2023

Scaling Strategy for Innovations Aimed at Data-driven School Improvement

LEARNINGS FROM THE KIX REGIONAL GRANT PROJECT: DATA-DRIVEN SCHOOL IMPROVEMENT (DSI) – CHALLENGES, OPPORTUNITIES AND SCALABLE SOLUTIONS (2021-2023)

SOCIETY FOR THE ADVANCEMENT OF EDUCATION (SAHE)

Primary authors: *Irfan Muzaffar and Amna Ansari*

Team contributions (in alphabetical order): *Sanaa Ahmed, Mashallah Ali, Sofia Uzma Alvi, Imran Haider, Salma Jafar, Sambedan Koirala, Marriam Musa, Abbas Rashid, Shankar Bahadur Thapa, Deependra Bickram Thapa, Sahar Shah, Naveed Saleh Siddique, Imran Uddin*

Disclaimer: *This work was supported by the Global Partnership for Education Knowledge and Innovation Exchange, a joint endeavour with the International Development Research Centre, Canada.*

The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

Contents

Executive Summary

This document outlines a strategy for scaling innovations in data-driven school improvement based on the findings of a KIX[1]-supported research for development (R4D) project implemented in Pakistan (Punjab, Khyber Pakhtunkhwa and Sindh provinces) and Nepal. The project titled *Data-driven School Improvement (DSI) – Challenges, Opportunities and Scalable Solutions* is an application of scaling science[2] to the shared policy challenge of sub-optimal use of education data and addresses how countries can adapt and scale innovations in data-driven school improvement to optimize the use of education data at all levels (school to central/federal).

Often, scaling is considered synonymous with expanding coverage of a program or service. However, such a description of scaling dwells more on operational changes in provision rather than the social benefit or impact of the act of scaling. Scaling *impact* or the positive good an innovation or intervention creates for people and the environment is more meaningful than simply increasing coverage or expansion which may or may not correlate with good change or positive impact. Thus, when scaling any innovation, a key concern is whether and to what extent can its *impact* be effectively and optimally scaled to address the challenges faced. Responding to this question, the DSI research demonstrates how the *scaling impact* approach[3] comprising the four guiding principles of Justification, Optimal Scaling, Coordination and Dynamic Evaluation offers unique value for scaling innovations in data-driven school improvement.

The conceptual underpinnings of DSI are provided by these guiding principles to achieve *a scale of impact important to people and the environment and contribute to a broader system of development change*[4]. The project also recognizes that successful scaling of innovations that aid data-driven school improvement within education systems involves navigating complex interactions among various elements situated in different ecological layers within these systems. These ecological layers include the microsystem, including the school and classroom culture; the mesosystem, including and encompassing departmental cultures, education managers, district-level practices, and reporting structures; the exo-system, involving a wide range of research partners, donors, technical assistance teams, and other external stakeholders; and the macro-system, including and encompassing national trends and policies, federal and/or provincial level dynamics, and policymakers. To effectively scale educational innovations, it is crucial to explore the extent to which the scaling principles (of Justification, Optimal Scale, Coordination and Dynamic Evaluation) are applied across these layers, and their interrelatedness.

Incorporating these elements in the study's conceptual framework, the DSI project adopted a qualitative methodology, drawing data from in-depth interviews, focus group discussions and observational notes, and combining them with secondary/documentary analysis. The data

[1] KIX is a joint endeavor of the Global Partnership for Education (GPE) and International Development Research Centre (IDRC) which aims to strengthen national education systems and accelerate educational progress in the Global South by filling knowledge gaps, increasing access to evidence, and strengthening systems to support the generation and uptake of evidence and innovations in low- and middle- income countries.

[2] Price-Kelly, van Haeren & McLean. 2020. *The Scaling Playbook*. International Development Research Centre, Ottawa, Canada.

[3] McLean, R., & Gargani, J. 2019. *Scaling Impact: Innovation for The Public Good*. Routledge, p. 189.

[4] For a fuller account of these principles, please see McLean & Gargani (2019) and Price-Kelly & McLean (2020).

collection instruments, comprising interview items and focus group discussion protocols, were developed in close alignment with the four guiding principles. In Punjab and KP, where a particular innovation in data-driven school improvement – namely, the School Improvement Framework (SIF) – has already scaled, data collection goals revolved around the need to generate information on the challenges, opportunities, and potential solutions in the scaled implementation of the SIF. In Sindh and Nepal, where the project adopted the course of feasibility research for potential scaling of a similar innovation, the guiding principles for data collection and subsequent instrument development were tailored to focus on the potential scaling of a proposed innovation in data-driven school improvement. Data was collected from approximately over 350 individuals (including pre-pilot and pilot activities) across all ecological layers of stakeholders in the education systems, and all four research sites, and coded and analysed using the Dedoose online platform for qualitative data analysis.

The learnings from all four contexts demonstrate that it is possible to maintain the core elements or fundamental characteristics of data-driven school improvement while adapting innovations for the same to local contexts. In contexts like Nepal, where utilisation of data is reported to be sub-optimal, the clear need to introduce an innovation in data-driven school improvement also provides a strong justification for scaling. However, strategically, scaling must become a more shared choice. Such ***moral justification*** necessitates tapping the unique constellation of stakeholders likely to be impacted by the innovation in every context, assessing impact risks and establishing endorsement accordingly. For instance, when excavating moral justification for a potential DSI innovation in Sindh, some officials in Sindh alluded that the generation and use of Gender, Equity and Social Inclusion (GESI)-sensitive data such as disability or minority/ethnicity could potentially lead to discrimination against these groups. This contrasted with the findings in Punjab and KP, partly indicating differences in the everyday lived experiences of individuals across the provinces. Nevertheless, this perspective suggests a critical need for awareness-building and policy discussion regarding the role of data in supporting inclusion, addressing the dual challenge of ensuring that data collection supports inclusion, establishing safeguards to prevent misuse of sensitive information, as well as seeking endorsement from a diverse set of stakeholders, who are directly or indirectly likely to be impacted by the innovation.

Research in all four contexts also indicates that, in addition to technical and moral considerations, existing or potential *windows of opportunity* can amplify justification. For instance, in both Punjab and KP, the existence of Large Scale Monitoring Systems (LSMSs) and the increased availability of regular data had created a 'window of opportunity' for the implementation of the SIF. This window of opportunity was reinforced by a confluence of circumstances including renewed political emphasis on education reforms, the need for data, and the time-bound nature of support from international development partners. However, since such windows are transient, sustained, inclusive efforts to strengthen justification through moral and technical lenses are requisite.

Data-driven school improvement, by way of its fundamental characteristic(s), must be based in data – hence data collection from schools is a crucial first step. However, the DSI research revealed that, where the collection of data is being compromised due to its dependence on a large number of external monitors – such as in Punjab with its fleet of external monitors – a hybrid data

collection model involving both external data collection as well as self-reporting by schools – offers a practical strategy towards **scaling optimally**. Delivering on the potential for cross-country learning by undertaking a comparative analysis of data collection and utilization models in Punjab and Nepal, the research further shows that there not need always be a cost-quality trade off in data collection. For instance, the data collection model in Nepal, alongside being more cost-effective, involving checks and balances for quality and accuracy of data including (but not limited to): (i) in-built, software checks (ii) restricting permission for Head Teachers to make changes in reported data beyond a limited period of time (iii) a comprehensive system for reporting and correcting errors in data at the level of the local government, and (iv) an expanding set of fields against which Head Teachers need to enter data, making it increasingly difficult to fabricate or report fake data.

More broadly, the principle of optimal scaling acknowledges that scaling creates a collection of anticipated and unanticipated and desirable and undesirable impacts, bringing with it several trade-offs. Given this realization, it is important to name and consider the trade-offs in all the different dimensions of optimal scale, including the magnitude, variety, equity and sustainability of impacts[5]. Optimality then, is about reaching a level where the magnitude, variety, sustainability, and equity of impacts are balanced in a way that is widely endorsed by stakeholders. The DSI research revealed several other such tradeoffs in the scaling of innovations in data-driven school improvement including the need to: (i) ensure equitable impacts, for instance, by making concerted efforts to address gender-based impediments in processes of school improvement, (ii) leverage technological innovation and adaptation such that systems remain robust and efficient for greater magnitude and sustainability of impact, and (ii) buffer against policy changes by creating mechanisms that allow for quicker adaptation to policy changes and involve a wider constituency of stakeholders in planning processes to assess impacts (and risks) better.

Indeed, the plurality of people, places and things that affect and/or are affected by the scaling of an innovation come to comprise the scaling environment or 'scaling system' in which any innovation scales. It is therefore important to **coordinate** the scaling environment such that its evolving set of actors are aligned around creating impact at optimal scale. This includes creating champions, building effective partnerships, aligning incentives, and so on. The DSI project team's experience and collaboration with provincial and federal counterparts in Pakistan like the Pakistan Institute of Education (PIE) outlined in this report lend weight to this claim. For instance, the series of policy dialogues conducted by the DSI research team at the provincial and federal levels, including with PIE, helped frame the policy debate around scaling a contextualized innovation in data-driven school improvement in Sindh. Allowing cross-fertilization of ideas across provinces, early and ongoing coordination among relevant education stakeholders in Sindh in such a manner has helped create a more enabling environment for a DSI innovation to be launched and scaled. Further, federal level institutions like PIE are also able to experiment with the innovations in data-driven school improvement within a much smaller jurisdiction to further explore the sustainability of such innovations at scale. Broadly speaking, then, the sustainability of the optimal scaling of innovations in data-driven school improvement requires involvement of and championing at all

[5] These dimensions are discussed in greater detail under Chapter 4.

levels, including the central/federal level institutions, even in countries like Pakistan where education is decentralized, in addition to the existing engagement with provinces.

In addition to coordination across actors, coordination across multiple innovations can also be helpful in avoiding duplication of efforts and achieving synergies, thus potentially enhancing overall impact. The DSI R4D establishes positive reception of the SIF in Punjab and KP. However, new programming by development partners such as under the FCDO Data and Research in Education (DARE) program promotes a renewed emphasis on data standardization and improved reporting on the SDGs, promising enhanced quality in education statistics from EMISs[6]. This focus on better reporting has accompanied diminished attention to interventions like SIF that fostered data utilization. In such situations, in addition to coordinating across actors, it is useful to coordinate across innovations or multiple/ solutions offered by a particular innovation using a *portfolio approach*[7]. This strategy can be applied to broaden the objectives of SIF data collection. For instance, the framework can be repurposed to better align its constituent data collection, utilisation and action management processes with reporting on and meeting the Sustainable Development Goals for education so that, instead of curtailing or replacing the framework, provincial governments and development partners are able to sustain its impact(s), albeit with additional benefits or purposes. For example, certain SIF indicators like student attendance and facilities conditions overlap with SDG 4 reporting needs. Rather than separate data collection processes, the existing SIF methodology could be leveraged and repurposed to serve the additional goal of SDG monitoring. Countries undertaking scaling of innovations in data-driven school improvement for the first time can similarly ensure flexibility in the adopted/adapted design so that coordinating across an evolving set of actors and potentially competing innovations can become easier.

It is also important to continuously and **dynamically evaluate** scaling and adapt strategies to mitigate unanticipated and undesirable effects of scaling, such as the local social and cultural gender barriers in processes of school improvement in Punjab and KP mentioned in this report. Also, based on the existing learnings from DSI, determining and meeting needs for future R4D, for instance, around specific measures that can align responsibility with authority at various levels of the education systems in Punjab and KP as well as conducting impact evaluations that assess the extent to which the observed improvements in school outcomes can be attributed to the SIF are imperative. For scaling in new contexts, the demonstrated effectiveness of SIF and similar innovations in this way can help justify the decision to scale, as well as anticipate factors for success alongside assessing its unintended and/or undesirable effects, and strategies for mitigation. This is the ultimate goal of this strategy document – that is, in addition to promoting the long-term efficacy and sustainability of innovations in data-driven school improvement and accelerate progress to this end in Pakistan and Nepal, it aims to offer cross-learnings in the form of a roadmap for effectively scaling innovations in data-driven school improvement for GPE member and other countries at large.

[6] DARE is an approximately GBP 22.9 million initiative that seeks, among other things, strengthened data systems in Pakistan. Source: https://devtracker.fcdo.gov.uk/projects/GB-GOV-1-300575/summary

[7] McLean, R., & Gargani, J. 2019. *Scaling Impact: Innovation for The Public Good*. Routledge, p. 78-80.

Acknowledgements

The Society for the Advancement of Education (SAHE) acknowledges with gratitude numerous stakeholders in Pakistan and Nepal whose contributions were instrumental in the successful execution of the Data-driven School Improvement (DSI) project.

We express our sincere appreciation to the federal and provincial governments of Pakistan and Nepal for their unwavering support. This includes the Ministry of Federal Education and Professional Training (MoFEPT), Pakistan, the Ministry of Education, Science & Technology (MoEST), Nepal, the Elementary & Secondary Education Department (ESED), and Khyber Pakhtunkhwa Education Monitoring Authority (KPEMA), Khyber Pakhtunkhwa, the School Education Department (SED), and the Programme Monitoring and Implementation Unit (PMIU), Punjab, and the School Education & Literacy Department (SELD), Monitoring & Evaluation (SELD), and the Directorate of Curriculum, Assessment and Research (DCAR), Sindh.

We are indebted to the countless officers at the district and sub-district levels in both Nepal and Pakistan, including teachers and head teachers from all the research sites. Their diligence and cooperation significantly enriched the project, even though their names cannot all be individually mentioned due to space constraints.

Specifically, we want to acknowledge the invaluable insights and review contributions from Dr. Margarita Lopez, Program Officer, GPE-KIX, who provided continuous technical support throughout the project. Her assistance was particularly valuable during multiple challenges we faced in the course of this research.

We are especially thankful to Mr. Deependra Bickram Thapa, Mr. Sambedan Koirala, and Mr. Shankar Bahadur Thapa for their tireless efforts to solidify the research partnership between Nepal and Pakistan. Their roles were instrumental in fostering a collaborative research environment that significantly benefited our project.

We also wish to highlight the special efforts of Mr. Sohail Raza, a data professional who offered invaluable expertise, Mr. Aftab Alam, Director EMIS Sindh, for his insightful guidance on data-driven school improvement in Sindh, and Mr. Shahzad Arif, Team Leader of the KP Education Sector Programme, for his insights about data-driven school improvement in KP.

We also extend our thanks to the Brookings Institution for their collaboration under the Research on Scaling the Impact of Innovations in Education (ROSIE) initiative. Their input was crucial in helping us understand the nuances and complexities of scaling processes.

We appreciate the numerous partners in knowledge sharing and learning exchange, as well as the DSI project team who exhibited resilience in delivering the project despite the challenges presented by the COVID-19 pandemic and the 2022 floods in Pakistan.

Our heartfelt gratitude goes to all those involved in this project for their vital contribution to the pursuit of education reform in Pakistan, Nepal, and beyond.

List of Abbreviations

AEO	Assistant Education Officer
ASC	Annual School Census
ASDEO	Assistant Sub-Divisional Education Officer
CEHRD	Center for Education and Human Resource Development
CEO	Chief Executive Officer
CMO	Chief Monitoring Officer
COT	Classroom Observation Tool
COVID-19	Coronavirus Disease of 2019
CPD	Continuous Professional Development
DCMA	Data Collection and Monitoring Assistant
DEO	District Education Officer
DPS	District Performance Scorecard
DSI	Data-driven School Improvement
EAP	Europe, Asia and Pacific
ECED	Early Childhood Education and Development
EFA	Education for All
EMIS	Education Management Information System
ESED	Elementary and Secondary Education Department
FCDO	Foreign, Commonwealth and Development Office
GESI	Gender, Equity, and Social Inclusion
GPE	Global Partnership for Education
HCI	Human Capital Index
HT	Head Teacher
ICT	Islamabad Capital Territory
IDRC	International Development Research Centre
IEMIS	Integrated Education Management Information System

KIX	Knowledge and Innovation Exchange
KP	Khyber Pakhtunkhwa
KPEMA	Khyber Pakhtunkhwa Education Monitoring Authority
LG	Local Government
LSMS	Large Scale Monitoring System
LSU	Local Support Unit
MA	Monitoring Assistant
M&E	Monitoring & Evaluation
MEA	Monitoring & Evaluation Assistant
MICS	Multiple Indicator Cluster Survey
MoEST	Ministry of Education, Science & Technology
MoFAGA	Ministry of Federal Affairs and General Administration
MoFEPT	Ministry of Federal Education & Professional Training
PBE	Pre-lower Basic Education
PIE	Pakistan Institute of Education
PITB	Punjab Information Technology Board
PMIU	Programme Monitoring & Implementation Unit
R4D	Research for Development
RCT	Randomized Control Trial
RSU	Rural Support Unit
SAHE	Society for the Advancement of Education
SC	School Council
SED	School Education Department
SELD	School Education & Literacy Department
SEMIS	Sindh Education Management Information System
SIF	School Improvement Framework
SIS	School Information System
SLO	Student Learning Outcome
SSI	School Status Index

SSMS	Sindh School Management System
TA	Technical Assistance
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children’s Emergency Fund

1 Introduction and Purpose

This document outlines a strategy for scaling innovations in data-driven school improvement based on the findings of a KIX[8]-supported research for development (R4D) project implemented in Pakistan and Nepal. The project titled *Data-driven School Improvement (DSI) – Challenges, Opportunities and Scalable Solutions* is an application of scaling science[9] to the shared policy challenge of sub-optimal use of education data. In other words, the project employs a 'systematic, principle-based science of scaling that can increase the likelihood of such innovations in data-driven school improvement benefitting society'[10]. While the strategy borrows from and contributes to learnings on scalability challenges and responses in the specific project countries and contexts under DSI, its usability extends to the wider set of GPE member (and other) countries wishing to scale innovations in data-driven school improvement.

1.1 A Brief Introduction to the DSI Project

Awarded in May 2021, DSI was a two-year R4D project designed to respond to a specific priority for countries in the KIX EAP region, namely optimizing the use of Education Management Information Systems (EMIS), implemented in Pakistan and Nepal. Almost all countries have established EMISs but their use is often more directed towards project reporting than planning at the school, provincial and central or federal levels. And while existing data collection, processing, and reporting practices help provide snapshots of education systems, they do not help policy- and decision- makers see and address improvement needs of individual schools. This can be achieved by scaling innovations that improve the organization, interpretation and use of education data for school improvement. The DSI project aimed to address how countries can adapt and scale such innovations to optimize the use of education data at all levels (school to central/federal).

The general objective of DSI has been to generate knowledge based on the scaling of a promising innovation in data-driven school improvement, namely the School Improvement Framework (SIF)[11] being implemented in Pakistan (and its adaptations), that optimizes the use of data produced by schools to improve school management and results as well as enhance the support schools receive from other levels/tiers of government. In line with this broader objective, the specific objectives of the project include:

[8] A joint endeavor of the Global Partnership for Education (GPE) and International Development Research Centre (IDRC), KIX aims to strengthen national education systems and accelerate educational progress in the Global South by filling knowledge gaps, increasing access to evidence, and strengthening systems to support the generation and uptake of evidence and innovations in low- and middle- income countries.

[9] Price-Kelly, van Haeren & McLean. 2020. *The Scaling Playbook*. International Development Research Centre, Ottawa, Canada.

[10] Price-Kelly, van Haeren & McLean. 2020. *The Scaling Playbook*. International Development Research Centre, Ottawa, Canada, p.2.

[11] The SIF is a conceptual and methodological tool developed to use EMIS data to identify and address schools' needs by actors within the education system. For a detailed note on the SIF, please see Annex A.

i. identifying the nature, scope, and extent of challenges to the scalability of the SIF, including specific challenges relating to gender equality by way of participation and leadership in school improvement processes.
ii. identifying conditions for success and effective scaling of the SIF and other EMIS-led innovations to improve school-level outcomes for all children within and across the three countries.
iii. mobilizing knowledge on optimal use of school-based data for education decision-making, policy and management amongst education stakeholders, and
iv. strengthening capacities of education managers at all levels to implement SIF and be able to identify and address bottlenecks in the delivery of equitable and quality education at the school level.

The project has been implemented in Pakistan (Punjab, Khyber Pakhtunkhwa and Sindh provinces) and Nepal. In the provinces of Punjab and KP, where the SIF has already been implemented and scaled, research under the project explores the opportunities, challenges and solutions in effective scaling of the SIF and similar innovations in data-driven school improvement. In Sindh and Nepal on the other hand, where no such innovation exists, the project examines the possibility or feasibility of scaling a contextually relevant innovation in data-driven school improvement. By doing so, the project has endeavored to achieve (i) an enhanced understanding of the frameworks for scaling innovations aimed at data-driven school improvement for actors across the wider education system, and (ii) communities of practice within and across the GPE member countries, which are more knowledgeable of data-driven school improvement. It is pertinent to mention that the project adopts a unique approach to scaling on which this strategy is based, and the defining tenets of this scaling science are explicated below.

1.2 What We Mean by Scaling

Often, scaling is considered synonymous with expanding coverage of a program or service. However, such a description of scaling dwells more on operational changes in provision rather than the social benefit or impact of the act of scaling. Scaling *impact* or the positive good an innovation or intervention creates for people and the environment is more meaningful than simply increasing coverage or expansion, for instance, which may or may not correlate with good change or positive impact. This concept lies at the heart of scaling science – Robert McLean and John Gargani argue that when thinking about scaling, one must think about scaling impact and seeking an optimal result where the impacts that are proven effective and desirable to stakeholders are cultivated and encouraged, and those that may cause harm, lead to waste, or are not desired by the impacted community are inhibited[12]. They outline a principled approach to scaling comprising the following guiding principles which, as detailed in subsequent sections, also inspire the DSI research[13]:

[12] McLean, R., & Gargani, J. 2019. *Scaling Impact: Innovation for The Public Good*. Routledge, p. 189.

[13] For a fuller account of these principles, please also see McLean & Gargani (2019) and Price-Kelly & McLean (2020) referenced above.

1.2.1 Justification

This principle is based on the premise that scaling is a choice that must be justified. It provokes a response to the question: *'why scale?'* and emphasizes that mere *technical* or scientific evidence that an innovation *can* scale is not enough – whether an innovation *should* scale additionally involves a moral justification or consideration of the values of all those that are likely to be impacted by the scaling. In this sense, the choice to scale is shared by the innovators and the people impacted alike, requiring innovators to work with stakeholders to anticipate potential benefits and risks associated with scaling, as well as learn what risks they are willing to accept at each level of scale.

1.2.2 Optimal Scale

The principle of optimal scaling necessitates thinking around three key observations. Firstly, doing more of something is not necessarily better. Secondly, and relatedly, it is important to acknowledge that rarely does scaling create *the* impact it is intended for – rather, scaling creates a collection of anticipated and unanticipated and desirable and undesirable impacts, bringing with it several trade-offs. Finally, given this realization, it is important to name and consider the trade-offs in at least four different dimensions of optimal scale, including the magnitude, variety, equity and sustainability of impacts[14]. Optimality then, is about reaching a level where the magnitude, variety, sustainability, and equity of impacts are balanced in a way that is widely endorsed by stakeholders.

1.2.3 Coordination

This principle emphasizes that scaling of any innovation does not occur in a vacuum but rather a scaling environment or 'scaling system'. The scaling system includes a plurality of people, places and things that affect and/or are affected by the scaling such as the initiators and enablers of scaling, competitors, and those impacted. However, it is important to 'coordinate' the scaling environment such that its evolving set of actors are aligned around creating impact at optimal scale. This includes creating champions, building effective partnerships, aligning incentives, and so on. In addition to coordination across actors, coordination across innovations can also be helpful in avoiding duplication of efforts and achieving synergies, thus potentially enhancing overall impact.

[14] These dimensions are discussed in detail under Chapter 4.

1.2.4 Dynamic Evaluation

This principle establishes that scaling is an intervention itself, not an attribute of an intervention. It asks not only “What is the impact of an innovation at a given level of scale?” but also, “What is the impact of the scaling?”. In other words, it emphasizes that scaling contributes to the process by which innovations create impact, but, at the same time, scaling also changes impacts i.e., scaling actions trigger scaling effects. Dynamic evaluation guides us to anticipate and react to these changes – it emphasizes that scaling brings dynamic change, necessitating dynamic evaluation before, during and after scaling. This principle highlights the importance of continuous learning and adaptation through all stages of the scaling effort, encouraging constant monitoring, learning, and adaptation to achieve impact at optimal scale over time.

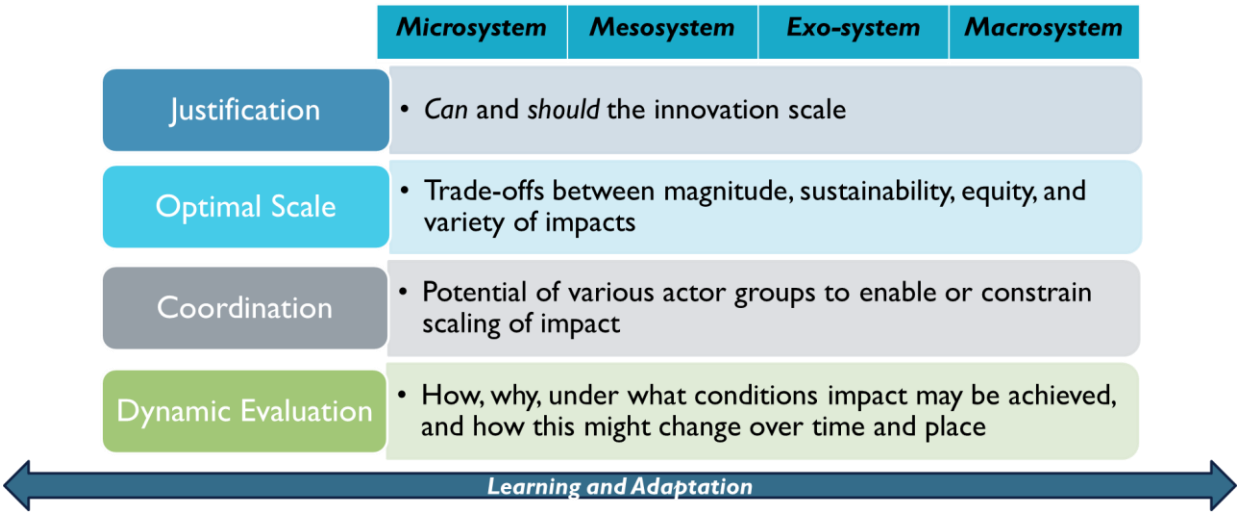
1.3 Scaling Considerations in DSI Research

The conceptual underpinnings of the DSI project are provided by the four guiding principles for scaling impact, to achieve *a scale of impact important to people and the environment and contribute to a broader system of development change*. The project recognizes that successful scaling of innovations that aid data-driven school improvement within education systems involves navigating complex interactions among various elements situated in different ecological layers within these education systems. These include:

- the microsystem, including the school and classroom culture;
- the mesosystem, including and encompassing departmental cultures, education managers, district-level practices, and reporting structures;
- the exo-system, involving a wide range of research partners, donors, technical assistance teams, and other external stakeholders; and
- the macro-system, including and encompassing national trends and policies, federal and/or provincial level dynamics, and policymakers.

To effectively scale educational innovations, it is crucial to explore the extent to which the scaling principles described above are met, and understand the interactions among these layers and their interrelatedness. Further, weaving back and forth scaling considerations across the different layers of the education ecosystem and the four principles of scaling requires constant learning and adaptation. These elements constitute the defining features of the DSI conceptual framework, as illustrated under Figure 1.

Figure 1: The DSI Conceptual Framework



As mentioned above, in Punjab and KP, where the SIF has already been implemented and scaled, operationalization of the DSI conceptual framework focused on exploring the opportunities, challenges and solutions in effective scaling of the SIF. In Sindh and Nepal on the other hand, where no such innovation exists, the project examined the possibility or feasibility of scaling the SIF or a similar contextually relevant innovation in data-driven school improvement[15]. The research followed a largely qualitative line of enquiry including interviews and focus groups with approximately 350 stakeholders across the different layers of the education eco-system in Pakistan and Nepal (see Annex B). The findings from this research exercise have helped shaped this scaling strategy, as outlined below.

1.4 Report Outline

Based on the research findings from Pakistan and Nepal, this document aims to provide a strategy or roadmap for effectively scaling innovations in data-driven school improvement in the project countries, but also GPE member countries and others at large. As mentioned above, the strategy is aligned with the four guiding principles for scaling outlined in *Scaling Impact* by John Gargani and Robert McLean[16]. The strategy also appropriately incorporates insights and recommendations from conducting dissemination activities under the project with relevant education stakeholders or the ultimate users of this knowledge, such as joint policy dialogues with government- and private- entities on scaling education initiatives. The dialogues have proved particularly useful in bringing together a wide constituency of stakeholders from the education

[15] The differentiated data collection goals for each of these two sets of contexts are further elaborated upon in Chapter 2.
[16] McLean, R., & Gargani, J. 2019. *Scaling Impact: Innovation for The Public Good*. Routledge. The principles are also explained in detail in Chapter 3.

sector to collectively explore the challenges, possible responses and opportunities associated with scaling educational initiatives in the region. The ultimate goal of this strategy is to promote the long-term efficacy and sustainability of innovations in data-driven school improvement and accelerate progress to this end in Pakistan and Nepal. However, in increasing stakeholders' awareness of the scaling challenges and practical approaches to overcoming them, the strategy is also envisioned to be useful for other GPE member countries at large, particularly in the KIX EAP region, where the sub-optimal use of education data is a shared policy challenge. For instance, each of the four chapters pertaining to the guiding principles of scaling summarize a set of strategies for both: (i) contexts where innovations in data-driven school improvement are already scaled, and (ii) contexts where the scaling of such innovations is yet to occur. The strategy therefore adopts a comprehensive, collaborative approach, emphasizing evidence-based action, adaptability, and sustainability.

The strategy is organized in the following manner: Chapter 2 briefly introduces the four research contexts: Punjab, Khyber Pakhtunkhwa, Sindh and Nepal, and delves into policy challenges and responses related to the use of data. In the light of these descriptions, chapters 3, 4, 5 and 6, discuss the research findings in relation to each of the four guiding principles of scaling – Justification, Optimal Scale, Coordination and Dynamic Evaluation – to propose appropriate scaling strategies for innovations aimed at data-driven school improvement. The final chapter, Chapter 7, aggregates the recommendations emerging from the application of these scaling principles.

2 Contexts, Challenges and Responses in Use of Education Data

This chapter briefly introduces the four research contexts: Punjab, Khyber Pakhtunkhwa, Sindh and Nepal, and delves into the policy challenges and responses related to the use of data.

2.1 Research Contexts

Preparing for anticipated elections in 2023, Pakistan is a fragile economy with a deepening political, economic and more importantly human development crisis. With a population of over 230 million[17], Pakistan's Human Capital Index (HCI) value of 0.41 is low both in absolute and relative terms, lower than the South Asia average and also Nepal (0.49)[18]. More recently, the COVID-19 pandemic has also worsened learning poverty in the country and the 2022 floods have further deepened the learning crisis, submerging over one-third of the country under water, affecting over 33 million people. At a broader level, education in Pakistan is a decentralized subject following the 18th Amendment to the Constitution of Pakistan which empowers provincial governments to lead on the delivery of equitable quality education. However, the country continues to comprise the second largest out of school population with an estimated 22.8 million children ages 5 – 16 (or 44 percent of the total children that age) out of school[19]. Educational outcomes also remain low or modest across all provinces, and are marred by gender, wealth and geographical disparities.

In Nepal, which is a federal republic comprising 7 provincial governments and 753 local governments, the Constitution adopted in 2015 similarly empowers Local Governments (LGs) to manage school education. Per the governance structure, the LGs fall under, and are directly accountable to, the Ministry of Federal Affairs and General Administration (MoFAGA) whilst being less accountable to other ministries such as the Ministry of Education, Science and Technology (MoEST) which, at the federal level, governs all the apex bodies and formulates, implements, and monitors educational plans and policies across the country. The Center for Education and Human Resource Development (CEHRD) is responsible for program level interventions at different layers of the government. Like Pakistan, several educational challenges continue to persist in Nepal, including among others poor quality and inequity in access, geographical remoteness, and gender, and socioeconomic and ethnic differences[20]. A brief description of the research context(s) within each of these two countries and their respective processes for data collection and use is provided below:

[17] The World Bank. 2023. *Population, Total - Pakistan*. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=PK.
[18] Amin, T. 2023. Human Capital Index: Pakistan's Value Lower than South Asia's Average: World Bank. *Business Recorder*. Available at: https://www.brecorder.com/news/40240051.
[19] UNICEF. 2023. *Education.* Available at: https://www.unicef.org/pakistan/education.
[20] UNICEF. 2023. *Education.* Available at: https://www.unicef.org/nepal/education.

2.1.1 Punjab

Located in the eastern part of Pakistan, Punjab is the country's most populous province with a population of approximately 110 million people (or 53 percent share of the country's total population)[21]. The Punjab School Education Department (SED) is one of the largest civil departments in Pakistan, responsible for the primary, elementary, secondary and higher secondary education of children (pre-school to Grade 12) in the province. Headed by the Secretary, it is responsible for ensuring maintenance of the prescribed educational standards across the province with the support of several attached departments, including the semi-autonomous Programme Monitoring and Implementation Unit (PMIU) providing dedicated support for the implementation of donor-supported programmes and independent monitoring in each of the province's 36 districts.

The province is home to 48,238 public schools – 22,731 (47%) boys' and 25,507 (53%) girls' schools – with an overall enrolment of 11.14 million (5.6 million or 51% male and 5.5 million or 49% female students) and teaching staff of 366,671 (45% male and 55% female)[22]. Although the province has made significant strides over the past decade, considerable work remains to be done to resolve issues of access, quality, and equity. For instance, per the Multiple Indicator Cluster Survey (MICS) data, completion rate at the primary level of education is only 66.3% in Punjab and declines sharply at the lower secondary and upper secondary levels, 56.1% and 38.6% respectively. Stark differences in completion rates also exist by location, wealth quintile, and gender.

In Punjab, data collection from schools is via a combination of external- and self- reporting mechanisms[23]. Data is collected from schools by independent monitors known as Monitoring and Evaluation Assistants (MEAs) reporting to the PMIU, as well as Assistants Education Officers (AEOs), reporting to District Education Officers (DEOs) and Chief Executive Officers (CEOs) in the government hierarchy. It is important to note that MEAs serve as external monitors who randomly visit schools on a monthly basis to report data on certain indicators using an MEA application while AEOs are responsible for providing academic leadership and school support to Head Teachers and teachers, and collect data thorough a Classroom Observation Tool (COT) application on indicators related to teaching and learning. In addition to these two sources of data, Head Teachers undertake self-reporting on a regular basis through the School Information System (SIS), an android-based application installed on tablets provided to schools, which has recently started feeding into the province's annual school census reports. The data collected from all three sources – MEAs, AEOs and the SIS app – is used by the Punjab Information & Technology Board (PITB), an autonomous body set by the Government of the Punjab, to feed the School Improvement Framework dashboard and generate school reports, which are then used for action management. With the support of the PITB, the SED continues to innovate and improve

[21] Population Welfare Department, Government of the Punjab. 2023. Available at: https://pwd.punjab.gov.pk/population_profile.
[22] PMIU-PESRP. 2022. *Annual School Census 2021*. Available at: https://www.pesrp.edu.pk/downloads/school_census/2020_21/School_Census_Report_2020_21.pdf.
[23] Annex C provides a visual summary of the data flows in each of the four research contexts.

upon ways of data collection to deal with challenges relating to access, quality, equity, and efficiency for improving educational outcomes and opportunities for all its children.

2.1.2 Khyber Pakhtunkhwa

Khyber Pakhtunkhwa, commonly known as KP, is located in the northern part of Pakistan with a unique cultural heritage and diverse topography including mountainous regions, plains and hills. Despite improvements in recent years, the region still confronts disparities in educational access, especially for girls and children living in rural and remote areas. For instance, like Punjab, education completion rates for KP also diminish as the level of education increases – as opposed to a completion rate of 52.3% at the primary level, only 46.6% and 34.8% of the children at middle and secondary levels complete education. There are also stark inequalities in completion rates by gender (e.g., 44% for females versus 60.3% for males at the primary level, and 24.7% for females versus 44.1% for males at the secondary level), as well as location and wealth.

The Elementary and Secondary Education Department (ESED) in KP, the province's largest civil department, is responsible for providing quality education to students from primary to secondary levels in the province. The department, headed by a Secretary, is responsible for the implementation of policies, programs, and initiatives to improve the standard of education in KP and is supported by eight attached departments. There are 27,524 functional government schools in the settled districts in KP, with an enrolment of 4.83 million and a teaching staff of approximately 155,898 (98,670 or 63% male and 57,228 or 37% female staff)[24].

In KP, the Khyber Pakhtunkhwa Education Monitoring Authority (KPEMA) is responsible for data collection from schools through Data Collection and Monitoring Assistants (DCMAs), who gather data through handheld devices/tablets. The data is collected in real-time and on regular basis, and directly uploaded to the KPEMA database. At the same time, similar to the AEOs in Punjab, data is also collected through Assistant Sub-Divisional Education Officers (ASDEOs) who report to district-level leadership such as the District Education Officers (DEOs) in the government hierarchy.

2.1.3 Sindh

Sindh, the second largest province in Pakistan, is situated in the south-eastern part of the country. Known for its rich cultural heritage, diverse population, and economic contribution as the home of Pakistan's largest city, Karachi, Sindh is a crucial component of Pakistan's national fabric. Like Punjab and KP, the education service delivery in Sindh is managed by the School Education & Literacy Department (SELD), headed by a Secretary, supervising a number of attached departments and institutions and providing provincial leadership on school education matters.

[24] KPEMA. 2022. Annual School Census Report for Settled Districts 2020-21. Available at: http://175.107.63.45/newimusite/images/reports/ASC_Report_2020-21_Final.pdf.

There are 49,103 schools in the public sector in Sindh from primary to higher secondary levels, with an enrolment of 4,561,140 (62% male and 38% female students) and a teaching strength of 133,076 government teachers (90,207 or 68% male and 42,869 or 32% female)[25]. There is a stark reduction in the number of public schools beyond the primary level in Sindh, contrary to Pakistan's constitutional provision for free and compulsory education for all children aged five to sixteen. Further, the comparatively lower enrolment rate for females, coupled with dwindling female participation reducing at higher education levels as well as lower female teaching staff hint at gender disparities in Sindh's education system.

Sindh has two parallel streams for school education data collection and reporting in the province. Both systems focus on different aspects of the educational infrastructure, like student enrolment, teacher qualifications, basic facilities, student-teacher ratios, etc. Firstly, the Sindh Education Management Information system (SEMIS) collects data annually (paper based) from all schools, employing the services of selected teachers and Head Teachers. The paper-based data is transferred on smart devices and shared with the district Local Support Unit (LSU) where it is cleaned and validated (for 10% of the data) before onward submission to the Reform Support Unit (RSU) in Karachi. The RSU consolidates data from all districts and publishes an Annual Census Report reporting numerical data against various indicators. All SEMIS data is gender disaggregated. Secondly, the Sindh School Monitoring System (SSMS) collects data on a monthly basis using smart devices and a mobile application covering all the schools. Data is collected by Monitoring Assistants (MAs) recruited and deployed at the taluka level (roughly 2 to 3 MAs per taluka). There are 29 Chief Monitoring Officers (CMOs), one per district who supervise the MAs and prepare a monthly data collection plan for them for the respective districts. After data is collected by the MAs, the CMOs review and validate the data for onward submission to the Provincial M&E Directorate. Data for all districts is consolidated at the level of the M&E Directorate as well as quality assured before being uploaded on the M&E Dashboard, which is updated every month. All data under SSMS is gender disaggregated, like the SEMIS. It is important to reiterate that there is no School Improvement Framework or similar innovation currently being implemented in Sindh, but existing processes for data collection and use hold promise for their introduction. The data gathered is used by various departments for planning purposes such as teacher recruitment, school construction, facilities improvement, and student enrolment strategies. The M&E data is also publicly accessible on the dashboard of M&E Directorate. However, issues with data quality have been reported.

2.1.4 Nepal

Nepal is a landlocked nation nestled among the towering peaks of the Himalayas. Geographically, the country is divided into three layers: mountains, hills and the Terai, each diverse in terms of population, language groups, religion, culture and climate. In Nepal, the transition from a unitary to a federal governmental system has brought new challenges and opportunities. For instance, as mentioned above, the country’s constitution empowers LGs to develop their own policies, programs and the guidelines, but this new responsibility also poses challenges for LGs, as many

[25] Sindh EMIS Report 2019-2020.

do not even have adequate human resources. Similarly, while the use of data in the education sector is evolving, concerns about its optimal utilization, especially at the school and local government levels, persist. Efforts continue to ensure the consistency, reliability, and validity of self-reported data, as these are crucial for evidence-based planning and implementation across all levels of the education system.

Education data management in Nepal has seen significant advancements since the inception of the Education for All (EFA) 2004 initiative. A key part of this journey has been the full-fledged use of the self-reported Education Management Information System (EMIS). This system, developed in collaboration with various development partners, including UNESCO, provides accurate, reliable, and timely information at all government levels, aiding in the efficient and effective delivery of educational services. Nepal's unique EMIS, known as the Flash Reporting System, is managed by the CEHRD. It collects data from all types of schools, covering students, teachers, infrastructure, examination details, and financial data. This system tracks student and teacher performance over 12 years of schooling, providing crucial support at all levels of government and particularly in schools.

With the transition to a federal governmental system, the manual EMIS was upgraded to a web based Integrated EMIS (IEMIS). This system enables self-reporting of school-level data, supporting informed planning, resource management, monitoring, and evaluation of school education. The decentralized nature of the IEMIS ensures equal ownership by schools, local governments, provincial government, and the central government. This balance enhances the system's sustainability and functionality while alleviating pressure on data producers and providers.

Despite the robustness of the IEMIS, issues concerning its best utilization at the school and local government levels persist. The reliability and consistency of self-reported data remain areas where the government is making consistent efforts for improvement – several validation rules and mechanisms are in place to ensure the reliability and validity of IEMIS data, essential elements for evidence-based planning and implementation across all levels of the education system.

2.2 Sub-optimal Use of Education Data: Challenges and Responses

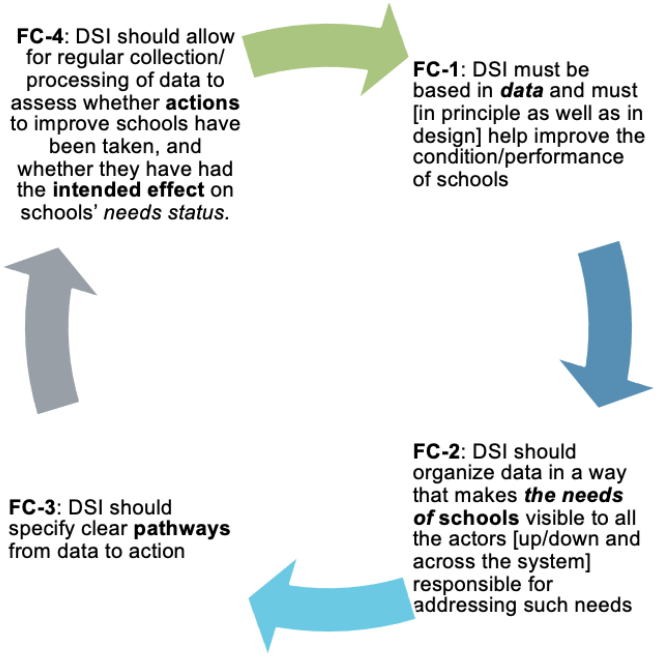
The underutilization of data for decision-making in educational policies presents a global challenge. This was demonstrated by a 2020 study by the Global Partnership for Education - Knowledge and Innovation Exchange (GPE-KIX), which surveyed 21 member countries in Europe, Asia, and the Pacific (EAP) region. The study revealed a significant underemployment of EMIS for internal planning at various levels, highlighting a lost opportunity for enhancing educational delivery efficacy[26].

[26] KIX EAP Hub. 2020. Thematic Priorities in 21 GPE Partner Countries of the Europe | Asia |Pacific Region. Available at: https://www.gpekix.org/knowledge-repository/thematic-priorities-21-gpe-partner-countries-europe-asia-pacific-region-volume.

Addressing this challenge requires transitioning from data collection to data-informed policymaking, for which two emerging strategies in Pakistan's context offer potential solutions: Large Scale Monitoring Systems and data-driven school improvement innovations (abbreviated as DSI innovations this point on[27]). The scaled implementation of the SIF falls under the latter, and Annex D traces the evolution of this paradigm shift in Pakistan from data as a reporting tool to data as an instrument of change in education, elucidating the SIF as an instance of data-driven school improvement in Punjab and KP. For the purposes of this chapter, it suffices to say that the SIF is based on a school-focused approach to organize, interpret, and integrate data from LSMS, effectively highlighting the needs of individual schools as well as supporting efforts to address those needs. Generically speaking, this process can be referred to as data-driven school improvement, and the SIF holds promise in addressing the challenge of sub-optimal use of data by meeting its fundamental conditions.

Figure 2 illustrates the fundamental characteristics of data-driven school improvement. As the name implies, data-driven school improvement must be based in data, which is organized such that it makes the needs of each school visible to those responsible for addressing them. Furthermore, it specifies clear pathways from data to action, and underscores the importance of regular data collection and processing in evaluating the impact of implemented actions on each school's needs as well as determining their effectiveness. This approach is indispensable in taking the necessary steps towards ongoing improvement of schools.

Figure 2: The Fundamental Characteristics of DSI Innovations



[27] Note the distinction between the use of the abbreviation 'DSI' (only) to refer to the (title of the) project, versus the use of the term 'DSI innovation(s)' to specifically refer to *innovations in data-driven school improvement*.

2.2.1 The Promise of Innovation, and the Complexity of Scale

The implementation of the SIF in Punjab and KP shows the promise of innovation in improving the organization, interpretation, and use of education data at different levels within a public education system. However, many effective innovations work at small scale but often do not translate into long term, systemic change needed to transform educational landscapes and improve learning around the world[28]. This is because scaling is a complex and non-linear process, the effects of which are rarely perfectly predictable. Consider the example of the global pandemic or the 2022 floods in Pakistan, for instance. Despite the pilot and launch and monitoring of the SIF in Punjab, it is unlikely that the provincial government would have reasonably anticipated the potential effects of such external shocks to the education system on the scaled implementation of the SIF, as well as the collection of impacts it continues to yield as a result. Even in a world where such shocks do not occur, it is possible and in fact likely that scaled implementation of the SIF across the 36 districts in Punjab would not yield perfectly even benefits across different types of schools (such as urban versus rural, boys' versus girls' etc.), different groups of stakeholders (e.g. school-level versus provincial- level actors), and so on. Similarly, it is safe to assume that introducing the existing form of SIF in Sindh and Nepal would not yield a collection of impacts exactly identical to Punjab and KP owing to contextual variations or complexities.

Thus, when scaling any innovation, a key concern is whether and to what extent can its impact be effectively and optimally scaled to address the challenge faced. Responding to this question, the DSI research demonstrates how the scaling *impact* approach comprising the four guiding principles of Justification, Optimal Scaling, Coordination and Dynamic Evaluation offers unique value for scaling innovations in data-driven school improvement. The broader determinants of data-driven school improvement, and subsequently SIF – such as a functional EMIS, organization of data to reflect school needs such that it is translatable into actionable and trackable information for education stakeholders – seem uniquely nurtured in the four research contexts. However, key questions such as *can, and should, the SIF and/or a similar DSI innovation (continue to) scale? What constitutes (or may constitute) optimal scale in each context?* and *how do DSI innovations bring about the desirable impacts at optimal scale in the various contexts?* need to be answered. Therefore, the operationalization of the DSI conceptual framework has constituted a qualitative research enquiry including interviews and focus groups with a wide constituency of approximately 350 stakeholders across all the different layers of the education eco-system in Pakistan and Nepal[29]. Since decisions about whether and how an innovation is scaled are reached in context, the data collection goals were uniquely adapted for each research site, as explained below.

2.2.1.1 Ex-Poste Analysis of Existing Innovation(s) For Scaled Impact in Punjab and KP

[28] KIX. 2021. *Scaling 101 Crash Course*.
[29] This included a careful mix of men and women stakeholders at each level for greater representation of diverse views.

Since the SIF has already scaled in Punjab and KP, the data collection goals in these contexts under DSI research were informed by the need to generate information on the barriers, potential solutions and opportunities in scaling the implementation of the SIF. In other words, the project team set out to determine optimality by investigating the collection of impacts and trade-offs in different dimensions of optimal scale. In particular, the guiding principle of dynamic evaluation came into play in these contexts, exploring how scaling actions have triggered scaling effects, and what further reactions may best respond to the same. The insights and findings from such an ex-poste analysis have helped identify strategies and recommendations for policy- and decision-makers on how to adjust scaling to amplify impact over time, as discussed in subsequent chapters.

2.2.1.2 Ex-Ante Analysis to Frame Scaling Debates of Potential Innovation(s) in Sindh and Nepal

On the other hand, in Sindh and Nepal, where the scaling of SIF or a similar innovation is yet to occur, the data collection goals under DSI research were focused more on the 'how of converting ideas into impacts'[30] i.e., conducting an ex-ante mapping of the expected pathway to scale illuminating the collection of, and trade-offs in, impacts that may come to define optimality. This included framing debates around, and exploring reactions to, the possible scaling of a DSI innovation across different layers of the education system in the two contexts. For instance, exploring the technical and moral justification for such an innovation, stakeholders in these contexts were asked whether and how a DSI innovation like the SIF may be a reliable solution to the problem of sub-optimal use of education data for school improvement, what are some of the foreseeable challenges, constraints, costs and benefits, what may be some of the unforeseen, unintended, undesirable impacts be of such a change, who all are likely to be impacted by such an innovation, and so on. The value- added of such an analysis has been exhibiting patterns of behavior that may likely improve the success of a promising DSI innovation in future, and at the same time, make individually small but collectively important changes in the same in the considered contexts.

The following chapters synthesize the findings from each of the four research sites in relation to each of the four guiding principles of scaling to suggest strategies for effectively and optimally scaling innovations in data-driven school improvement. The concluding section in each chapter summarizes these strategies for contexts in which: (i) a DSI innovation has already scaled, and (ii) a DSI innovation is to be implemented and scaled in the future.

[30] McLean, R. and Gargani, J., 2019. *Scaling Impact: Innovation for the Public Good*. Routledge.

3 Justifying the Decision to Scale

This chapter operationalizes the principle of Justification, revealing important insights and strategies in justifying the decision to scale the SIF or a similar DSI innovation in varied contexts.

3.1 Technical Justification – *Can* the Positive Impact of SIF Scale?

In their book Scaling Impact, McLean and Gargani emphasize that technical justification[31] is a fundamental prerequisite for social innovation. They argue that "without it our approach to development does not embrace science, can be tremendously risky, and will fail to progress[32]". However, this fundamental prerequisite may or may not be met when the adoption and scaling of an innovation represents an instance of what John Kingdon labels transient windows of opportunity[33] (see Annex E). Recent tools aimed at aiding practitioners in strengthening efforts to scale and sustain education initiatives such as the Adaptation Tracker by The Brookings Institution also render seizing windows of opportunity a key scaling driver. Contributing to an enabling environment for sustaining scale, this includes alignment with existing global, national, regional, and/or local government education priorities and timelines.

In both Punjab and KP, the existence of LSMS and the increased availability of regular data similarly created a "window of opportunity" for the implementation of the SIF. This window of opportunity was reinforced by a confluence of circumstances including renewed political emphasis on education reforms, the need for data, and the time-bound nature of support from international development partners. Small scale pilots were conducted, but they weren't designed as Randomized Control Trials (RCTs) to test the innovation's effectiveness. Rather, their aims were to test the logistical and technological operations, and to understand/resolve the challenges of implementation. Once the operational problems were presumably identified and their solutions found, the government mandated the implementation of the SIF at scale.

Therefore, a key insight emerging from the experience of scaling SIF in Punjab and KP is that where an innovation is being introduced for the first time, its technical justification may not borrow so much from the evidence that it *can* create specific impacts as from other factors such as the presence of an enabling environment for sustainable scale, including windows of opportunity. Nevertheless, the positive outcomes of the SIF implementation, as experienced and narrated by the microsystem (school staff) and the mesosystem (district education department functionaries), under the DSI research lend weight to its technical justification in existing and future contexts, as outlined below:

[31] McLean and Gargani describe technical justification as "Basing the decision to scale an innovation on evidence that suggests it *can* create specific impacts." McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p. 35.
[32] Ibid. p. 38.
[33] John W. Kingdon, Agendas, Alternatives, and Public Policies, 2nd ed. (New York: HarperCollins College Publishers, 1995), 165-169.

3.1.1 Student Participation

As described in Annex A, the SIF captures the needs of schools across four performance domains: student participation and personal development; teachers and teaching; leadership & school support; and school environment. During the course of DSI research, numerous remarks by respondents indicated the SIF's positive impact on student attendance, learning, and increased classroom engagement. Some Head Teachers also associated enhanced student performance in annual assessments with the implementation of the SIF. An example can be found in a statement from a Head Teacher in Punjab who acknowledged that:

"…following compliance with the SIF, we observed an improvement in children's attendance,", and *"the cleanliness system had also improved".*

Similar sentiments were noted in KP. Furthermore, in KP, the team documented perceptions about positive effects on girls' participation attributed to improvements in school environments, including the construction of boundary walls and provision of functional toilet facilities. These improvements were seen by stakeholders as emerging as a consequence of SIF implementation.

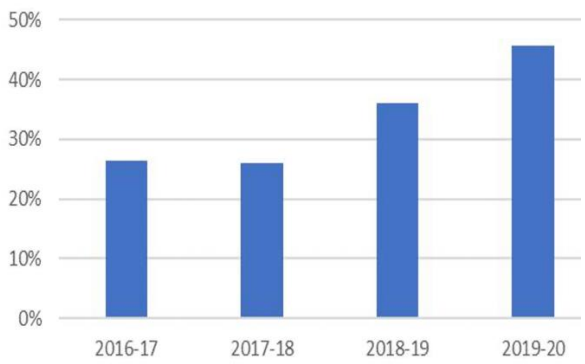
3.1.2 SIF Associated Improvement in CPD and Student Learning Outcomes

Although the information on student learning is anecdotal, it is vital to bear in mind that the SIF includes indicators pertinent to other initiatives that may directly impact student learning. One such initiative is the Continuous Professional Development program (CPD) in KP. The CPD activities in the KP were strategically designed to equip teachers with the skills to tackle specific learning difficulties identified in the preceding year's assessment. The designers of the CPD anticipated that such a targeted approach would help improve student learning outcomes.

The CPD was introduced in KP in 2016-17. Nevertheless, as indicated by the chart below, there was no significant change in the average learning outcomes between the academic years of 2016-17 and 2017-18. Monitoring data for 2017-18 demonstrated that absenteeism in CPD sessions was rampant, thus limiting the benefits of the program. To address this issue, the provincial education authorities decided to add CPD attendance as an indicator in the composite indices in 2018-2019. Linking participation in CPD with schools' and districts' performance and measuring it regularly created an incentive to reduce teacher absenteeism.

The subsequent improvement in student learning outcomes is generally attributed to the CPD. However, it is noteworthy that the success of the CPD program was in turn due to its inclusion as an indicator in the composite indices. This exemplifies how efficient use of data can influence positive change and improve educational outcomes.

Figure 3: Trends in Student Learning Outcomes at the Provincial Level in KP



Source: Khyber Pakhtunkhwa Education Sector Programme Data Analysis Report 2020

Like CPD, the Classroom Observation Tool (COT), which is also included as an indicator in the SIF in Punjab, is also helping develop a constructive relationship between the AEOs and the teachers, which, according to the Head Teachers and the teachers interviewed for this study, is helping the teachers teach better.

3.1.3 Emerging Relationship Between Local Academic Leaders and the Schools

In Punjab, AEOs are tasked with managing clusters of 10-12 schools, and providing academic leadership and school support to the same. As explained in Chapter 2, KP employs a similar model with ASDEOs providing pedagogical advice to teachers. However, an ASDEO's span encompasses a much larger number of schools (50-55), making it challenging to offer individualized attention. To mitigate this, the KP government has recently introduced *School Leaders* to oversee a more manageable number of 10-15 schools, similar to AEOs in Punjab[34].

Our field observations indicate an emerging relationship between the AEOs and school staff, expediting responses to school needs without dependence on formal, top-down action management. This dynamic was particularly apparent in our interviews with teachers, who regularly interact with the AEOs and expressed appreciation for their supportive role:

"The AEO utilizes a classroom observation tool. They observe and provide feedback using these tools. They spend approximately thirty to forty minutes in my classroom. Her attitude is immensely supportive".
"The AEOs have substantially helped us change some of our existing methods and improve our teaching approaches".

[34] Under the School Leaders Program (SLP), 2500 School Leaders (SLs) were to be recruited and made responsible for provision of mentoring and academic supervision. Their terms of references focus solely on supporting teachers and schools to improve quality and access to education. Each SL would be assigned approximately 10-12 schools. The number of schools under their span was deliberately kept small (compared with the ASDEOs span of 50-55 schools) to enable them to visit each primary school regularly. The SLs would also be given digital tablets and trained on the use of a custom-built, android-based SLP application to collect data and to support schools based on the information collected by them.

"If I encounter a teaching issue, I have the convenience of reaching out to the AEO via WhatsApp". "Furthermore, whenever a new concept or tool is introduced on the app, the AEO takes the time to explain it. When a new indicator is introduced, they detail its significance, such as the indicator related to student hygiene".

As described in the preceding section, the beneficial effects of Continuous Professional Development (CPD), although not directly linked to the SIF, were enhanced by using teachers' participation in the CPD as one of the indicators in the SIF. It is worth noting that while the SIF does not initiate the relationship between schools and AEOs/ASDEOs/School Leaders, it significantly nurtures it by creating concrete opportunities for them to use the information provided in the SSI to engage with the schools.

So far, the technical justification for scaling SIF in Punjab and KP borrows from the unique windows of opportunity on which the innovation capitalized in each context, as well as the promise of the innovation in delivering desirable or positive impacts at scale. The scientific evidence or technical justification to scale the SIF or a similar DSI innovation in other or new contexts such as Sindh and Nepal necessitates similar considerations. In Sindh and Nepal, where the SIF or a similar DSI innovation currently does not exist, the rationale for adoption revolves around its anticipated benefits. The DSI data points towards the likely justification for adopting and scaling DSI in these regions being influenced heavily by the presence of local windows of opportunity, and the R4D under the project also creates an opportunity for Sindh and Nepal to leverage insights gathered from impact risk assessments identified in other settings, as outlined below:

Sindh: The findings from Sindh underscore the possibility for a 'pragmatic' justification and application of a DSI innovation that aligns with the experiences of Punjab and KP. Essentially, the core conditions required for initiating a DSI innovation exist in Sindh, as evidenced by the ensuing description that pertains to the four defining characteristics of the DSI (outlined earlier):

The presence of an efficient system for data collection from schools (FC-1): As previously explained, the SIF in Punjab and KP is built upon an existing infrastructure for data collection and processing. In Sindh, such an infrastructure is provided by the Sindh School Monitoring System (SSMS). Established in 2015, the SSMS has augmented the Sindh Education Management Information System (SEMIS). The latter was only comprising an annual census covering all public sector schools, from early years to higher secondary education. Simultaneously, SSMS regularly tracks participation indicators related to teachers and students. It also collects data on the status of school facilities, the effectiveness of school management committees, and the monitoring of school expenditures.

An institutional mechanism to process the data in according with the DSI framework (FC-2): The Directorate of Monitoring and Evaluation (M&E) possesses the requisite technological capability to carry out regular data processing and generate consequential reports. As it stands, these reports generate district-level aggregate statistics. However, the existing technological infrastructure allows for easy reconfiguration to produce detailed reports for each individual school, facilitating the aggregation of data pertaining to the specific needs of these institutions.

This potential for reconfiguration enables a more targeted, efficient, and responsive approach to addressing school-level requirements.

Presence of clear pathways from information to action (FC-3): The reports generated by the SSMS are scrutinized in monthly data review meetings held at the district and taluka level, which are spearheaded by Assistant and Deputy Commissioners. According to feedback from senior officials from Sindh's Monitoring and Evaluation Directorate, who manage the SSMS, and field monitors, the data is deployed to address specific concerns, such as teacher and student absenteeism and the provision of basic amenities. Pertinent issues are identified by Monitoring Assistants (MAs) and presented in the monthly District Reforms Oversight Committee (DROC) meetings[35], presided over by the Deputy Commissioner. The DROC meetings also serve as platforms for the execution of action plans to address the issues identified in the reports, indicating the potential for data-driven actions in the province's context.

Iterative collection and processing of data (FC-4): Much like its counterparts in Punjab and KP, SSMS also aspires to collect data from each school on a monthly basis. However, it is noteworthy that this goal remains unrealized due to a shortage of monitors. For instance, in Mirpur Khas, a district under our study, there are 16 sanctioned positions for Monitoring Assistants (MAs), out of which 14 are currently occupied. These 14 monitors are tasked with collecting data from an estimated 2,000 schools in the district. This implies that each MA would need to gather data from roughly seven schools every working day, a feat that is practically unachievable given the current staffing and resource constraints. Despite the existing challenges, with a few strategic adjustments discussed in the following sections, Sindh has the potential to secure data from schools with a consistency that meets the operational needs of the DSI.

Despite the ongoing reforms, there are still gaps in educational outcomes in Sindh, making it evident that further improvement is needed. As part of the knowledge mobilization series of events conducted under DSI (referred in later chapters), successful policy dialogues with stakeholders in Sindh led the latter to agree with the need to refocus improvement efforts on the site of school. The participants also appeared convinced that the basic conditions for adoption of DSI existed in Sindh. However, the EMIS directorate in Sindh has expressed the need for technical support for development of the monitoring application and in the contextualised design of the composite index to be used in the DSI in Sindh. In the light of the above, a strong pragmatic justification exists for adoption of DSI in Sindh as a strategic move that can contribute significantly to the broader agenda of education reform in the province. The window of opportunity for DSI in Sindh consists of the systematic accumulation of data from all provincial schools, its subsequent conversion into district-specific reports, and the routine utilization of these reports to discern and rectify issues. This indicates that Sindh possesses all the requisite conditions for the inception of a data-driven school improvement system (or DSI innovation). Given that the data collection and reporting mechanisms are already operational on a grand scale, transitioning to such a system or innovation predominantly necessitates the formulation of indicators and the creation of a weighted

[35] School Education & Literacy Department, Government of Sindh. Available at: http://www.sindheducation.gov.pk/Contents/Notifications/Meeting%20of%20District%20Reform%20Oversight%20Committee%20(DROC).PDF.

index predicated on these indicators. This, in essence, means that the foundational groundwork for a DSI innovation is already laid down, and the next steps will involve refining the current data collection and analysis system to meet the specific requirements of DSI.

Nepal: Similarly, implementing a DSI innovation in Nepal, just as in Punjab, KP, and Sindh, potentially brings a collection of desired impacts. In Nepal, the focus is on the elements and practices of the Integrated Education Management Information System (IEMIS) that can help justify the adoption of DSI in the country.

The presence of an efficient system for data collection from schools (FC-1): Nepal predominantly relies on a self-reported data collection mechanism implemented in schools. Schools submit data through a web-based portal known as the Integrated Educational Management Information System (IEMIS). This comprehensive system facilitates the transmission of data from schools to local, provincial, and federal government entities. The IEMIS collects a wide range of educational data, spanning from early childhood education and development (ECED) and pre-lower basic education (PBE) through to secondary education (grades 11-12). It not only captures data across various educational levels, but also segregates the information based on the type of schools - community (public), institutional (private), and religious schools, among others. As detailed in the next chapter, the data collection system in Nepal is cost-effective and ensures quality data, offering possible learnings for optimally scaling the SIF in Punjab.

Processing of data (FC-2): In Nepal's school system, data collection is organized around three main reports, Flash reports I and II, and a Consolidated report. Flash I is prepared at the beginning of each academic year and focuses on input and process indicators. It provides information about the number of students enrolled, the number of teachers, student-teacher ratios, the punctuality of textbook delivery, and other related data. Flash II is prepared at the end of each academic year and is concerned with output/outcome and program indicators. It includes data on student retention, attendance, school performance (including information on School Improvement Plans, Financial and Social Audits), student achievement, and other indicators. The Consolidated report is prepared once each year and presents a five-year trend of intake and output data in the school education system, providing a broader view of shifts and changes over time. Although data collection in Nepal occurs at a lower frequency than contexts like Punjab and KP, the fact that there is a mechanism for processing and using data bears potential for a DSI innovation.

Action management (FC-3): A substantial amount of data is collected through the IEMIS in Nepal, but the utilization of this data is being recognised by most interviewees as a major issue, as it is primarily used for national and local purposes. This reinforces the findings from prior KIX research on shared education policy challenges in the KIX EAP region, where 'Nepal stands out as a country that utilizes EMIS for planning and monitoring purposes only at one level (the federal or central level)[36]. The implementation of a DSI innovation therefore encourages a more nuanced

[36]KIX EAP Hub. 2020. Thematic Priorities in 21 GPE Partner Countries of the Europe | Asia |Pacific Region. Available at: https://www.gpekix.org/knowledge-repository/thematic-priorities-21-gpe-partner-countries-europe-asia-pacific-region-volume.

usage of data at multiple levels, especially the school level, similar to practices in Punjab, KP, and Sindh, potentially enhancing data utilization and facilitate data-driven decision making.

Iterative collection and processing of data (FC-4): As mentioned above, the flash reports are produced twice in an academic year. Drawing upon the aforementioned discussion, Nepal also possesses the technological prowess and a cost-effective system for data collection from schools – it has implemented an integrated data processing system as part of its Integrated Educational Management Information System (IEMIS).

However, our study indicates that Nepal's flash reporting system does not exhibit a focused consideration of individual schools. The current system lacks a localized action management framework that assigns specific responsibilities to officials for addressing school-level issues and ensuring their accountability. Such a mechanism is crucial for maintaining high-quality education and addressing challenges proactively at the ground level. Nevertheless, the periodic use of School Report Cards can align well with the SIF by contextualising and using the domains of school performance and developing a composite index similar to the SSI in Punjab and KP. The IEMIS system in Nepal is capable of providing these tools. The participants in the study recognised that use of SSI can potentially help categorise school in terms of their needs, thus creating a robust response to the policy challenge associated with the use of data in Nepal.

3.2 Moral Justification – *Should* the SIF Scale?

McLean and Gargani argue that technical justification alone is insufficient and must be complemented by moral justification. *Scaling Impact* suggests that even though evidence might indicate scalability, it does not always imply that scaling should occur. The decision to scale should, therefore, consider various factors such as values, context, and urgency, because uniform evidence does not always provide the same level of justification for action in all circumstances. This highlights that scaling is a choice requiring justification through a balance between both values and evidence, and that this choice should be shared between the innovators and those who are impacted.

The DSI research respondents who were involved in the design of the SIF in Punjab and KP described consultations at every tier of the system to finalize the indicators of the DSI. These teams collaborated closely with civil servants at the micro and macro- levels, engaging in a consultative process in which each indicator and its weight in the index were deliberated upon and justified by the participants. Despite a rigorous design process for the innovation, its broad-scale adoption was guided by pragmatic considerations, including political imperatives[37], agreed-upon results frameworks, and, not least, the schedules tied to the release of development assistance. Thus, while the process of design and development assured participation of stakeholders from micro-, macro-, and exo- levels, once the highest level of each province had approved the framework, the justification process came to a halt. In other words, while actors at the micro-level participated in the discussions on the development and refinement of the

[37] The change of government in Punjab following the general elections of 2011 created incentives for the new political leadership to seize the opportunities for innovative initiatives.

indicators, they lacked the authority to modify the framework once it had been approved by the highest provincial authorities. Combined with the principle of dynamic evaluation, this indicates a clear gap in the need for innovators and provincial authorities to undertake continuous learning, adapting the design and scaling course of the innovation to minimize impact risk for all those involved.

Indeed, the DSI research findings based on the analysis of SIF domains/indicators as well as participant interviews and FGDs resonate with this position, revealing that there is room and need for including GESI-sensitive information in the computation of the School Status Index, and consequent representation of school improvement needs as places of learning for them. While data is collected on indicators such as student enrolment and attendance, social dimensions pertaining to children's disability, poverty, ethnicity etc. are often missing, limiting the framework's ability to identify and address potential linkages with student enrolment or drop out, attendance, and, consequently, performance. A related issue is the need to address gender beyond the binary of girls and boys. For instance, including sex-disaggregated information on enrolment and attendance of children in its existing form may not suffice as the framework may still be limited in its ability to distinguish between the reasons behind and consequent ways of improvement in poor performance on the same for boys and girls. The design of the framework thus needs to be altered such that subsequent action management appropriately incorporates the differentiated reasons behind absence from school and/or unenrollment (e.g. menstruation, distance from school, child labor etc.) and effectively addresses *gender* equality. Further, the intersectionality of children's identities – such as gender combined with disability, poverty, ethnicity and so on – and resulting vulnerabilities remain unexplored and warrant greater consideration in the framework[38].

Interestingly, some respondents in Sindh alluded to how generation and use of disaggregated data on aspects such as disability or minority/ethnicity may also potentially lead to discrimination against these groups. This fear underscores the need for broader discussions and capacity-building efforts on GESI data collection, informed by relevant literature and best practices, as well as provincial context(s) to address the dual challenge of ensuring that data collection supports inclusion, while also establishing safeguards to prevent misuse of sensitive information. Nevertheless, this example from Sindh further reinforces the point that the technical and moral justification for an innovation is likely to be different in different contexts, and would need involvement and participation of those uniquely relevant in each context for wider stakeholder endorsement. This context dependence implies that there is no one-size-fits-all, universally applicable list of indicators against which effective data collection and utilization ought to take place in diverse contexts. Rather, each country/context should employ a participatory approach to build consensus with relevant stakeholders across the different ecological layers within its education system to arrive at indicators against which data collection and reporting will be beneficial in its unique context. The avoidance of prescribing a set of indicators for SIF adoption or adaptation is therefore deliberate – nevertheless, the provincial examples from Pakistan above

[38] This aspect of refinement in justification is considered in more detail in Chapter 6 in explicating the linkage between dynamic evaluation (the fourth guiding principle) and justification. Here, it suffices to note that (in the spirit of dynamic evaluation), the DSI research noted inadequacies in the design of the SIF which need to be addressed to make the innovation more Gender, Equity and Social Inclusion (GESI)- responsive.

highlight the need to further explore how existing systems of data collection and use may be expanded to accommodate additional concerns, such as gender, equity and social inclusion considerations.

Box 1: Moral Justification for DSI in Sindh

When excavating the moral justification for a potential DSI innovation in Sindh, the DSI team found that the most noteworthy (and somewhat troubling) aspect of data and its use that some officials alluded to was that the generation and use of disaggregated data on aspects such as disability or minority/ethnicity could potentially lead to discrimination against these groups. Interestingly, this contrasts the comparatively overwhelmingly response by research participants in Punjab and KP in favor of the same, partly indicating a difference in the everyday, lived experiences of individuals across provinces. Nevertheless, this perspective suggests a critical need for awareness-building and policy discussion regarding the role of data in supporting inclusion. In fact, there would be serious risk in scaling a new innovation that involves collection of such data if indeed its use can harm or further disadvantage those already marginalized. This fear or risk underscores the urgent need for broader discussions and capacity-building efforts on GESI data collection, informed by relevant literature and best practices in Sindh's context as well as an encouraging and participatory approach to establishing endorsement for a DSI innovation. These efforts should address the dual challenge of ensuring that data collection supports inclusion, establishing safeguards to prevent misuse of sensitive information, as well as seeking endorsement from a diverse set of stakeholders, who are directly or indirectly likely to be impacted by the innovation.

3.3 Summarizing Scaling Strategies based on the Principle of Justification

The above analyses demonstrate that upholding justification as a guiding principle remains essential for achieving optimal scale in both contexts where the SIF or a similar DSI innovation has scaled, as well as contexts where scaled implementation of such an innovation is yet to occur.

Regarding the former, where the SIF has already scaled such as in Punjab and KP, Justification urges the innovators (in this case, the provincial authorities and associated technical assistance teams) to adopt scaling strategies such as:

- Continuing to identify the universe of stakeholders directly or indirectly impacted by scaling of the SIF and seeking their endorsement in further scaling the innovation. In particular, the impacted people's values, desires and interests 'influence what they perceive to be a problem, the urgency with which it must be resolved, and the merit of competing solutions'[39]. Respondents under the DSI research who are directly involved in and impacted by the scaling of SIF, for instance, indicate the need to make the SIF more gender and equity focused by deliberating indicator inadequacies and undertaking dynamic evaluation and adjustment to better assess impact risk. Relatedly, innovators must also consider the extent to which the

[39] Robert McLean and John Gargani, Scaling Impact: Innovations for the Public Good (New York: Routledge, 2019). p. 50.

design of the innovation permits flexibility in making adjustments and adaptations (especially when scaling out or further expanding coverage of the innovation).

Regarding the latter, where the SIF or a similar DSI innovation is yet to scale:

- The learnings from all four contexts demonstrate how it is possible to maintain the core elements or fundamental characteristics of data-driven school improvement while adapting innovations for the same to local contexts[40], thus indicating technical or scientific justification that the innovation *can* scale. In contexts like Nepal, where utilisation of data is particularly lower than optimal, the clear need to introduce an innovation in data-driven school improvement also provides a strong justification for scaling[41]. Both Sindh and Nepal have opportunities to initiate small-scale experiments to generate evidence on the usefulness of indices for school improvement, while also incorporating lessons from current implementation in Punjab and KP.
- Moral justification, however, necessitates tapping the unique constellation of stakeholders likely to be impacted by the innovation in each of the two contexts, assessing impact risks and establishing endorsement accordingly. The concerns regarding data collection on GESI considerations in Sindh is a case in point.
- Nevertheless, research in all four contexts indicates that in addition to technical and moral considerations, existing or potential windows of opportunity can amplify justification. However, since such windows are transient, sustained, inclusive efforts to strengthen justification through moral and technical lenses are requisite.

In summary, strategically, scaling must become a more shared choice. This necessitates identifying the multiple impacts of innovations in data-driven school improvement and their scaling processes, as well as how different stakeholders perceive them. The scaling impacts of the SIF, particularly the unanticipated and/or undesirable impacts, are discussed in the next chapter to drive strategic recommendations for optimal scale.

[40] Robert McLean and John Gargani, Scaling Impact: Innovations for the Public Good (New York: Routledge, 2019). p. 117.
[41] Robert McLean and John Gargani, Scaling Impact: Innovations for the Public Good (New York: Routledge, 2019). p. 102.

4 Towards Optimality

This chapter, relying on the insights gathered from the DSI project, contemplates the guiding principle of Optimal Scale. It explores its application in the expansion of DSI innovations in both Punjab and KP and extends the lessons learnt to regions like Sindh and Nepal where such innovations have not yet been introduced.

In Punjab and KP, the overarching assumption that guided the widespread implementation of the SIF was embedded in the conventional belief that 'more' would invariably equate to 'better' outcomes. In other words, the underlying expectation was that the district and provincial education systems, along with individual schools, would uniformly respond to the SIF's implementation, optimizing the benefits accrued at all levels. However, the DSI research reveals that the scaling of the SIF produced varying effects on stakeholders across all hierarchical levels and implementation stages, engendering a complex array of both intended and unintended, beneficial but also some potentially deleterious impacts. The following sections delve deeper into this collection of impacts, highlighting the need to balance trade-offs in the magnitude, variety, equity and sustainability of the impacts of scaling the SIF or a similar DSI innovation.

4.1 Optimal Scale and the Balance of Impacts

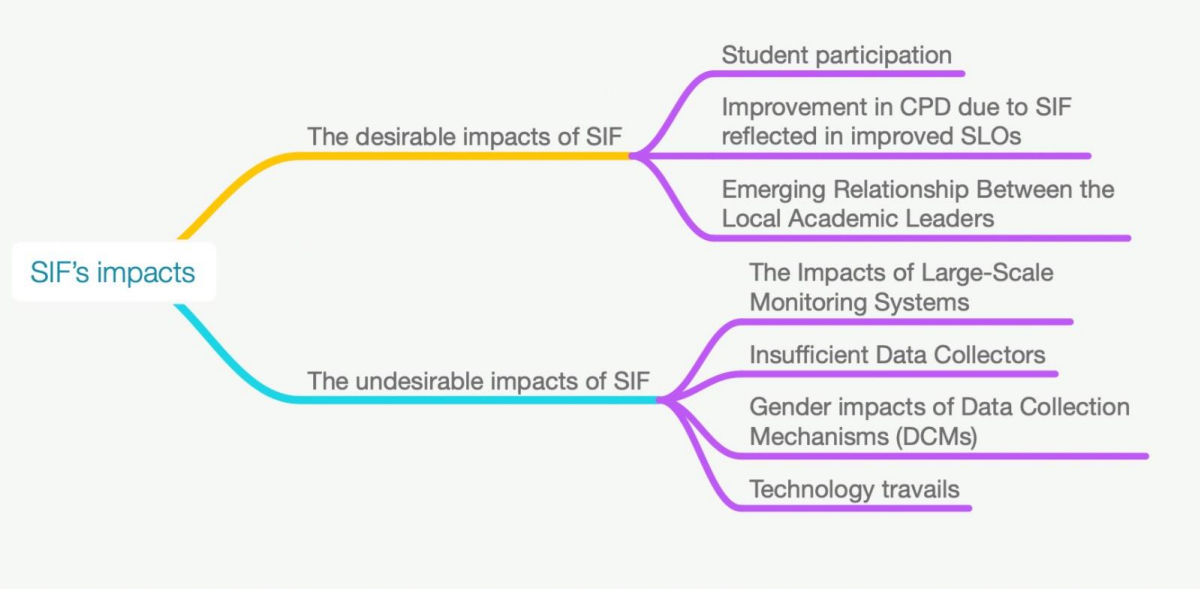
Optimal scale is about reaching a level in scaling an initiative where the magnitude, variety, sustainability, and equity of impacts are balanced in a way that is widely endorsed by stakeholders[42]. Optimality calls for acknowledging the fact that scaling up may indeed bring trade-offs, and that more is not necessarily better. McLean and Gargani emphasize this worldview where scaling is not just a process of replication or expansion, but an active pursuit of optimal scale as described above. The scaling of the SIF, in this sense, would ensure that it promotes the achievement of desired impacts while mitigating potential undesirable outcomes.

The deployment of the SIF in Punjab and KP has yielded several positive outcomes, which have also been discussed in the previous chapter. However, in line with the scaling premise above, this strategy acknowledges that achieving optimal scale entails maximizing an innovation's positive effects while minimizing potential adverse impacts. Therefore, in keeping with this understanding, this chapter (and consequently, strategy) will concentrate on the elements of the data-driven school improvement initiative that yielded unintended and undesirable effects in the case of Punjab and KP (see Figure 4), in the hope that the strategic recommendations based on

[42] For a full account of what these dimensions entail, please see McLean & Gargani (2019), p. 53. Magnitude of impact in relation to SIF or a similar innovation, for instance, entails whether scaling the innovation brings about more data-driven school improvement for all geographies/schools in a way that people value more. Variety of impact entails the full account of the positive/negative, intended/unintended, anticipated/unanticipated impacts that such an innovation is likely to result in. Sustainability considerations may include evidence on the potential duration and reliability of impacts of such an innovation over space and time, while equity concerns the granularity of impacts, whether some groups are or will be affected more than others.

the same can help mitigate them in the considered contexts as well as prevent them in other countries opting for data-driven school improvement innovations. The DSI research identified several negative effects that emerged during the scaling of the SIF in Punjab and KP. The origins of these impacts can be traced back to the complex interplay of the ways large-scale monitoring systems operate, gender-related cultural sensitivities, and power dynamics within the hierarchical structure of educational departments. Understanding these intersecting factors is also critical for dynamic evaluation and optimization of the SIF's impact at a larger scale, in accordance with the fourth guiding principle.

Figure 4: Desirable, Undesirable, Intended, And Unintended Impacts of SIF in Punjab and KP



4.1.1.1 Magnitude, Variety and Sustainability of Impact

The operation of the SIF depends on accurate and regular flow of data from the schools. It is for this reasons that the existing LSMS in Punjab and KP were deemed the prerequisite for the development of the SIF. Ironically, as DSI research shows, some of the unanticipated characteristics of the LSMS also turned out to be the most significant risks to the SIF's effectiveness[43]. Furthermore, there is evidence that some changes in the government policies also implied changes in the availability of data which adversely impacted the operation of SIF. At

[43] For readers who may have skipped Chapter 2 of the report, it is pertinent to revisit the data collection systems in Punjab and KP. As previously noted, both provinces utilize comprehensive monitoring mechanisms that include a large number of monitors. In Punjab, these individuals are known as Monitoring and Evaluation Assistants (MEAs), while in KP, they are referred to as Data Collection and Monitoring Assistants (DCMAs). Their primary responsibility involves ensuring each school within their respective province is visited at least once a month. The responsibility of data collection does not rest solely on MEAs and DCMAs. Schools in Punjab, for example, are also tasked with self-reporting data on certain indicators. Assistant Education Officers (AEOs) in Punjab and Assistant Sub-District Education Officers (ASDEOs) in KP also play a significant role in gathering data from schools on select indicators.

least two examples uncovered by the DSI research in each of the two contexts where SIF has scaled – Punjab and KP – exhibit the undesirable and unanticipated effects of the innovation.

In Punjab, during the initial phase of data collection (December 2021), initial DSI findings indicated that an unexpectedly high percentage of schools - over half (56%) - were categorized as having high or critical needs. The data collected from the micro- and macro- level respondents indicated that a high percentage of schools were erroneously being placed in the high or critical needs category. This finding from the interviews was corroborated by the actual data from the Government of Punjab, which showed that 56% of all schools were placed in the high- and critical-needs category, which, according to research participants at the macro-level, was not an accurate representation of the actual situation. Upon probing, it was found that these statistics did not portray the actual scenario. Further investigation revealed that out of a sanctioned strength of 1,068, only 610 Monitoring and Evaluation Assistants (MEAs) were currently active (at the time of data collection)[44]. *This represented more than a doubling of the data collection burden on field MEAs, resulting in an inability to collect data from all schools with the frequency stipulated by the original School Information Framework (SIF) model*. The following excerpt from an MEA interview represents this effect on the MEAs' ability to collect data:

Figure *5:* Incidence *of Too Many Critical and High-Needs Schools in Punjab*

Improvement Needs	% Of Schools
Low Needs	6.98%
Moderate Needs	36.65%
High Needs	38.85%
Critical Needs	17.52%

Source: SIF Dashboard (Punjab)

"*I manage to visit approximately fifty schools in a month, regardless of whether the target is a hundred or more. I can't go to more than forty to fifty schools*".

Additionally, the initial probing of the monitoring budgets showed that there was need for an annual budget of over PKR 700 million for MEA salaries and allowances alone. As such, the financial implications of data collection from the schools by external monitors surfaced as a significant scaling challenge during the time of the research. The COVID-19 induced financial crisis implied further restricted ability on part of the provincial government to fill the vacant MEA posts. Thus, while there exists a robust framework to organise data in the form of SIF, the expensive and resource intensive data-collection mechanisms undermine its optimal use.

On the other hand, however, Nepal has a cost-effective self-reporting mechanism to acquire data from schools, but this data is not yet organised in ways that help directly identify and address the

[44] At the time of completion of the study, we are informed that GoPB has recruited the MEAs against most, but not all, vacancies. GoPB has also expressed a renewed resolve to improve the implementation of the SIF.

school needs. The comparative analysis of data collection and utilisation models in the two project countries under DSI research provides potential strategies for undertaking cost-effective yet reliable and quality data collection across all schools in Pakistan as well as undertaking better data utilisation in Nepal (see Box 2).

4.1.1.2 Technology Related Impacts

Relatedly, the undesirable impacts produced by the technology are experienced as a result of both hardware and software related challenges. Software challenges arise in conjunction with the data collection issues mentioned in the preceding section. Lack of timely data collection due to the above-mentioned issues revealed a weakness in the data processing and reporting system of the SIF in Punjab and KP. As mentioned above, the software module in Punjab was ill-equipped to appropriately manage the calculations of the School Status Index (SSI) when data was not available. It erroneously interpreted instances of missing data as reflecting critical needs. These issues were compounded by natural disasters such as the COVID-19 pandemic and the 2022 floods. These disasters affected student and teacher participation in ways that the current data collection practices were not equipped to capture, and the ensuing challenges raised questions about the capacity of the SIF model to sustain schools' resilience in the face of emergencies or shocks.

In KP, the student learning outcomes (SLOs) were included in the SIF due to the existing government policy to administer centrally administered universal assessments at the end of the 5th grade in all government administered schools. In 2018, KP changed this policy, resulting in loss of data against this indicator. A considerable number of schools were similarly downgraded from the 'good' to the 'fair' category after the new policy went into effect. The findings from DSI key informants in KP suggest that these reclassifications did not represent the actual performance levels of the schools and were merely due to loss of data against an indicator in the SIF. Like in the case of Punjab, the downgrading of schools in KP also happened due to missing data. However, unlike Punjab, the lack of data was not due to shortage of monitors but due to a policy decision to discontinue the centrally administered universal assessments at the end of grade 5[45].

4.1.1.3 Equity of Impact

In addition to GESI-related challenges in the design of the SIF framework (highlighted in the previous chapter), the DSI research excavated several GESI-related impediments in processes leading to data-driven school improvement. While scaled province-wide and across a range of

[45] In 2018, schools in KP that achieved high scores in the Student Learning Outcomes (SLOs) category often found themselves classified as 'good'. This result was largely influenced by the substantial weight attributed to SLOs in the index (40 points), leading to a scenario where a school with high SLOs scores could excel despite potentially lower performance in other domains. Conversely, schools with suboptimal SLO scores would fare poorly even if they performed admirably in other domains. However, with the transition to a sample-based assessment in 2019 and the ensuing data shortfall for SLOs, those schools previously benefiting from high SLO scores but lagging in other domains experienced a downgrade in their standing. Thus, policy shifts in KP, influencing the availability of assessment data, have effectively rendered the SSI only partially reflective of the needs across three of the original four domains.

actors, the research uncovered specific challenges facing female education officers such as AEOs and ASDEOs responsible for data collection. The data collection procedure by external monitors in schools necessitates their on-site presence, leading to a set of unique impacts, particularly in the context of female monitors. Their mobility is restricted not just by the logistical

Box 2. Learning from Nepal – No Trade-offs Between Sustainability or Cost Effectiveness and Quality in Data Collection

The DSI team delivered on the potential for cross-country learning by undertaking a comparative analysis of data collection and utilization models in Punjab (Pakistan) and Nepal. A specific aspect in this analysis included the relatively cost-effective model of self-reporting of data by Head Teachers/teachers in Nepal at the school level, versus the data collection system in Punjab which relies on a large number of field monitors. A related point of investigation to this end has been the risks that accompany self-reported data, potentially compromising the quality and effective use of data for improvement purposes. A key finding is that the data collection model in Nepal, alongside being more cost-effective, includes checks and balances for quality and accuracy of data including (but not limited to): (i) in-built, software checks (ii) restricting permission for Head Teachers to make changes in reported data beyond a limited period of time (iii) a comprehensive system for reporting and correcting errors in data at the level of the local government, and (iv) an expanding set of fields against which Head Teachers need to enter data, making it increasingly difficult to fabricate or report fake data. Resultantly, as part of the Nepal School Sector Development Plan which necessitates the validation of the Integrated EMIS in Nepal every two years, three independent verifications of school level data since 2016 have shown over 95% accuracy of data. The DSI research team's engagements with government counterparts to explore the possibilities of self-reporting data suggest that the monitoring establishments consisting of a large number of monitors are backed by existing laws, and self-reporting of data by schools – even if demonstrated as beneficial – will face legal and political changes in its implementation. Further engagement with stakeholders on these challenges, however, in line with the core principle of dynamic evaluation, led to an insight regarding the use of a hybrid model of data collection with both self- reporting and external data collection targeting different sets of schools based on well-defined criteria.

Scaling Insight: Data-driven school improvement, by way of its fundamental characteristic(s), must be based in data – hence data collection from schools is a crucial first step. However, where the collection of data by external monitors is being compromised in the face of fiscal pressures, a hybrid data collection model involving both external data collection as well as self-reporting by schools—offers a practical strategy to address this challenge.

challenges of often difficult terrains but also by societal norms that govern their movement. Female monitors candidly shared:

"We are provided a monthly allowance to hire a vehicle for visits, but we face many difficulties as we cannot go alone. Data collection poses problems, but we collect data just as our male counterparts do".

For many monitors, the path to school is not just a physical journey but one dotted with social obstacles as well. Women are restricted from travelling to remote areas due to societal expectations, a problem exacerbated by the physically demanding nature of the region's

mountainous topography. Despite these limitations, women carry out their duties. One monitor explains:

"We arrange our own transport. The male monitors use motorbikes. Despite difficulties, the district office and the head office provide us all kinds of support".

However, while men are provided with motorcycles and allowances, women face a significant problem, as expressed by a female respondent below:

"There is a transport issue for females, and visiting hilly areas is difficult. Males are given bikes, and females get money for transport which is not enough".

Female ASDEOs grapple with the same transport-related challenges when they undertake school visits. They receive a meagre transportation allowance from the Elementary and Secondary Education Department (ESED), which provides motorbikes to their male colleagues. However, the rising cost of transportation and living expenses has rendered this allowance inadequate, forcing many female ASDEOs to resort to using their personal funds to cover travel expenses. With the cost of living and transportation soaring unremittingly, it has become progressively more challenging for them to sustain their data collection efforts. Sharing her experience, a female ASDEO stated:

"The multitude of challenges in data collection is significantly compounded by transportation issues".

Elaborating on her struggle, she expressed:

"For us women, the lack of personal vehicles combined with escalating fuel costs and an insufficient mobility allowance constitute a persistent problem".

The limited public transportation options in these areas only exacerbate these difficulties. As one female monitor shared:

"We face a lot of stress due to mobility issues, such as who will go to the field, the expense of travel, and reaching schools is often a struggle due to bad weather."

These realities underline the gender-disparity and the accompanying challenges faced by these women, as well as the complex interplay of factors such as geography or terrain, social and cultural barriers and government policy.

Furthermore, when female monitors take maternity leaves, their colleagues have to shoulder their workload.

"This issue also arises that if women are on maternity leave, the remaining staff has to do their work", a monitor explained.

These maternity leaves also lead to deductions in their performance-based allowances, with no compensation offered for the additional work undertaken by others. In conclusion, societal norms, logistical challenges, and insufficient transportation allowances combine to produce undesirable impacts on the women involved in the data collection from schools.

4.2 Summarizing Scaling Strategies Based on the Principle of Optimal Scale

The rapid scaling of the SIF in Punjab and KP has provided varied impacts, both desirable and undesirable. At the school level, stakeholders perceived the use of SIF as behind some positive changes in student attendance, classroom engagement, and learning performance, linking SIF adoption to increased participation. The SIF also created opportunities for teacher development, as the inclusion of Continuous Professional Development and Classroom Observation Tool indicators motivated teacher attendance and instructional improvements. Another benefit was the emerging relationships between cluster leaders like AEOs and schools, which enabled rapid localized academic leadership support and advice. The SIF provided concrete reasons for these leaders to actively collaborate with schools on addressing needs. Additionally, some actions on school environmental conditions like boundary walls and functional toilets were partly attributed to the SIF needs identification process. In essence, the beneficial impacts spanned domains of improved student participation, teacher growth, enhanced school-cluster leader connections, and select facilities advancements - all contributing to a more conducive teaching and learning environment.

Simultaneously, the SIF also produced some undesirable impacts on stakeholders. Data collector shortages coupled with mobility and cultural constraints for women monitors resulted in data deficits, undermining composite indicator accuracy and overall SIF performance. Technologically, errors in SSI calculations due to missing data impacted perceptions about the operational credibility of the SIF. The absence of mechanisms to account for contextual shifts like policy changes or disasters such as COVID-19 and the 2022 floods also made the impact of the SIF uneven across regions with different levels of crises. For district officials, these data distortions reduced SIF monitoring accuracy. Rigid centralized action workflows hindered localized support, constraining cluster leaders despite emerging relationships with schools. In summary, while expanding rapidly, unintended negative consequences emerged across the dimensions of data collection, technology, action management, and iterative review.

As mentioned earlier, the optimal scale is defined as the “point at which magnitude, variety, sustainability, and equity of impacts are balanced in a way that is widely endorsed”[46]. From the perspective of this strategy, the achievement of this balance should remain a strategic direction to be followed with the help of the guiding principles of *coordination* and *dynamic evaluation* discussed in the next two chapters. Achieving the point in the scaling of SIF where optimality is reached would require a set of strategic activities, outlined below:

For contexts where the SIF has already scaled, such as in Punjab and KP, scaling strategies include:

- Optimizing scale by balancing the different dimensions of optimal scaling. Punjab is a case in point where a potential solution to the trade-off in sustainability versus magnitude and quality of impact is offered by a hybrid data collection model involving lesser reliance and financial

[46] McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p. 54

burden in the form of monitoring assistants, whilst increased self-reporting from Head Teachers in schools. The following table summarizes how this and other recommendations for optimizing scale may map over time in the two contexts (Punjab and KP):

Table 1: Strategic Recommendations to Establish Optimal Scale of the SIF in Punjab and KP

Strategic Activities	Short-Term Actions	Medium-Term Actions	Long-Term Actions
1. Optimize Data Collection Systems by ensuring sustainability: Introduce innovations in data collection system, providing more support to data collectors and establishing a hybrid system of data collection involving both self-reporting and external validation, thus reducing the burden on the monitoring system	Explore the benefits and possible risks in reduction in external monitoring and its replacement by a hybrid system of data collection from all schools that relies largely on self-reporting of data by schools.	Pilot hybrid system of data collection to generate local evidence on ways it can reduce negative impacts of current data collection mechanisms on monitors as well as schools.	Institutionalize a hybrid system of data collection, making them a part of standard operations.
2. Ensure equitable impacts: Make concerted efforts to ensure equity in the impact of SIF across the system.	Identify areas of unequal impact and develop targeted strategies to address them.	Monitor the effectiveness of these strategies and adjust as necessary.	Institutionalize the consideration of equity in all operations and policies.
Actively work towards addressing women's mobility issues and cultural barriers that prevent their active participation.	Identify and implement immediate measures that can help women navigate cultural norms and geographical constraints.	Develop and execute programs and policies that encourage and ensure women's participation.	Mainstream gender inclusion strategies across all activities and policies.
3. Leverage Technology Optimally: Ensure the technology systems are robust, fully tested, and free from bugs before implementation.	Prioritize the debugging and stress testing of the existing system.	Regularly update and maintain the systems to ensure they stay robust and efficient.	Build a culture of technological innovation and adaptation.
4. Buffer against Policy Changes: Create mechanisms that allow for quicker adaptation to policy changes and involve people in planning processes to assess impacts better.	Conduct a policy impact analysis to understand and prepare for potential changes.	Implement strategies to ensure flexibility and adaptability to policy changes.	Institutionalize the continuous evaluation of policy impacts on SIF.

For contexts where the SIF or a DSI innovation is yet to scale, scaling strategies include:

- Anticipating as many (of the potentially unanticipated and/or undesirable) impacts as possible and determining the optimal balance in trade-offs. The findings from Punjab and KP on the undesirable impacts of the SIF offer useful learnings in this regard. The key point for

consideration is that rarely ever does scaling produce *the* intended impact – instead, it creates a collection of impacts that will present a mix of benefits and costs to intended and unintended stakeholders in the scaling process.

- Relatedly, engaging people who are or will be affected by the innovation in planning and decision-making processes around scaling, so that uncertainty and the 'unanticipated ness' of impacts is considerably reduced[47].

[47] McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p. 64.

5 Coordinating the Scaling Environment

The discussion so far on the principles of Justification and Optimal Scale has underscored that scaling is not a simple replication or expansion in coverage of a well-conceived innovation; rather, it is a process that engenders diverse consequences for different stakeholders. As detailed in the previous chapter on Optimal Scale, the scaling of the SIF in Punjab and KP has created a collection of impacts, both desirable and undesirable, and anticipated and unanticipated. These findings accentuate the nuanced nature of scaling, reinforcing that effective scaling involves adapting and overcoming challenges that emerge in the collection of impacts it produces. As such, then, an effective scaling strategy must: (i) identify innovation impacts, and (ii) help address them through appropriate steps. While the previous chapter discussed innovation impacts and hinted at useful remedies, this chapter operationalizes the third guiding principle of Coordination to suggest strategies for creating an enabling scaling environment for the SIF or a similar DSI innovation. In part, it attempts to translate the stipulations of this principle to identify a set of strategic leverage points to overcome existing and/or potential challenges to optimal scaling of the SIF enumerated earlier.

5.1 Coordination Among an Evolving Set of Actors in a Scaling System

The R4D conducted under DSI re-emphasizes the need for a flexible scaling process, encouraging engagement and adaptation in emerging relationships with a dynamic or evolving set of actors in the scaling system. For instance, during the course of DSI research in Pakistan in November 2021, a new entity, the Pakistan Institute of Education (PIE) became established as an attached department of the Federal Ministry of Education and Professional Training (MoFEPT). And although education is a decentralized subject in Pakistan and each of the country's four provinces manage their own EMISs, PIE became mandated with aspects such as improved education data collection, consolidation and dissemination including at and between the federal and provincial levels; education assessment; and basic and applied research on various aspects of education for evidence-based policy and decision making. However, as McLean & Gargani note, people, places and things exist for their own purposes in a scaling system and are rarely agents of change in a scaling process until engaged or affected. Noting the emergence of a new, prominent actor, the DSI research team ensured early and ongoing engagement with PIE, which then officially confirmed interest in the use of DSI research to inform its mandate of improved data integration and standardization at the federal and provincial levels.

Resultantly, the DSI research team and PIE jointly conducted a policy dialogue, where lessons and best practices in successful scaling of SIF and similar innovations were shared with development partners and government counterparts from all across Pakistan. The Institute used the event as an opportunity to sensitize all relevant education stakeholders, including development partners, to the research findings, and encourage them to incorporate feasible recommendations in their work portfolios to improve data integration between federal and provincial levels. Further, the Institute is also considering piloting an index-based innovation for data-driven school improvement similar to the SIF at the federal level, in Islamabad Capital

Territory (ICT). Broadly speaking, then, the sustainability of the optimal scaling of innovations in data-driven school improvement requires involvement of and championing at all levels, including the central/federal level institutions, even in countries like Pakistan where education is decentralized, in addition to the existing engagement with provinces. The DSI project team's experience and collaboration with provincial and federal counterparts like PIE lends weight to this claim. For instance, the series of policy dialogues conducted by the DSI research team at the provincial and federal levels helped frame the policy debate around scaling a contextualized innovation in data-driven school improvement in Sindh. Allowing cross-fertilization of ideas across provinces, early and ongoing coordination among relevant education stakeholders in Sindh in such a manner has helped create a more enabling environment for a DSI innovation to be launched and scaled. Further, as mentioned, federal level institutions are also able to experiment with the innovations in data-driven school improvement within a much smaller jurisdiction to further explore the sustainability of such innovations at scale.

5.2 Balancing Power in the Scaling System

A related consideration in coordinating across an evolving set of actors is the way in which they enable or constrain the scaling environment. For instance, people, places and things may begin to function as *initiators, enablers, competitors,* or *the impacted*[48], and these roles may not be mutually exclusive[49]. In the case of Punjab, for instance, borrowing support from the Technical Assistance (TA) team, the provincial School Education Department both *initiated* the School Improvement Framework and *enabled* it. At the same time, it is important to remember that these roles may change or shift over time. For example, the *enablers* of an innovation may, over time, become its *competitors*, offering a next-best or better than alternative to scaling the innovation (in this case, for data-driven school improvement). A central feature of the scaling system, therefore, is the interaction between the different actors in their roles and the scaling process, and the extent to which they are able to influence the scaling process is determined by the power they have. A key purpose of coordination then is to balance the power in the scaling system such that it serves the greater (public) good.

The DSI research findings from Punjab and KP illuminate at least two examples where coordinating the scaling system to balance the flow of power in the scaling system over time is needed:

5.2.1 Example 1: Power Differentials Within Governance Structures

One of the unanticipated effects or challenges associated with the scaling of the SIF in Punjab and KP pertains to the (im)balance between authority, accountability and responsibility. The

[48] McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p. 69-70.
[49] Ibid. pp. 69-70. Emphasis added.

project found that power differentials within a hierarchical governance structure can have a strong mediating effect on the implementation of the system-wide SIF. The DSI research findings revealed that mere assignment of actions to individual actors is a necessary but not a sufficient condition for producing SIF's desired impact on schools. This is because often, the actors at the micro- and meso- levels such as Head Teachers within schools and AEOs and ASDEOs at the cluster level do not have the authority to take actions for school improvement for which they are entrusted with responsibility, as well as held accountable. Additionally, the burden of performance seems to pivot towards the education officers as both the actors above (such as District- and Deputy District Education Officers) as well as below them (such as Head Teachers and teachers) in the hierarchical governance structure expect them to deliver results on school improvement the most.

On the other hand, AEOs and ASDEOs explicate the need for capacitating them, and further supporting them in their role to provide effective school support. For instance, in KP, the DSI research revealed that ASDEOs did not even have access to and/or were not using information provided on the SIF dashboard to build into school-level improvement plans. The evolution of DSI in the province of Khyber Pakhtunkhwa tacitly assumed that actions needed to improve schools (based on the data) can only be taken by the governing tiers above the school. Thus, according to the respondents in the field, the access to the School Status Index data in KP was limited to the Secretary and other provincial level actors and District Education Officers (DEOs), with no access for ASDEOs and lower levels in the hierarchical chain. Thus, as part of its R4D, the DSI project team initiated engagement with the provincial government on the subject, and resultantly, the ASDEOs were not only granted access to the application for School Status Index data to better support and strategize school improvements but also urged to use SIF and all its systems, which is key to moving towards improved utilization of education data. The project also recommended the provincial government to undertake regular and detailed professional development and other training activities for AEOs and ASDEOs, better supporting them in their role of providing academic leadership and school support. For optimal scaling, therefore, it is necessary that governments invest adequately in the development of business processes and workflows as well as requisite trainings for staff.

5.2.2 Example 2: Shifting Priorities of Governments and Development Partners

The DSI R4D establishes positive reception of the SIF in Punjab and KP. In KP, as mentioned earlier, the SIF has assumed the form of three separate indices, each measuring the outcomes at the district, intra-district (cluster of schools), and school level. Introduction of both forms of SIF have produced positive reactions from the stakeholders in general. At the time of writing this report, however, the DSI team has learnt that the use of composite indices in KP has been paused due to reallocation of resources to the competing alternatives. Similar thoughts were expressed by a participant from Punjab about SIF in a policy dialogue conducted by the DSI research team and PIE. The DSI research potentially reveals two reasons why this may be the case:

(i) Reduced attention and priority by provincial governments: The focus groups conducted with meso- and micro- level actors hinted that, with the withdrawal of Technical Assistance

earlier available to provincial governments (such as the TA support available under KESP to the KP education department), attention and priority to the scaled implementation of educational innovations like the SIF has diminished.

(ii) Shifting focus on *competitors*: Relatedly, new programming by development partners such as under the FCDO Data and Research in Education (DARE) program promotes a renewed emphasis on data standardization and improved reporting on the SDGs, promising enhanced quality in education statistics from EMISs[50]. This focus on better reporting has accompanied diminished attention on interventions like SIF that fostered data utilization.

In such situations, in addition to coordinating across actors, it is useful to coordinate across innovations or multiple/ solutions offered by a particular innovation using a portfolio approach, as suggested by McLean and Gargani[51]. Part of the portfolio approach suggests coordinating innovations by repurposing the same innovation to serve additional goals. This strategy can be applied to broaden the objectives of SIF data collection. In the case of the SIF, for instance, the framework can be repurposed to better align its constituent data collection, utilisation and action management processes with reporting on and meeting the Sustainable Development Goals for education so that, instead of curtailing or replacing the framework, provincial governments and development partners are able to sustain its impact(s), albeit with additional benefits or purposes. For example, certain SIF indicators like student attendance and facilities conditions overlap with SDG 4 reporting needs. Rather than separate data collection processes, the existing SIF methodology could be leveraged and repurposed to serve the additional goal of SDG monitoring[52]. Specific steps in the process of repurposing may include:

a. Mapping points of intersection between SIF indicators and SDG reporting needs.
b. Identifying opportunities for coordinated data gathering to meet both purposes.
c. Validating integrated approaches to fulfil divergent needs with one aligned process.
d. Implementing synchronized data collection cycles that feed both SIF and reporting databases.

In essence, repurposing SIF data collection through a portfolio approach could enable efficient, regular local level data to inform both school improvement efforts and national reporting. This coordination has potential to create synergies between apparently competing alternatives, thus improving the sustainability of the SIF. That is to say, coordinating the scaling system such that emerging actors and their evolving priorities cohere around the goal of data-driven school improvement is crucial for sustaining impact of data-driven school improvement innovations.

[50] DARE is an approximately GBP 22.9 million initiative that seeks, among other things, strengthened data systems in Pakistan. Source: https://devtracker.fcdo.gov.uk/projects/GB-GOV-1-300575/summary
[51] Ibid. p.78-80
[52] Alternatively, the current investments in data standardisation and reporting could be repurposed to generate, in addition to the reports on achievements against SDGs, composite indices similar to those employed in the SIF.

5.3 Summarizing Scaling Strategies Based on the Principle of Coordination

The application of the principle of Coordination to the scaling of SIF reinforces the importance of coordinating across both actors and innovations in a scaling system. The illustrations from Punjab and KP have demonstrated how issues in coordination are relational and, in a wide sense, political, which can be addressed through a multi-level, collective perspective on scaling innovations, especially by those likely to be affected by the SIF.

From the standpoint of Coordination, scaling strategies for contexts like Punjab and KP where the SIF or a similar DSI innovation has already scaled include:

- Correcting imbalances in the varying levels of interest and readiness that relevant stakeholders have in participating in the scaling process. The scaling of the SIF demonstrates a wide array of impacts across various stakeholder groups. Increasing the reach of the SIF is unlikely to address the scaling impacts because they are produced by the scaling itself. It necessitates strategic adaptability, informed by an understanding of the innovation's effects at each layer of the ecosystem and at each level of engagement. The examples illuminated above which highlight changing attention of provincial governments as well as lack of capacity and authority at the micro- and meso- levels (for AEOs and ASDEOs) speak to this point.
- Identifying and mapping networks and organizations that believe in the innovation to help avoid competition and duplication as well as negotiating new partnerships with them over the course of scaling. In the case of Pakistan, the DSI knowledge mobilization activities across governments and development partners like the FCDO and WB has helped build and foster such new partnerships, which can eventually aid a repurposing of the SIF to suit the interests of a wide constituency of stakeholders. Here, a portfolio approach can provide a pathway to reconcile seemingly competing objectives, such as SDG reporting and local-level school improvement. By repurposing the SIF to serve these additional objectives, innovators can ensure the system's sustainability while enhancing its range and impact.
- Beyond new partnerships, continuously aligning the scaling of the SIF with other new or emerging initiatives to similarly coordinate and facilitate sustained scale.

In contexts like Sindh and Nepal, where a DSI innovation is yet to be scaled, scaling strategies based on Coordination include:

- Mapping the education ecosystem, and identifying potential initiators, enablers, competitors and the impacted within the scaling system for a DSI innovation. As mentioned above, much of the groundwork to this end has already been achieved under the DSI R4D exploring the potential scaling of a DSI innovation in contexts like Sindh and Nepal.
- Creating buy-in among relevant stakeholders adopting a participatory, multi-stakeholder approach suited to the local context. The knowledge mobilization activities under DSI evidenced how, in the case of Sindh, such a collaborative approach facilitated acceptance and technical (and moral) justification for scaling the SIF or a similar innovation in Sindh. The chapter on Justification also highlighted the need to collaborate even with stakeholders that

hold varying or even opposite perspectives on moral justification for a DSI innovation in Sindh, as such pragmatism can eventually aid or strengthen local ownership of the innovation.

- Ensuring flexibility in the design of the Framework or a similar innovation so that coordinating across an evolving set of actors and potentially competing innovations can become easier. Greater flexibility implies that relevant actors and stakeholders in a local context have greater leverage to adapt and make the innovation their own – such responsiveness makes for a successful scaling strategy.
- As mentioned above, continuously identifying and coordinating across an evolving set of actors and other interventions and innovations to sustain scaling of the DSI innovation.

6 Dynamically Evaluating the Scaling Intervention

The findings from DSI research and ensuing scaling strategies outlined thus far repeatedly emphasize scaling as a dynamically evolving process, rather than a static one. Generating dynamic change, the evaluation of this process therefore also ought to mirror its dynamism. This chapter centers on the guiding principle of Dynamic Evaluation which encourages that learning underpins scaling, from start to finish and, contrary to popular belief, dynamic evaluation is a stance held *before, during*, and *after* scaling[53]. The examples below from the four DSI research contexts – Punjab, KP, Sindh and Nepal – further elucidate this point.

6.1 Evaluating *Scaling* as an Intervention *Before, During* and *After*

Often, impact evaluations treat scale as a given attribute of any intervention – when evaluating the impact of the intervention, a stable cause and effect relationship is assumed at a given level of scale. McLean & Gargani (2019) argue that scaling is not an attribute of interventions, rather it is an intervention itself. And because we scale, or change the level of scale to change the nature of impacts, scaling should be evaluated as an intervention in its own right. This position renders dynamic evaluation as distinct and different from normal evaluation approaches.

The principle of dynamic evaluation further calls for making evaluations before, during and after scaling. This is because while some of the scaling effects may be predictable, expected or planned for, others are only uncovered once the process of scaling happens, as also suggested by the evidence presented in earlier chapters. Thus, in order to effectively scale impact, it is important to dynamically evaluate scaling actions and their effects before, during and after the process of scaling and work to adjust learning strategies in light of the same. The following sections explain how this may be done across the four research contexts under DSI.

6.1.1 Dynamic Evaluation *Before* Scaling

The evidence presented in earlier chapters, especially Justification, suggests that the scaling of SIF in Punjab and KP was founded more on the basis of enabling conditions such as accessible data from large-scale monitoring systems, conducive political conditions for data-driven accountability, and the readiness of international development partners to back data-informed decision-making within the education sector, and less on scientific evidence of its effectiveness as well as moral justification.

Operationalizing dynamic evaluation before scaling and gauging moral justification, innovators in the two contexts could have posed questions to explore the values and perspectives of those likely to be directly or indirectly by the SIF. For example:

[53] McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p.82

a. How do teachers, students, parents, and the community perceive the SIF's impact on school improvement?
b. To what extent do they believe that the SIF aligns with their values and aspirations for quality education?
c. Does the use of composite indices as a measure of school performance align with their view of what constitutes a 'good school'?
d. Do they perceive the effects of the SIF as equitable and beneficial across different geographical settings, socio-economic contexts, and types of schools?

The DSI R4D delved into these questions, but it did so only after the scaling of the SIF. Doing so beforehand, however, would not only allow better justification and wider endorsement, but also help estimate scaling effects and influence the scaling system in light of the same to scale impact. The key takeaway for scaling SIF or a similar DSI innovation in new contexts such as Sindh and Nepal, therefore, is to address these questions beforehand, including in the form of small-scale pilots before the launch of the innovation. For instance, in response to question b, c, and d above, the DSI R4D found that the indicators used for measurement of school needs were narrow and could not improve the outcomes for some disadvantaged groups, such as the girls and disabled. To do so would require adjustments in the design of the indices used in the SIF. The responses to these kinds of questions could help the innovators understand whether the SIF resonates with stakeholders' values and meets their expectations for quality education. Such an understanding and consequent improvements in the design and scaling of innovations can strengthen justification, help anticipate and address scaling effects, and coordinate the scaling system better.

New contexts like Sindh and Nepal can also benefit further from unique learnings from the experience of Punjab and KP before scaling a DSI innovation, including the need for:

a. Grounding the scaling in established technical and moral justification for the innovation.
b. Pre-empting or anticipating some of the (undesirable and unanticipated) scaling effects and planning strategies for mitigating them.
c. Mapping relevant actors and coordinating the system such that it enables scaling of the innovation, even before it is launched.
d. As discussed below, planning for and delivering continuous and ongoing dynamic evaluation of scaling the innovation, during and after the process of scaling.

6.1.2 Dynamic Evaluation *During* and *After* Scaling

The experience of scaling SIF in Punjab and KP demonstrates how scaling actions trigger scaling effects, producing varied impacts, both desirable and undesirable on stakeholders at every step of implementation. Dynamic evaluation can establish the positive outcomes of the innovation, as it did in the case of DSI R4D. In addition, it can also uncover undesirable effects during the scaling process. This is important for the success of the innovation. Thus, while quantitative data might be adequate to narrowly establish the production and attribution of positive impacts, qualitative design is required to understand the ways in which the innovation impacts the people involved in and affected by its implementation. Learning from the R4D conducted under DSI research,

dynamically evaluating the scaling of SIF during/after the scaling process by asking the following questions can aid a better understanding of how the SIF may optimally be scaled:

a. What are the unexpected challenges or side-effects faced by stakeholders in the course of SIF implementation?
b. How do these difficulties relate to the context-specific characteristics of the operating environment?
c. What impact do these difficulties have on the stakeholders, particularly the impacted?
d. What measures can be taken to mitigate these difficulties and minimize their negative impact?

By exploring these questions, the innovators can grasp the complexities of scaling interventions, uncover hidden obstacles, and identify opportunities for improvement. For instance, the complexities and challenges encountering female education officers in the implementation of SIF have been illuminated as a critical issue as a consequence of such enquiry. Insights like these from dynamic evaluation can help the innovators refine the scaling strategies, such as, in this case, the need to provide suitable transportation or reducing the travel distances by replacing the regular data collection by well-regulated self-reporting of data by the schools.

Similarly, the DSI R4D identified the issue of imbalance between authority, responsibility and accountability stemming from largely rigid hierarchical structures. While changing rigid or inflexible governing structures may not be an option, there may be other possible ways in which application of coordination between various actors may help balance authority and responsibility in the medium and long term. Further probing the following aspects across the different layers of education stakeholders, for instance, can help provide alternative strategies:

a. What measures can be adopted to ensure that the responsibility vested in school-level actors matches the actual authority they hold?
b. How might the SIF be adapted to better accommodate the realities of authority distribution within the school system?
c. What systemic changes are required to empower teachers, head teachers, and local management groups such as School Management Councils and Parent Teacher Committees?
d. What possible impacts might arise from altering authority-responsibility dynamics, and how can these be pre-empted and addressed?

A systemic response to these questions can help guide coordination strategies to engage various actors that empower the micro-level actors by aligning responsibility with appropriate authority. Such alignment may not only prevent undue stress or adverse effects on relevant personnel but also foster a more conducive environment for the effective implementation of the SIF.

The DSI research also evidences how the scaling of SIF has followed different pathways in Punjab and KP. Its implementation has also remained uneven due to the external shocks. In both provinces, the DSI R4D was conducted at a time when education system was heavily disrupted by the COVID-19. The school closures, inability of the system to regularly generate data,

restrictions on travel, impacted the implementation of the SIF making it difficult to determine its direct effects on the schools. Some improvements in the school outcomes documented in this R4D, such as improvement in the teachers' participation in the continuous professional development and the subsequent improvement in student learning outcomes in KP are produced after the introduction of the composite indices. However, all other things being equal, can they be attributed to the SIF? This question remains unanswered and calls for an impact evaluation after some of the undesirable impacts on the system and stakeholders have been addressed through adjustments in the data collection and processing mechanisms. Impact evaluations that determine the overall worth or significance of the composite indices in driving school improvements and enhancing educational outcomes can become an important tool in determining the extent to which innovations like the SIF deliver on their promise. The starting point of this dynamic evaluation could include the following questions:

a. How effective are these indices in representing school performance?
b. How well do they drive decision-making processes in schools?
c. How well do they facilitate targeted interventions in schools that need them most?
d. Do stakeholders feel that the SIF addressed urgent educational needs in their context?

By exploring these questions in detail, the evaluation process could assign demonstrable value to this aspect of the SIF based on its effectiveness and impact on the intended outcomes, thus adding an element of *technical justification* to future launches of the innovation, for instance, in Nepal and Sindh.

6.2 Summarizing Scaling Strategies Based on the Principle of Dynamic Evaluation

This chapter suggests that implementors of DSI innovations in the current contexts, as well as initiators in the new contexts, should incorporate dynamic evaluation as a stance throughout the scaling process. This entails a thorough assessment of the SIF and similar DSI innovations before implementation, meticulous fine-tuning during its scaling, and a rigorous examination of its outcomes after the optimal scale is achieved. As shown in the chapter, this approach not only potentially enhances the technical and moral justification of the innovation but also helps optimize its scale and coordinate solutions amongst all involved actors.

From a strategic perspective, the chapter also shows the recourse to dynamic evaluation as a tool to identify and address potential and actual undesirable impacts of scaling. The strategy of active problem-solving during scaling, informed by dynamic evaluation, is crucial to mitigate these negative effects and incremental progress towards achieving the optimal scale. The dynamic evaluation can also reveal disparities in power dynamics within the system and guide the adoption of measures to rectify the balance between responsibility and authority among the involved actors. Moreover, dynamic evaluation can help innovators identify the elements of the innovation that

need to be adjusted, maintained, or discarded, offering strategic directions for further refinement and potential replication of the innovation in similar or different contexts.

To summarize, the scaling strategies stemming from the principle of dynamic evaluation for contexts where DSI innovations are already scaled include:

- Continuously evaluating scaling, accepting that even apparent failures provide learning opportunities as well as adapting strategies to mitigate unanticipated and undesirable effects of scaling. In the case of Punjab and KP, this includes among other things, considering and addressing local social and cultural gender barriers in processes of school improvement, and employing gender responsive or transformative strategies to transform this dynamic and empower females (as well as the disabled and other disadvantaged groups).
- Based on the existing learnings from DSI, determining and meeting future R4D needs. For instance, determining the specific measures that can align responsibility with authority at various levels of the education systems in Punjab and KP may be a key area where further R4D may be helpful in providing practical solutions that are also endorsed by relevant stakeholders. This would involve a deeper understanding of how the SIF can be adapted to better accommodate the realities of authority distribution within the school system and the systemic changes required to empower the micro-level actors. Similarly, the attribution of the observed improvements in school outcomes directly to the SIF remains a question that future research may address. This calls for an impact evaluation that controls for external factors and seeks to isolate the unique contribution of SIF towards improved outcomes.

Additionally, scaling strategies for contexts where the SIF or a DSI innovation is yet to scale include:

- Justifying scaling by the demonstrated effectiveness of SIF or a similar DSI innovation in local contexts at a smaller scale intervention, alongside its external validity[54]. Paying heed to contextual considerations such as local gender barriers by constantly considering, investigating and assessing them is key.
- As much as possible, anticipating factors for success of the SIF or a similar DSI innovation in context alongside assessing its unintended and/or undesirable effects, and strategies for mitigating them.
- Considering the apparent failures that become evident through dynamic evaluation of scaling as learning opportunities for effective scaling (e.g., through adjustments to innovation design and implementation processes).

[54] McLean, R., & Gargani, J. (2019). *Scaling Impact: Innovation for the Public Good*. Routledge. p.141.

7 Conclusion

This strategy is an outcome of the application of scaling science to the scaling of a promising innovation in data-driven school improvement, the SIF, being implemented in Pakistan. By operationalizing the guiding principles of scaling, it demonstrates how scaling the impact of the SIF or a similar DSI innovation involves 'a *coordinated* effort to achieve a collection of impacts at *optimal scale* that is both morally *justified* and warranted by the *dynamic evaluation* of evidence', both in contexts in where it has already scaled, as well as other, new contexts. It is important to remember that although distinct, the scaling principles work in tandem to help achieve optimality in scaling (e.g., as demonstrated in the links between justification and dynamic evaluation above). Table 2 summarizes the recommendations stemming from this research exercise, as well as suggestive responsibilities. The principles also overarchingly emphasize flexibility in the design and scaling processes of innovations, as well as the need to continuously identify and engage with relevant stakeholders, their values, needs and interests to coordinate scaling better. The strategy, however, is only a starting point – as highlighted by the principle of dynamic evaluation, learning is a continuous journey in the scaling process, the needs of which must be met with constant research for development.

Table 2: Summary of Recommendations for Scaling the SIF or a Similar DSI Innovation

Scaling Principles	**Strategies for scaling SIF in existing contexts**	**Strategies for scaling SIF or a similar DSI innovation in new contexts**	**Suggestive responsibilities**
Justification	❖ Continue to identify the universe of stakeholders directly or indirectly impacted by scaling of the SIF and seek their endorsement in further scaling the innovation.	❖ Initiate small-scale experiments *in context* to generate evidence on the technical justification for scaling the use of indices for school improvement, while also incorporating lessons from current implementation in Punjab and KP. ❖ Establish moral justification for scaling, tapping the unique constellation of stakeholders likely to be impacted by the innovation in the given context, assessing impact risks and	❖ Initiators, who make it possible to start a subsequent stage of the scaling process (e.g. governments), and enablers, who implement or support the scaling (e.g. development partners, technical assistance teams etc.), to capitalize on windows of opportunity. ❖ Initiators and enablers in collaboration with competitors that offer better or worse alternatives and those impacted by the scaling process (such as micro- and meso-

		seeking endorsement accordingly. ❖ In addition to technical and moral considerations, tap existing or potential windows of opportunity to amplify justification.	level stakeholders e.g. teachers, Head Teachers, education officers etc.) to establish technical and moral justification for scaling.
Optimal scale	❖ Optimize scale by balancing different dimensions of scaling, ensuring sustainability alongside magnitude of impact (e.g., through a cost-effective data collection mechanism and technologically sound design) as well as balancing equity of impact in the design and processes of the innovation.	❖ Anticipate as many (of the potentially unanticipated and/or undesirable) impacts of scaling as possible and determine the optimal balance in trade-offs. ❖ Engage people who are or will be affected by the innovation in planning and decision-making processes around scaling, so that uncertainty and the 'unanticipated ness' of impacts[55] is considerably reduced.	❖ Initiators and enablers, engaging those impacted by the scaling in the planning process.
Coordination	❖ Correct imbalances in the varying levels of interest and readiness that relevant stakeholders have in participating in the scaling process e.g., changing attention of provincial governments as well as lack of capacity and authority at the micro- and meso-levels. ❖ Identify and map networks and organizations that	❖ Map the education ecosystem, and identify potential initiators, enablers, competitors and the impacted within the scaling system for a DSI innovation. ❖ Create buy-in among relevant stakeholders adopting a participatory, multi-stakeholder approach suited to the local context. ❖ Ensure flexibility in the adopted/adapted design of the SIF or a	❖ Initiators, enablers, competitors and those impacted to cohere or align around the goal of creating impact at optimal scale (it is pertinent to note that actors may shift over time, for instance, enablers may become competitors over time – coordinating includes taking account of all the evolving actors, and initiators and enablers may take the lead in this process).

[55] McLean, R., & Gargani, J. (2019). *Scaling impact: innovation for the public good*. Routledge. p. 61.

	believe in the SIF to help avoid competition and duplication as well as negotiating new partnerships with them over the course of scaling. In the case of Pakistan, the DSI knowledge mobilization activities across governments and development partners like the FCDO and WB has helped build and foster such new partnerships, which can eventually aid a repurposing of the SIF to suit the interests of a wide constituency of stakeholders. ❖ Beyond new partnerships, continuously align the scaling of the SIF with other new or emerging initiatives to similarly coordinate and facilitate sustained scale.	similar innovation so that coordinating across an evolving set of actors and potentially competing innovations can become easier. ❖ Identify and coordinate across an evolving set of actors and other interventions and innovations to sustain scaling of the DSI innovation.	

Dynamic evaluation	❖ Continuously evaluate scaling, and adapt strategies to mitigate unanticipated and undesirable effects of scaling e.g., in the case of Punjab and KP, local social and cultural gender barriers in processes of school improvement. ❖ Based on the existing learnings from DSI, determine and meet needs for future R4D, for instance, around specific measures that can align responsibility with authority at various levels of the education systems in Punjab and KP as well as impact evaluations assessing the extent to which the observed improvements in school outcomes can be attributed to the SIF.	❖ Justify scaling by the demonstrated effectiveness of SIF or a similar DSI innovation in local contexts at a smaller scale intervention (as mentioned above). ❖ As mentioned above, as much as possible, anticipate factors for success of the SIF or a similar DSI innovation in context alongside assessing its unintended and/or undesirable effects, and strategies for mitigating them. ❖ Consider the apparent failures that become evident through dynamic evaluation of scaling as learning opportunities for effective scaling (e.g., through adjustments to innovation design and implementation processes).	❖ Initiators and enablers, engaging those impacted by the scaling process.

8 Annexures

Annex A: A Note on The School Improvement Framework

Effective use of EMISs requires more than just data collection. In addition to relevant information on key education indicators, it requires a coherent structure that meaningfully organizes and translates data into actionable information for stakeholders at different levels of government. This is particularly relevant for schools from where a lot of information is collected, but very little is channeled back throughout the education system to drive school-level improvements. The School Improvement Framework (SIF) is a conceptual and methodological tool developed to use EMIS data to identify and address schools' needs by actors within the education system by doing the following:

(i) *Organizing indicators into key domains of school performance*. In Punjab (Pakistan), for instance, the SIF organizes 24 key education indicators into the following domains:

(1) Student participation and personal development
(e.g., student attendance, achievement, graduation rate)
(2) Teachers and teaching
(e.g., teacher adequacy, presence, professional development, teaching practices)
(3) Leadership and school support
(e.g., headteacher availability/presence, instructional leadership to teaching staff)
(4) School environment
(e.g., adequacy, safety and quality of school infrastructure).

(ii) *Creating a coherent basis for determining school performance*. Once indicators are assigned to domains, the SIF estimates a weighted composite index called the School Status Index (SSI).

$$\text{School Score} = I_{\_(\textit{Student Participation})} + I_{\_(\textit{Teachers \& Teaching})} + I_{\_(\textit{Leadership and Support})} + I_{\_(\textit{School Environment})}$$

(iii) *Categorizing schools by level of need for improvement*. The SSI is computed for each school. The SIF categorizes schools on the basis of the SSI into various bands, according to their needs for improvement (e.g. Needing Improvement, Satisfactory, Good and Outstanding).

(iv) *Providing feedback to schools for self-appraisal*. Given the nature and level of their needs, schools can develop and implement plans to improve performance as well as compare their progress over time.

(v) *Generating and managing actions at each level of the system*. Based on the reports generated using SSI, the SIF identifies actions that different actors at different levels of the education system need to take. Consider the challenge of teacher shortages in Punjab (Pakistan), for instance. Since this factor is beyond the control of schools or School Heads, the action management system resulting from SIF nudges the concerned authorities at the provincial level of government to fill

vacant teacher posts. Not only this, by categorizing schools in terms of their needs, the SIF helps governments prioritize allocations of limited resources to schools that need them the most.

Annex B: DSI Research Methods, Data Collection and Analysis

Research under DSI was largely qualitative, comprising and combining methods such as targeted, in-depth interviews, focus group discussions and observations/observation-al notes with secondary/documentary analysis. The data collection instruments, comprising interview items and focus group discussion protocols, were developed in close alignment with the four guiding principles of scaling – justification, optimal scaling, coordination and dynamic evaluation. In Punjab and KP, where the SIF has already scaled, data collection goals revolved around the need to generate information on the challenges, opportunities, and potential solutions in the scaled implementation of the SIF. In Sindh and Nepal, where the project adopted the course of feasibility research, the guiding principles for data collection and subsequent instrument development were tailored to focus on the potential scaling of a proposed innovation in data-driven school improvement.

It is pertinent to note that the process of finalizing data collection instruments included pre-piloting and piloting activities inviting useful insights from relevant stakeholders such as Head Teachers, teachers, School Council members and sub-district, district- and provincial- level education officers. By kindling the researcher-practitioner duo in such a manner, the project successfully increased the relevance of the research for the ultimate users or beneficiaries of the research effort themselves. A Gender Equity and Social Inclusion-lens review was conducted for all research instruments to ensure that they integrated contextually relevant GESI-specific fields, and Institutional Review Board (IRB) approval was granted by the Lahore University of Management Sciences (LUMS). All instruments were developed in English and translated into local languages to ensure quality data collection.

Data collection

To further ensure quality data collection, Field Assistants in each research context were selected following a rigorous recruitment process and were provided 2-3 day in-person trainings on the scope of DSI research, the specific objectives of data collection, GESI and ethical considerations, and a detailed item-by-item review of the research instruments. The trainings also included a small pilot or mock activity to provide field assistants further guidance on effective notetaking and data collection. Further, Field Assistants were also provided brief refresher trainings immediately prior to the commencement of field activities to ensure good quality of data. Data was collected from approximately over 350 individuals (including pre-pilot and pilot activities) across all four research sites per the sample details provided below:

Punjab

Microsystem		Mesosystem		Exosystem		Macrosystem		Total
Stakeholders	**Number**	**Stakeholders**	**Number**	**Stakeholders**	**Number**	**Stakeholders**	**Number**	
Teachers	31	MEAs	19	WB	[Engagements at/through	Secretary Education	1	

Head Teachers	9	AEOs	19	FCDO	sensemaking meetings and research dissemination events]	PMIU	1	
School Councils	30	CEOs	-	UNICEF		PITB	2	
Total	70		38		-		4	112

Stakeholders		Lahore		Multan		Rawalpindi		Total
School Council		**M**	**F**	**M**	**F**	**M**	**F**	
13 M	17 F	4	9	7	2	2	6	30
Teachers								
5 M	26 F	0	9	3	8	2	9	31
Head Teachers								
3 M	6 F	0	3	2	1	1	2	9
MEAs								
19 M	0 F	7	0	5	0	7	0	19
AEOs								
10 M	9 F	4	4	4	2	2	3	19
CEOs		-	-	-	-	-	-	-
Total								
50 M	58 F	15	25	21	13	14	20	108

Khyber Pakhtunkhwa

Stakeholders		Swabi		Swat		Haripur		Total
ASDEOs		**M**	**F**	**M**	**F**	**M**	**F**	
8 M	9 F	2	3	3	3	3	3	17

DCMAs								
11 M	11 F	3	3	4	4	4	4	22
DEOs								
3 M	3 F	1	1	1	1	1	1	6
Total								
22 M	23 F	6	7	8	8	8	8	45

Sindh

Stakeholders		**Mirpur Khas**		**Shikarpur**		**Karachi**		**Total**
School Council		**M**	**F**	**M**	**F**	**M**	**F**	
25 M	28 F	11	24	14	4	-	-	53
Teachers								
15 M	25 F	4	17	11	8	-	-	40
Head Teachers								
6 M	5 F	3	3	3	2	-	-	11
MAs								
16 M	0 F	7	0	4	0	5	0	16
DEOs/TEOs								
4 M	0 F	2	0	2	0	-	-	04
CMOs (3 M-0F)								
3 M- 0F		1	0	1	0	1	0	03
DG M&E (1M-0F)		-	-	-	-	1	-	01
Total								

70 M	**58 F**	**28**	**44**	**35**	**14**	**7**	**0**	**128**

Nepal

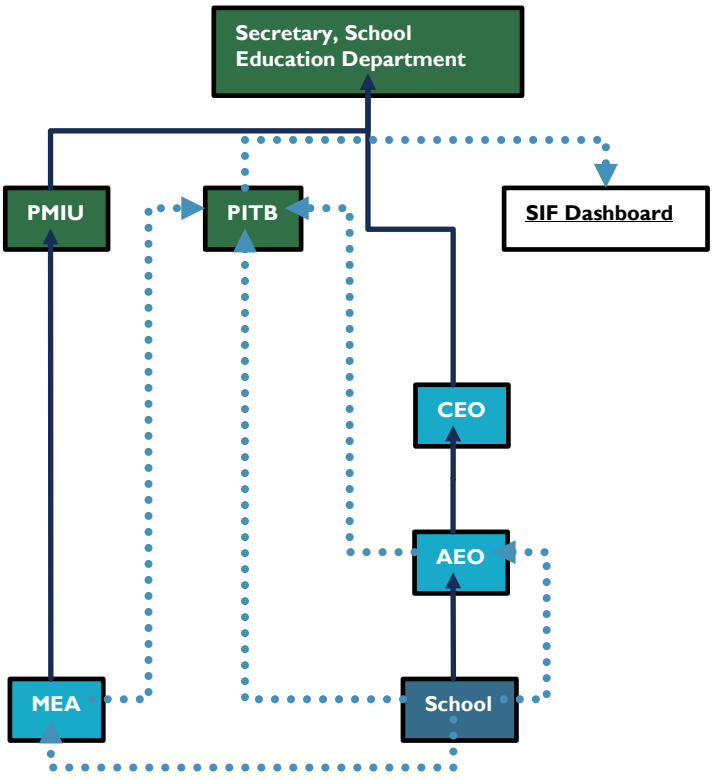
Stakeholders			
School Council		**M**	**F**
		7	3
Teachers			
		14	9
Head Teachers			
		2	8
EMIS Personnel			
		10	0
Total (53)		**23**	**30**

Data analysis

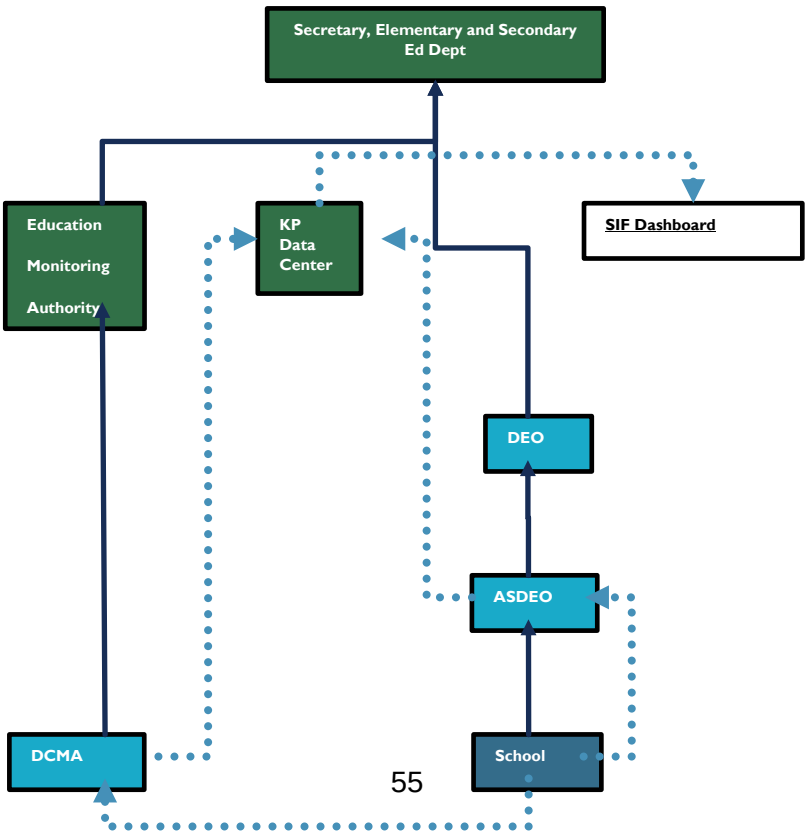
The data collected in each research site was transcribed by the field teams, and quality-assured by respective field supervisors. The transcribed data files were uploaded into Dedoose software for coding and analysis. A hybrid coding process – including both deductive coding based on the research objectives and conceptual framework of DSI, and inductive coding based on the readings and interpretations of raw data – was employed to analyze the data. A sample transcript was also coded by the relevant research team members at the start of the coding process for inter-rater reliability. To avoid loss in translation, the transcripts were coded in native languages such as Urdu and Nepalese. Approximately 91 transcripts were inputted into Dedoose software for analysis (36 from interviews and focus group discussions in Punjab, 12 from interviews and focus group discussions in KP, 35 from interviews and focus group discussions in Sindh, and 8 from interviews and focus group discussions in Nepal). Analytical memos were also developed alongside the coding process capturing researchers' reflections on findings and indicating emergent themes in the process of analysing information.

Annex C: Data Flows in the Four Research Contexts – A Visual Summary

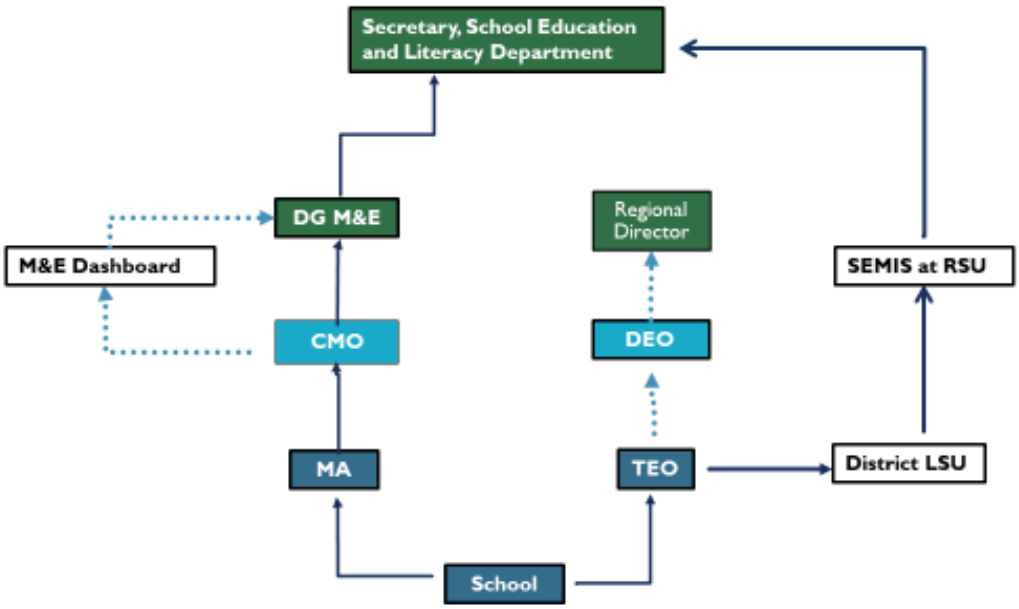
Data Flow in Punjab



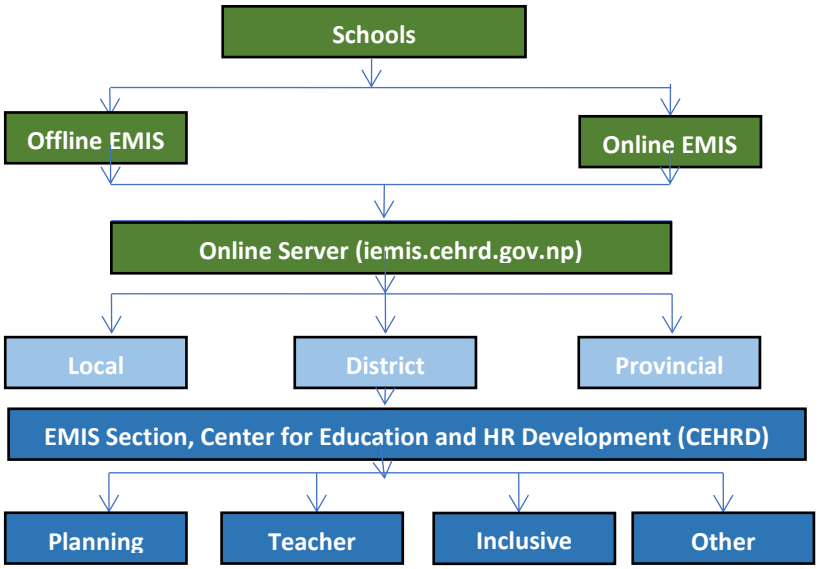
Data Flow in KP



Data Flow in Sindh



Data Flow in Nepal



Annex D: Evolution of Data-Driven Education Reform in Pakistan: Tracing the Journey from EMIS to LSMS, and LSMS to SIF

In the landscape of Pakistan's educational policy, two significant parallel developments can be discerned, each pioneering a different approach to data collection and utilization. The first traces the evolution of the Education Management Information Systems (EMIS), a framework rooted in an annual activity of data gathering about the educational system, resulting in the production of a comprehensive yearly school census report. The second strand of development leads to the establishment of the Large-Scale Monitoring Systems (LSMS), a dynamic initiative that revolves around regular and frequent school monitoring, purposed for continual accountability and course correction.

Since the early 1990s, EMIS, including the National EMIS (NEMIS) and Provincial EMISs, have been operational in Pakistan. Their annual reports outline a thorough analysis of the education landscape, presenting significant details such as the number and types of educational institutions, and a variety of participation indicators like Gross Enrolment Ratio (GER), Net Enrolment Ratio (NER), survival and transition rates, and adult literacy. However, the utilization of EMIS data for improving school quality remains noticeably minimal.

In contrast, 2005 marked the inception of the Large-Scale Monitoring System (LSMS) in Punjab supervised by the PMIU. This development set a precedent for analogous systems across other regions. The Khyber Pakhtunkhwa (KP) province saw the emergence of the Independent Monitoring Unit (IMU) and LSMS in 2012-13, with the IMU subsequently transforming into the Education Monitoring Authority (KPEMA). By 2015, Sindh had also implemented a similar system, and Balochistan joined the trend with the introduction of the Real Time Monitoring System in 2017. These monitoring systems were deliberately designed to facilitate accountability and enable course correction, with a focus on consistently gathering data on student and teacher participation, as well as school environments. Notably, these LSMSs played a pivotal role in supporting accountability drives such as the Chief Minister’s Roadmap in Punjab. Later, they also provided the fundamental conditions for the development of data-driven school improvement initiatives like the School Information Framework (SIF) in Punjab and several composite indicators, including the District Performance Score, Intra-District Performance Score, and School Status Index.

DSI Innovations

Innovation in public administration involves transforming the organization of government elements (such as LSMS), by identifying problems, creating new processes, and implementing innovative solutions to address existing issues. The issue of not being able to effectively identify and respond to individual schools’ needs prompted the development of a new process for organizing data to address this issue in the Punjab and KP provincial contexts.

As mentioned above, the Large-scale Monitoring Systems (LSMS) established by the PMIU in Punjab provided a robust system for tracking the performance of education across the province. This intricate network of nearly over 1,000 Monitoring and Evaluation Assistants (MEAs), equipped with a state-of-the-art android-based application, regularly collected data on sixteen carefully selected indicators from over 52,000 public schools each month. This detailed information informed district-level reviews and facilitated inter-district comparisons, offering a wealth of insights into student and teacher participation, school infrastructure, and students' essential competencies in literacy and numeracy.

Figure: 16 Indicators Used for Regular Monitoring in Punjab

Domains	Indicators[1]	Definitions
Quality indicators	1 LND English	Total correct answers as % of total questions in English
	2 LND Maths	Total correct answers as % of total questions in Maths
	3 LND Urdu	Total correct answers as % of total questions in Urdu
Core Indicators	4 Student Attendance (1-12)	Students (1-12) present as % of total students (1-12) enrolled
	5 Teacher Presence	Teachers present as a % of total teachers
	6 Admin Visits	% of schools visited by district administrators
	7 Functioning of Facilities	Functioning & available facilities[2] as a % of total required facilities
Retention Indicators	8 Student Attendance (K)	Kachi students present as % of total Kachi students enrolled
	9 Retention (K)	Current enrolment as % of baseline Kachi enrolment retained
	10 Student Attendance (1-5)	Primary student present as % of total Primary students enrolled
	11 Retention (1-5)	Current enrolment as % of baseline Primary enrolment retained
Infrastructure Indicators	12 Dangerous buildings	Schools with dangerous buildings as a % of total schools
	13 OCMG Classrooms	Schools with inadequate[4] number of classrooms as a % of total schools
	14 OCMG Teachers	Schools with inadequate[3] number of primary teachers as a % of total schools
Monitoring and Management Indicators	15 AEO Span of Control	Average number of schools handled by each AEO
	16 DTE visits[5]	% of schools with adequate number of coaching visits by DTEs

Despite its innovative design and wide-ranging coverage, the LSMS carried notable limitations. For instance, the statistics produced by LSMS buried them under district and sub-district rankings, thus obfuscating the distinct needs of schools by a focus on high-level comparisons. Further, the system overlooked the significant variability in performance of schools within the same district, treating all schools within a given markaz (center), tehsil, or district as if they were identical entities. Therefore, to counter these shortcomings, the School Improvement Framework (SIF) was conceived, piloted and scaled across Punjab in 2020-21. The SIF, an innovation centred around individual schools, was designed to better structure, interpret, and amalgamate data, thereby focusing on the particular needs of each school. It positioned schools not merely as data points in a broader district or provincial study, but as autonomous entities capable of assessing their

own provisions and outcomes. Thus, the SIF augmented the existing LSMS by adding a more sophisticated, context-sensitive mechanism for school improvement. It was designed to indicate the diverse needs of individual schools and also provided an action management system to help respond to these needs.

The Use of Composite Indices in KP as an Illustrative Case of a DSI Innovation

In KP, the SIF has taken the form of the use of three composite indicators at different levels:

- The Inter District Performance Score, used by provincial leadership to drive improvements in the districts.
- The Intra District Performance Score (covering a cluster of schools, instead of individual schools), used by the district leadership to identify the needs and take actions to address them at the level of clusters.
- The School Status Index, used primarily to gauge the improvement in the status of schools based on the actions taken as a result of conversations driven by the Inter and Intra district measures.

Annex E: Technical Justification and the Concept of Windows of Opportunity

John Kingdon argues that agenda-setting and policy formation do not follow a linear process, but rather emerge from the confluence of three fluid “streams” - problems, policies, and politics. The problem stream consists of issues that gain attention as problems requiring government action. The policy stream encompasses the constant generation of policy ideas and alternatives by communities of specialists. The political stream includes factors like national mood and interest group campaigns that determine if the climate is right for a proposal. When these streams align favorably, a transient window of opportunity opens allowing advocates to couple solutions to recognized problems and find receptive policymakers. The window of opportunity for a policy solution opens only transiently and must be seized before the above-mentioned convergence disappears.

The emergence and use of LSMS (described earlier in Annex D) in Punjab is a case in point, originally driven by the issues faced by the governments in monitoring several initiatives in education sector. The policy advocates within the government as well as the international development partners advocated the establishment of a large-scale system to collect data from all schools in Punjab. Given the need to monitor and show performance, the politicians and civil servants viewed these proposals favorably. The existence of the technology, human resources, and domestic and international support further facilitated the establishment of the monitoring systems. To summarize, the practical challenge of lack of data on education reforms, local political ambition, and international support converged to open a window of opportunity to establish the LSMSs. As mentioned in Annex D, data collections systems were established in KP and Sindh in 2012 and 2015 respectively under similar circumstances. The circumstances under which SIF was justified were similar to the LSMS as mentioned above.