2023 LEAP CHALLENGE

LEAP Final Deliverable(s)

Project Host:

Teach the World Foundation



Fellows:

Christopher Geary, Team Lead, Social Entrepreneur Fellow

Catherine Lebel, Research Fellow

Francis Bizoza Bigirimana, Social Entrepreneur Fellow

Lea Moersdorf, Research Fellow



TABLE OF CONTENTS

Executive Summary	4
Introduction	4
Organisation's role & strength	5
Need summary	5
Solution summary & next steps	6
Deliverable 1: Scaling the TTWF research roadmap	8
Introduction	8
Approach	9
Impact measurement as a conduit for growth	9
Return on Investment	10
Compelling Outcomes	10
Stakeholders	11
What to measure and when	12
Defining societal impacts that can be measured	12
The interdependency of outcomes	19
When to measure outcomes	20
Measuring outcomes scalably alongside product development	22
Takeaways and Recommendations	23
Needs based outcomes	23
Clear definition	24
Increase the scope of measurement incrementally	24
Deliverable 2: Insights from existing data	25
Introduction	25
Overview of existing data	25
Standards/benchmarks	25
Insights from Existing Data	27
Microschools data	27
In-School Data	29
Takeaways and Recommendations	32
Benchmarks	32
Domains to assess	32
The value of longitudinal study design	32
Choosing a control group	33
Feasibility of Automated Assessments	34



Deliverable 3: A comprehensive and inclusive a	pproach to assessing
impact and ensuring sustainability	35
Introduction	35
Community Engagement	36
Takeaways and Recommendations	37
Approaches, Timeline and Tools	37
A. The Reactive Approach	37
B. Proactive Approaches	38
Bibliography / References	40
Appendix	41
Brief Introduction to R	41
R Scripts Used for Graphics	41



Executive Summary

Introduction

Teach the World Foundation (TTWF) is a non-profit headquartered in Pakistan, that designs, develops and operates MicroSchool, In-School and SmartPhone programmes to enable young people to learn basic and functional literacies. Aside from broader corollary societal benefits of basic literacies, the programme enables students in Pakistan to gain access to the formal education system and consequently embark on their journey of lifelong learning and all the accompanying benefits that affords. TTWF has established evidence, through existing studies, that students achieve intended learning gains within the TTWF programmes and through social proof that students are able to transfer their basic literacies in their communities to functional literacies.

The MicroSchool programme is a ground-breaking initiative that enables the affordable setup and implementation of a school for 100 students led by a facilitator in a community within a week. The programme occupies existing available buildings. Learning by students is self directed and delivered through educational apps on tablets that are able to function in online and offline environments. Data about learning is available through an evolving set of app dashboards and is also collected through periodic EGRA, EGMA and EGRA-Urdu testing.

TTWF are now embarking on a first step in scaling their models through an initial partnership with the Sindh Government to establish 100 MicroSchools and 25 InSchool programmes against a schedule of operational and educational KPIs.

As TTWF progresses into the execution of the Sindh project, with an eye on the future, it now seeks to establish longer term KPIs that can demonstrate meaningful and robustly evidenced outcomes for learners and communities as well as meaningful, confidently defined and compelling operational and educational outcomes for future partners and stakeholders that will be key to the long term growth and impact of its initiatives.

TTWF has been paired with a team of four LEAP Fellows with experience in social entrepreneurship (2 Fellows) and research (2 Fellows) and whose experience in data analysis, scaling educational programmes and data collection will support the aims of TTWF in this project.



Organisation's role & strength

With an existing base of evidence and research, TTWF came into the 2023 LEAP programme with an eye on scaling the breadth of impact that can be measured in their programmes. Organisationally TTWF is well organised with dedicated and experienced team members focused upon the evaluative, educational, technological, operational, business development and strategic aspects of programme implementation, measurement and growth. There is an existing culture of seeking to understand the efficacy and evidence of programmes that is pervasive across the organisation and drives decision making. TTWF has begun to consider the operational realities as well as the practicalities of a broader spectrum of impact measurement and is seeking to implement an enhanced technology architecture to support this.

Need summary

The first four weeks of the project with TTWF began with Fellows examining existing data collection practices, available data and its corresponding analysis from across prior programmes and assessment exercises that had been previously conducted. Fellows met with TTWF team members from strategy, education, evaluation, technology, operations and business development teams. This gave a clearer understanding of the objectives and vision behind an expanded impact measurement, considering the goal through the lens of different groups within TTWF who are tasked with meeting different immediate and longer term programme requirements (e.g. operational considerations for communities accepting a Microschool implementation), as well as existing stakeholder KPIs (e.g. Sindh Government KPIs for operational outcomes of the Microschool and In-School interventions).

It became apparent through this discovery period that there was a need to establish clear, consistent and robust benchmarks for the analysis of existing learning progress data within the programmes. TTWF uses EGRA, EGRA-Urdu and EGMA assessments, which whilst being well established across literacy development, do not come with a consistently comparable group of international benchmarks from peer initiatives, or that apply to the learning of Urdu. Therefore, it would be important to understand how to establish these during the LEAP project and to set scalable practices and benchmarks for the collection and analysis of future learning progress assessment going forward.



Many of the broader spectrum of impacts that are often cited from educational programmes quickly convert into economic measurements. However, as we developed an understanding of the specificities of the stakeholders that are and will be instrumental in the sustainable scaling of the TTWF programme in the long term, it became apparent that the societal impact must relate to expressed needs and therefore those perceived as compelling outcomes by learners, communities, governments, donors and future non-profit partners. This also serves the opportunity to focus the scope of outcomes to be considered, their potential for measurement and required interdependencies thereof.

Solution summary & next steps

At the conclusion of the discovery phase of the project the deliverables were concluded as follows:

- How can TTWF scale its research roadmap to capture both specific and broad impacts in line with identified needs of current and future stakeholders, including learner competencies, community impact, and overall societal benefit? How and when can these be measured in a way that is scalable and alongside product development? How can other comparable programmes be used to better understand potential benefits of TTWF?
- 2) In line with #1, how can existing data on learning outcomes be used to make a compelling case for learning gains associated with TTWF programmes? How can other existing data (not learning outcomes) be used to help persuade communities and governments of the value of TTWF programmes?
- 3) What further data collection possibilities (concerning the learner and beyond) will be required to help respond and be compelling to the needs of the different stakeholders?

The outputs of these deliverables will seek to place TTWF in a position for its teams to proceed in existing programmes with a consistent and robust means of assessing and analysing learner outcomes as well as establishing a data collection and evidentiary framework for sustainably measuring and demonstrating the broader impacts of the TTWF programmes on society at large.

The benchmarking of learner outcome assessments, particularly within the context of an organisation assessing progress through EGRA and EGMA, will serve as a valuable reference for other programmes seeking to establish similar consistency and robustness in evaluating



learning progress.

The broader societal impact framework begins from a consideration of societal impacts that can have relevance to a cross section of educational initiatives. The process through which its measurements can be focused to be implementable and relevant to an individual organisation's programmes and stakeholder considerations is something that other organisations can implement as a transferable approach to demonstrating a broader spectrum of societal impact themselves.



Deliverable 1: Scaling the TTWF research roadmap

Introduction

THEORY OF CHANGE

The Teach the World Foundation Microschools programme has embarked on its first steps to scale its model in partnership with the Sindh government in Pakistan. As it looks towards the future, the team from Teach the World Foundation sees potential governmental and international organisation partnerships being established to bring its model to further countries.

In the original theory of change that we reviewed with Teach the World Foundation at the beginning of the LEAP project, there was a significant interest in understanding the more broadly attributable positive societal impacts that occur, over time, as a result of Teach the World Foundation Microschools being established within individual communities, and then more broadly on a national or international level.



Define your goals and how you will achieve them



As we pursued this thread of discussion, we found that across all teams within the Teach the World Foundation organisation, while there was a confidence in existing practice around the data collection to establish learning gains, as well as students self-learning being core to the learning model of the intervention, there is also a significant ambition to establish a broader sense of the "Return on Investment" of the programme, with this being seen as crucial to the long term ambitions for the growth of the organisation.

It was with this in mind that we developed the question(s) for this deliverable:

How can TTWF scale its research roadmap to capture both specific and broad impacts in line with identified needs of current and future stakeholders, including learner competencies, community impact, and overall societal benefit?

How and when can these be measured in a way that is scalable and alongside product development?

How can other comparable programmes be used to better understand potential benefits of TTWF?

Approach

Impact measurement as a conduit for growth

Student learning outcomes and learning gains have been core to the established evidentiary practice of the Teach the World Foundation Microschool programme since its inception. In any context of impact measurement for the programme, these will remain foundational to its ongoing evaluation.

We learned in early discussion with Shafiq Khan, President and Founding Partner of Teach the World Foundation, that central to his vision to contribute to ending global illiteracy, is long-term growth through partnerships and strategic relationships such as other existing non-profit organisations, as well as international organisations such as the World Bank.

It is understood that as the range of partners and stakeholders to a programme's outcomes grows, so do the different lenses through which desired impacts and outcomes are sought and perceived as being compelling.



As our discussions progressed, we found that it is important to establish what we mean by the broader impacts that might be measured and in what way it will be possible to understand whether they are compelling to future partners and stakeholders.

Return on Investment

Thinking initially on the concept of educational return on investment, the idea of a long term measurable financial return immediately came to mind. Indeed, there is much established research into the connection between educational interventions and economic growth across both developed and developing economies. Tanzi and Chu (1998) argue that public expenditure allocations for education in developing economies can improve economic growth while promoting equity. Mussida, Sciulli and Signorelli (2018) find that secondary school dropout and increased years of compulsory secondary education in developing economies is a key element in economic development and growth.

These human capital centred approaches tend to focus on the effects of educational impacts at a point in learning where the learner is at a proximate age to that in which the learner will enter the workforce, whereas the focus of the Teach the World Foundation MicroSchool intervention is predominantly serving children from 7-12 years old, leaving a potentially significant gap in years until such a relationship could be established beyond research involving comparable programmes.

Psacharopoulos and Patrinos (2002) identify that financial measures often fall short of taking into account the true realisation of return between economic gains that return to the public versus to the private purse of individuals. Social rates of return are actually shown to be highest from studies of primary school education.

Compelling Outcomes

Reflecting upon the economic nature of conceptualising broader societal impact under the banner of "Return on Investment", led us to focus subsequent discussions around the outcomes of the programme to determine what might most motivate stakeholders and partners by being most "compelling". We might therefore guide the consideration of what to measure in a way that balances the growth ambitions of the Teach the World Microschool programme with the aspiration to measure a broader scope of impact beyond direct educational outcomes.



Discovery conversations and data shared by teams from the Teach the World Foundation covering operations, technology, curriculum, evaluation and business development led us to understand more about the expectations of existing stakeholders to the programme e.g. the Sindh Government and their established programme key performance indicators that centre as much upon operational programme outcomes as learning outcomes; as well as e.g. community feedback from the operations teams where there might be resistance to an educational intervention in a community - this might be perceived as taking away from a family's capacity to earn income, due to impact on daily life of children going to school and what that might mean for disruption to a family's established routines.

These understandings, drawn directly from the lived experience of the people establishing and growing the Teach the World Foundation Microschool programme across the range of stakeholders and partners that have enabled its success to the current day, led to the determination that the outcomes to be measured, whilst being identifiable from existing literature and academic writing, must be prioritised based on the needs of the stakeholders being served by the programme.

Stakeholders

We were able to identify, working with the different teams from Teach the World Foundation, as well as studying existing literature concerning the societal measurements of educational programmes, the key stakeholders that are most relevant are:

- The Learner
- The Community
- Government (Including Ministries of Education, Health, Commerce or Economy)
- Non-Profit partner organisations and programme operators

Therefore, the means through which the research roadmap can be most effectively scaled to capture both specific and broad impacts to reflect the needs of current and future stakeholders, will be through social returns of the programme, identified through consultation with stakeholders on the perceived priority of their needs, to establish which outcomes are most compelling to them.

This will consequently convey a mutually societally beneficial measurement of outcomes, that reflects the most motivating case for the key stakeholders of the Teach the World Foundation Microschool programme, and that therefore ensures their motivation to proactively support and participate in its growth.



What to measure and when

Defining societal impacts that can be measured

As a means to establish attributable societal impacts that it would be possible to measure, we initially considered researching comparable programmes in countries demonstrating contextual similarities to Teach the World's intervention environments in Pakistan.

In order to do this, we sought published research and programme outcomes directly from educational programmes and non profits. We found that such insight was primarily available through independent published academic studies of comparable educational interventions, rather than through interventions themselves. It is worth noting that this creates a significant barrier to achieving a holistic overview of the insights of comparable programmes because many of the publications are not made freely publicly available in full form.

However, through this process, we were able to determine 69 societal outcomes relating to educational, social, human capital, health, programme operations and economics that could be deemed to be attributable and perceived as compelling by the stakeholders of the Teach the World Foundation Microschool programme. These are detailed in Table 1 below and within the linked <u>Societal Outcome Workbook</u>.

We do not consider that this list of outcomes is in any way exhaustive, and will inevitably be added to or reduced in the future as the longitudinal outcomes of the programme are better understood, however within the scope of this LEAP sprint and the time and resources available we see this list as providing a range of outcomes that apply to each stakeholder.

In order to determine that the outcomes are attributable in the context of the Teach the World Foundation Microschool programme, we identified relevant academic studies that documented the relationship between the outcomes and educational interventions. We also referenced these studies in order to provide a clear definition for each of the outcomes that we identified and listed. The full list of references is included aligned to each outcome and definition in the <u>Societal Outcome Workbook</u>.



Table 1: Outcomes identified and defined, with related stakeholders

Outcome	Stakeholder	Definition
Learning Progress	Learner	The demonstrable improvements within a topic as evidenced by practice
Literacy	Learner	The ability to read, write and calculate simple symbols, digits, words, sentences, texts, with understanding at normal speed, and tackle everyday life problems related to citizenship, gender sensitivity, health, ethics and technical know-how to improve life and ultimately society
Basic Literacy	Learner	Understand the alphabet and digraphs, write letters correctly, spell words and write comprehensible simple phrases correctly or not. Listen to and comment upon stories.
Basic Numeracy	Learner	Understanding of numbers 1-10 and the patterns to count beyond 20. Recognise visual quantities up to 5, understand greater and less than and distributions.
Computer and Mobile Device Literacy	Learner	Ability to use computers and related devices with an understanding of how they operate.
Digital Literacy	Learner	Ability to find, evaluate, and communicate information within digital platforms



Education Mainstreaming	Learner	The placement of a child in a general education classroom (Note this typically applies to SEN students but is illiteracy a special but addressable need)
Functional Literacy	Learner	The capacity of a person to engage in all those activities in which literacy is required for effective function of his or her group and community and also for enabling him or her to continue to use reading, writing and calculation for his or her own and the community's development.
Functional Numeracy	Learner	A person having all the numerical and mathematical skills to function in their society, community, or group.
Employability (Individual)	Learner	The ability to gain and maintain new employment
Personal Wellbeing	Learner	The physical, emotional and mental health of a person and their ability to form positive personal relationships with others around them.
Quality of Life (Individual)	Learner	The capability of an individual to meet their basic human needs, such as food, water, shelter, freedom, access to education, healthcare, or employment.
Instrumental Activities of Daily Living	Learner	Use of the phone, shopping, preparing food, housekeeping and laundry, transport and moving around, taking medicine, handling finances.
Earning Potential	Learner	The amount of money that can be earned within a defined period of time
Job Improvement	Community	The ability for an individual to realise an increased earning



		potential within a defined period of time
Disposable Income	Community	Available cash of an individual, family or community within a defined time period after tax
Better Informed	Community	Able to consume and understand a broader range of communicated information from a wider range of sources
Child Wellbeing	Community	Combined measure of material, physical, cognitive, educational and social emotional outcomes in the life of a child.
Social Capital	Community	A set of shared values or resources that allows individuals to work together in a group to effectively achieve a common purpose
Standard of Living: Employment Availability relevant to experience	Community	Availability of employment relevant to experience
Standard of Living: Hours of work to purchase necessities	Community	The hours that need to be worked to purchase necessities
Standard of Living: Access to healthcare	Community	Ability to obtain healthcare services including prevention, diagnosis and disease management or other health related conditions in a way that is both affordable and convenient
Standard of Living: Quality healthcare accessible	Community	Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes
Standard of Living: Quality and availability of education	Community	An education system that combines equity and quality
Standard of Living: Access to and affordability of transportation	Community	A household's ability to purchase basic and available mobility within their budget
Standard of Living: Environment quality	Community	The physical environment and environmental factors where the



		community lives
Standard of Living: Climate resilience	Community	The ability of a community to to anticipate, prepare for, and respond to hazardous events, trends, or disturbances related to climate
Quality of Life (Community)	Community	The capability of an individual or community to meet their basic human needs, such as food, water, shelter, freedom, access to education, healthcare or employment.
Employability (Community)	Community	The ability to gain and maintain new employment
Employment opportunities from the programme	Community	The opportunities that are achievable after developing basic or functional literacy skills that were not previously attainable.
Social security savings	Government - General	State financial assistance to subsidise low incomes
Food assistance savings	Government - General	Cash and in-kind assistance to communities for the alleviation of hunger and malnutrition
Utility assistance savings	Government - General	Cash benefits for the provision of fuel, electricity and water in low income communities. In some communities can include subsidies for telephone and internet services.Cash or property provision on a subsidised or free basis
Housing assistance savings	Government - General	Cash or property provision on a subsidised or free basis
Unemployment assistance savings	Government - General	Cash payments from a government to an unemployed individual
Childcare assistance savings	Government - General	Cash payments to the specified carer of a child
Health benefit savings	Government - General	State funded programmes for medical care, the prevention of ill



		health and protection from threats to health
Social mobility (Intergenerational)	Government - General	The ability for learners to earn more income than their parents' generation
Civic Participation	Government - General	Capacity to participate in formal or informal individual or group activities that benefit either a community or society
Public safety and security	Government - General	Security, stability, predictability, protection, freedom from fear, structure, order, law and limits
Reduced digital divide	Government - General	Increased access and creative, critical or strategic use of digital tools and connectivity.
Learning progress	Government - Ministry of Education	The improvement of human capital formation as identified by the proportion of students achieving the knowledge prescribed by curriculum in literacy and numeracy on completion of primary and secondary cycles of education.
Schools opened	Government - Ministry of Education	Number of programme relevant schools opened
programmes completed	Government - Ministry of Education	Number of programmes completed within a defined time period
Learners enrolled	Government - Ministry of Education	Number of unique learners enrolled in school programmes
Inclusive education	Government - Ministry of Education	The ability of the education system to welcome all learners.
Student attendance	Government - Ministry of Education	Average number of students or % of students that attend each programme session
Average study hours	Government - Ministry of Education	The average number of hours students spend taking part in the programme within a defined period



Student retention	Government - Ministry of Education	Number or percentage of of students who complete programme
Expenditure and cost effectiveness	Government - Ministry of Education	Cost per student and average cost vs average budget
Later grade learning benefits	Government - Ministry of Education	Better educational outcomes later in education including grades, family income and personal success.
School dropout rates	Government - Ministry of Education	The % of students who enrol and then become no longer enrolled in school
Ratio in cost of education to eventual returns	Government - Ministry of Education	The individual, social and labour productivity returns of education.
Efficient consumer choices	Government - Ministry of Commerce	Being in a position to obtain the greatest level of consumption from the available resources
GDP Growth	Government - Ministry of Commerce	Income earned from the production of goods and services in a country during a defined period.
Employment rate	Government - Ministry of Commerce	Ratio of employed people to working age population
Improved health	Government - Ministry of Health	Quantitative and qualitative analysis of hospitalisation, outpatient service usage and WHO-5 wellbeing items.
Improved decision making related to health (See social capital)	Government - Ministry of Health	Informed decision making related to health (See Social Capital)
Setup Time	NGO and NGO partners	The time taken to establish an intervention
Setup Cost	NGO and NGO partners	The cost to establish an intervention
Setup process	NGO and NGO partners	The number of steps and people involved to establish an intervention
Training Requirement	NGO and NGO partners	The amount of time to train partners or operators of an intervention



Operating Cost	NGO and NGO partners	The ongoing cost across a defined period(s) of time to operate an intervention
Staffing requirement	NGO and NGO partners	The number of people and cost of recruitment to establish and operate an intervention
Technical requirement	NGO and NGO partners	The educational or experiential level of the individuals required to establish and operate an intervention
Licence requirement	NGO and NGO partners	The regulatory barriers and time requirements of regulation to establish and operate an intervention
Marketing trends	NGO and NGO partners	The frequency of like for like interventions and keywords being mentioned within social and traditional media.
Donor trends	Funder	The amount of capital or capital increase in donation or investment into a given intervention type within a period of time across multiple donors
Geographical priorities	Funder	The primary geographical focus for donation or investment into a particular intervention type.

The interdependency of outcomes

To determine when outcomes can be measured, as well as longitudinal considerations as to their actual measurement and data collection, it is necessary first to consider the interdependencies of the outcomes themselves. This will help to establish the order in which measurement takes place, the prioritisation of resources, and an understanding of when the constituent ingredients of outcomes are being or have been measured.

To make the consideration of interdependency manageable within the time and resources available and pursuant to the abovementioned principle of first establishing these outcomes as compelling based on the needs of the identified stakeholders being served, we <u>surveyed</u> the



Teach the World Foundation teams across business development, education, human resources and administration, leadership, monitoring and evaluation, operations and research as teams of individuals who interact directly with the stakeholders on a regular basis.

In the survey, the respondents were asked to prioritise the outcomes giving a score of 0-5 based on their understanding and experience of the perceived needs of the stakeholders. Scores across respondents were then averaged by team and by outcome to establish an initial prioritisation of the outcomes to be measured from those identified.

Where more time is available, and subject to the survey being calibrated to be answerable inclusively across the range of intellectual and educational abilities of the individuals within different stakeholder groups, this prioritisation can be refined with the input of a larger number of direct primary stakeholders.

When to measure outcomes

From 69 initial outcomes, the prioritisation identified the top 20 outcomes across the different stakeholder groups. Given that many of the outcomes would be difficult to establish in the absence of interdependencies that were deemed lower in priority, it was necessary to expand the consideration of the interdependency of the outcomes to 40 outcomes.

This is not to say that for broader compelling societal outcomes to be measured, all of these must be measured. However in considering this broader range of outcomes from the outset, visualising their interdependency can serve to inform Teach the World Foundation of opportunities for a smaller number to be measured and expanded incrementally over time as the method and capacity for such measurement becomes ingrained in the normal working practice of the Teach the World Foundation and its Microschool operations.

The interdependency of outcomes was established by researching the associated studies related to each defined outcome. Such interdependencies can be further directly solidified from the Teach the World programme itself as data is collected in relation to the outcomes in the future.

For the relationships between the interdependencies established to be understood, it is necessary for them to be visualised, following which they can then be placed in a suggested order of measurement, set in this case against unquantified units of time, to be quantified in the future as their data collection methodologies are determined, that can be established as the actual measurements and data collection practices themselves are put in place.







This figure can be seen within the Societal Outcome Workbook



Figure 2: Outcome timeline



This figure can be seen within the Societal Outcome Workbook

Measuring outcomes scalably alongside product development

Having a broad visualisation of prioritised outcomes, their interdependencies and where they might sit in a timeline serves as the foundational step in embarking on a broader societal measurement of outcomes that can be developed scalably over time. Being able to see the relationship between outcomes provides an ability to make a balanced decision not only of what to measure, but to consider the contributing factors in establishing its impact.

 Many of the broader societal outcomes will need to be measured through qualitative methods. These will need to take into account the capability of facilitators to administer their collection as well as that of respondents to answer questions and be provided through a means of consistently, and without bias, measuring and recording the information collected.



- 2) The introduction of any data collection process and practice to measure outcomes from stakeholder groups will have an inevitable effect on the normal working process of existing Teach the World Foundation operations, therefore it will be necessary to consider the implications of this and ensure that the measurement is both practicable and affordable within the resources of the organisation at that time.
- 3) Where data is collected from different stakeholder groups using different formats of collection e.g. online or offline collection in relation to the same outcome it will be important to ensure that the capability exists for the recorded information from both sources to be integrated and stored while maintaining its integrity. There may be considerations as to the readiness and availability of an organisational management information system as a potential prerequisite to this being achieved.
- 4) Where an outcome is to be measured with reference to it being of a compelling nature to a future partner or supporting organisation, it will be important to consider the time required to establish the evidence of the outcome, whilst also being clear as the readiness of the product in educational, technological and operational terms, to also support the same developmental step in subsequent growth. The balance of these considerations will establish whether or not the outcome has sufficient priority at that time to merit the resources involved in its measurement.

It is notable that evidentiary practices are intrinsic to the working practice of Teach the World Foundation. As the demands of product development and data collection expand, it would be advisable to produce a unified roadmap of each to keep clear ongoing consideration of the mutual priorities of each.

Takeaways and Recommendations

As a programme that establishes both a physical and educational presence in communities, Teach the World Foundation develops programme operations with on the ground human resources that have the potential to sustain qualitative societal measurements of its programme impact.

Needs based outcomes

It will be important to maintain a consistent process for the establishment of an outcome's need from a stakeholder perspective to ensure that the most compelling outcomes are being measured to deliver a symbiosis between measurement and growth.



Clear definition

Clearly defined outcomes being maintained as further outcome measurement possibilities are considered is the foundation for determining their need and being able to prioritise and organise their measurement.

Increase the scope of measurement incrementally

It is important to grow the scope of outcome measurement step by step with a clear understanding and consensus across internal Teach the World Foundation teams as to the mutual priority between its social value and its potential to influence the future of the successful and sustainable growth of Teach the World Foundation initiatives.

Figure 3: A process for mapping and prioritising societal outcomes





Deliverable 2: Insights from existing data

Introduction

TTWF has collected baseline, midline, and endline data from children to evaluate two programmes: the Microschools programme, and the In-School programme. This data has been analysed to provide insights into (1) learning gains within the programmes, and (2) a future data collection strategy.

Overview of existing data

Assessments were conducted one-on-one with students to measure English, Mathematics, and Urdu competency. Assessments were based on the Early Grade Reading Assessments (EGRA) and Early Grade Mathematics Assessments (EGMA), which are assessments developed by the United States Agency for International Development.

Assessments were conducted by a local facilitator, one-on-one with students at baseline, midline, and endline during the programmes in paper-pencil format.

Raw, individual-level data was provided for some assessments, and aggregated data for others.

Standards/benchmarks

Scores in individual children can be compared across time (e.g., baseline to endline) using a paired t-test approach to test whether the difference between time points is significant. Change can also be quantified (e.g., 50% improvement), either as a percentage increase over baseline levels, or as a percentage of the possible room to improve (i.e., difference between baseline score and the maximum value). This analysis can help you infer whether endline scores are better than baseline scores, but it is difficult to interpret without context. For this, some sort of standard, benchmark, or target is required.



These benchmarks provide an external reference so you can understand, for example, how many children were "fluent readers" before and after the intervention. This gives important context for interpreting the results, because a 25% change or 13-point increase is not necessarily meaningful without understanding how that impacts a person's abilities and learning.

We searched the literature for standards for EGRA and EGMA. While there are no standards for Pakistan, we found benchmarks which were developed during benchmark-setting exercises in each country. For example, see Tanzania's report <u>here</u>. See Table 1.

Table 2: Benchmarks for reading and mathematics on EGRA/EGMA assessments from other countries.

	English Reading			Mathematics			
	Nonword reading	Oral reading fluency	Reading comprehension	Letter ID	Missing number identification	Addition & subtraction (Level 2)	Problem solving
Egypt (gr 3)	25 cwpm	60 cwpm	80% correct				
Ghana (gr 2)	20 cwpm	45 cwpm	80% correct	35 correct sounds pm	70% correct	80% correct	80% correct
Kenya		65 cwpm					
Liberia		35-40 cwpm					
Papua New Guinea		45 cwpm					
Tanzania (gr 2)	40 cwpm	50 cwpm	80% correct (oral comprehension)		60% correct	80% correct	
Vanuatu (gr 2)		45 cwpm					
Zambia (gr 2)	15 cwpm (emergent) 30 cwpm (fluent)	20 cwpm (emergent) 45 cwpm (fluent)	40% correct (emergent) 80% correct (fluent)		30% correct 60% correct	40% correct 70% correct	

cwpm=correct words per minute



To understand the relevance of these scores in the context of Pakistan, some socioeconomic comparators are shown in Table 2.

Table 3: Socioeconomic data for the countries that provided EGRA/EGMA benchmarks, in comparison to the same metrics for Pakistan.

	Human Development Index (2021)	GDP (USD, Billions)	GDP (PPP), IMF 2023, billions international \$	GDP per capita (PPP), thousands international \$, 2020
Pakistan	0.544	376.5	1568	5.00
Egypt	0.731	476.7	1809	11.99
Ghana	0.632	72.8	227	5.26
Kenya	0.575	113.4	339	4.50
Liberia	0.481	4.0	9.7	1.38
Papua New Guinea	0.558	30.6	41.7	3.74
Tanzania	0.549	75.7	228	2.55
Vanuatu	0.607	1.0	0.98	2.85
Zambia	0.565	29.8	84	3.18

GDP=gross domestic product; PPP=purchasing power parity

Pakistan has established benchmarks for Urdu reading (available <u>here</u>). However, there are no benchmarks (yet) for EGRA or EGMA reading / maths standards. Pakistan has provided mean EGRA scores for the <u>Pakistan Reading Project</u>, which we use in the following analysis.

Insights from Existing Data

Microschools data

Individual data was available for 97 children (54 females, 41 males) in the microschool programme (37 Badin, 52 Karachi, 8 Lahore) on the following subtests:

- Letter Identification



- Reading (percent correct)
- Letter Writing
- Pictorial Identification

Mean values and baseline and endline for these data are shown in Table 3. A paired t-test between baseline and endline scores across individual students showed that improvement in all domains was significant (p<0.001).

Table 4: Mean scores on assessments. All assessments showed significant improvements from baseline to endline.

	Baseline	Endline	Improvement (points)	Effect size (Cohens d)	P-value (paired t-test)
Letter ID	11.9 +/- 21.0	38.3 +/- 31.5	26.4	0.904	<0.001
Reading	8.9 +/- 21.1	28.7 +/- 24.8	19.8	0.569	<0.001
Letter Writing	38.0 +/- 38.0	73.3 +/- 36.7	35.3	0.938	<0.001
Pictorial Intelligence	36.2 +/- 30.6	68.1 +/- 29.4	32.0	1.01	<0.001





Figure 4: Microschool data is shown for Letter ID (percent correct) in four different ways. In all figures, the grey box indicates the Pakistan average for Grade 2 (scores within the grey are lower than average; scores in the white area are above average). The upper left plot shows mean scores at baseline and endline. The upper right plot shows individual children's data as black lines (and still has the orange line indicating mean scores, including error bars). The lower left plot indicates mean scores by city, showing larger increases for children in Karachi than in the other two cities. The lower right plot shows individual children's data by city (and still has the thicker lines indicating mean scores, including error bars).

In-School Data

Individual data were provided for children in Kindergarten, Grade 1, and Grade 2. Children were either part of an intervention group (received TTWF in-school programme) or a control group (received regular instruction, but no intervention). A variety of assessments were administered to each grade (see Table 4) at two measurement points (baseline and endline).

Table 5: Overview of available data for the in-school programme



	Kindergarten	Grade 1	Grade 2
Total # of children	131 (102 girls, 28 boys, 1 NA)	141 (100 girls, 40 boys, 1 NA)	153 (104 girls, 49 boys)
# Control Group	50 (39 girls, 11 boys)	46 (32 girls, 14 boys)	51 (32 girls, 19 boys)
# Intervention Group	81 (63 girls, 17 boys, 1 NA)	95 (68 girls, 26 boys, 1 NA)	102 (72 girls, 30 boys)
Subtests	Letter ID, letter sound ID, reading, pictorial intelligence, general knowledge	Letter ID, letter writing, letter sound ID, reading, letter sound discrimination, reading comprehension, listening comprehension, pictorial intelligence, general knowledge, paragraph reading, spelling	Letter ID, letter writing, letter sound ID, reading, letter sound discrimination, reading comprehension, listening comprehension, pictorial intelligence, paragraph reading, grammar

Overall, it was difficult to draw clear conclusions regarding the impact of the in-school programme. In many subtests, the intervention and control group did not have the same performance level at baseline, making it difficult to compare the change over time of the two groups. Also, in many cases both groups showed similar changes and therefore this change cannot be attributed to the intervention programme. Based on the insights from the existing data, some takeaways and recommendations are presented in the following.





Figure 5: In-school data is shown for Letter ID (percent correct) of the first grade in four different ways. In all figures, the grey box indicates the Pakistan average for Grade 2 (scores within the grey are lower than average; scores in the white area are above average). The upper left plot shows mean scores at baseline and endline for the intervention group. The upper right plot shows individual children's data as black lines (and still has the orange line indicating mean scores, including error bars), again for the intervention group. The lower left plot indicates mean scores by treatment (intervention vs. control). The lower right plot shows individual children's data still has the thicker lines indicating mean scores, including error bars).

Note: All plots were created in R. More explanations and example scripts are provided in the appendix - <u>Brief Introduction to R</u> and <u>R Scripts Used for Graphics</u>.



Takeaways and Recommendations

Benchmarks

As mentioned above, benchmarks provide context within which scores (and improvements over time) can be more easily interpreted. Benchmarks are given in Table 1 for a variety of countries. Some domains (e.g., Reading Comprehension) show very good alignment across countries; other domains do not. We recommend using some sort of benchmark to interpret your change scores, whether that be averaging across other countries, choosing the most socio-demographically similar country, or using the Pakistani average scores. This will help provide more concrete details on your learner outcomes (e.g., "only 10% of students were fluent readers at baseline, but after our microschools programme, 60% of students were fluent readers").

Domains to assess

There are many possible domains to assess (e.g., reading fluency, letter identification, addition, subtraction, etc.). The more domains assessed, the more data you have, but also the more time consuming the assessments. To save time and personnel power, we would recommend focusing on the most salient domains to assess, and ensuring that they line up with available benchmarks. For example, reading fluency is very commonly measured and has the most extensive set of countries with fluency benchmarks. Note that it is oftentimes reported as correct words per minute (cwpm).

The value of longitudinal study design

Longitudinal designs (i.e., measuring the same children multiple times) are more difficult than cross-sectional studies (i.e., measuring different children at each time point) because you have to measure the *same* children at each time point. Attrition can be a problem, and it is more time consuming to ensure the same children are measured at each time point. However, longitudinal designs are much more powerful for examining the effects of an intervention.

Longitudinal designs control for individual differences. If different sets of children are measured at each time point, it is possible, for example, that the random sample of children measured at endline just happened to be better readers than the random sample measured at baseline. In this way, you could not tell whether differences were due to the intervention, or to random differences between groups. Longitudinal studies control for this, and by doing so reduce variability in the measures.

Longitudinal studies have better sensitivity to change. Because longitudinal studies inherently control for individual differences, they are much more sensitive to change. They can



pick up more subtle changes over time, and/or require a much smaller group of children to detect a significant difference between baseline and endline.

Longitudinal studies require fewer participants. Because of their better sensitivity, longitudinal studies require fewer participants to detect change. For example, from the microschool data presented above, there is an effect size of Cohen's d=0.569 for Reading (percent correct) and d=0.904 for Letter ID. At p<0.05 and 95% power (95% chance of detecting a change if it is present), you need to measure 35 participants to detect changes in Reading (70 assessments total) and only 15 to detect changes in Letter ID (30 assessments total) using a paired t-test and longitudinal data (see Table 5).

However, if you are using a two-sample t-test (cross-sectional design with different students at baseline and endline), you need 68 students at each time point (136 assessments total) for Reading and 28 students at each time point (56 assessments total) for Letter ID to detect the same effects. Therefore, you need more than twice as many assessments in a cross sectional design compared to longitudinal to detect the same effects.

	Longitudinal design (pre-post measures in the same individuals)	Cross-sectional design (pre, post measures in different groups)
Reading (d=0.569)	35 participants, 2 assessments each	68 students at baseline 68 students at endline
Letter ID (d=0.904)	15 students, 2 assessments each	28 students at baseline 28 students at endline

Table 6: Number of participants required for each type of study design

Choosing a control group

For the in-school programme, schools chose which classrooms received the intervention and which classrooms acted as controls. For an experiment to appropriately test the effectiveness of an intervention, the intervention and control groups must be similar on all relevant characteristics. Ideally, students (or classes) are randomised at the beginning of the study to either intervention or control, ensuring that there is no bias in the selection. In this case, this was not possible. Instead, schools chose which classes would be part of each group. This introduces bias and means that the control group is not necessarily a good reference from which to evaluate the effects of the intervention. For example, schools may have selected classes that would be "easier" to implement the intervention, or they may have picked children who needed it more. Therefore, we advise caution while interpreting this data.



Feasibility of Automated Assessments

For assessments to be useful, they must be valid, reliable, feasible to implement, and useful (i.e., measuring something you want to know about).

EGRA and EGMA have already been established to be **reliable and valid** tools for assessing <u>early reading</u> and <u>mathematics</u> skills. They are also **useful**, since they are measuring relevant reading and mathematics skills that TTWF is trying to improve with their programmes.

However, **feasibility** has been more challenging with some of the EGRA and EGMA assessments, as the assessments require juggling multiple things (e.g., timer, tablet, etc) and facilitators have limited training. Furthermore, these assessments require facilitators to be one-on-one with children, meaning that assessing large groups of children is time consuming. These logistical constraints are common to any in-person assessment.

Therefore, automated (digital) self-assessment tools offer an attractive alternative, given that they do not require one-on-one facilitation. Compared to facilitators juggling stopwatches and tablets, they may also offer a more objective way of assessing progress, as they do not require a facilitator to record timing or evaluate correctness of responses. However, digital self-assessments can be limited in their scope (i.e., they cannot measure the full range of skills relevant to reading), and there is high potential for user error (or manipulation). It is important to ensure that any digital self-assessments are **reliable, valid, and useful**, in addition to their feasibility.

There is some data on the utility of using automated assessments of children's reading that shows promise. For an article about automatic assessment of oral reading fluency, see <u>Bolaños</u> <u>et al., 2013</u>. Their programme showed very good agreement with scores from human raters for reading fluency in children from Grade 1-4 in Colorado, USA. A different assessment also showed good agreement in terms of identifying Dutch children with reading difficulties in Grade 2 (summary <u>here</u>; paper <u>here</u>).

To determine if automated assessments are useful in a TTWF context, it could be useful to compare them to the facilitated assessments using EGRA/EGMA. Strong correlations would indicate good agreement between the two assessment methods. However, a normal distribution of scores is important. Further, it can be difficult to assess correspondence when there are ceiling or floor effects (i.e., lots of children scoring at the very bottom or very top of the range). Finally, it should be ensured at the beginning that the comparisons are between assessments trying to measure the same thing (e.g., reading fluency). The more the two assessments differ, the less likely it is to encounter high correlations.



Deliverable 3: A comprehensive and inclusive approach to assessing impact and ensuring sustainability

What further data collection possibilities (concerning the learner and beyond) will be required to help respond and be compelling to the needs of the different stakeholders?

Introduction

Solving global challenges is a shared responsibility, necessitating a collective effort from all stakeholders dedicated to enhancing the well-being of communities.

Organisations often invest in rigorous monitoring and evaluation methods, including baseline, midline, and endline surveys, to gauge the impact of their interventions. However, it is equally crucial to gather insights directly from the communities where programme beneficiaries reside.

Teach The World Foundation (TTWF), in its pursuit to enhance foundational skills for children in Pakistani communities through its microschool programme, recognises the importance of community feedback. Beyond quantitative data, the foundation acknowledges the significance of qualitative information obtained from parents, community leaders, religious figures, partner organisations, and government representatives. This community-centric approach not only validates the intervention's impact on individuals but also fosters ownership within the community, promoting sustainability.

Engaging with the community is a linchpin in this process, providing a comprehensive understanding of the cultural, social, and political dynamics at play. Such engagement establishes trust, allows for adaptive programming based on real-time feedback, and ensures that the intervention aligns with the broader development goals of the community. In essence, Teach The World Foundation is committed to a holistic evaluation strategy that reflects the true, multifaceted impact of its microschool programme for the betterment of children and communities in Pakistan.



Community Engagement

Engaging with the community is indeed vital for several reasons:

Holistic Understanding:

The community has a holistic understanding of the context in which the intervention is taking place. They can provide insights into various aspects of the beneficiaries' lives that may not be captured through traditional monitoring and evaluation methods.

Cultural Sensitivity:

Community members can offer insights into cultural nuances that may impact the success or failure of the intervention. This understanding is crucial for tailoring programmes to the specific needs and values of the community.

Long-term Sustainability:

Involving the community in the process creates a sense of ownership. When people feel a sense of ownership, they are more likely to actively participate in, and sustain the changes brought about by the intervention.

Feedback Loop:

Continuous feedback from the community allows for adaptive programming. If there are aspects of the intervention that are not working as intended or if there are unanticipated challenges, the community can provide valuable feedback for mid-course corrections.

Building Trust:

Engaging the community builds trust between the implementing organisation and the beneficiaries. Trust is essential for the success of any intervention, and it takes time to establish. Regular communication and involvement in decision-making processes help in building and maintaining trust.

Social and Political Dynamics:



Communities are embedded in wider social and political structures. Understanding these dynamics is critical for ensuring that the intervention aligns with broader development goals and policies.

Measuring Real-World Impact:

While quantitative data from surveys is important, qualitative data from the community provides a richer understanding of the real-world impact of the intervention on individuals and the community as a whole.

For Teach The World Foundation's microschool programme, involving parents, community leaders, religious leaders, other partner organisations, and government officials in the assessment process would provide a more comprehensive and accurate picture of the programme's impact.

This collaborative approach not only enhances the credibility of the evaluation but also contributes to the long-term success and sustainability of the intervention.

Takeaways and Recommendations

Approaches, Timeline and Tools

We have two categories of approach, Reactive and Proactive approaches.

A. The Reactive Approach

Involves creating avenues where clients / community can share feedback and concerns, interests, priorities, urgencies with ease.

These avenues may include:

- 1. Open door policy, / walk ins
- 2. Suggestion boxes
- 3. Help desks



B. Proactive Approaches

TTWF can take the initiative to identify issues of concern, interest, priority needs through the following activities. The approaches can be used at different phases as well as integrated in the organisational pre-planned activities as described below.

Community Data Collection Checkpoint 1:

After the community identification stage, the field team can conduct small corner meetings with individuals from the community, which can come off as informal focus group discussions for interviews.

Community Data Collection Checkpoint 2:

At the recruitment camps that take place before the official launch of the Microschool, TTWF can use this opportunity to conduct a baseline survey. This will serve an important role in establishing the actual level at which the prospective learners are in foundational literacy competencies, as well as collect data from the parents about the learners' backgrounds, what has hindered their enrollment into formal school, and any other relevant information.

Community Data Collection Checkpoint 3:

While conducting formative assessments in the schools, at the point where parents are called to witness their children receiving their performance reports, this is an opportunity that can be utilised to gather information from parents on what changes they are observing in learners at home. As an interaction that happens at regular intervals throughout the year, with consistent parent and student representation, it will form a potential longitudinal foundational practice in the data collection practice.



Table 7: Approaches to the collection of data in the community

Suggested community data collection point	Approach	Stakeholder	Resource
Community Data Collection Checkpoint 1	Focus group discussions	1. Parents	Guiding Questions
	1 on 1 interviews	 Local/community leaders Religious leaders 	Guiding Questions
Community Data Collection Checkpoint 2	Home visits or school visits	1. Parents	Question Rubric
	Facility / Office visits by TTWF STAFF	 Religious leaders Government officials 	Question Rubric
Community Data Collection Checkpoint 3	Conferences / Webinars/Seminars, all school events search Graduation or passing out events.	1. NGOs	These are events organised by TTWF, they could be seminars or webinars or an event to graduate all learners who have successfully completed the programme. A printed certificate indicating what the learners have achieved through the programme would be a great incentive.
	Graduation / passing out event	All the above can be invited	

These approaches above have been informed by a <u>survey</u> filled in by 11 respondents TTWF and rated these approaches to be applicable in the programme as organised in the table above. Each approach corresponds with a target group of people and the survey and rubric questions are designed based on the type of feedback expected from each category of people.



Bibliography / References

Tanzi, Vito, and Ke-young Chu, eds., 1998, Income Distribution and High-Quality growth

Mussida, Sciulli, and Signorelli, 2018, Secondary school dropout and work outcomes in ten developing countries

Psacharopoulos, Patrinos, 2002, Returns to Investment in Education A Further Update

Note: References considered within Deliverable 1 are contained within the <u>Societal Impact</u> <u>Workbook</u>



Appendix

Brief Introduction to R

R is an open-source programming language, which can be used for statistical computations and graphics. The software environment developed as the "R project" runs on a variety of systems (see here). Apart from R itself, it is possible to download R Studio, which provides a comfortable user interface (first steps see here). To run analyses or generate graphics in R, you need to write an R script. By running the script, you will receive graphics or numerical results of your analyses, dependent on the specific script. The second ingredient you need is your data. R can import many different formats, including .csv and .xlsx files.

R Scripts Used for Graphics

The R scripts that were used to generate the graphics (in R called plots) presented in this report can be found <u>here (Microschool)</u> and <u>here (InSchool)</u>. These scripts serve as examples to create further scripts for other tests (e.g., Mathematics instead of English) or grades (second graders instead of first graders). They start with some general preparations for your R session, followed by importing the data file, preparing the data for plotting, and creating the plots. The plots are created through for-loops because each data set contains multiple outcome variables that might be of interest. In total four different types of plots are created, including one-group versus multiple-group plots, and providing only lines for mean scores versus including individual children's lines (in line with the examples provided in deliverable 2). For these four types separate for-loops are provided.