7 (2).

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., and Wuetherick, B. (2015). "Strategies and Best Practices for Data Literacy Education (Knowledge Synthesis Report". DOI: 10.13140/RG.2.1.1922.5044.

Roberts, J. (2018). "From the Editor: The Possibilities and Limitations of Experiential Learning Research in Higher Education". *Journal of Experiential Education* 41 (1), 3-7.

Sachs, J. D. (2012). "From Millennium Development Goals to Sustainable Development Goals". *Lancet* 379, 2206-2211.

Schield, M. (2004). "Information Literacy, Statistical Literacy and Data Literacy". *IASSIST Quarterly, Summer/Fall 2004*, 6-11.

Schield, M. (2010). "Assessing Statistical Literacy: Take CARE". In *Assessment Methods in Statistical Education: An International Perspective* (eds. P. Bidgood, N. Hunt and F. Jollife). Wiley and Sons.

Stafford-Smith, M., Griggs, D., Gaffney, O., Ullah, F., Reyers, B., Kanie, N., Stigson, B., Shrivastava, P., Leach M., and O'Connell, D. (2017). "Integration: the key to implementing the Sustainable Development Goals". *Sustainability Science* 12, 911-919.

Stuart, E., Samman, E., Avis, W. and Berliner, T. (2015). "The Data Revolution: finding the missing millions". ODI Research Report 03. Retrieved from https://www.odi.org/publications/9476-data-revolution-finding-missing-millions (accessed 17/12/2020).

UN (2015). "Transforming Our World: The 2030 Agenda for Sustainable Development". Retrieved from https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf (accessed 06/08/2020).

UN (2020). "Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development". [Online: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202020%20review_Eng.pdf]

UN IEAG (2014). "A World That Counts: Mobilising the data revolution for sustainable development". United Nations Independent Expert Advisory Group on a Data Revolution for Sustainable Development. Retrieved from https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf (accessed 06/08/2020).

Wild, C. J. (2017). "Statistical Literacy as the Earth Moves". *Statistics Education Research Journal* 16 (1), 31-37.

World Bank (n.d.). Data on Statistical Capacity. Webpage. [Online: http://datatopics.worldbank.org/statisticalcapacity/]

World Bank (2020). *Statistical Capacity Dashboard* [online resource available at https://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx]. Accessed 17/12/2020.

Yayboke, E., Nealer, E. and Rice, C. (2017). "Harnessing the Data Revolution to Achieve the Sustainable Development Goals: Enabling Frogs to Leap". A Report of the CSIS Project on Prosperity and Development with the JICA Research Institute.