A ROADMAP FOR MEASURING DISTANCE LEARNING

A REVIEW OF EVIDENCE AND EMERGING PRACTICES

January 2021

This publication was produced for review by the United States Agency for International Development (USAID). It was prepared by Emily Morris, Anna Farrell, Emma Venetis, EnCompass LLC, and MSI, a Tetra Tech company for the DEEP (Contract No. GS-10F-0245M). The views expressed in this publication do not necessarily reflect the views of USAID or the United States Government.

CONTENTS

ACKNOWLEDGEMENTS

Thank you to the staff of the Center for Education in USAID's Bureau for Development, Democracy, and Innovation for their guidance and support in developing this review. Specific contributors included: Dr. Saima S. Malik, Rebecca Rhodes, and Dr. Elena Walls. Reviewers included: Dr. Nina Papadopoulos, Robert Burch, Anjuli Shivshanker, Olga Merchan, Kayla Boisvert, Leah Maxson, and Beth Johnson.

We also thank all those who participated in the key informant interviews that provided valuable insights and examples to inform this review. Key informant participants included:

Anthony Udeh (EDC/Northern Nigeria's Education Initiative Plus); Bjoern Hassler (EdTech Hub); Carmen Strigel (RTI); Chris McBurnie (EdTech Hub); Christine Beggs (Room to Read); Daniel Lavan (EDC); Devika Rai (People in Need, Nepal); Elisheva Cohen (Indiana University Bloomington); Gaëlle Simon (Management Systems International); Heather Simpson (Room to Read); Hetal Thrukal (researcher); Laura Harrington (Chemonics DRC); Lisa Hartenberger (EDC); Mark LeClair (Farm Radio International); Molly Farrell (Worcester Public Schools, United States); Moitshepi Matsheng (Young 1ove, Botswana); Nikhit D'Sa (University of Notre Dame); Nisha Ligon (Ubongo Kids); Noam Angrist (Young 1ove, Oxford University); Rachel Heavner (Worldreader); Scott Isbrandt (mEducation Alliance); and Wayan Vota (ICTWorks Publisher).

Thank you to Mary Burns for her guidance, review, and input.

Thank you to Amy Mulcahy-Dunn for her review and Natalya Barker for her editorial support. Thank you to Dexis Consulting Group-led Education Support Initiative, as well as Kirsten Bub, A'ame Joslin, and Karla Giuliano Sarr for their collaboration across this series of USAID/DEEP reports.

Thank you to Denisse Leatxe (Txiki Txoko) for graphic development. All icon graphics by Txiki Txoko, with some modified icons from The Noun Project.

Thank you to the Inter-agency Network for Education in Emergencies (INEE) and to the Global Reading Network (GRN) for organizing informative panels and conversations about monitoring and evaluating distance learning. These conversations also helped to enrich this review.

ACRONYMS AND ABBREVIATIONS

ASER	Annual Status of Education Report
CATI	Computer-Assisted Telephone Interviewing
EDC	Education Development Center, Inc.
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Mathematics Assessment
FRI	Farm Radio International
GAGE	Gender and Adolescence Global Evidence
IAI	Interactive Audio Instruction
ICT	Information and Communications Technology
INEE	Inter-agency Network for Education in Emergencies
IRI	Interactive Radio Instruction
IVR	Interactive Voice Response
LMS	Learning Management System
MOOC	Massive Open Online Course
M&E	Monitoring and Evaluation
MEL	Monitoring, Evaluation, and Learning
MMS	Multimedia Message Service
NGO	Non-Governmental Organization
OER	Open Educational Resources
SES	Socioeconomic Status
SRGBV	School-Related Gender-Based Violence
SMS	Short Message Service
TV	Television
UDL	Universal Design for Learning
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

As countries around the world have closed learning institutions in response to the COVID-19 pandemic, teaching and learning have pivoted from in-person instruction to distance learning. Simultaneously, there has been a surge in efforts to promote access to distance learning programming. Distance learning is commonly used to reach learners who need flexible learning opportunities, as well as to reach groups excluded from formal education institutions.[i] It can serve as the main form of instruction or can complement or supplement in-person learning. As countries and education agencies take up distance learning, it is important to design and implement evidence-based strategies for monitoring and evaluation to measure whether distance learning efforts are serving the intended communities and achieving intended objectives.

The purpose of this review is to support education practitioners, host country government representatives, donors, implementers, non-governmental organizations (NGOs), civil society organizations, and other stakeholders in applying best practices to monitor and evaluate distance learning initiatives designed for diverse learners and implemented both within and outside of learning institutions. This review covers the four key distance learning modalities: radio/audio, television/video, mobile phone, and online learning. Printed texts, which are often developed to accompany these first four modalities, can also be a fifth modality in contexts where technology is not used.

Most of the data sources were drawn from work in the primary education sub-sector. However, much of the guidance can be applied to secondary and tertiary-level distance learning. This review is also applicable to data collection in both crisis and non-crisis contexts.

This review presents a roadmap that guides users through four steps of planning and designing how distance learning delivered through any of these modalities can be monitored and evaluated.

Step 1: Determine the Objectives of Monitoring and Evaluating Distance Learning
Step 2: Determine What Will Be Measured (Reach, Engagement, and Outcomes)
Step 3: Determine How Data Will Be Collected (In-Person or Remotely)
Step 4: Determine the Methods and Approaches for Measurement

Based on emerging global evidence, this review guides users through the process of measuring the reach, engagement, and outcomes of distance learning initiatives.

In addition to providing step-by-step guidance, this review provides three overarching recommendations for developing and implementing evidence-based monitoring, evaluation, and learning (MEL) plans for distance learning initiatives. The recommendations are to:

---

[i] See Morris, E. and Farrell, A. (2020) Delivering Distance Learning in Emergencies: A Review of Evidence and Best Practices. Nontraditional learners include working adults, educators, or homeschooled children and youth (Burns 2011). Historically excluded communities include members of ethnic, indigenous, and linguistic minority groups; women; people with disabilities; communities living in remote areas and/or poverty: and communities in crisis and conflict settings.

1. **Integrate in-person and remote approaches, use multi-modal interfaces, and employ mixed methods to measure distance learning.** Integrating in-person and remote data collection (e.g., in-person testing and remote interviews), using multi-modal interfaces (e.g., phone calls and SMS surveys), and mixed-methods approaches (e.g., interviews, surveys, and photographs) helps promote greater participation and leads to more accurate results.
2. **Encourage innovative solutions to measure reach, engagement, and outcomes during a quick pivot to distance learning, while also developing high-quality MEL strategies for the longer term.** The guidance in this review helps teams think about short-term MEL needs while working toward longer-term strategies for assessing the effectiveness of distance learning.
3. **Design equitable monitoring and evaluation approaches and conduct systematic equity analyses of distance learning initiatives.** Evaluative approaches to distance learning must attempt to measure and analyze whether marginalized individuals and groups are being systematically included or excluded through distance learning programming as well as during in-person and remote data collection.[ii]

This review complements other knowledge products and practical guides produced by USAID to support high-quality distance learning program development and delivery. A complete listing of these documents may be found on USAID's COVID-19 resources page on Education Links. These include: *Delivering Distance Learning in Emergencies: A Review of Evidence and Best Practice*, *Return to Learning During Crises: Decision-making and Planning Tools Toolkit,* and *USAID's Guidance Note on Measuring Reach for Distance Learning*. It is also a key component of the upcoming USAID Toolkit for Designing and Planning a Comprehensive Distance Learning Strategy that provides guidance and support to governments, USAID Missions, partners, and other education stakeholders in the design, implementation, and Monitoring, Evaluation and Learning (MEL) processes. As with all of USAID's resources, this review is not specific to any one context, event, or distance learning initiative. It is meant to be useful to users both in responding to short-term learning needs during the COVID-19 pandemic and long-term distance learning strategies.

[ii] Individuals and groups may be marginalized because of their disability; gender identity; sexual orientation; minority ethnic, racial, indigenous, language, or religious group identity; socioeconomic status or caste; residency in rural or high-density areas; residency in a crisis or conflict zone; health status; age; and/or lack of social, economic, and/or political protections.

INTRODUCTION

Measuring distance learning requires cooperation across various partners, including learners, educators, caregivers, implementers, technology providers, and donors, to name a few. Whether distance learning programming[1] is designed for preschoolers, primary school children, out-of-school youth, educators, or beyond,[2] it is essential to determine from the beginning of the design phase how the programming will be monitored and evaluated. Tracking and assessing reach, engagement, and outcomes is critical for ensuring that educational initiatives are effective and inclusive.[3] In this review, reach, engagement, and outcomes are defined as:

- **Reach** captures access to technology (devices and software); infrastructure for technology (electricity, internet, or phone connectivity); and distance learning programming and content.
- **Engagement** assesses the extent to which users participate as intended in the programming, including the degree to which users perceive the content to be relevant, captivating, and of high quality.
- **Outcomes**[4] measure changes in learning of content knowledge as well as social and emotional learning (SEL).

Text Box 1

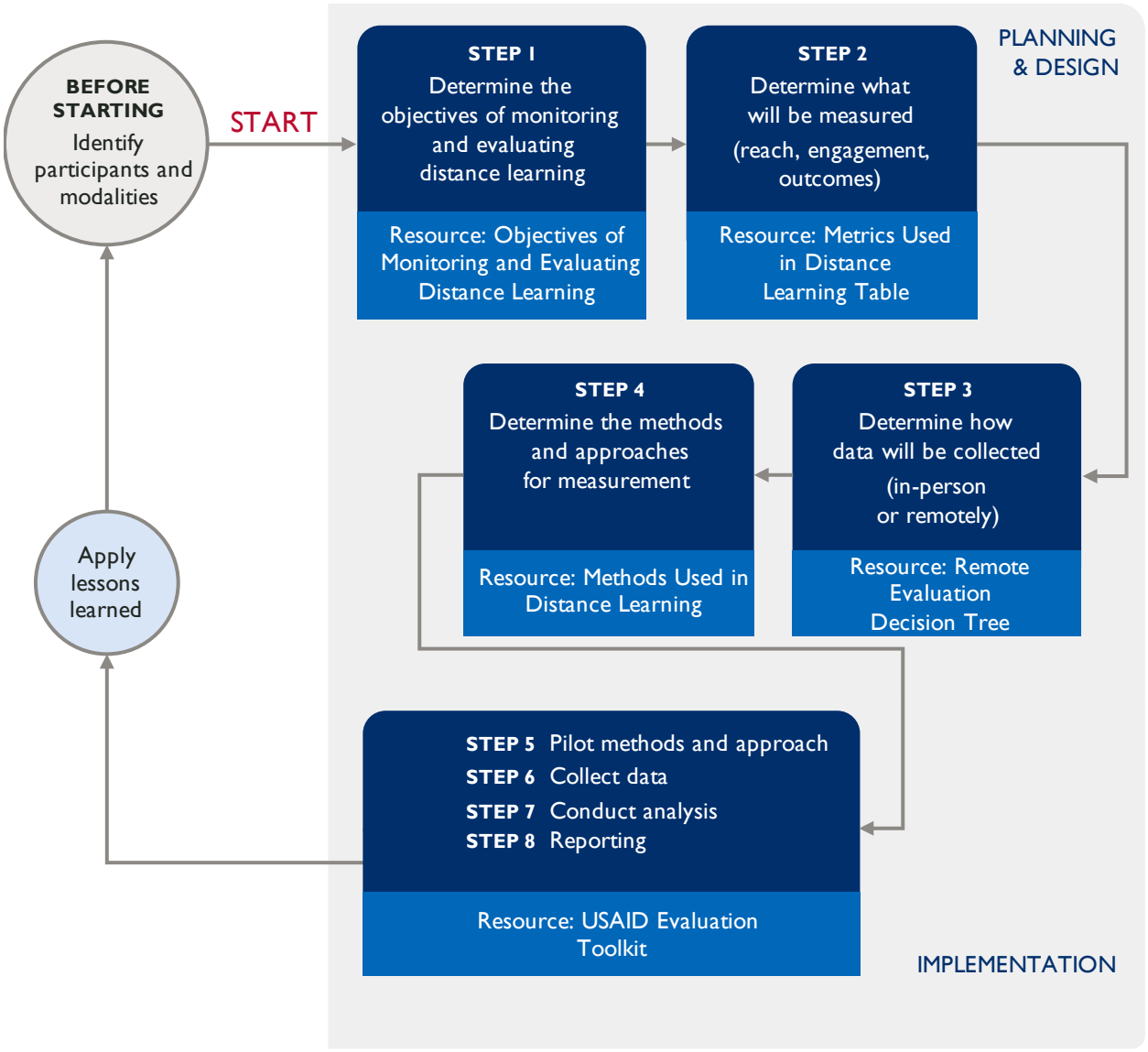
USAID focuses on measuring the percent of learners regularly participating in distance learning programming funded with USG assistance (See USAID indicator SUPP-17). The indicator offers a clear definition of a distance learning program and participation in such a program. USAID defines participation as attending 60 percent or more of the program sessions. USAID recommends this indicator for assessing reach and engagement of learners in USG-funded distance learning programming. Additionally, custom indicators that measure reach, engagement, and outcomes can be used.

This review employed a qualitative methodology drawing on document analysis and key informant interviews with 23 ICT and/or distance learning experts and practitioners. Documents included: peer-reviewed journal articles, evaluation reports, and implementer documents. Key informant interviews were conducted to supplement existing literature. During this review process, the Inter-agency Network for Education in Emergencies (INEE) and the Basic Education Coalition/Global Reading Network hosted virtual panels that also informed the recommendations. There is extensive evidence on how to conduct in-person evaluations, and, as such, this review concentrates on promising remote data collection approaches while still advocating for in-person data collection when feasible. Insights about best practices that emerged from these documents, interviews, and panels are presented within each of the steps and informed the design of all tools and guidance. Full case studies are included in Annex D. This review also follows the principles of USAID's *Universal Design for Learning (UDL) Toolkit* and USAID's *Collaborating, Learning, and Adapting (CLA)* approach to MEL. It also uses USAID's *Guidance for Monitoring, Evaluation, and Learning During the COVID-19 Pandemic* and is aligned with the distance learning performance indicator outlined in Text Box 1.

The roadmap in Figure 1 lays out eight essential steps to plan, design, and implement monitoring and evaluation (M&E) of distance learning activities as well as relevant USAID resources that can support

each of the steps. This review focuses on the first four planning steps. A detailed explanation of each of these steps is presented in the sections below.

Figure 1: Roadmap for measuring distance learning



For ease of reference, the table below provides a summary of the guiding questions and main recommendations associated with each step detailed in the roadmap. The chapters then elaborate on these steps.

STEP 1. DETERMINE THE OBJECTIVES OF MONITORING AND EVALUATING DISTANCE LEARNING

Guiding Questions	Main Recommendations
• Why is distance learning being measured? • How will the data be used? • Who is the audience for the evaluation?	1. Ensure the reason distance learning is being measured matches what is being measured. 2. Be clear how the data will be used and who the audience is for the evaluation. 3. Summative assessment approaches need to follow established best practices in evaluation.

STEP 2. DETERMINE WHAT WILL BE MEASURED (REACH, ENGAGEMENT AND OUTCOMES)

Guiding Questions	Main Recommendations
• How are reach, engagement, and outcomes measured in distance learning programs? • What are some examples of these measures? • How can teams build these measures systematically into MEL designs? • What kinds of equity analyses should be considered?	1. Where feasible, coverage should be measured alongside listenership and viewership. 2. Where feasible and ethical, reach of educational apps and online learning management systems should be tracked automatically. 3. Identify who is being reached and who is not being reached. 4. Use additional metrics to assess engagement and completion. 5. Ensure there is a process for formative evaluation in place. 6. Include perspectives of the most marginalized. 7. Use in-person assessments for measuring outcomes, when feasible. 8. Emphasize low-stakes formative assessments to inform teaching and learning. 9. Check the emotional wellbeing of learners before conducting assessments.

STEP 3. DETERMINE HOW DATA WILL BE COLLECTED (IN PERSON OR REMOTELY)

Guiding Questions	Main Recommendations
• Should data be collected in person or remotely? • What key considerations-safety of teams, access to technology, infrastructure, feasibility of capturing reach, engagement, and outcomes-should be considered? • What equity considerations should be taken into account (e.g., geographical reach, socioeconomic status, gender, disability?) • What technologies should be used (e.g., paper, mobile phone, tablet, computer)?	1. Use an integrated (in-person and remote) approach to data collection. 2. Collect mobile phone numbers of caregivers and families at the beginning of an intervention or school year. 3. Create MEL platforms for basic phones or support families in acquiring smartphones to ensure marginalized individuals and households can be included in distance learning and data collection activities. 4. Assume that others are listening in during remote data collection.

STEP 4. DETERMINE THE METHODS AND APPROACHES FOR MEASUREMENT (QUANTITATIVE AND QUALITATIVE)

Guiding Questions	Main Recommendations
• What quantitative and qualitative methods can be used to measure distance learning? • What technologies (e.g., paper, mobile phone, tablet, computer) and interfaces (e.g., SMS, survey software) are used to gather data • What sampling strategies (e.g., census, representative, purposive) can be used? • What kinds of equity analyses should be considered? • What is the strength of the evidence for these evaluative approaches and where is there a need for more evidence?	1. Use mixed methods to collect data. 2. Match the evaluation purpose to who is collecting the data. 3. Treat language as a right and a resource. 4. Acknowledge that even simple assessments may feel like a big deal to a learner or caregiver. 5. Ensure data collection efforts are not further marginalizing participants. 6. Give the data back to the participants. 7. Plan early and plan ethically.

THE ROADMAP FOR MEASURING DISTANCE LEARNING

STEP 1: DETERMINE THE OBJECTIVES OF MONITORING AND EVALUATING DISTANCE LEARNING

Text Box 2

Step 1 Guiding Questions

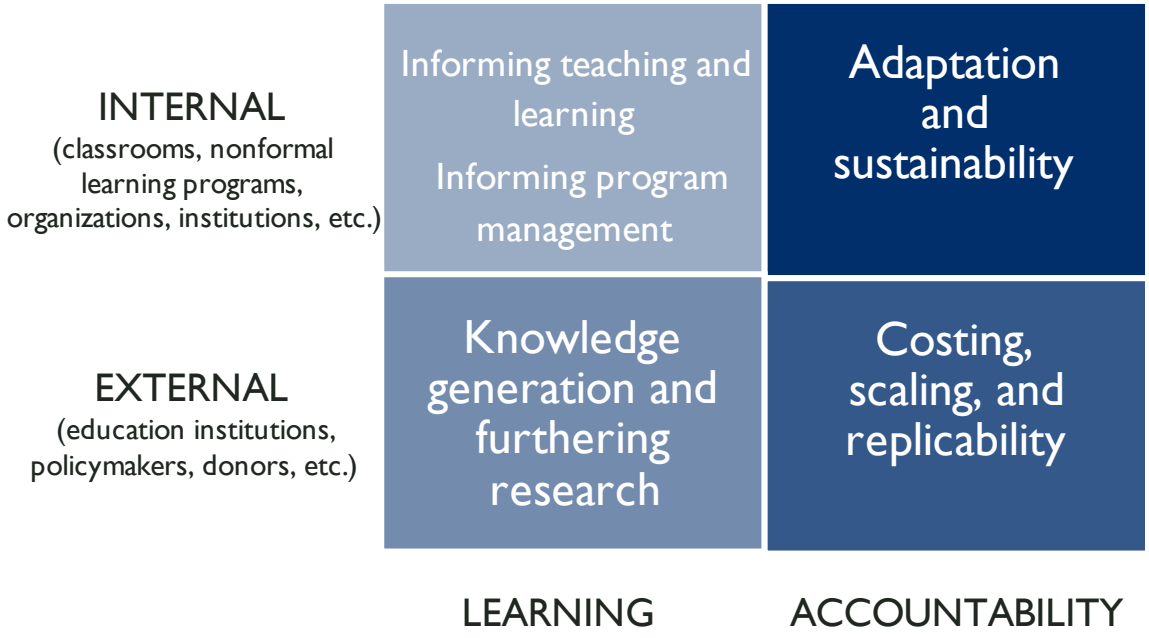
- Why is distance learning being measured?
- How will the data be used?
- Who is the audience for the evaluation?

It is critical to define clearly *why* distance learning is being measured and *how* the data will be used before starting the process of designing MEL activities. Evaluations can be used for many purposes and generally fit into two main categories: formative and summative. Formative assessments and evaluations in distance learning examine how programming is being received and used (e.g., users are able to access and use as intended), how content and programming can be improved (e.g., gauging radio listener feedback and recall before broadcasting occurs), and/or how to inform teaching and learning internally (e.g., learner check-ins and low-stakes measures). Formative and internal monitoring also helps teams plan for the sustainability of programming and make adjustments as necessary during implementation.

Summative evaluations examine the effects of distance learning on the target learners, educators, or other participants. Examples of summative evaluations include literacy assessments (e.g., Early Grade Reading Assessment [EGRA] or Annual Status of Education Reports [ASER]), numeracy assessments (e.g., Early Grade Mathematics Assessment [EGMA] or ASER), and high-stakes exams at the end of grade levels or learning cycles. Summative data can be used for learning and accountability as well as informing scaling and replication (see Annex B for the different kinds of formative and summative evaluations).

Figure 2 highlights **learning** and **accountability** as the two key aims for measuring distance learning. Within these aims, there are internal and external objectives that should be considered when creating a MEL plan for distance learning to ensure appropriate data is collected. Internally (within programs, classrooms, learning institutions, etc.) data can be used to inform program content and management (learning) or to guide adaptation and sustainability of programming (accountability). Externally, MEL can be used to generate knowledge and further the evidence base (learning) and inform costing, scaling, and replicability at different levels (e.g., district, regions, national) (accountability).

Figure 2: Aims and objectives of monitoring and evaluating distance learning



Adapted from Hempel and Fiala, 2012.[5]

Text Box 3

Recommendations for Determining the Objectives of Monitoring and Evaluating Distance Learning

1. **Ensure the reason distance learning is being measured matches what is being measured.** For example, if the objective is to inform teaching and learning (formative), then measures should be formative (e.g., simple knowledge check-ins) and not high-stakes summative tests (e.g., validated math exams).
2. **Be clear how the data will be used and who the audience is for the evaluation.** If the main audience is educators, then the data need to be useful and understandable to educators. For example, aggregate numeracy data for a whole intervention of learners will not help an educator understand where their specific learners are during school closures—this will require a formative assessment of learners in a particular classroom.
3. **Summative assessment approaches need to follow established best practices in evaluation.** For example, if using USG funds, make sure to follow USAID's CLA guidelines and guidance, whether assessments are conducted in person or remotely.

STEP 2: DETERMINE WHAT WILL BE MEASURED (REACH, ENGAGEMENT, AND OUTCOMES)

Text Box 4

Step 2 Guiding Questions

- How are reach, engagement, and outcomes measured in distance learning programs?
- What are some examples of these measures?
- How can teams build these measures systematically into MEL designs?
- What kinds of equity analyses should be considered?

Distance learning measures can be grouped under three main domains: reach, engagement, and outcomes. Within each of these domains, there are a number of quantitative and qualitative metrics that can be used to capture essential data. Implementing teams should discuss and determine early on which metrics are critical to meeting the monitoring and evaluation activities' objectives. However, MEL designs should include measures from all three domains, and be sequenced logically. For example, outcome measures cannot be determined without clear reach and engagement data. Performance indicators and targets should be developed simultaneously per USAID MEL guidance.[6] Figure 3 outlines the main questions that each of these three domains addresses.

Figure 3: Distance learning domains of measurement

Table 1 shows illustrative metrics critical to planning and developing distance learning programming and materials.[7] These metrics have been grouped into the three main domains (reach, engagement, and outcomes) with outcomes divided into two types: a) content knowledge and learning and b) social and emotional learning (SEL).[8]

Table 1: Metrics for distance learning program planning and development, by domain

	REACH	ENGAGEMENT	OUTCOMES	
			CONTENT KNOWLEDGE LEARNING	SOCIAL AND EMOTIONAL LEARNING
Infrastructure	Infrastructure and connectivity (internet and phone data) coverage			
Technology & Accessibility	Access to technology devices (hardware) and software; access to assistive technology	Utilization of technology devices and software as intended; access to assistive technology	Change in technological literacy	
Programming	Basic listenership, viewership, and usership by target audiences (frequency and duration of use)	Participation in programming as intended and completion	Change in subject matter, content knowledge, and skills acquisition and retention	Change in social and emotional and soft skills, attitudes, and beliefs Change in behaviors
Accompanying Materials	Access to accompanying materials	Utilization of accompanying materials as intended		
Cost	Unit cost of reaching learners	Unit cost of engaging learners	Cost of improving outcomes, in dollars per unit of measure of an outcome	

Table 2 provides illustrative metrics for measuring the quality of distance learning programming and materials, which are commonly captured through formative evaluation activities (e.g., gathering user feedback during program piloting).

Table 2: Metrics for measuring quality of distance learning programming and materials through formative evaluation, by domain

	REACH	ENGAGEMENT	OUTCOMES	
			CONTENT KNOWLEDGE LEARNING	SOCIAL AND EMOTIONAL LEARNING
INTENDED AUDIENCE				
Interaction	Opportunities for interaction built into the programming	Extent to which users interact in programming as intended	Change in subject matter, content knowledge, and skills acquisition and retention	Change in social and emotional and soft skills, attitudes, and beliefs Change in behavior
Quality and Relevance	Adherence to curriculum, grade level, scheduling, geographical reach, and other parameters facilitating use	Level of quality; relevance to developmental stage and age, gender, language, etc.		
Response	Number of users who share feedback by demographics	Level of interest in programs; popularity of programming		
UNINTENDED AUDIENCE				
Shadow Audience	Shadow audience access	Shadow audience participation and response	Shadow audience unintended changes in knowledge, skills, attitudes, and behaviors	

Text Box 5

Equity Analyses

When examining equity, it is critical to understand the breadth of reach, specifically considering who is able to access distance learning and who is left behind. Likewise, it is necessary to understand how the depth of engagement, interest in the programming and materials, knowledge and skills acquisition, and the attitude and behavior change differ among different sub-populations of learners.

In order for equity analyses to be conducted, groups historically excluded from and marginalized through distance learning interventions need to be identified in the initial distance learning planning and design phase. At the onset of designing a MEL strategy, demographic data need to be collected from users in sufficient detail in order to identify who falls into these marginalized groups. While schools, learning centers, and educational initiatives may collect demographic data as part of their regular MEL practices, ensuring data on grade level; disability; gender; geographical residence (e.g., urban, rural); socioeconomic status; and language, racial, ethnic, or minority group status is also collected is critical. Where feasible, reach, engagement, and outcomes should then be assessed across these different groups. Furthermore, samples should be analyzed to see who participated and who did not in order to identify individuals or groups that are systematically excluded (sample bias).

Reach metrics capture who is accessing the content and materials relative to the intended or targeted users.[9] This metric can also measure unintended users, or shadow audiences, who utilize the programming even if they are not the originally targeted users (e.g., a younger sibling who watches a program with their older sibling for whom it is designed).

Text Box 6

Measuring shadow audiences in radio and television

While data analytics through educational apps and web-based programming allow for identifying unintended audiences, reach is more difficult to measure in radio and television broadcasts. Most radio and audio programs, when broadcast on public radio stations, do reach a shadow audience.[10] These shadow audiences include caregivers, out-of-school children and youth, and other adults in the community. Open broadcasting of radio and video programs provides educational opportunities to those not participating in a school or nonformal learning program, and can also "demystify" education for caregivers and community members, making learning content and objectives more transparent.[11,12,13]

While GeoPoll and other nation-wide surveys can capture data on all listeners and viewers, there is very little evidence of reach, engagement, and outcomes of shadow audiences in interactive audio instruction program interventions. Most of what has been written is anecdotal, such as that from the Somali Interactive Radio Instruction Program and the Radio Instruction for Strengthening Education in Tanzania, which discovered that their programs were reaching adult learners and marginalized groups, such as internally displaced persons and caregivers, as well as adults who had low education levels.[14,15] The South Sudan Interactive Radio Instruction program conducted a large listenership survey that determined they had a large shadow audience. This survey used multi-stage sampling to survey over 2,000 people from the 14 counties where their programs aired. Through stratified and random sampling, one rural and one urban Payam (Payam is the equivalent of a district) was selected from each county. Within these Payams, villages were randomly selected, and then households were systematically sampled. The data showed that out-of-school respondents made up a substantial portion of the listeners to the program supporting formal education.[16,17]

The metrics for gathering reach data across the four distance learning modalities are largely the same, with some nuances described below. The methods and technologies for measuring reach and engagement in radio/audio and television/video content typically use in-person or mobile phone surveys. Mobile phone apps and online programming generally rely on embedded user analytics.

The key metrics for capturing reach are:

- **Access to technology, programming, and content:** This includes access to technology devices (e.g., phones, computers, tablets), software (e.g., apps and learning management systems (LMSs)), infrastructure (e.g., electricity and connectivity), programming and content (e.g., basic listenership, viewership, downloads, log-ins, and clicks), and distribution of accompanying or stand-alone print materials.
- **Accessibility:** This includes the ability of marginalized groups and young learners who rely on caregivers to access content as well as assistive technology devices for persons with disabilities.[18]

Text Box 7

Recommendations for Measuring Reach

1. **Where feasible, coverage should be measured alongside actual listenership and viewership.** The most robust way to measure actual listenership and viewership is a representative sample of the target population conducted through a combination of mobile phone and in-person household surveys (for those who do not own mobile phones). Ongoing monitoring of technology and materials distributions should also be conducted.
2. **Where feasible and ethical, reach of educational apps and online learning management systems should be tracked automatically using cookies and analytics.** Automatic tracking requires programming analytics into apps and websites in the development phase and eliminates the need for ongoing data entry.
3. **Identify who is being reached and who is not being reached.** This requires accessing initial demographic data of all targeted learners or users. In newly designed distance learning initiatives, resources and time should be allocated for collecting comprehensive data (e.g., in-person or remote household surveys) and ensuring existing demographics (e.g., school data) includes mobile phone contacts.

ACCESS AND ACCESSIBILITY

As shown in Table 1, multiple metrics can be used to measure access to infrastructure for technology and coverage (electricity, internet, or phone connectivity), access to technology devices and software, and access to distance learning programming and content. Additional analyses need to be conducted as to whether programming and content are accessible to marginalized groups (e.g., learners in remote areas; younger learners who rely on their caregivers to access programs). Accessibility analyses should also include tracking whether people with disabilities have access to assistive technology.

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. Access to technology devices, connectivity, programming, and content should be measured throughout an intervention.[19] While there is no simple metric for capturing who has actual access to educational radio and television broadcasts, a number of organizations have created systems to estimate **coverage.** Farm Radio International (FRI) has the most comprehensive attempts to measure radio coverage potential of all studies in this review. FRI uses a customized system that creates a physical map of areas with radio signal coverage using Geographical Information System data and gauges the approximated reach of broadcasts using census data of radio owners in this coverage area. They then create a formula to determine who is actually accessing the program based on coverage and radio ownership (see the Case Study on FRI in Annex D).

In radio and television, **listenership and viewership** are typically measured through surveys conducted in-person or via mobile phone with household representatives. Ubongo television programming measures viewership by contracting nationally representative studies that collect in-person and remote data using household surveys (see the Case Study on Ubongo in Annex D). Mobile phone surveys are either conducted with all registered mobile phone users or a sample of participants from an educational initiative. Nielsen Audio Surveys and GeoPoll Surveys are two well-known remote data

collection companies that collect daily audience measurement data for radio and television across all registered mobile phone users (see the Case Study on GeoPoll in Annex D). Drawing on registered mobile users has equity considerations, as mobile ownership is often linked to gender, socioeconomic status, age, and other factors.[20] While GeoPoll and Nielsen have a comprehensive approach to measuring reach and engagement of radio or television users, they are also costly and out of reach for smaller education organizations and initiatives.

The majority of radio and television studies included in this review gather data on **access to technology, programming, and content** through surveys conducted with either all or a sample of their participants. For distance learning projects, data collectors use forms via paper, tablets, mobile phones, or computers to track who receives technology devices (e.g., radios or tablets), who receives accompanying materials (if developed), who regularly listens to or views programming (e.g., connectivity and frequency of listenership or viewership), and how frequently users access programming.

When listenership or viewership takes place in groups (e.g., interactive audio instruction listener groups), tracking whether the group is able to access the broadcast or program, in addition to tracking individual users' participation (attendance, attrition, and whether or not technology was accessible) is critical. In the Read Haiti/Strong Beginnings and the DRC ACCELERE! I projects, data collectors called a sample of potential listeners to ask them reach, engagement, and simple knowledge outcome questions. (These methods will be discussed in detail under Step 4 below. Both case studies are also included in the Case Studies in Annex D).

MOBILE PHONE AND ONLINE PROGRAMMING. Backend tracking software is the most common and effective way to obtain basic user information on reach and engagement when audio and video content is pre-loaded onto a device such as a mobile phone, tablet, or computer, or when content is accessed through a podcast, video, educational app, or learning management system (LMS) like Canvas or Blackboard.[21] This tracking occurs when users **log into** an app or LMS, **click** on links, and **download** content from apps and websites. These actions are tracked using cookies and logged in databases that programs can analyze. For example, Worldreader automatically aggregates approximately one million lines of data on reach and engagement per day from both their app and online library, including users' location, book views, book completions, time on page, and overall time spent reading (see the Case Study on Worldreader in Annex D). Interactive Voice Response (IVR) systems, such as the one used by FRI across sub-Saharan Africa, also automatically track reach by recording callers' phone numbers to a database. Ubongo, as detailed in Annex D, tracks data through its educational apps.

Educational apps and LMSs nearly always have the potential to automatically collect reach and engagement data for users who access the educational platforms. As discussed in Step 4, this tracking must take into account a number of technical and ethical considerations in the design stage.

PRINT MATERIALS. Programs often develop print materials to accompany radio/audio, television/video, mobile phone, and online programming. They can also serve as a fifth modality if technology is not available. Reach of print materials is tracked in the same ways as distribution of technology—through paper-based forms, mobile phone questionnaires, or tablet or computer-based programs (see Step 4 for more details). Tracking who receives materials (reach) as well as determining how they are being used (engagement) is essential for program management as well as accountability. When print materials are designed to accompany technology-based distance learning programming, the effectiveness of the design depends on whether or not the user has the accompanying materials available to them.

Engagement captures the extent to which educators and learners utilize the distance learning programming and materials, and whether programming is used as intended. For example, listener engagement in a radio program is often determined by the proportion of target learners using the series on a weekly basis to build literacy and numeracy skills. Engagement metrics may also gauge whether users perceive the content to be relevant and captivating, whether users stay interested over time (e.g., reading a full digital book versus abandoning it partway through), and which content is most popular across users (e.g., most listened to, watched, or read). Self-reported metrics can also be used to assess whether the content is interesting and relevant to users, and how satisfied learners, educators, and caregivers are with the distance learning programming content or experience. Engagement measures can be collected alongside reach, and both domains have a long history of being measured through in-person data collection.[22,23]

The key metrics for capturing engagement are:

- **Extent to which users participate as intended in programming:** Analysis of the use of technology, programming, and materials by users (educators, learners, and caregivers) per the design. This also includes users' level of participation in programs per scheduled interactions and touchpoints, and whether learners complete the radio/audio, television/video, mobile phone application, or online content.
- **Quality and relevance of the programming:** Whether programs are of high quality and reflect the developmental, cultural, social, or other needs of users.
- **Response to the content:** Users' reactions, opinions, and perspectives on the distance learning content and whether they feel the content is captivating.

Text Box 8

Recommendations for Measuring Engagement

1. **Do not stop at measuring reach (access); use additional metrics to assess engagement and completion.** Distance learning only leads to supporting educational retention or acquisition of learners and educators if users they are successfully accessing the content AND utilizing it as intended. Engagement data can lead to suggestions on how to support learners through modifying, adjusting, and adapting for their circumstances.
2. **Ensure there is a process for formatively evaluating each radio/audio, television/video, mobile phone, or online episode/module, as well as print materials.** Criteria for formative evaluation of content should assess the relevance of language used; developmental appropriateness of activities; representation of and responsiveness to different genders, ethnicities, races, classes, etc.; and level of engagement and interaction. Ideally, users are observed and are surveyed or interviewed for feedback. As a number of respondents for this review reiterated, "bad distance learning programs have the potential to do more harm than good," so evaluating quality is paramount.
3. **Include perspectives of the most marginalized in measures of engagement.** In order to ensure relevance and accessibility of content, and following principles of UDL, those who are often the most marginalized and excluded should be part of efforts to formatively evaluate content.

Quality and relevance of programming are typically measured through formative evaluation of content in episodes, modules, and materials, as is whether users take part in scripted interactions built into the different modalities. These formative evaluations of content are often conducted in-person but can be conducted remotely if users have access to appropriate technology. Data can be collected through observations, surveys, interviews, or focus group discussions (FGDs).

PARTICIPATION IN PROGRAMMING AS INTENDED

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. The most common way to capture **whether users are utilizing the technology devices, programming, and materials that are part of a distance learning strategy** is to ask them through a brief survey (self-reported metric). For example, a GeoPoll mobile phone survey of FRI in Ethiopia, Malawi, Tanzania, and Uganda asked listeners how often and in what ways they used the educational content from a radio series. This helped gauge whether a sample of users was following the design and intended use of the programs.

Many radio and television instructional methodologies incorporate interaction into the programming. For example, Education Development Center (EDC) builds in pauses and guidance for interactions into interactive radio/audio instruction (IRI/IAI) scripts (see USAID's *Delivering Distance Learning in Emergencies: A Review of Evidence and Best Practice* for more details on IRI/IAI). FRI builds listener interaction into all of their series as a general practice. While observing the use of IRI or IAI in

classrooms and learning centers has been the most common way **to measure whether listeners are engaging in the interactions,** feedback loops are increasingly being built into each episode. For example, in each FRI episode, listeners "flash" (call and hang up) a broadcast number, and an IVR system automatically calls the user back and walks them through a sequence of questions. These questions cover reach (e.g., Did they listen? Did they open an app?), engagement (e.g., What part did they like best? Did they participate in interactive elements?), outcomes (e.g., Can they answer a question about something covered in the program?), and an opportunity to leave voice-recorded feedback. Callers' numbers are automatically recorded into a database.

In Pakistan, Tabadlab proposed using a Short Messaging Service (SMS) portal for measuring television lesson summaries, basic knowledge, and tips and tricks for its television programming. A learner watching the program follows instructions at the end of the video to text a keyword and code to a phone number. They then receive a sequence of lesson-specific questions via an SMS loop (chain of questions). These feedback loops are built into the programming and systematically capture basic reach and engagement (and sometimes knowledge/outcome) data. While the sample of users who respond to these programmed feedback loops are not representative of the overall target population, and learners who do not have a technology device and connectivity rarely participate, these programmed interactions are still a useful way to capture some engagement data.

Nearly all education initiatives included in this review had some metric of **completion** in place. In education initiatives where learners meet in groups per the distance learning design, attendance data are typically collected either through paper forms, SMS texts, survey software, spreadsheets, or sent as a photo in WhatsApp. For self-directed (asynchronous) learning, users are sampled and asked in questionnaires or interviews to self-report the number of programs they listened to or viewed. The more intentional the data flow is in the initial distance learning design, the more accurately and efficiently teams can process the data and use the results.

MOBILE PHONE AND ONLINE PROGRAMMING. When content and programming is accessible through a pre-loaded device, app, website, or LMS, **the extent to which learners engage in** and **complete** the programming is automatically tracked through analytics. For example, Worldreader continuously captures a sequence of engagement analytics for their BookSmart app. They can measure the number of readers, the number of digital books distributed, the percent of readers reading at or above the time spent reading per day benchmark, and the number of books that readers have completed or abandoned. Likewise, online platforms (e.g., LMSs and Massive Open Online Courses [MOOCs]) have the potential to track how far users have progressed in online courses.

Educational phone apps, IVR systems, and online content have built-in features for facilitating **interactions** (interactive activities and assessments). Both Education Initiative Plus and Worldreader use photo messages sent through a text message or WhatsApp to measure whether learners **participated as intended** in their programs. Education Initiative Plus held a writing competition for learners and asked caregivers to send photos of the completed projects. Similarly, Worldreader held a 14-day reading challenge and asked caregivers and educators in a WhatsApp group to send photos of learners completing activities related to the daily book. Online LMSs typically include functions for polls, chat rooms, breakout rooms, discussion threads, collaborative writing, and document sharing, all of which can be used to capture engagement data.

QUALITY AND RELEVANCE OF PROGRAMMING AND USERS' RESPONSE

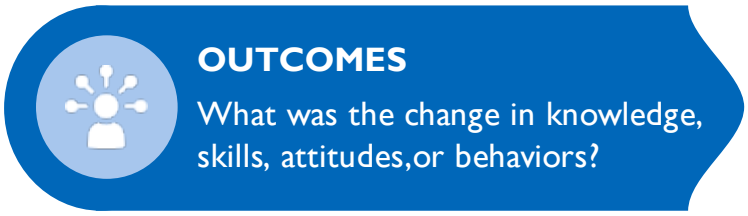
Regardless of the modality, it is essential to ensure that the content is captivating, interesting, and of high quality, and that the format, language, and accompanying materials are appropriate for age, literacy level, language, and cultural, social, and political context. Before content is broadcast, programmed into software, or otherwise disseminated, a formative evaluation of content should be built into the design and production phase. Subsequent post-production evaluations are also important for gauging responses from a wider audience.

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. When assessing the quality and relevance of programming, it is important to draw on what is known to be developmentally- and age-appropriate for promoting learning. For example, among preschoolers, the appropriate IRI length is no more than 20 minutes.[24,25] Studies of IRI also indicates that including music or rhythms, a storyline, variation in voices, and frequent pauses for movement and interaction is critical for keeping young learners' attention.[26] As described in Annex D, Ubongo conducts comprehensive formative evaluations of content and processes for their television programs to measure relevance to their target audiences. They ensure that the programs aired are of high quality by conducting extensive in-person psychological and cognitive studies with children.[27,28] Similarly, evaluations of distance learning programs for low literacy, out-of-school youth reviewed for this study indicated that language needs to be carefully evaluated to ensure that it is at the right literacy level and pace for learners who do not speak the language of instruction as a first language.

Formative evaluations of programming should also **measure listeners' or viewers' responses** to the radio/audio or television/video content. In a recent formative evaluation of a radio series on inclusive education (Soma Nami) in Tanzania, educators, caregivers, and community members participated in phone interviews in which they recalled scenes in the episode, described the extent to which they felt that the episode was engaging, explained which parts stood out, shared messages they felt were not clear, and identified language they found problematic. [29] These data can be used to revise or re-design programming. These data can also help gauge the extent to which the user understood the content and language. FRI, EDC, and Ubongo use similar processes after each draft episode and again after the programs have been in circulation with a wider audience.

MOBILE PHONE AND ONLINE PROGRAMMING. Engagement with mobile phones and online programming is often gauged through formative evaluation of the content. **Assessing the quality and relevance** of mobile and online programming for pre-primary and primary school learners is commonly carried out through surveys, interviews, or FGDs with caregivers since they often oversee their child's use of technology devices and programming. For example, Young 1ove connects with grade 3-5 learners and their caregivers via a mobile phone survey during their remote data collection activities. They ask caregivers about their perspective of their learners' needs, provide coaching techniques, and then ask to speak with the learner directly to conduct a modified ASER test to assess their numeracy skills.[30]

The Education Initiative Plus in Northern Nigeria uses an IVR system to present literacy and health programming via mobile phone. After a lesson, staff ask caregivers or learners to **respond to the content** by sharing whether or not they found the lesson useful. Caregivers can request additional help from a coach through the IVR system and receive a coaching phone call. In addition to supporting caregivers, the calls from Young 1ove and Education Initiative Plus are used to assess whether learners and caregivers are participating in the numeracy and literacy lessons as intended.



Outcome metrics assess change in knowledge, skills, attitudes, or behaviors.[31] Outcomes are separated into two types: those that measure content knowledge and learning (e.g., history, learning how to read); and those that capture SEL (e.g., confidence in learning to read and getting into a habit of reading). Learning outcomes can be based on curricular objectives, as in the case of formal school or a structured nonformal program, or they can extend beyond the curriculum (e.g., acquisition of technical skills through apprenticeships). They can be measured directly (e.g., on a test), through self-reporting (e.g., an interview or survey), or through observation (e.g., watching in person or through a remote video call). Outcomes from radio/audio, television/video, and mobile phone programming have been the most evaluated. There is a growing body of evidence on online learning for upper primary through higher education, but there has been relatively little written on online learning outcomes for pre-primary, primary, and adults and youth with low literacy skills in the Global South.

Outcomes should be considered in the design of the programming and materials (e.g., promoting confidence in reading was built into program design), but can be unintended and unplanned (e.g., promoting appreciation for non-fiction materials). While it is hard to anticipate unintended outcomes, triangulating with multiple methods and using both quantitative and qualitative methodologies creates more opportunities for unintended outcomes to be measured.[32]

Most of the research on distance learning outcomes among pre-primary and primary-level learners employs in-person methods. This is in large part because in-person methods help mitigate varied technology access and connectivity issues among learners, educators, caregivers, and other participants. In-person data collection allows for rapport building (especially with young children and marginalized youth), validity testing, and controlling interferences and disruptions, among other factors (to be discussed in further detail under Steps 3 and 4).

There is currently limited evaluative evidence of remote measurement of distance learning outcomes in the Global South. The consensus across the interviews conducted for this review is that remote collection of distance learning outcomes is useful for low-stakes *formative assessment* objectives, including gauging what content knowledge learners have retained during school and learning program closures, monitoring learners' socioemotional and physical well-being, and determining what additional teaching and learning support learners may need when schooling resumes. Observing educators' well-being and connecting them to critical resources, knowledge, and skills is also an important objective of monitoring distance learning outcomes.

Practitioners interviewed for this review cautioned against using remote high stakes exams for *summative* objectives during quick pivots to distance learning, such as tests that determine the extent to which television/video programming increases literacy and numeracy outcomes during school closures. Summative evaluation of learning outcomes may be feasible in a situation where distance learning has been intentionally planned from the start or when there is a test or survey built into an online or educational app platform. However, many equity factors impact who can be accessed remotely using phone calls, SMS, and IVR surveys during a quick pivot to distance learning.

Text Box 9

Testing During Crises and Quick Pivots to Distance Learning

During the Ebola outbreak, closures of learning institutions in Sierra Leone led to a loss of literacy knowledge among learners (ACACPS Assessment April 2016). Limited social interaction with classmates, peers, and educators also caused learners to lose self-esteem and confidence (ACACPS April 2016). Furthermore, pushout and dropout rates increased, as did teenage pregnancies and children's involvement in labor and illicit and criminal activities (ACACPS April 2016). During program pivots, it is critical to capture retention (or loss) and acquisition of knowledge, as well as changes in social and emotional and soft skills, attitudes, and behaviors. However, low-stakes formative assessments (e.g., learner check-ins) are emphasized as the best way to track learner knowledge and social and emotional and soft skills during learning institution closures.

There is mounting concern that remote summative outcome assessments during a quick pivot may place additional stress on learners, educators, and families during a rapid shift to distance learning. In a recent study of 1,784 children (grades 2 to 6) in China (Hubei Province) during COVID-19, 22.6 percent of children reported experiencing depression and 18.9 percent reported increased anxiety.[33] In India and China, studies have recorded an increase in teen anxiety, fragile mental health, and suicide.[34,35] Research on high-stakes testing in Nigeria, Sri Lanka, Tanzania, and the United States indicates that high-stakes testing does increase physiological stress on learners of different ages and can impact performance, retention, and well-being, especially for marginalized learners.[36,37,38,39] Less is known about learners in conflict and crisis settings. As a key principle of data collection, protecting learners and other participants and guaranteeing their well-being is more important than determining outcomes (see *Guidance for USAID Educator Sector Implementing Partners: Monitoring, Evaluation, and Learning (MEL) During the COVID-19 Pandemic*).

The key metrics for capturing outcomes are:

- **Content knowledge learning.** Formative or summative assessments of whether skills and/or content (e.g., numeracy, literacy) were acquired or retained during distance learning.
- **Social and emotional learning.** Formative or summative assessments of how learners' social and emotional and soft skills, and/or attitudes, perceptions, or beliefs have changed through distance learning programming and content (see *Best Practices on Effective Remote SEL/Soft Skills Interventions* for more details on these areas).
- **Behaviors and practices.** Formative or summative assessments of how learners', educators', or caregivers' behaviors or practices have changed through distance learning programming and content.

Although cost-effectiveness was not analyzed for this review, it is an important outcome metric in Table 1. There is limited evidence on the comparative cost-effectiveness of radio/audio, television/video, mobile phone, or online distance learning modalities. The most extensive cost-effectiveness study of a distance learning initiative was conducted on Ubongo's math programming.[40]

Text Box 10

Recommendations for Measuring Outcomes

1. **Use in-person assessments for measuring outcomes, when feasible.** In-person assessment of summative outcomes allows for controlling reliability, validity, and precision, as well as ensuring ethics and confidentiality. If using a summative knowledge-based assessment remotely, reliability needs to be measured for remote administration. In-person reliability does not transfer to remote assessments. Other additional analyses would need to be conducted.
2. **Emphasize low-stakes formative assessments to inform teaching and learning.** These assessments should not be tied to the assessment of educators (e.g., teacher promotion).
3. **Check the emotional well-being of learners before conducting assessments (high or low stakes), and be careful not to add additional stress.** This practice is critical to ensuring that learners are protected.
4. **Include perspectives of the most marginalized in outcome measures.** In order to ensure that marginalized learners are also benefiting from distance learning (e.g., experiencing the same positive growth in knowledge, skills, attitudes, and behaviors as learners who are not from marginalized groups), those who are most often marginalized should be included in sampling frames for outcome and impact evaluations (following principles of UDL).

CONTENT KNOWLEDGE AND LEARNING

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. A number of the initiatives (Read Haiti; People in Need, Nepal) have been conducting rapid literacy, numeracy, or other content knowledge check-ins to maintain contact with learners, to inquire about their well-being, and/or to inform program management. Following learning institution closures during a recent wave of unrest in Haiti, Read Haiti used an assessment to measure potential literacy loss. The program continued this assessment in an abbreviated form via mobile phone calls during COVID-19. Young 1ove was the only team interviewed that tested a remote, content knowledge assessment for validity (through test-retest, a statistical test, of in-person and remote assessment). Their intervention provided math problems of the day via an SMS to caregivers' and learners' phones, then gave a follow-up phone call to learners to help them solve the math problems. These data were primarily designed to a) tailor the SMS interventions to the learners' programmatic and numeracy level (as a low-stakes formative assessment), and b) inform program management on what was working.

MOBILE PHONE AND ONLINE PROGRAMMING. The ability to measure outcomes for mobile phone and online programming is often pre-programmed into the software. Educators can use these functions (e.g., Google Forms) to create their own quizzes and assessments in LMSs. Multiple-choice and short-answer assessments can be set up to be graded automatically. Essays should be read and graded directly by educators.[41] Assessments can also be built in as part of an app. For example, the Ubongo Kids interactive e-book *Kibena and the Math Rats* has a basic numeracy quiz question to answer before learners can proceed to the next page. Ubongo can automatically capture these data to assess learners'

numeracy skills. Whether through mobile phones or online platforms, these assessments should be kept to 15-20 questions. Multiple-choice or short-answer options are preferable on mobile phones because typing longer answers on a phone can be challenging. Change in knowledge and content-based skills for secondary school (or equivalency) and tertiary level learners may be measured differently from pre-primary and primary age learners. Numerous country-level examples of LMSs for different education levels are included in The World Bank's brief *How Countries are Using Edtech (Including Online Learning, Radio, Television, Texting) to Support Access to Remote Learning During the COVID-19 Pandemic*.

SOCIAL AND EMOTIONAL LEARNING

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. Interactive radio/audio instruction and video programming content often covers social and emotional and soft skills and integrates attitudes, perceptions, and beliefs into programming content as part of the pedagogical and instructional design. Most teams that measure SEL do so through a variety of in-person assessments (see assessments of Ubongo's television programs and EDC's IRI series for noted practices). FRI was the only team interviewed for this review that regularly builds attitude and perception feedback metrics directly into their distance learning radio series for remote data collection. As discussed in the case study in Annex D, FRI draws on social and behavior change communication (SBCC) approaches in its interactive programming. In addition to reaching smallholder farmers, with a particular commitment to women agriculturalists, they have produced programming on the mental health and well-being of youth in Malawi and Tanzania, among other types of programming.[42] In one study across Ethiopia, Tanzania, and Uganda, FRI included five questions on gendered beliefs and norms in their 18-question mobile phone survey and created questions within each episode that gauged attitudes and beliefs through an IVR infrastructure. They also created WhatsApp groups in which participants could record their stories and experiences, thus capturing qualitative data on self-esteem, self-efficacy, and other metrics of well-being.

MOBILE PHONE AND ONLINE PROGRAMMING. No remotely administered studies of children's SEL skills and outcomes developed through mobile phones and online programming were found for the Global South as a part of this review. However, Room to Read and People in Need conducted rapid assessments of socioemotional well-being through phone calls with the socially marginalized young women in their respective programs. People in Need works with young women in Nepal through an accelerated learning program that covers both content knowledge and SEL skills. These women have either never been to formal school or have been pushed out; nearly all are married and have children. Before the pandemic, the women met through in-person learning groups with a facilitator from their community. When in-person meetings ceased, the social mobilizer shifted to calling the women one to two times per week. During the 15 to 20-minute calls, the facilitator covers some of the content knowledge lessons and inquiries about the women's socioemotional and physical well-being. They use these data to ensure that the women continue to work toward their learning goals and have the requisite socioemotional support to do so.

Room to Read's Girls' Education Program works with 11 to 18-year-old girls across eight countries on a life skills and mentoring curriculum. The curriculum covers topics like self-confidence, expressing and managing emotions, empathy, self-control, decision-making, perseverance, communication, creative problem-solving, and relationship-building. Social mobilizers, who are trained women from the same communities, integrate a one-minute, three-question survey into their regular mentoring calls. YouthHubAfrica offers another example of an initiative through which social and emotional skills development is being addressed and measured remotely during COVID-19. In Nigeria, the organization is training young women between the ages of 13 and 18 years to understand how the COVID-19

pandemic and related lockdown of educational institutions have affected girls and their communities, while also building confidence and skills in photo storytelling.[43]

BEHAVIORS AND PRACTICES

RADIO/AUDIO AND TELEVISION/VIDEO PROGRAMMING. The most robust studies of change in behavior and practice have been conducted in-person. (See Ubongo's evaluations of television, EDC and FRI's evaluations of change in behaviors and practices, and Sesame's longitudinal study of labor market outcomes). The only remote evaluation of behaviors or practices noted in this review was an assessment conducted by FRI, which regularly includes questions in their radio series to measure the extent to which farmers have been able to practice techniques they have learned through radio programming (responses are then captured through IVR surveys). While there may be other examples of measuring change in behaviors and practices for radio/audio and television/video, none were identified for pre-primary, primary, or equivalent learners in this review.

MOBILE PHONE AND ONLINE PROGRAMMING. Only a few organizations in this study analyzed behaviors for pre-primary or primary learners or equivalent using remote technologies. These studies draw on cross-sectional data, which do not allow for a study of individual behavior over time but rather group behavior at specific points in time. Worldreader, as discussed earlier in this review, measures a combination of reading engagement and behaviors, including how frequently users read and if they complete their books. Another example of assessment of behavioral outcomes is a small pilot of 37 Tanzanian educators conducted by the Tusome Pamoja activity. The activity team uses WhatsApp to support educators in teaching about school-related gender-based violence (SRGBV) in their classrooms. The team has also been using WhatsApp as a mechanism for mentoring and building a community of practice to ultimately encourage changes in teaching. At the same time, they are using the conversations to observe changes in practice in teaching and thinking about SRGBV.

STEP 3: DETERMINE HOW DATA WILL BE COLLECTED (IN PERSON OR REMOTELY)

Text Box 11

Step 3 Guiding Questions for Implementation

- Should data be collected in person or remotely?
- What key considerations—safety of teams, access to technology, infrastructure, feasibility of capturing reach, engagement, and outcomes—should be considered?
- What equity considerations should be taken into account (e.g., geographical reach, socioeconomic status, gender, disability)?
- What technologies should be used (e.g., paper, mobile phone, tablet, computer)?

As discussed in Step 2, data collection conducted in person or through mailed or distributed paper-based surveys (correspondence) has the longest precedent for measuring reach, engagement, and outcomes of distance learning. Technology has only been used for remote data collection in the past few decades as mobile phone and Internet access has grown. For example, GeoPoll, a leader in mobile phone surveying, was only established in 2012. Consequently, the literature on the efficacy of distance learning as measured through in-person methods (e.g., tests, surveys, interviews, and observations) is stronger and more conclusive than the literature on distance learning measured remotely. However, this is changing, especially in the context of COVID-19 and in other crises and conflict settings where in-person data collection is either not safe or not possible. There are now a number of well-constructed studies on measuring reach and engagement, and, to a lesser extent, outcomes. In order to reach a greater number of participants, to collect data faster, and to reduce costs, a combination of remote data collection and in-person approaches are recommended, as laid out in the section below.

Text Box 12

Recommendations for Determining how the Data will be Collected (in-person or remotely)

1. **Use an integrated (in-person and remote) approach to data collection.** Be clear how the data will be used, whether for formative or summative purposes. Collect summative outcome data in-person when it is feasible (for reasons of reliability, validity, precision, confidentiality, and building rapport). Use remote methods to collect more frequent and systematic reach and engagement data to allow for real-time data collection and to ensure the data are used formatively to inform teaching, learning, and program management. Remote data collection should also be used when in-person data collection is not safe or feasible.
2. **Collect mobile phone numbers of caregivers and families at the beginning of an intervention or school year.** This enables remote data collection to reach as many participants as possible and aids data collection in the event of a sudden pivot.
3. **Create MEL platforms for basic phones or support families in acquiring smartphones to ensure marginalized individuals and households can be included in distance learning and data collection activities.** Consider subsidizing rather than distributing free mobile phones so users assume ownership of the device. This avoids having to build in costs and resources for managing replacement, repairs, and upgrades.[44]
4. **Assume that others are listening in during remote data collection.** When collecting remote data orally via a phone call survey, plan as if a caregiver, spouse, or other family member is listening along with the learner. In a study in India of self-employed women, 65 percent of women had their phones on speakerphone, so it was not appropriate to ask questions that should not be heard by other family members.[45]

DECIDING WHETHER DATA COLLECTION SHOULD BE IN-PERSON OR REMOTE

While the studies included in this review revealed the potential for, and importance of, expanding remote data collection approaches, they also reiterated four reasons why in-person data collection is still the preferred means of measuring distance learning outcomes. One exception may be online learning, where quantitative measures can be captured through online quizzes, tests, and other assessments, or where video or audio conferencing can be used to facilitate oral assessments. In-person data collection of outcomes is preferable in order to:

- **Reach the most marginalized communities**, which tend to have less access to technology devices, software, and infrastructure (electricity and connectivity). These include geographically remote learners, families living in poverty, women, and persons with disabilities. People in Need Nepal indicated that 25 percent of the girls they worked with in their in-person intervention were not reachable by mobile phone, and, therefore, the organization had no way of monitoring

these girls' socioemotional and physical well-being, content knowledge retention, and engagement following the pivot to distance learning during COVID-19.

- **Account for a range of technological literacy, reading levels, and language proficiencies**. Distance learning programming can favor users with stronger technological and reading skills, those in higher grades, and those who are proficient in the language(s) being used in the distance learning programming and assessment if not considered in the design. For example, Rising Academies in Liberia found that learners who did not speak Standard English had a harder time engaging in SMS messaging and assessments when Standard English was used versus when Liberian English was used.
- **Control for environmental factors.** Mitigating interference of a caregiver or sibling, noise and distractions, and safety when data are collected remotely can be difficult, which can affect the reliability, validity, and accuracy of the data. Young 1ove attempted to control for such factors and analyzed reliability and test-retest (a statistical test) for in-person and remote administration. However, they used the data predominantly for formative purposes (e.g., informing activity design and teaching and learning). When conducting in-person assessment, it may be possible to take step to mitigate threats to reliability and validity. This is more difficult during remote data collection.
- **Establish a level of trust and familiarity between data providers (learners, educators, and caregivers) and the data collectors**. In-person assessments allow for observation of behaviors, and can be particularly important when asking questions about socioemotional well-being and environmental safety. It is often easier to ensure the well-being and safety of participants when collecting data in-person. Unfortunately, none of the studies included in this review studied learners' and educators' well-being and experience with in-person data collected compared to remote data collection.

Key considerations for remote data collection. The decision to collect data remotely or in person depends on the factors described below:

Safety and socioemotional well-being of learners, educators, caregivers, and data collectors are paramount. In contexts such as the COVID-19 pandemic, disease outbreaks like Ebola, and some ongoing conflicts, remote data collection may be the safest option. However, in circumstances where technology is limited, data collection may need to occur partially or fully in-person, with careful attention to safety (e.g., maintaining social distancing and/or providing personal protective equipment). Physical and socioemotional well-being is placed as the first consideration, as the principle of doing no harm comes before all other considerations.

Access to technology and means of charging technology, particularly mobile phones, is required for all but paper-based data collection. For audio, video, and mobile programming, almost all data are collected via mobile phones, as phone coverage is much higher in most contexts than Internet availability. Gathering data on who has mobile phone access before starting data collection is essential.

Infrastructure, such as access to reliable electricity, Internet, and phone service is a crucial consideration, as is the technology device used and the method selected.

Feasibility of capturing reach, engagement, and outcomes remotely. The few educational initiatives that have assessed distance learning outcomes remotely for pre-primary or primary level learners in the Global South caution that the data should be used for low-stakes formative assessments as opposed to high-stakes summative assessments.

Text Box 13

Equity Considerations

One of the biggest challenges with remote data collection is identifying who is not being reached. Even though approximately 80 percent of people in the Global South own mobile phones, the most marginalized individuals and communities are often excluded during remote assessments.[46] For example, FRI found that women were often the least represented among beneficiaries using a call-in line to provide feedback on the reach of the FRI radio programming, or on their engagement with the content. This was in part because they were less likely to have their own phone and/or because they did not have the time to wait on hold to provide their perspective. While the program did not provide phones to users, they did create a separate call-in line for women. This reduced their waiting time.

People with disabilities are also commonly excluded in remote data collection either because data collection methods are not accessible or do not offer reasonable accommodations (e.g., phone-based surveys, interviews in contexts where video relay is nonexistent). When methods are accessible (e.g., using accessible software for surveys, mailing paper surveys in accessible formats) the required technologies and ancillary devices are often not available to individuals with disabilities to make use of the software. As discussed under Step 4, when deciding on remote versus in-person data collection, it is critical to keep Universal Design for Learning (UDL) principles in mind and choose options that are inclusive of marginalized users.

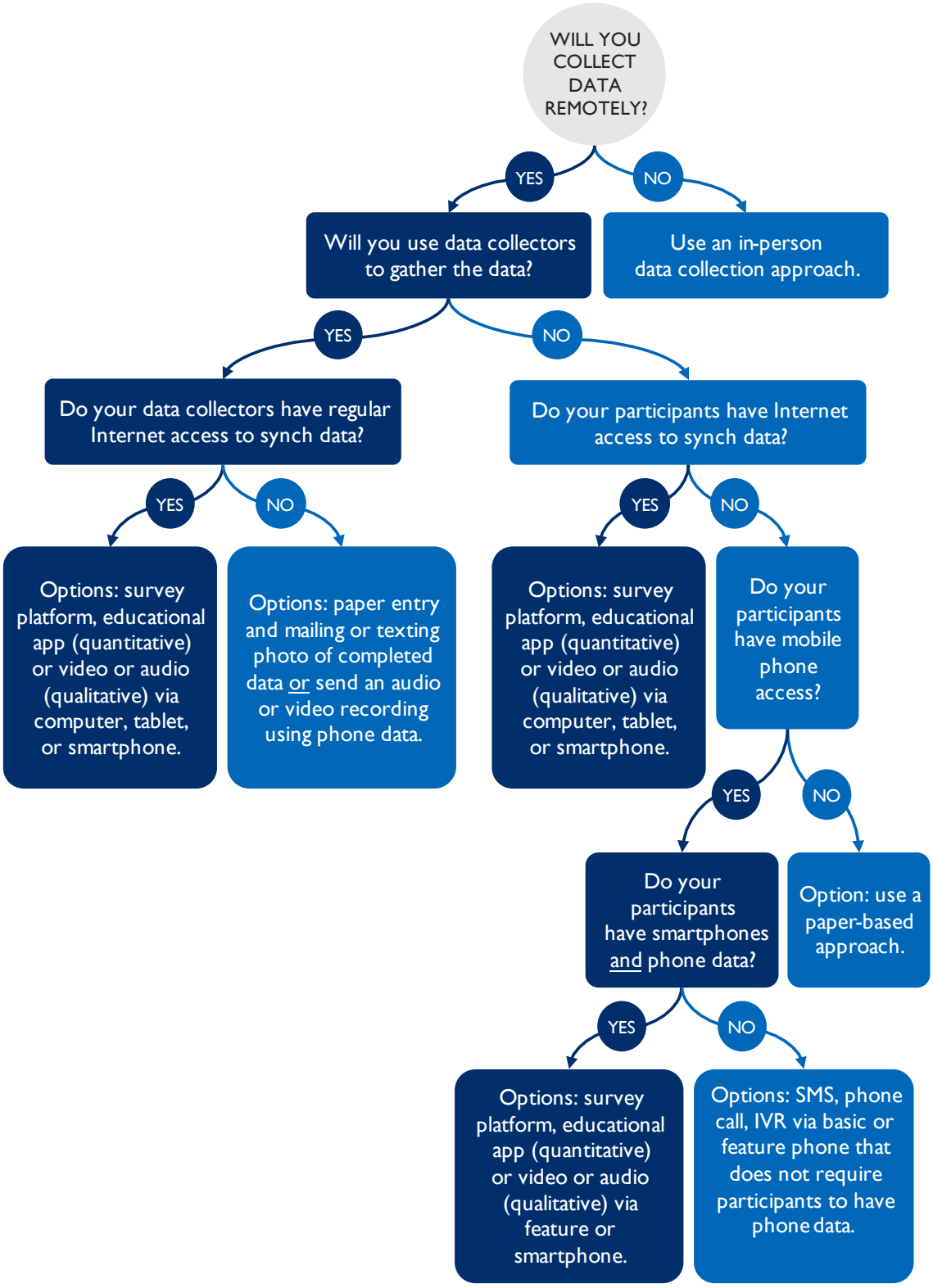
Across all the findings, the most marginalized are commonly excluded during remote data collection for a variety of reasons: a) they lack the mobile phone or technology device and connectivity, b) if they are using a family or community member's phone, they have limited time and access to the phone (e.g., the owner may not accept the call or may limit the times when the learner can use it), c) the phone calls may be placed on speaker and so the information is not confidential, or d) the participant may be coached by a family member. See Step 4 for further discussion.

SELECTING REMOTE DATA COLLECTION TECHNOLOGIES

Mobile phones are currently the most frequently used technology for collecting reach, engagement, and outcome-level data. At a minimum, all mobile phones, whether basic, feature, or smartphones, can send and receive phone calls and SMS messages for short questionnaires, surveys, and interviews. Feature and smartphones open up the possibility of video calls, photo messages, backend app/website analytics, and more. Data collectors may consider using a wider range of technologies to collect and input data, including computers, tablets, mobile phones, and, when necessary, paper. In all cases, a needs assessment of which technologies beneficiaries and data collectors have access to must be conducted *before* selecting devices to use for data collection. This, in turn, will inform the types of data that can be collected and the methods and approaches that can be used to measure, as discussed further in Step 4.

The decision tree below walks users through the different decision points for determining which technology to use for collecting data, whether in-person or remotely.

Figure 4: Decision tree for determining data collection technology



While the decision tree outlines how to choose appropriate remote data collection technologies, integration of in-person and remote strategies is recommended where feasible. Remote technologies allow reach and engagement data to be collected and analyzed in real-time. They can also engage participants who may not be able to attend in-person data collection activities (e.g., a young, nursing mother who cannot travel to a learning center with her newborn). However, the strongest evidence from evaluations and the consensus across interviews is that measuring outcomes in-person is still critical and effective, as in-person evaluations allow for rapport building, validity testing, controlling for interferences and disruptions, and does not rely on access to technology. A number of the MEL practices analyzed indicate that using a combination of in-person and remote techniques enables a balance of obtaining real-time data with deep and meaningful assessment in both emergencies and non-emergencies.[47] For example, a distance learning initiative may collect data on reach through an SMS audience survey, level of engagement through phone interviews or a social media group, and outcome metrics of knowledge, attitudes, and behaviors through an in-person assessment where data are entered into a tablet software and later synched to a central database when an Internet connection is available.

STEP 4: DETERMINE THE METHODS AND APPROACHES FOR MEASUREMENT (QUANTITATIVE AND QUALITATIVE)

Text Box 14

Step 4 Guiding Questions for Implementation

- What quantitative and qualitative methods can be used to measure distance learning?
- What technologies (e.g., paper, mobile phone, tablet, computer) and interfaces (e.g., SMS, survey software) are used to gather the data?
- What sampling strategies (e.g., census, representative, purposive) can be used?
- What kinds of equity analyses should be considered?
- What is the strength of the evidence for these evaluative approaches and where is there a need for more evidence?

A wide range of quantitative and qualitative methods can be used to measure distance learning reach, engagement, and outcomes. Quantitative methods—surveys, questionnaires, and tests—are the most frequently used, but qualitative methods, such as interviews and focus group discussions (FGDs) can also be used. Participatory and arts-based research methods (e.g., photography, videos, drawings, theater) are less frequently used but can capture important data on learner, educator, and caregiver perspectives and experiences.[48]

Text Box 15

Recommendations for Determining the Methods and Approaches of Measurement

1. **Use mixed methods to collect data,** including a combination of surveys, interviews, FGDs, and other methods to better capture different aspects of reach, engagement, and outcomes, and to capture any unintended outcomes. Likewise, use multiple interfaces (e.g., IVR, SMS, phone surveys) to be more accessible, and design them to be short and engaging.
2. **Match the evaluation purpose to who is collecting the data.** When conducting formative evaluation or assessment, involve the educators, social mobilizers, and facilitators in the process, as they often know the learners best and are typically known to caregivers. While data collectors may be desirable for summative evaluations where reliability and validity are critical, as well as to provide anonymity, for formative assessments, having the people who are using the data be the collectors can be beneficial. Regardless of who is collecting the data, communicate with participants about why the data are being collected, how they will be used, and how participants' privacy will be protected, even for formative data that does not go through a formal institutional review board process.
3. **Treat language as a right and a resource.**[49] Know what the first languages of the participants are (both spoken and signed), and account for multilingualism in the design. For example, the person collecting the data should use a participant's first language for explaining directions and allow the participant to answer in this first language if the intent is not to measure language acquisition in another language. If an evaluation is conducted in a language other than the participants' first language, this will affect the data and the inclusivity of the participant in the process. To the extent possible, SMS messages, IVR systems, and surveys should also be developed in local languages.
4. **Acknowledge that even simple assessments may feel like a big deal to a learner or caregiver.** No matter how well the person collecting the data may explain the intent, caregivers want their children to succeed and may influence their answers. Find culturally and socially acceptable ways to allow participants to speak for themselves, whether this involves planning simultaneous activities for the caregiver or designing the assessment for both caregiver and child to have a role. In cases of marginalized youth populations, familiarity with the person collecting the data may play into whether the family allows the youth to speak in private. In the case of young, married girls in Nepal, conversations were designed with the expectation that they would be overheard.[50]
5. **Ensure data collection efforts are not further marginalizing participants.** The most marginalized learners, educators, and caregivers often have the most limited access to technology and are often excluded from mobile phone data collection. Design M&E strategies to be inclusive and equitable according to the principles of UDL. For example, for rural youth without access to phones, data collectors may need to do in-person, socially-distanced monitoring and evaluation. For people with low literacy levels and with disabilities, evaluations should be designed to be conducted in plain language and in formats that are accessible depending on the needs of the participants.

6. **Give the data back to the participants**. Ensure that there is a feedback loop in place for all participants to receive a summary of the findings.
7. **Plan early and plan ethically.** Distance learning is often designed to reach the most marginalized communities. Planning early and ethically helps to ensure that monitoring and evaluating are inclusive as well as protective.

Table 3 outlines the common quantitative and qualitative methods used in monitoring and evaluating distance learning. The table shows the domains of measurement, sampling frame, and associated technologies for each method, all of which are discussed in detail following the table.

Table 3: Methods used in measuring distance learning

METHODS	DESCRIPTION	DOMAINS	SAMPLE	TECHNOLOGIES
Quantitative				
Survey	A quantitative measure that captures data through close-ended questions that are used to collect and analyze data and show trends in a targeted population.	Reach Engagement Outcomes (knowledge, skills, attitudes, or behaviors)	Purposive or representative	Paper (mail) Phone (IVR, SMS, phone call) Tablet Computer
Questionnaire	A questionnaire is like a survey but usually shorter in questions and for a general population.	Reach Engagement Outcomes (knowledge)	Census (all users), purposive, or representative	Paper (mail) Phone (IVR, SMS, phone call) Tablet Computer
Test	A tool, technique, or method that is intended to measure learners' knowledge or their ability to complete a particular task. Testing is a form of assessment.[51]	Outcomes (knowledge or skills)	Representative	Paper Phone (phone call, video conference) Tablet Computer
Analytics	Digital analytics encompasses the collection, measurement, analysis, visualization, and interpretation of digital data, illustrating user behavior on websites, mobile sites, and mobile applications.	Reach Engagement	Census	Radio/TV (meters at stations) Phone (backend app access) Tablet (backend app/website access) Computer (backend app/website access)
Qualitative				
Interview	Structured conversation with an individual that includes open-ended questions and probes to yield in-depth responses.	Reach Engagement Outcomes (knowledge, skills, attitudes, or behaviors)	Purposive or representative	Phone (phone call, video conference) Tablet (video conference) Computer (video conference)
Group Interview or Focus Group	Structured conversation with a small group, includes open-ended questions and probes to yield in-depth responses.	Reach Engagement Outcomes (knowledge, skills, attitudes, or behaviors)	Purposive	Phone (phone call) Tablet Computer
Participatory and Arts-Based Research Methods	Participatory approaches are developed together with participants. Arts-based methods include photography, video, visual art, theater, dance, and music.	Engagement Outcomes (knowledge, skills, attitudes, or behaviors)	Purposive or representative	Paper Phone (MMS, video or audio record) Tablet Computer
Quantitative and Qualitative				
Observations	Observable human interactions through activities, behaviors, actions, conversations, etc.	Engagement Outcomes (knowledge, skills, attitudes, or behaviors)	Purposive or representative	Phone (video call or recording) Tablet (video call or recording) Computer (video call or recording)
Document Review (quantitative or qualitative)	Review of relevant documents including records, policies, reports, publications, correspondence, and artifacts.	Reach Engagement Outcomes	Purposive or representative	Paper Digital

METHODS, TECHNOLOGIES, AND INTERFACES

Paper has been used for data collection for the longest time and continues to be useful in areas with limited or no electricity and/or Internet. Surveys and questionnaires can be printed on paper or programmed into survey software that is synched with a cloud-based database.[52] Survey software can be used online or downloaded to a phone, tablet, or computer for offline use, and then synched when the Internet is available. Surveys can then be administered on a mobile phone, tablet, or computer where electricity and connectivity are accessible. Data can either be entered by data collectors as they ask participants the questions on a phone call or by participants themselves if they have the requisite literacy level and ability to write, eliminating the need for an additional data entry step. Regardless of who enters the data, implementers and researchers recommend keeping surveys to 20 minutes or less to keep people engaged, limit potential interruptions, and account for poor phone or Internet connectivity.

Short surveys can also be administered on any type of phone (basic, feature, or smartphone) through **SMS messages, Computer-Assisted Telephone Interviewing (CATI), and Interactive Voice Response (IVR) systems.** SMS messages are received by the participant, and they use the number keypad to answer the questions. Through CATI, "interviewers use specialized software to dial phone numbers, record the answers they are given, and read the next question based on any skip logic that is included."[53] **IVR systems** use prerecorded messages to walk users through a series of information and questions. SMS, CATI, and IVR surveys should be kept to under 20 questions maximum to promote completion. Surveys can also be integrated into radio or video programs, educational apps, or online module content directly, as described under Step 2. Organizations are increasingly using IVR systems because they have the potential for being more inclusive of low vision and low literacy users. However, IVR systems are limited in what they can ask, are most accessible in major languages, and require an E1 or T1 line or other infrastructure that enables multiple callers at once.[54]

While surveys can capture a broader range of data, **analytics** are useful for automatically capturing reach and engagement data from educational apps, websites, and offline-capable programs that can sync with the Internet when it is available.[55] In theory, the Internet protocol (IP) address automatically tracks the geographical location as well as the Internet provider. When users open an educational app or website to access educational programming, the administrators of this app or site can track unique page views, activity (e.g., what content is accessed), and engagement (e.g., length and duration of use). As long as users do not reject cookies or tracking data from being collected, this opens a range of reach and engagement data that can be captured.

A number of organizations interviewed considered developing **educational apps** as a remote way to capture reach, engagement, and outcome information with backend analytics; only Ubongo and Worldreader already had extensive experience with this endeavor. The Ubongo team raised several important considerations beyond the challenge of reaching only mobile phone owners. Educational apps require certain operating systems and memory space, and often the cheaper smartphones cannot accommodate these apps. Users may also have to register their information to be able to access the app through Play Store, Google Play, Apple Store, or other proprietary software. This requires that an educator, caregiver, or another family member have an email address and sufficient technological and reading literacy to download apps. Ubongo found that families that did not have an email user used a mobile phone repair kiosk to help them download the apps, but subsequently, the apps were registered

in the name of the repair person and not the actual user. Another challenge that emerged was geographical tracking through IP addresses. The geographical location detection may not be accurate in countries in the Global South that rely on website hosts and providers from the Global North. For example, an education team in the Democratic Republic of Congo trying to map the IP addresses of users found that the users' IP addresses were showing up in Europe (Gaëlle Simon, personal communication, September 22, 2020).[56] Furthermore, if apps are not all accessed through one login, users must register for each new app. Worldreader has circumvented this issue by partnering with Opera Mini, which is the primary browser loaded on feature phones in sub-Saharan Africa. When users click on "free books" in the Opera Mini browser, they are taken directly to the Worldreader app.

If users access an app offline, the data analytics will vary. For example, Ubongo, with one of their more popular apps, is able to capture some basic reach data (how long users accessed the content and when) but not necessarily all of the built-in outcome functions. Therefore, while this is an emerging interface with great promise, until more people across the Global South have the requisite technology devices, technological literacy, and connectivity, this option serves a very small proportion of the communities that distance learning efforts are trying to reach. These same challenges and considerations can be applied to online LMSs.

Qualitative data from interviews, FGDs, and live observations can be collected via **phone calls** or **video calls** on a phone, tablet, or computer. Data from recorded observations and arts-based research methods (e.g., video and audio recordings and photographs—see the Worldreader case study in Annex D) can be gathered through **multimedia messaging services (MMS) messages** and/or Wi-Fi-based messaging services such as WhatsApp. MMS systems are more advanced than basic SMS, which only allows for text messaging. Qualitative data on paper, such as drawings and documents for review, can also be sent as picture messages over MMS or WhatsApp.

Evaluative approaches can be **mixed-methods or multi-modal** (i.e., employing some or all of the methods and interfaces discussed in this review). Mixed-methods (quantitative and qualitative) MEL approaches capture different kinds of data, triangulate data, and better evaluate equity as well as unintended reach, engagement, and outcomes. Multi-modal, or using different interfaces, expands the reach of who can participate in data collection. For example, in the case of surveys, participants with low reading levels or low vision may be better served by IVR and CATI, but SMS messaging is more accessible to participants who are hard of hearing.

Text Box 16

Principles of Phone-Based Learning Assessments

Young 1ove developed the following set of principles for conducting phone-based assessments.

Protect children and participants (do no harm).

1. Test the reliability and validity of summative assessments.
2. Keep instructions simple and use practice items to ensure that respondents understand them.
3. Some assessments will be more conducive to phones than others.
4. Keep assessments short.
5. Track the speed of responses.
6. Experiment with how to get people to pick up the phone.
7. Establish rapport with adult phone owners and youth respondents.

SAMPLING

There are three different sampling techniques covered in this review:

- **Census** samples are used to collect data on or from all participants in an initiative or the population exposed or, in the case of radio and television, reached. Not everyone will respond to phone calls, messages, or give permission for tracking, but they are included in the sampling frame.
- **Representative**, also known as probability, samples are a subset of the total population (e.g., participants in an intervention, targeted viewers, or listeners). Participants are randomly selected using statistical methods and following procedures that ensure different segments of the population have equal probabilities of being chosen to be in the sample.[57]
- **Purposive** samples are also used to gather data from a subset of the total population but are not randomly selected. Rather, the researcher uses their knowledge of the learners, educators, and caregivers to select specific participants per the research or evaluation questions. Purposive sampling can help with equity and inclusion, as researchers can target participants who might otherwise be excluded from data collection.

With all three types of sampling, it is necessary to have information on all participants in advance, including sufficient demographic data to know who is being reached and not reached. The most robust way to collect data on demographics is through in-person household surveys, though these are often costly and time-consuming. While the widely used household survey, Demographic and Health Survey, is in the process of being developed for a remote data collection system, at the time of this review no remote household surveys have been fully implemented into practice.[58]

For representative, quantitative samples, define the target learner population and, if feasible, construct a sampling frame with as many demographic data about individuals and households as possible. It is essential to consider which sub-populations within the representative sampling frame are marginalized and may be difficult to reach. When determining the sampling frame, it is important to stratify to ensure these marginalized sub-populations are included in data collection and analysis and analyses can be reported by different marginalized groups.

HOW INTERFACES ARE USED

In addition to technology choices regarding paper-based formats, mobile phones, tablets, and computers, there are many interface options to be considered. Table 4 outlines the major categories of software as well as the data requirements, considerations, and strength of the evidence for their use. The strongest evaluative evidence that currently exists is for paper-based, in-person data collection, and backend data analytics programmed into apps and websites. Moderate to strong evidence exists for survey software, which has been increasingly deployed over the past decade. Messaging services and IVR are more recent technologies but are quite promising for collecting reach and engagement data. They also accommodate participants with low reading levels and low vision. The strength of evidence is based on the definitions in the USAID *Landscape Report on Early Literacy* (2016) (see Annex C for the definitions).

Table 4: Interfaces by considerations and strength of evidence

	INTERFACE	TECHNOLOGY FOR COLLECTING DATA	MOBILE OR INTERNET DATA	CONSIDERATIONS	STRENGTH OF EVIDENCE
1	Paper	Paper Oral *Data can be entered into paper, mobile, tablet, or computer by data collector	None (for participant)	Can serve areas with limited Internet and electricity Can be a challenge to distribute in remote areas when in-person contact is not feasible Visually impaired and low literacy users require text to be read orally	Strong In-person, paper-based assessment has been widely used and evaluated
2	Survey Softwares (Tangerine, KoBo Toolbox, Survey to Go, Survey Monkey, Qualtrics)	Mobile Tablet Computer	2-3G Downloaded through app and used offline	Visually impaired and low literacy users require text to be read orally Knowledge of how to complete a survey if participants (not data collectors) are doing data entry	Moderate-Strong Remote surveys increasingly used and evaluated
3	SMS (Short Message Service) MMS (Multimedia Messaging Service)	Mobile	Phone data for SMS and MMS, Internet connection or phone data for WhatsApp	Visually impaired and low literacy users require text to be read orally	Emerging Increasingly used, but limited in evaluation evidence
4	Computer-Assisted Telephone Interviewing (CATI)	Mobile	Phone call	Hearing-impaired users would require text-based access	Emerging Increasingly used, but limited in evaluation evidence (for education)
5	Interactive Voice Response (IVR)	Mobile (pre-recorded)	Phone data	Hearing-impaired users would require text-based access	Emerging Increasingly used, but limited in evaluation evidence (for education)
6	Cookies, Analytics, and Logfiles	Phone Tablet Computer	2-3G Downloaded through App and offline	Visually impaired and low literacy users require text to be read orally Ethical considerations of tracking data without full awareness of users Clearing the cache or changing browser settings causes all the cookies to be lost	Strong (for smartphones, tablets, and computers) Emerging (for basic phones)

FEEDBACK LOOPS AND COMMUNITIES OF PRACTICE FOR TROUBLESHOOTING TEACHING, LEARNING, AND TECHNOLOGY

When learning suddenly pivots from in-person to remote teaching and learning, educators, learners, and caregivers are often left without adequate resources to successfully take up distance learning. A number of interfaces, outlined below, have been developed to support users in this process. The list below is not exclusive. In addition to those mentioned, Facebook, Instagram, Snapchat, and TikTok are widely used as feedback loops, and new options are continually being developed. These practices help facilitate better data collection practices.

IVR SUPPORT SYSTEM. The Education Initiative Plus in Northern Nigeria uses an IVR system to provide health messaging, psychosocial support, caregiver support, and 200 basic literacy lessons. The system automatically collects data on who called in (reach), measures satisfaction with the lessons (reach/engagement), and provides an opportunity to request a coaching phone call for extra help. When caregivers request assistance through the IVR system, they are automatically sent an SMS with phone numbers for several coaches. Simultaneously, a coach receives a message that a caregiver has requested help. This two-pronged approach means that either the caregiver or the coach can initiate a help call.

CALL-IN LINES. EdTech Hub has been working with BRAC and the Swedish Committee for Afghanistan (Svenska Afghanistankommittén) to pilot a free call-in line between 50 educators and their existing learners. These phone calls cover socioemotional and well-being support and encourage learners to tune into radio and television programming.[59]

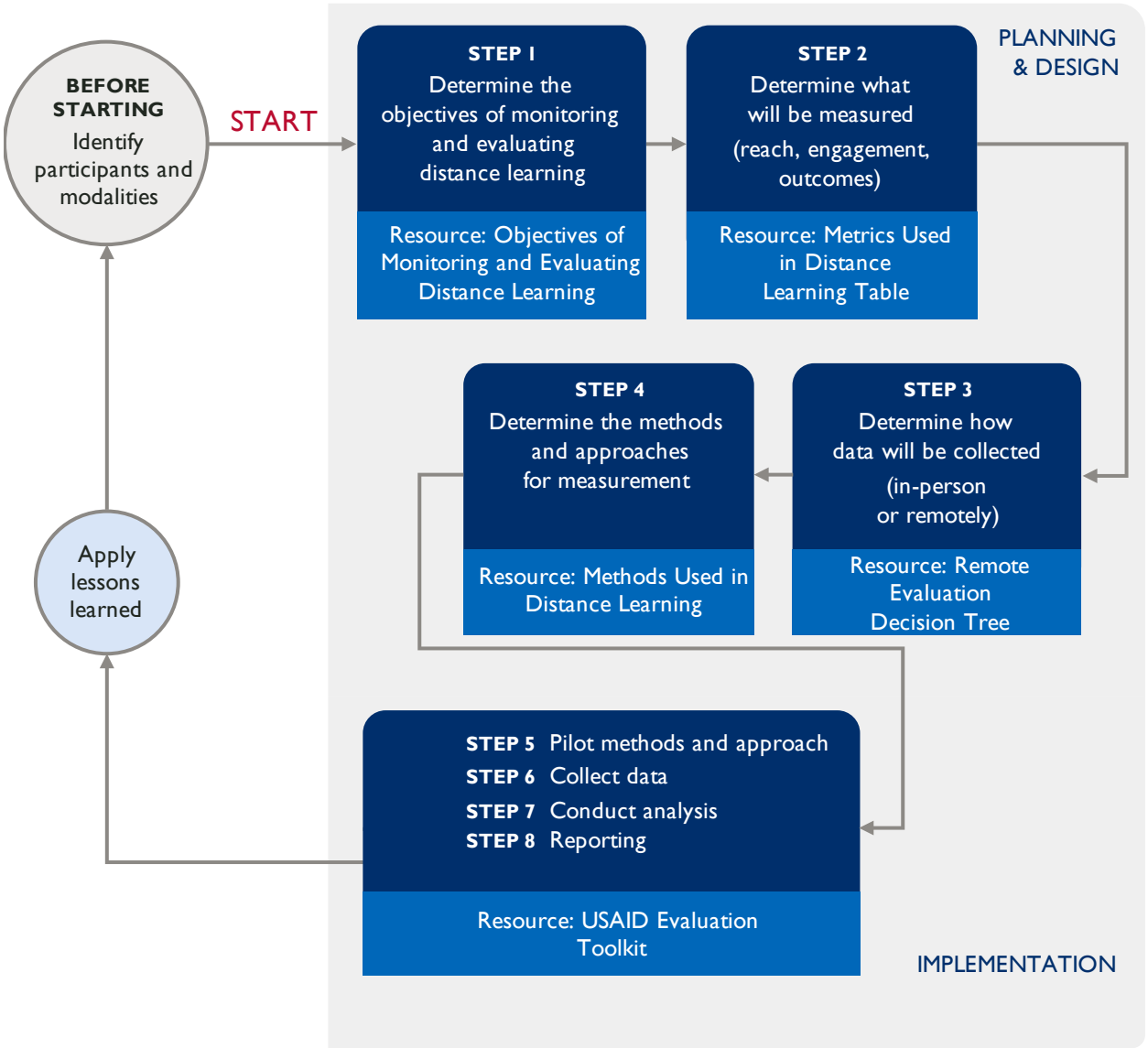
WHATSAPP GROUPS. Young 1ove used WhatsApp groups to train data collectors for remote data collection by sending voice notes, video clips, and pictures, as well as answering questions. In addition to training data collectors, some teacher training has also been moved to WhatsApp. RTI has been working with educators in Tanzania to improve SEL competencies in educational institutions using PowerPoint, video recordings, and asking and answering questions. Now that educational institutions have reopened in Tanzania, the WhatsApp group is being used to answer educators' questions and collect preliminary data on their classroom experiences. Worldreader set up WhatsApp groups with library partners including the Kenyan National Library. The librarians used the groups to share examples of resources, such as training tips and pictures of activities being conducted in the libraries. They also supported each other with solutions to technical issues they came across and shared stories of outreach to community partners.

CONCLUSION

This review provides guidance for teams in the process of designing and planning how to measure distance learning reach, engagement, and outcomes. A comprehensive MEL approach to measuring distance learning includes long-term and short-term strategies. Distance learning can be used to enhance in-person education, extend learning to those not reached through formal education systems, and provide children, youth, and adults with critical skills and knowledge.

This review is structured around three domains of measuring distance learning: reach, engagement, and outcomes. Comprehensive MEL strategies for distance learning should follow the roadmap. This review focuses specifically on the first four steps.

Roadmap for measuring distance learning



Step 1: Determine the Objectives of Monitoring and Evaluating Distance Learning

As distance learning modalities are planned and designed, the objectives of monitoring and evaluating distance learning must be determined. This includes deciding if an assessment is for *formative* purposes—to inform program content, teaching and learning, or program management—or *summative* purposes—to draw conclusions on change in knowledge, skills, attitudes, and behaviors.

Step 2: Determine What Will Be Measured (Reach, Engagement, and Outcomes)

Measuring reach, engagement, and outcomes is important for understanding the effectiveness of distance learning, but this review reveals that the actual metrics for gathering this data can vary greatly by modality and population.

Step 3: Determine How Data Will Be Collected (In-Person or Remotely)

In-person data collection has been the precedent for measuring the efficacy and quality of distance learning, and is still the preferred approach for measuring distance learning outcomes. However, Ebola, COVID-19, natural disasters, and other emergencies have pushed teams to think of new and innovative methods for capturing data remotely.

Step 4: Determine the Methods and Approaches for Measurement

The methods, technology, and data collection interfaces must be determined at the onset of planning to ensure the content reaches intended audiences, is captivating, and helps learners and educators achieve greater socioemotional and physical well-being, skills, and knowledge.

In all four steps, ethics in data collection need to take priority. All data collection teams interviewed for this review, as well as documents analyzed, emphasized the principle of protecting participants and doing no harm.

This review presents three overarching recommendations for teams developing MEL strategies for distance learning.

1. **Integrate in-person and remote approaches, use multi-modal interfaces, and employ mixed methods to measure distance learning.** Reach and engagement can be increasingly measured remotely, using either mobile phone surveys or calls (especially for radio/audio and television/video programming) or through pre-programmed analytics in mobile phone apps and online learning management systems (LMSs). In-person data collection is still preferable for measuring outcomes as it allows for controlling reliability, validity, and precision, as well as ensuring ethics and confidentiality. In addition to a mix of in-person and remote approaches, using multi-modal interfaces (e.g., phone calls and SMS surveys), and mixed methods (e.g., interviews, surveys, and photographs) can help increase who is being reached, potentially capture unintended (shadow) audiences, and promote better analyses.
2. **Encourage innovative solutions to measure reach, engagement, and outcomes during a quick pivot to distance learning, while also developing high-quality MEL strategies for the longer term.** As the case studies and interviews for this review revealed, teams around the world are using creative solutions to collect critical data during sudden learning institution closures. However, best practices in measuring distance learning require intentional and long-term planning. The guidance in this review helps teams address short-term MEL needs while working toward long-term MEL strategies.
3. **Design equitable monitoring and evaluation approaches and conduct systematic equity analyses of distance learning initiatives.** Distance learning has the potential to include marginalized learners, but it can also create digital and learning divides and exclude

learners who do not have access to technological infrastructure and devices or programming and content. Evaluative approaches to distance learning must attempt to measure and analyze where marginalized individuals and groups are being systematically included or excluded through distance learning programming as well as in-person and remote data collection. Measuring reach and engagement through remote methods can be more time- and cost-efficient than in-person tracking and provide real-time data, but strategies to ensure that marginalized individuals and groups who cannot be reached through mobile phones and other remote data collection approaches must be put in place.

Distance learning is now, more than ever, a part of the education landscape. As the international education community continues to learn lessons about monitoring and evaluating distance learning, implementing teams will continue to refine how they measure reach, engagement, and outcomes. It is important that initiatives achieve reach and engagement before expecting them to attain outcomes. It is also of paramount importance to build measures to identify who is not being reached and why in order to design or redesign content and programming to be fully inclusive. This review and roadmap contribute to efforts to build more comprehensive distance learning MEL strategies that support short- and long-term improvements in the quality and impact of distance learning worldwide.

EVALUATIONS REVIEWED

RADIO EVALUATIONS REVIEWED

Barnett Sarah, Jetske Van Dijk, Abdulai Swaray, Tamba Amara, and Patricia Young. "Redesigning an education project for child friendly radio: a multisectoral collaboration to promote children's health, education, and human rights after a humanitarian crisis in Sierra Leone." *BMJ,* 2018; 363:k4667. https://doi.org/10.1136/bmj.k4667

Bosch, Adrea, Lisa Hartenberger Toby, and Abdul Rahman Alhamzy. "In a world of exploding possibilities in distance learning, don't forget about the light bulb." *Quarterly Review of Distance Education*, 16, no. 2 (2017): 129-138.

Carfax Projects and AET International. *Final evaluation report: evaluation of Africa Educational Trust's Speak Up II in South Sudan.* Carfax Projects, 2019.

Christina, Rachel and Nathalie Louge. *Expanding access to early childhood development using interactive audio instruction*. World Bank Group and Education Development Center, 2015.

Creative Associates International, Inc. *An Evaluation of the FQEL Education Program in Guinea*. Prepared for the United States Agency for International Development, 2002.

Deloitte. *Evaluation of Meena Radio Programme*. UNICEF, 2015.

Dock, Alan., and John Helwig, eds. *Interactive Radio Instruction: Impact, sustainability, and future directions.* Washington, DC: World Bank and United States Agency for International Development, and Education Development Center, Inc., 1999.

Early Childhood and Family Education Activity (EDIFAM). *Proyecto IRI en El Salvador: Una Propuesta Aplicada a la Educacion Inicial y Parvularia.* Prepared for the United States Agency for International Development, 2005.

Education Development Center. *Final report of the Somali Interactive Radio Instruction Program*. Prepared for the United States Agency for International Development. Education Development Center. n.d.

Education Development Center. *Literacy, Language, and Learning initiative (L3): national fluency and mathematics assessment of Rwandan schools.* Prepared for the United States Agency for International Development. Education Development Center, 2017.

Evans, Norma, and Daniel Pier. *Interactive radio usage and its impact on grades 1 and 2 teachers and students – midterm study of the Appui Technique aux Educateurs et Communautes (ATEC) Program, Madagascar.* Education Development Center, 2008.

Farm Radio International. *Final report: Her Farm Radio in Ethiopia, Malawi, Tanzania and Uganda. Prepared for the International Fund for Agricultural Development.* Ottawa, Ontario: Farm Radio International, 2017a.

Farm Radio International. *GeoPoll technical report: Farm Radio International Uganda - Listenership of Bushenyi FM radio in Bushenyi and Kasese.* Farm Radio International, 2017b.

Gilberds, Heather, Adena Brown, and Mark Leclair. *Interactive Radio program report: an integrated approach to addressing the issue of youth depression in Malawi and Tanzania.* Farm Radio International, 2016.

GlobalEdTechHub. "Over the last 2 months, we've been working with @BRACworld & @SAK_Sweden to test a free telephone helpline connecting students and teachers in rural #Afghanistan Here are 6 things we've learnt from our #MVP - testing the concept with 50 teachers + their existing students." Twitter, July 23, 2020. https://twitter.com/GlobalEdTechHub/status/1286282087348174849

Hartenberger, Lisa and Andrea Bosch. *Making Interactive Radio Instruction even better for girls: the data, the scripts and the potential.* United States Agency for International Development, 1996.

Ho, Jennifer and Hetal Thukral. *Tuned in to student success: assessing the impact of Interactive Radio Instruction for the hardest to reach.* Washington, DC: Education Development Center, 2009.

Leigh, Stuart. *Changing times in South Africa: remodeling interactive learning.* LearnTech case study series no. 8. Prepared for the United States Agency for International Development. Washington, D.C.: Education Development Center, 1995.

Leigh, Stuart and Andrew Epstein. *South Sudan Interactive Radio Instruction performance evaluation report.* Prepared for the United States Agency for International Development. Washington, DC: Management Systems International, 2012. https://www.eccnetwork.net/resources/south-sudan-interactive-radio-instruction-performance-evaluation-report

Leigh, Stuart and Frances P. Cash. "Effectiveness and Methodology of Interactive Radio Instruction." In *Interactive Radio Instruction: Impact, Sustainability, and Future Directions* by Alan Dock and John Helwig, eds., 27-35. Education and Technology Technical Notes Series 4, no. 1 (1999). The World Bank and the United States Agency for International Development.

Letshabo, Kathleen. *Somali Interactive Radio Instruction Program (SIRP) – Grade 1 Evaluation 2006/2007.* Education Development Center, Inc, 2007.

Letshabo, Kathleen. Multiple reports. *Learning at Taonga Market – Evaluation of Interactive Radio Instruction.* Grades 1, 2, and 3. Education Development Center, Inc., 2000-2007.

Midling, Michael, Louise Filion, Emmanuel M. David-Gnahoui, Mbarou Gassama-Mbaye, Amadou T. Diallo, and Abdoul K. Diallo. *Program evaluation for USAID/Guinea basic education program portfolio.* Prepared for the United States Agency for International Development. Arlington, VA: DevTech Systems, Inc., 2006.

Morris, Emily, Abraham Faki Othman, James Mitchell, and Miriam Philip. *Radio Instruction to Strengthen Education (RISE) in Zanzibar. Learning gains assessment in Zanzibar: more than child's play. Prepared for the United States Agency for International Development.* Washington, DC: Education Development Center, 2009. https://www.sciencedirect.com/science/article/abs/pii/S0193397313000026

Naslund-Hadley, Emma, Juan Manuel Hernandez-Agramonte, and Susan Parker. "Math in my school: improving basic math skills in Paraguay." Innovations for Poverty Action, 2015.

Naslund-Hadley, Emma, S. W. Parker, and Juan Manuel Hernandez-Agramonte. "Fostering early math comprehension: experimental evidence from Paraguay." *Global Education Review* 1, no. 4 (2014): 135-154. https://www.edc.org/fostering-early-math-comprehension-experimental-evidence-paraguay

Royer, James. *EDA/Edikasyon a Distans pou Ayiti Project – Final Report for the 2003-2004 Project.* Education Development Center, Inc., 2004.

Royer, James. Multiple reports. *Pakistan Education Sector Reform Assessment Program – IRI Student Assessment Report.* Education Development Center, Inc., 2004-2006.

Ruiz, Richard. "Orientations in language planning." *NABE Journal* 8, no. 2 (1984): 15-34.

Rumie. "Social Impact." Rumie. n.d. https://rumie.org/impact

Sarrassat Sophie, Nicholas Meda, Hermann Badolo, Moctar Ouedraogo, Henri Some, Robert Bambara, Joanna Murray et al. "Effect of a mass radio campaign on family behaviours and child survival in Burkina Faso: a repeated cross-sectional, cluster-randomised trial." *The Lancet Global Health* 6 no. 3 (2018): E330-E341. https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30004-4/fulltext

Simpson, Robert. *Can Interactive Radio Instruction turn post-conflict educational challenges into opportunities? A case study of the 'Speak Up!" English language programme, South Sudan.* Liverpool School of Tropical Medicine, 2013.

South Sudan Interactive Radio Instruction Program and Directorate of Alternative Education, MoE. *Effectiveness of primary four Interactive Radio Instruction in Southern Sudan, Summative Evaluation*. 2011.

Thrukal, Hetal. "Examining the impact of teacher practice on student learning in Interactive Radio Instruction (IRI) classrooms: lessons from New Delhi and Rajasthan (India)." PhD diss., University of Maryland, College Park, 2016.

Tilson, Thomas, D. T. Jamison, M. Fryer, P. Godoy-Kain, and M. Imhoof. "The cost effectiveness of Interactive Radio Instruction for improving primary school instruction in Honduras, Bolivia and Lesotho." Paper presented at the CIES Annual Conference in Pittsburgh, Pennsylvania, March 15, 1991.

VIDEO AND TELEVISION EVALUATIONS REVIEWED

Akhter, Nasreen. "Evaluation of educational television programs for distance learning." *The Turkish Online Journal of Educational Technology* 10, no. 4: 188-194.

Alade, Fashina., and Amy I. Nathanson. "What preschoolers bring to the show: the relation between viewer characteristics and children's learning from educational television." *Media Psychology* 19 (2016): 406-430.

Bayan Academy for Social Entrepreneurship and Human Resource Development, Inc. *Impact evaluation of Puno Ng Buhay program videos on environment*. Prepared for Knowledge Channel Foundation, Inc., 2014.

Baydar, Nazli, Cigdem Kagitcibasi, Aylin C. Kuntay, and Fatos Goksen. "Effects of an educational television program on preschoolers: variability in benefits." *Journal of Applied Developmental Psychology* 29 (2008): 349-360.

Bogatz, Gerry Ann, and Sameul Ball. *The second year of Sesame Street: a continuing evaluation*. Volume 1. Princeton, NJ: Educational Testing Service, 1971.

Borzekowski, Dina L. G. "A quasi-experiment examining the impact of educational cartoons on Tanzanian children." *Journal of Applied Developmental Psychology* 54 (2018): 53-59.

Borzekowski, Dina L. G. and Holly K. Henry. "The impact of Jalan Sesama on the educational and healthy development of Indonesian preschool children: an experimental study." *International Journal of Behavioral Development* 35, no. 2 (2010): 169-179.

Borzekowski, Dina L. G., Agnes Lucy Lando, Sara H. Olsen, and Lauren Giffen. "The impact of an educational media intervention to support children's early learning in Rwanda." *International Journal of Early Childhood*, 51 (2019): 109-126. https://doi.org/10.1007/s13158-019-00237-4

Borzekowski, Dina L. G., and Jacob E. Macha. "The role of Kilimani Sesame in the healthy development of Tanzanian preschool children." *Journal of Applied Developmental Psychology* 31 (2010): 298-305.

Borzekowski, Dina L. G., L. E. Kauffman, E. A. Dura, M. Chale, and C. B. Bauer. *Examining the impact of Akili and Me's new content on Tanzanian children*. Prepared for UBONGO and the Human Development Innovation Fund. College Park, MD: University of Marylandl, 2020.

Broadbent, Emma, Leonora Dowley, and Muniratu Issifu. "Interactive distance instruction in girls' education: evidence from Varkey Foundation Ghana's Making Ghanaian Girls Great! project." Conference on Girls' Education in Africa: Paper Submission, 2017.

Chesterton, Paul. *Evaluation of the Meena Communication Initiative.* UNICEF Regional Office for South Asia Kathmandu, 2014.

Cruz, Priscila, Flavia Goulart, Christina Kwauk, and Jenny Perlman Robinson. *Media center: innovating with distance learning in Amazonas, Brazil.* Center for Universal Education at Brookings, 2016.

Ince, Deborah. "'Tsehai Loves Learning' speaks to young readers in Ethiopia." USAID Frontlines, 2014. https://2012-2017.usaid.gov/news-information/frontlines/afghanistan/tsehai-loves-learning-speaks-young-readers-ethiopia

Johnston, Jamie, and Christopher Ksoll. "Effectiveness of interactive satellite-transmitted instruction: experimental evidence from Ghanaian primary schools." CEPA Working Paper No. 17-08. Stanford Center for Education Policy Analysis, 2017. http://cepa.stanford.edu/wp17-08

Kearney, Melissa S. and Phillip B. Levine. "Early childhood education by television: Lessons from Sesame Street." *American Economic Journal: Applied Economic 2019,* 11, no. 1 (2019): 318-350.

Mapa, Dennis S. *Impact study for the Proficient Measures for Quality Education (PMQE) Project of the Knowledge Center.* University of the Philippines Diliman, 2008.

Mapa, Dennis S., and Kristine Joy Briones. *Evaluation of the impact of presence of Knowledge Channel on students' perception in ten (10) selected schools supported by the British Embassy in the Philippines*. UP Statistical Center Research Foundation, Inc., 2007.

Mapa, Dennis S., Kristine Briones, and Albert Kirby A. Tardeo. *Outcome assessment study of Knowledge Channel's Television Education for the Advancement of Muslim Mindanaeo (TEAM-M) project*. UP Statistical Center Research Foundation, Inc., 2008.

Mares, Marie-Louise and Zhongdang Pan. "Effects of Sesame Street: a meta-analysis of children's learning in 15 countries." *Journal of Applied Developmental Psychology* 34 (2013): 140-151. https://doi.org/10.1016/j.appdev.2013.01.001

Parmanik, Rupali, Shakti Prakash, and Gayan Prakash. "An empirical 'KAP' study of students' DD-gyan Darshan Channel programmes: District Gautam Buddha Nagar." *Global Journal of Enterprise Information System* 2, no. 2 (2010): 89-93.

Patrinos, Harry, Joseph Shapiro, and Jorge Moreno. *Education for all: compensating for disadvantage in Mexico*. The World Bank, 2004.

Tadesse, Michael Emru. *Tsehai Loves Learning: an exemplary practice from Ethiopia promoting child rights to education.* 2018. https://www.researchgate.net/publication/336799697_Tsehai_Loves_Learning_An_Exemplary_Practice_from_Ethiopia_Promoting_Child_Right_to_Education

Wang, Hua, and Arvind Singhal. "Audience-centered discourses in communication and social change: the 'Voicebook' of Main Kuch Bhi Kar Sakti Hoon, an entertainment-education initiative in India." *Journal of Multicultural Discourses*, 13, no. 2 (2018): 176-191. https://doi.org/10.1080/17447143.2018.1481857

Zaidi, Mosharraf and Umar Nadeem. "Pakistan: tackling Covid-19 in education." Tabadlab, 2020. https://www.tabadlab.com/publications/covid-19-response-briefing/pakistan-tackling-covid-19-in-education/

MOBILE PHONE EVALUATIONS REVIEWED

Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert. "Can mobile phones improve learning? Evidence from a field experiment in Niger." *American Economic Journal: Applied Economics* 4 no. 4 (2012): 94-120.

Angirst, N., P. Bergman, D. K. Evans, S. Hares, M. C. H. Jukes, and T. Letsomo. *Practical lessons for phone-based assessments of learning*. BMJ Global Health, 2020.

Angrist, Noam, Peter Bergman, Caton Brewster, and Moitshepi Matsheng. "Stemming learning loss during the pandemic: a rapid randomized trial of a low-tech intervention in Botswana." Centre for the Study of African Economics Working Paper WPS/2020-13, 2020.

Butgereit, Laurie. *Math on MXit: using MXit as a Medium for Mathematics Education*. Meraka Institute, 2007. http://hdl.handle.net/10204/1614

EdTechHub. "Afghanistan Helpline Sandbox: Sprint One Review". *EdTechHub,* 2020.

EdTechHub. "Afghanistan Helpline Sandbox - Sprint 2 Review". *EdTechHub,* 2020.

GlobalEdTechHub. "Over the last 2 months, we've been working with @BRACworld & @SAK_Sweden to test a free telephone helpline connecting students and teachers in rural #Afghanistan Here are 6

things we've learnt from our #MVP - testing the concept with 50 teachers + their existing students." Twitter, July 23, 2020. https://twitter.com/GlobalEdTechHub/status/1286282087348174849

Gultig, John. *African Storybook Initiative External 'accountability' evaluation: 2013 – 2016.* African Storybook Initiative, 2017. https://www.saide.org.za/documents/2017_03_29_ASb_accountability_evaluation.pdf

Hanemann, Ulrike, ed. *Mobile-based post literacy programme, Pakistan*. UNESCO Institute for Lifelong Learning, 2013. https://uil.unesco.org/case-study/effective-practices-database-litbase-0/mobile-based-post-literacy-programme-pakistan

Jantjies, Mmaki and Mike Joy. "Teaching through mobile technology: a reflection from high school studies in South Africa." In *Handbook of Research on Instructional Systems and Educational Technology*, ed., Terry T. Kidd and Lonnie R. Morris, Hershey, PA: IGI Global, 2017, 299-312. http://wrap.warwick.ac.uk/91929/

Kaleebu, Nasiib, Alison Gee, Nick Maybanks, Richard Jones, Marshall Jauk, and Amanda H. A. Watson. "SMS story: Early results of an innovative education trial." *Contemporary PNG Studies: DWU Research Journal* 19 (2013): 50-62.

Kam, Matthew, Anuj Kumar, Shirley Jain, Akhil Mathur, and John Canny. *Improving literacy in rural India: cellphone games in an after-school program.* 2008.

Pouezevara, Sarah and Simon King. *MobiLiteracy-Uganda program. Phase 1: Endline Report*. Prepared for Urban Planet. RTI International. 2014.

Raftree, Linda. *Landscape review: mobiles for Youth Workforce Development.* Prepared for The Mastercard Foundation. Rockville, MD: JBS International, 2013.

Ramos, Angelo Juan O., Felix Liberero, Jerome P. Triñona, and Adelina I. Ranga. "Using a ubiquitous technology for m-Learning in Asia: Project MIND in the Philippines." 2nd International Conference on e-Learning, 2007.

Roberts, Nicky. and Riitta Vänskä. "Challenging assumptions: mobile learning for mathematics project in South Africa." *Distance Education* 32, no. 2 (2011): 243-259. http://dx.doi.org/10.1080/01587919.2011.584850

Roberts, Nicky, Garth Spencer-Smith, and Neil Butcher. *An implementation evaluation of the ukuFUNda virtual school.* Johannesburg: Kelello, 2016. https://www.researchgate.net/publication/318902728_An_implementation_evaluation_of_the_ukuFUNda_virtual_school

School-to-School International. *Your child, reading, and you*. Prepared for All Children Reading: A Grand Challenge for Development. School-to-School International, 2017.

ONLINE AND COMPUTER EVALUATIONS REVIEWED

Hennessy, Sarah, Kenneth Ruthven, and Sue Brindley. "Teacher perspectives on integrating ICT into subject teaching: commitment, constraints, caution, and change." *Journal of Curriculum Studies* 37, no. 2 (2005): 155-192. https://doi.org/10.1080/0022027032000276961

Jugee, Sarvesh, and Mohammad Santally. "The tablet PC Initiative in Mauritius: a situational analysis." *International* e-*Journal of Advances in Education* 2, no. 4 (2016): 14–22.

Moon, Chris J., Allison Kavanagh, Jackie Jeffrey, and Joseph N. Gebbels. "Social entrepreneurship and disruptive innovation: evaluating the use of Rumie's free educational software in seven developing economies.' European Conference on Innovation and Entrepreneurship Education 2016. 2017.

Piper, Benjamin, Evelyn Jepkemei, Dunston Kwayumba, and Kennedy Kibukho. "Kenya's ICT policy in practice: the effectiveness of tablets and e-reader in improving student outcomes." *Forum for International Research in Education* 2, no. 1 (2015): 3-18. http://dx.doi.org/10.18275/fire201502011025

Scoular, Claire, Esther Care, and Nafisa Awwal. "An approach to scoring collaboration in online game environments." *The Electronic Journal of e-Learning* 15, no. 4 (2017): 335-342.

Wanjiru, Benson, Miheso O'Connor, and S. Ndethiu. "Effects of mathematical vocabulary instruction on students' achievement in mathematics in secondary schools of Murang'a county, Kenya." Paper presented at the 4th International Conference on Education, Nairobi, Kenya, 2015.

OTHER EVALUATIONS AND SOURCES REVIEWED

Please see endnotes for a complete list of all sources cited in the text.

Almasri, Nada, Luay Tahat, Sawsan Skaf, and Aman Al Masri. "A digital platform for supervised self-directed learning in emergencies: the case of the Syrian crisis." *Technology, Pedagogy, and Education* 28, no. 1 (2019): 91-113. https://doi.org/10.1080/1475939X.2019.1568293

Ampofo, Samuel Yaw, Benjamin Bizimana, John Mbuthi, Irenee Ndayambaje, Norbert Ogeta, and John Aluko Orodho. "Information communication technology penetration and its impact on education: lessons of experience from selected Africa countries of Ghana, Kenya, and Rwanda." *Journal of Information Engineering and Applications* 4, no. 11 (2014): 84-96.

Amporo, Angelica Towne and Hawah Nabbuye. "Taking distance learning 'offline': Lessons learned from navigating the digital divide during COVID-19." *Brookings* (blog), August 7, 2020. https://www.brookings.edu/blog/education-plus-development/2020/08/07/taking-distance-learning-offline-lessons-learned-from-navigating-the-digital-divide-during-covid-19/

Amuko, Sheila. "Pedagogical practices in integration of ICT in teaching and learning mathematics, in secondary schools in Nairobi County, Kenya." *ISOR Journal of Mathematics* 11, no. 5 (2015): 20-23.

Asadullah, Niaz. "COVID-19, schooling, and learning." PowerPoint. 2020. https://bigd.bracu.ac.bd/wp-content/uploads/2020/06/COVID-19-Schooling-and-Learning_June-25-2020.pdf

Asego, Chris. "Eneza Education: Meet the entrepreneurs." Eneza Education, 2013.

Barakabitze, Alcardo Alex. "The context of education initiative, importance and inhibitors of ICTs towards improving teaching and learning in Tanzania: a critical literature review." *Information and Knowledge Management* 4, no. 10 (2014): 83-97.

Barry, Barbara and Landon Newby. *Use of technology in emergency and post-crisis situations*. Global Education Cluster Working Group and IIEP-UNESCO, 2012.

Boulay, Marc, J. Douglas Storey, and Suruchi Sood. "Indirect exposure to a family planning mass media campaign in Nepal." *Journal of Health Communication* 7 (2002): 379-399.

Brinkel, Johanna, Jürgen May, R. Krumkamp, M. Lamshoft, Benno Kreuels, Ellis Owusu-Dabo, Aliyu Mohammed, Phyllis Dako-Gyeke, Alexander Krämer, Julius N. Fobil. "Mobile phone-based interactive voice response as a tool for improving access to healthcare in remote areas in Ghana - an evaluation of user experiences." *Tropical Medicine and International Health* 22 no. 5 (2017): 622-630.

Broadband Commission for Sustainable Development. *The state of broadband 2018: Broadband catalyzing sustainable development*. Broadband Commission for Sustainable Development, 2018.

Brown, Lydia X. Z. "How to center disability in the tech response to COVID-19." *Brookings* (blog), July 20, 2020. https://www.brookings.edu/techstream/how-to-center-disability-in-the-tech-response-to-covid-19/?utm_campaign=brookings-comm&utm_medium=email&utm_content=91771230&utm_source=hs_email

Buabeng-Andoh, Charles. "Factors influencing teachers' adoption and integration of information and communication technology into teaching: a review of the literature." *International Journal of Education and Development using Information and Communication Technology* 8, no. 1 (2012): 136-155.

Bukhari, Maqsud Alam. *Broadcast media in distance education.* UNESCO, 1997.

Burde, Dana, Ozen Guven, Heddy Lahmann, and Khaled Al-Abbadi. *What works to promote children's educational access, quality of learning, and wellbeing in conflict-affected contexts*. Department for International Development, London, UK, 2015.

Burns, Mary. *Distance education for teacher training: modes, models and methods.* Education Development Center, 2011.

Burns, Mary. *Technology teaching and learning: Research, experience, & global lessons learned*. Prepared for the United States Agency for International Development. Education Development Center, 2012. https://www.edc.org/technology-teaching-and-learning-research-experience-and-global-lessons-learned

Burns, Mary. (Unpublished). "For want of a good theory: Considerations for technology integration in well-resourced schools."

Carlson, Sam. *Using technology to deliver educational services to children and youth in environments affected by crisis and/or conflict*. Washington, DC: United States Agency for International Development, 2013.

CISCO. *Cisco visual networking index: global mobile data traffic forecast update, 2016–2021*. Cisco white paper, 2017.

Chemonics. "ACCELERE! I Project." *Chemonics.* n.d. https://chemonics.com/projects/accelerating-access-learning-democratic-republic-congo/

Cho, Hyunyi and Charles T. Salmon. "Unintended effects of health communication campaigns." *Journal of Communication* 57 (2007): 293-317. https://doi.org/10.1111/j.1460-2466.2007.00344.x

Corkrey, Ross and Lynne Parkinson. "Interactive voice response: review of studies 1989-2000." *Behavior Research Methods, Instruments, & Computers* 34, no. 3 (2002): 342-353.

David, Raluca, Arnaldo Pellini, Katy Jordan, and Toby Phillips. Education during the COVID-19 crisis: Opportunities and constraints of using EdTech in low-income countries. *The EdTech Hub,* 2020. https://edtechhub.org/coronavirus/edtech-low-income-countries/

Economist, The. "In much of sub-Saharan Africa, mobile phones are more common than access to electricity." *The Economist*, November 8,2017. https://www.economist.com/graphic-detail/2017/11/08/in-much-of-sub-saharan-africa-mobile-phones-are-more-common-than-access-to-electricity

Evans, David. "Teach literacy by text message. Really." *World Bank Blogs*, July 7, 2014. https://blogs.worldbank.org/impactevaluations/teach-literacy-text-message-really

Evans, David. "Can monitoring teachers and students - with no incentive or punishment attached - improve test scores? Yes." *World Bank Blogs*, September 23, 2015. https://blogs.worldbank.org/impactevaluations/can-monitoring-teachers-and-students-no-incentive-or-punishment-attached-improve-test-scores-yes

Fuchs, Thomas and Ludger Woessmann. "Computers and student learning: bivariate and multivariate evidence on the availability and use of computers at home and at school." Ifo Institute for Economic Research at the University of Munich, Working Paper No. 8, 2005.

GeoPoll. n.d. https://www.geopoll.com/

Global Center for the Development of the Whole Child. "Radio Distance Learning." University of Notre Dame. n.d. https://iei.nd.edu/gc-dwc/strong-beginnings-radio

Hallgarten, Joe. "Evidence on efforts to mitigate the negative educational impact of past disease outbreaks." K4D Helpdesk, 2020. https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15202/793_mitigating_education_effects_of_disease_outbreaks.pdf?sequence=6&isAllowed=y

Hayes, Anne, Norma Moran, and Ann Turnbull. *Universal Design for Learning to Help All Children Read: Promoting Literacy for Learners with Disabilities*. REACH and the Global Reading Network. Chevy Chase: MD: University Research Co. LLC, 2019.

Hoyt, Brian. "Haiti's Private Schools." *World Bank Blogs* (blog). *World Bank,* January 22, 2010. https://blogs.worldbank.org/psd/haitis-private-schools#:~:text=The%20country%20has%20a%20total,private%2C%20fee%2Dbased%20schools

Kabir, Fatima S. and Abdullahi Kadage Tukur. "ICTS and educational development: the utilization of mobile phones in distance education in Nigeria." *Turkish Online Journal of Distance Education* 18 no. 1 (2017): 63-76. https://doi.org/10.17718/tojde.285716

Karsenti, Thierry, Toby Harper-Merrett, Djeneba Traoré, Moses Mbangwana, K. Touré. *The PanAfrica research agenda on the pedagogical integration of ICTs.* OECD World Forum, 2009.
Khan, Salman. "I started Khan Academy. We can still avoid an education catastrophe." *The New York Times,* August 13, 2020. https://www.nytimes.com/2020/08/13/opinion/coronavirus-school-digital.html

Kim, Young-Suk Grace, Helen N. Boyle, Stephanie Simmons Zuilkowski, and Pooja Nakamura. *Landscape report on early grade literacy.* Washington, DC: USAID, 2016.

Lewis, Kent and Simon Thacker. *ICT and the education of refugees: A stocktaking of innovative approaches in the MENA region.* SABER-ICT Technical Paper Series, No. 17. Washington, DC: World Bank, 2016. https://openknowledge.worldbank.org/handle/10986/26522

Mader, Jackie. "Can simple text messages for parents boost reading scores for kids?" The Hechinger Report. 2020. https://hechingerreport.org/can-simple-text-messages-for-parents-boost-reading-scores-for-kids/?utm_source=The+Hechinger+Report&utm_campaign=3047a2b627-weekly_2020_07_28_02_14&utm_medium=email&utm_term=0_d3ee4c3e04-3047a2b627-322583341

Małachowska, Agnieszka, N. Jones, B. Abu Hamad, T. Al Abbadi, W. Al Almaireh, S. Alheiwidi, K. Bani Odeh, et. al. *GAGE virtual research toolkit: qualitative research with young people on their COVID-19 experiences.* London: Gender & Adolescence: Global Evidence 2020.

Marchant, Eleanor R. *Interactive Voice Response and radio for peacebuilding: a macro view of the literature and experiences from the field* (CGCS Report). Center for Global Communication Studies, 2016.

McBurnie, Chris and Björn Haßler. "Is there learning continuity during the COVID-19 pandemic?" Six lessons. Open Development & Education (blog), July 12, 2020. https://opendeved.net/2020/07/12/is-there-learning-continuity/

Ministry of Federal Education and Professional Training, Government of Pakistan. "Teleschool." *Government of Pakistan,* 2020. http://mofept.gov.pk/Detail/ZjcxNzllM2QtMmIwNC00ZjhiLWE5NTgtNmM5MzM1MjM0OWQw

Morris, Emily and Anna Farrell. *Delivering Distance Learning in Emergencies: A Review of the Evidence and Best Practice.* Prepared for USAID. EnCompass LLC and MSI, 2020.

Nielsen. "Nielsen Audio." *Nielsen.* n.d. https://www.nielsen.com/us/en/solutions/capabilities/audio/

NISSEM. "Responding to COVID-19 in the Bahamas: interview with Marcellus C. Taylor." UKFIET (blog), July 30, 2020. https://www.ukfiet.org/2020/responding-to-covid-19-in-the-bahamas-interview-with-marcellus-c-taylor/

OECD. *Students, computers and learning: making the connection*. PISA, OECD Publishing, 2015.
Powers, Shawn and Kaliope Azzi-Huck. "The impact of Ebola on education in Sierra Leone." *World Bank Blogs* May 4, 2016. https://blogs.worldbank.org/education/impact-ebola-education-sierra-leone

Randolph, Elizabeth. "Co-creation of teaching activities during COVID-19." *RTI Shared* (blog), May 19, 2020. https://shared.rti.org/content/co-creation-teaching-activities-during-covid-19

Rising Academy Network. (Unpublished). "Rising on air: insights from Sierra Leone. Rising Academy Network."

Room to Read. *2018 research, monitoring, & evaluation report: literacy program*. Room to Read, 2018.

Rothman, Emily F., Timothy Heeren, Michael Winter, David Dorfman, Allyson Baughman, and Gregory Stuart. "Collecting self-reported data on dating abuse perpetration from a sample of primarily black and

hispanic, urban-residing, young adults: A comparison of timeline followback interview and interactive voice response methods." *Journal of Interpersonal Violence* 35, no. 1-2 (2020): 100-126. https://doi.org/10.1177%2F0886260516681154

RTI International. "Tangerine - Mobile learning assessments made easy." RTI International. n.d. https://www.rti.org/impact/tangerine-mobile-learning-assessments-made-easy

Shivshanker, Anjuli, Elena Walls, Saima Malik, Rebecca Pagel, and Chris Ying. *Guidance for USAID Education Sector Implementing Partners: monitoring, evaluation, and learning during the COVID-19 pandemic.* United States Agency for International Development, 2020.

Smith, Tovia. "More states opting to 'Robo-Grade' student essays by computer." *NPR,* June 30, 2018. https://www.npr.org/2018/06/30/624373367/more-states-opting-to-robo-grade-student-essays-by-computer

Strigel, C. "Supporting assessment and family outreaching & considering access, user engagement, and content in making the right technology choice for your audience." *Basic Education Coalition, EdTech Working Group Webinar,* April 27, 2020. https://shared.rti.org/content/virtual-assessment-and-making-right-technology-choices-presentation

Strigel, Carmen. "Looking beyond school closures: considering teacher wellbeing and support." *RTI Shared* (blog), July 21, 2020. https://taleemghar.punjab.gov.pk/#tg-b03

Stritzke, Werner, Justine Dandy, Kevin Durkin, and Stephen Houghton. "Use of Interactive Voice Response (IVR) technology in health research with children." *Behavior Research Methods*, 37 no. 1 (2005): 119-126. https://doi.org/10.3758%2FBF03206405

Taleem Ghar. *Taleem Ghar*. n.d. https://taleemghar.punjab.gov.pk/#tg-b03

Tauson, Michaelle and Luke Stannard. *EdTech for learning in emergencies and displaced settings: a rigorous review and narrative synthesis.* Save the Children, 2018.

Tourangeau, Roger, Darby Miller Steiger, and David Wilson. "Self-administered questions by telephone: evaluating interactive voice response." *Public Opinion Quarterly* 66 (2002): 265-278. https://doi.org/10.1086/340029

Trucano, Michael. "The Matthew Effect in educational technology." *World Bank Blogs*, June 7, 2013. https://blogs.worldbank.org/edutech/matthew-effect-educational-technology

Trucano, Michael. "Evaluating the Khan Academy." *World Bank Blogs,* June 6, 2014. https://blogs.worldbank.org/edutech/evaluating-khan-academy

UBONGO Kids. *UBONGO*. n.d.https://ubongokids.com/apps/

UBONGO Kids. "Kibena and the Math Rats." *UBONGO* n.d http://ubongokids.com/math_rats/

UNESCO. Global Education Monitoring Report 2020: Inclusion and education: all means all. Paris, UNESCO. 2020. https://en.unesco.org/gem-report/report/2020/inclusion

USAID. *USAID Evaluation Policy*. Washington, DC: USAID, 2011.

USAID. "Introduction to Gender-Responsive Teaching Methods." PowerPoint. Washington. DC: United States Agency for International Development, 2018.

USAID Learning Lab. "CLA Toolkit." *USAID.* n.d. https://usaidlearninglab.org/qrg/understanding-cla-0

USAID Learning Lab. "The USAID Program Cycle." *USAID.* n.d. https://usaidlearninglab.org/program-cycle-overview-page

Wang, Hua. and Arvind Singhal. "Audience-centered discourses in communication and social change: the 'Voicebook' of Main Kuch Bhi Kar Sakti Hoon, an entertainment-education initiative in India." *Journal of Multicultural Discourses* 13, no. 2 (2018): 176-191. https://doi.org/10.1080/17447143.2018.1481857

Watson, Joe. "Learning through television in low-income contexts: mitigating the impact of coronavirus (COVID-19)." *The EdTech Hub* (blog), March 31, 2020. https://edtechhub.org/2020/03/31/learning-through-television-in-low-income-contexts-mitigating-the-impact-of-covid-19/

World Bank. "How countries are using edtech (including online learning, radio, television, texting) to support access to remote learning during the COVID-10 pandemic." The World Bank, 2020a. https://www.worldbank.org/en/topic/edutech/brief/how-countries-are-using-edtech-to-support-remote-learning-during-the-covid-19-pandemic

World Bank. "Remote learning and COVID-19." The World Bank, 2020b.

World Bank. "Remote learning, distance education and online learning during the COVID19 pandemic: a resource list by the World Bank's EdTech Team." The World Bank, 2020c.

World Bank. "Remote learning, EdTech, & COVID-19." The World Bank, 2020d. https://www.worldbank.org/en/topic/edutech/brief/edtech-covid-19

Worldreader. *2019 Annual Report.* Worldreader, 2019. https://www.worldreader.org/2019-annual-report/

Worldreader. "BookSmart." *Worldreader.* n.d. https://www.worldreader.org/our-work/technology/booksmart/

Young 1ove. *Young 1ove.* n.d. https://www.young1ove.org/tarl

ANNEXES

ANNEX A: EVIDENCE BASE

All studies or evaluations were analyzed according to their methodological design (e.g., comparison group design, non-comparison group design), sampling, and methods utilized (e.g., surveys, interviews, tests).

Program/Project Evaluations by Instructional Modality

MODALITY	# OF REFERENCES	# OF COUNTRIES	% OF REFERENCES IN EIE CONTEXTS	% OF EXTERNAL EVALUATIONS
Radio	33	24	42%	46%
Video	24	22	21%	83%
Phone	21	48	29%	43%
Online	10	29	10%	90%
Not specific to a program	22	N/A	N/A	N/A

ANNEX B: EXAMPLES OF FORMATIVE AND SUMMATIVE EVALUATIONS

FORMATIVE EVALUATION:	
Needs assessment or landscape analysis	Assessing the context, needs of the community and users, and existing efforts and data (descriptive).
Formative evaluation of a component	Assessing whether a specific component of programming is relevant, responsive, and engaging (normative).
Process or performance evaluation	Assessing the process of program or technology delivery, quality and satisfaction with implementation, and progress towards achieving overarching objectives and goals (normative).
SUMMATIVE EVALUATION:	
Outcome	Measures the extent to which an intervention has led to intended outcomes. See USAID's Evaluation Policy on outcome evaluation.
Impact	Measures the extent to which an intervention has caused intended outcomes (per USAID's Evaluation Policy, impact evaluations must be quasi-experimental or experimental in design, ideally with a comparison group).
Cost-effectiveness and cost-benefit analysis	"Cost-efficiency analysis compares the costs of an intervention to the outputs derived from that intervention. Such analysis is useful when choosing among alternative delivery models for a given output. For instance, cost-efficiency analysis would reveal how much it costs per teacher per year if the intervention provides professional coaching through different delivery modes, or to compare unit costs of books produced using different procurement processes."[60]
Secondary analysis	Re-examines existing data to investigate new questions using approaches and questions not explored previously.
Meta-analysis	Analyses the outcomes or impacts across several evaluations and interventions per an overarching evaluation question.

Meta-analysis and secondary analysis definitions adapted from Trochim n.d.[61]

ANNEX C: DEFINITIONS OF THE STRENGTH OF EVIDENCE

Strength of the body of evidence was adapted from the definitions in USAID's *Landscape Report on Early Literacy* (2016). Note that USAID also has an *Assessing the Quality of Education Evaluations Tool* (2017) that can be used for individual evaluation tools and evaluation designs.

STRENGTH OF EVIDENCE	
Strong	Quantitative data are consistent/reliable, precise/accurate, and allow for clear conclusions to be drawn (including but not limited to establishing correlation or causality). Strong qualitative data are backed by theoretical evidence and allow for an in-depth look at key literacy acquisition domains.
Moderate	Although data exists, for quantitative data there are no strong causal conclusions that can be drawn, and the consistency/reliability and precision/accuracy of measures is not enough to draw clear conclusions. Moderate evidence for qualitative data is informed to some extent by theoretical evidence, but there is a lack of depth and complexity to the data, which makes it hard to come to clear conclusions.
Limited / No Data	There is either no data available, or no analyses can be made.
Emerging	There is very little data collected and/or the quality of the data is low. For quantitative data, no correlations or associations can be made, and data are limited to descriptive information. For qualitative data, there is no clear underlying theoretical evidence to support the data.

ANNEX D: CASE STUDIES

CASE	IMPLEMENTER	DISTANCE LEARNING MODALITIES	MEASURES
Measuring Well-being of Girls in Eight Countries During COVID-19	Room to Read	Mobile	Reach Engagement Outcomes
GeoPoll Phone Surveys	GeoPoll	Radio Television Print Online	Reach Engagement Outcomes
Mobile Phone Monitoring of Adolescent Girls' Well-being in Nepal	People in Need	Mobile	Reach Engagement Outcomes
Monitoring Reach and Engagement through Phone Surveys and Sharing through Dashboard	EDC	Radio	Reach Engagement
Monitoring Engagement and Literacy Outcomes through Phone Call Surveys in Haiti	Alliance for Catholic Education	Radio Paper	Reach Engagement Outcomes
Measuring Reach and Engagement through Analytics on a Reading App	Worldreader	Mobile Online	Reach Engagement
Measuring Formative Numeracy Outcomes through Phone Surveys	Young 1ove	Mobile	Reach Engagement Outcomes
Measuring Reach, Engagement, Knowledge, and Practices of Radio Programs Through a Multi-Modal Interface	Farm Radio International	Radio	Reach Engagement Outcomes
Measuring Formative Literacy and Health Outcomes through Phone Surveys in the DRC	Chemonics	Radio	Reach Engagement Outcomes
Integration of Software on Monitoring Plus Video Access in Pakistan	Government of Pakistan and Government of Punjab	Television/video	In development
Capturing Reach, Engagement, and Outcomes for Radio, Television, and Educational Apps: A Multi-modal and Mixed-Methods Approach	Ubongo	Radio Television/radio Mobile Paper	Reach Engagement Outcomes
Monitoring and Supporting Teachers During Distance Learning in Tanzania	RTI	Mobile	Engagement Outcomes

MEASURING WELL-BEING OF GIRLS IN EIGHT COUNTRIES DURING COVID-19

Initiative: Girls' Education Program	**Implementers:** Room to Read	**Location/s:** Bangladesh, Cambodia, India, Laos, Nepal, Sri Lanka, Tanzania, Vietnam
Distance learning modality(s): Mobile phone, limited online	**Technology for collecting data:** Mobile phones, computers	**Interface for collecting data:** Phone call or video call, voice survey

Program description: Room to Read's Girls' Education Program works with 11–18 year-old girls across eight countries (Bangladesh, Cambodia, India, Laos, Nepal, Sri Lanka, Tanzania, Vietnam) on a life skills and mentoring curriculum that covers topics like self-confidence, expressing and managing emotions, empathy, self-control, decision-making, perseverance, communication, creative problem-solving, and relationship building. Social mobilizers, who are trained women from local communities that facilitate the programming, lead classes and in-person mentoring, both individually with girls, as well as in a small group setting. Prior to COVID-19, all sessions were delivered on the school premises, either during school hours or after class. The program also has a robust family, school, and community engagement component. Girls are part of a cohort that stays together over the course of the 6-7-year program.

In response to COVID-19, Room to Read has had to pivot to remote mentoring as well as shifting to collecting data remotely, predominantly through short phone check-ins (questionnaires). Through June 2020, social mobilizers have conducted more than 72,000 remote individual mentoring sessions with 28,000 girls across eight countries. This represents 70 percent of girls who participate in the in-person mentoring program (weighted average across countries). The remote mentoring sessions are primarily taking place via mobile phone, with video calls being the preferred method if the girls have access to a smartphone. A smaller proportion of girls are accessing remote mentoring using a computer.

M&E approach: Historically, Room to Read has used a risk and response protocol with four indicators to assess, as early as possible, the risk of a girl dropping out of school. The indicators used to flag potential pushout factors are not passing a school exam, absence from life skills sessions, low school attendance, and caregivers who don't attend program meetings. Room to Read also conducts an extensive household survey when participants enter the program—which includes collecting household contact information.

Currently, social mobilizers are integrating a one-minute, three-question survey into their mentoring calls. Each mentoring/monitoring call is approximately 15 minutes long. The three survey questions are: "Are you currently self-studying/keeping up with academic learning at home?", "Has anyone in your household lost a job or a source of income as a result of COVID-19?", and "Are you concerned about being able to return to school once schools reopen?". For girls who indicate they are concerned about not returning to school, there is a follow-up question to explore the reasons why. The data are captured by a survey that the social mobilizers complete on their phones following the session. Survey results are available to country office teams on a weekly basis and are analyzed and reported on in full by the Global RM&E team on a monthly basis.

Intended data use: Data have typically been collected on a weekly basis, and girls who are identified as being at risk of dropping out are promptly flagged for additional support and intervention by the social mobilizer. These data also inform programmatic offerings, helps social mobilizers prioritize girls who have expressed concerns regarding their ability to return to school, and provides important insights into how the COVID-19 crisis is affecting adolescent girls in developing countries.

Measuring reach: How many girls were contacted by social mobilizers.

Measuring engagement: How girls answered the three survey questions.

Measuring outcomes: Socioemotional well-being of girls during the COVID-19 pandemic.

Equity and inclusion considerations and findings: Approximately 30 percent of the girls have not yet been reached via mobile phone. This is likely due to lack of access to a phone, particularly in rural areas. It is also well-documented that women and girls have less access to mobile phones globally.

Lessons learned: Mobile access is critical.

References:
2018 Research, Monitoring & Evaluation Report: Girls' Education Program

GEOPOLL PHONE SURVEYS

Implementers: GeoPoll	**Location/s:** 72 countries in Africa, the Middle East, and the Caribbean	
Distance learning modality(s): Measures radio, TV, print/online	**Technology for collecting data:** Mobile phone	**Interface for collecting data:** Text message (SMS), voice surveys, and mobile app
Program description: "GeoPoll is a mobile survey platform and leading global survey provider, reaching a growing network of over 300 million mobile phone subscribers in emerging markets."[62] GeoPoll works in and is increasingly contracted by educational initiatives despite the limitation that the survey only tracks access by registered mobile phone owners, therefore excluding non-mobile phone holders. In their methodology, GeoPoll invites all mobile phone users to participate in their GeoPoll Audience Measurement interface (SMS messages, voice surveys, and mobile app).[62] Through daily surveys and electronic diaries, their data informs media houses, agencies, and companies on radio or television access (e.g., how they access, with whom they listen or watch), behavior (e.g., peak listening times, popular stations, preferred content), and preferences (e.g., programing they like the best). Respondents are randomly assigned to one of four different time blocks each day. Every four hours, GeoPoll surveys participants about their activities in the preceding four hours. Survey questions can be general or tailored to an initiative's needs.		
Intended data use: Predominantly for measuring reach and engagement for marketing, but increasingly used for simple evaluative purposes (basic outcome measures).		
Measuring reach: Questions are tailored to programs, but might include: Who accesses, when they access, how they access (e.g., technology and station), whether they watched with others, and other customized questions.		
Measuring engagement: How many programs watched, what parts were interesting, recall of programming or characters, how engaging they thought the programs were.		
Measuring outcomes: Basic knowledge acquired, attitudes or beliefs, behaviors.		
Equity and inclusion considerations and findings: Collects gender and geography; measures registered mobile phone holders only, so does not measure those without phones or other data.		
Challenges and other considerations: Expensive and requires substantial financial resources. Frequent data useful for marking programs or can be tailored to need.		
References: GeoPoll Audience Measurement (GAM) Methodology and Interface Overview GeoPoll website		

MOBILE PHONE MONITORING OF ADOLESCENT GIRLS' WELL-BEING IN NEPAL

Initiative: Aarambha- Her Turn	**Implementers:** People in Need	**Location/s:** Nepal (southern)
Distance learning modality(s): In-person, pivoted to phone call mentoring	**Technology for collecting data:** Mobile phones	**Interface for collecting data:** Phone calls

Program description: Our Turn works with married adolescent girls (aged 12-19) who have left school before or after marriage. This accelerated training program covers the grade 1-5 national curriculum in a 10-month program. Girls attend in-person community learning center classes 2-3 hours three days a week. Trained facilitators, who are young women from the community, lead the teaching. "The goal of the project is to improve the quality of married girls' lives by equipping them with essential life skills, numeracy and literacy, to improve their and their families' health, increase their resilience to violence and ability to negotiate important life decisions such as childbearing and spacing. The project is based on a model where empowerment is based on three pillars: access to information, skills and agency, and self-value." This in-person program shifted to 20-minute mentoring sessions and check-ins 1 to 2 times a week. The program was only in the first year when COVID-19 affected their classes.

M&E approach: Paper-based data collection until COVID-19, then pivoted to check-ins via mobile phone to collect formative data that monitors reach, engagement, learning, and socioemotional and physical well-being. Facilitators compile data and submit reports to the project team, who conducts monthly spot checks on a sample of girls to triangulate the information they are receiving in the reports.

Intended data use: To ensure girls' socioemotional and physical well-being and to provide resources as needed. To help understand whether remote mentoring is working. To check in on learning taking place.

Measuring reach: Which girls can be reached through mobile phone (75 percent of girls could be reached) and the reasons other girls are not accessible (e.g., no phone in home or at neighbors, migrated for work). Who has "access" vs. "real access" to the phone—"access" means someone in their household or neighboring house has a phone, but "real access" is whether they can use for the full mentoring sessions without interruptions.

Measuring engagement: Whether the girls are getting time to study.

Measuring outcomes: Learning of content knowledge takes place through formative assessments conducted over the phone. Socioemotional and physical well-being are monitored through phone calls.

Equity and inclusion considerations and findings: While all the participants have been marginalized—young mothers in marriages, with little to no school, and from economically and geographical marginalized communities—there are girls within this population that are even more marginalized and difficult to reach during center closures. Because the girls are generally living with their in-laws and do not own their own phones, their conversations are often put on speakerphone and overheard by others. Building trust with the entire family has been critical to allowing girls' participation. The facilitators are from the community and speak the languages of the girls.

Challenges and other considerations: Because the girls do not own their phones in most cases, they are borrowing from a family member or neighbor and are frequently interrupted during calls or cut short. Girls have children in many cases and are also caregiving. Most of the project team members (central office) do not speak local languages and therefore spot checks are linguistically challenging because the girls often do not speak Nepali.

Lessons learned: While maintaining learning has been a challenge, some of the girls have said that they appreciate the bi-weekly calls, which are easier for them to attend than in-person sessions. As when classes are in-person, they have to walk to a center, which can be a challenge for girls that have newborns or are pregnant. Using facilitators to capture formative data has helped ensure the girls have continuity and support, and helps the facilitator keep track of the socioemotional and physical well-being of their learners.

References:
Our Turn website

MONITORING REACH AND ENGAGEMENT THROUGH PHONE SURVEYS AND SHARING THROUGH DASHBOARD

Initiative: M&E Dashboard	**Implementers:** Education Development Center (EDC)	**Location/s:** 6 countries (DRC, Honduras, Liberia, Mali, Rwanda, Zambia)
Distance learning modality(s): Radio	**Technology for collecting data:** Mobile phone	**Interface for collecting data:** Phone call

Program description: Education Development Center (EDC) "pioneered the use of radio to bring curriculum and teacher training to classrooms in some of the world's least-developed countries. Throughout Africa and South America, radio has been a tremendous resource for learning and dissemination. Interactive Radio Instruction (IRI) is encompassed under the more general term Interactive Audio Instruction (IAI), as evolving technology has allowed different forms of delivering these audio programs, including CD players and MP3 players. IAI, which only requires an audio device and an adult facilitator, reaches large numbers of teachers and learners who are isolated by distance and poor infrastructure. It can be used in almost any setting, from formal classrooms to community learning centers to outdoor venues."

"EDC uses Interactive Audio Instruction to deliver education to students, train teachers, and mobilize communities in some of the hardest-to-reach places in the world, establishing sustainable systems with consistently positive results."[63] EDC has launched a Dashboard to share monitoring data collected, including reach and engagement information to "better understand **the effects of COVID-19 on project beneficiaries**" [emphasis original].[64]

M&E approach: EDC has piloted many different approaches to monitoring and evaluating their IAI/IRI programming. They are currently collecting remote data in Mali, Liberia, and Zambia. While the majority of EDC's evaluations rely on in-person assessments (e.g., literacy, numeracy, life skills, work readiness), they have increasingly tried remote data collection. For the current data used for the dashboard, EDC is speaking with caregivers through mobile phone surveys, using existing phone number lists for participants (captured before COVID-19 pandemic or obtained through parent teacher associations, school management committees, and community mobilization facilitators'). They are using data collectors to capture these data. In Northern Nigeria, they have also piloted IVR technology to provide audio and support. Reach and engagement data for IVR can be measured by spoken or texted response rates to questions following the recording.

Intended data use: Survey data are used primarily to provide project management and stakeholders with information to help shape short-, medium-, and long-term responses to challenges to receiving distance learning. EDC is also conducting complexity aware monitoring to inform the design of both the distance education programs and the plans for returning to school in-person (e.g., questions on the COVID-19 survey on knowledge of prevention behaviors, access to basic services, concerns about health, livelihoods and resiliency).

Measuring reach: Questions include: access to technology and lessons (type of technology and connectivity), awareness of radio programs, stations where radio is heard, frequency heard, who children hear lessons with, receipt of learning materials.

Measuring engagement: Engagement measures include who supports educational activities at home, how many hours a week caregivers spend supporting educational activities, more or less time in educational activities post-COVID-19, how engaging are radio programs.

Measuring outcomes: Not collecting remotely. These are embedded in to IAI/IRI curricula and assessed in-person in contexts where that is possible.

Equity and inclusion considerations and findings: EDC disaggregates all data by gender, and integrates girl and boy characters and those with disabilities into the programs.

Challenges and other considerations: Piloted learning assessments over the phone during Ebola but with limited success—there were too many distractions and confounding factors. In Zambia, the team had to rely on PTA phone numbers, which had many errors so their sample size was more limited. The reach of radio broadcast signals was also an issue.

Lessons learned: EDC has many lessons learned through implementing IAI during Ebola and in crises and conflict settings. They have not been able to capture learning outcomes remotely but have focused on capturing reach and engagement through multiple metrics and in making these data transparent. They also reiterated the importance of capturing phone numbers prior to pivoting to remote monitoring.

References:
EDC website on Interactive Audio Instruction
Audio Now!: Responding to COVID-19

MONITORING ENGAGEMENT AND LITERACY OUTCOMES THROUGH PHONE CALL SURVEYS IN HAITI

Initiative: Read Haiti	**Implementers:** Alliance for Catholic Education	**Location/s:** Haiti
Distance learning modality(s): Radio, paper	**Technology for collecting data:** Paper, survey software	**Interface for collecting data:** Phone calls

Program description: After nation-wide school closures due to civil protests, the COVID-19 pandemic meant that a majority of school children in Haiti could effectively miss an entire academic year. Recognizing radio as the most democratic and effective means of disseminating learning materials, the University of Notre Dame's Alliance for Catholic Education (ACE) Haiti and Global Center for the Development of the Whole Child's Read Haiti (Ayiti Li) program responded to school closures in Haiti by creating three radio programs—a literacy program for first and second grade students, a reading/story hour, and a pre-primary social and emotional (SEL) and parent engagement program. Supported through a Global Development Alliance composed of USAID, the W.K. Kellogg Foundation, and an anonymous foundation, the team distributed supplementary worksheets, locally-purchased radios, and solar panels to 15,200 students within five departments across the country.

M&E approach: To monitor and evaluate the Haitian Creole and French literacy radio education program, the Read Haiti team remotely trained program supervisors (assessors) to collect data through monitoring phone calls. With a subset of students throughout Read Haiti's network of 340 schools, assessors conducted pre- and post-tests at the beginning and end of the literacy radio education program to measure learning outcomes. In addition, assessors conducted weekly monitoring calls to track engagement and progress.

Data were originally intended to be gathered digitally through tablets that program supervisors normally use during the course of the school year. However, due to a lack of connectivity, ability of assessors to sync their tablets, and challenges with firmware/system updates, the team changed course. Assessors recorded data on a paper form, thereafter sending a photo of the paper form on a weekly basis to the person managing data entry. Data entry tables were then sent to the remote monitoring and evaluation team for data validation and analysis. To foster better communication among all teams included in this study, assessors used region-specific WhatsApp groups to ask questions and discuss logistics.

The team obtained caregiver phone numbers from schools before closures, and consequently were able to get in touch with caregivers quickly before the radio program was set into motion. When the radios, solar panels, and supplementary learning materials were distributed (in-person with appropriate PPE and other safety measures), the team took the opportunity to provide caregivers with an overview of the program content, broadcast times and respective community radio stations, and guidance to remain healthy during this global pandemic.

Intended data use: The data collected from over 1,200 learners will be used in aggregate to formatively assess change in learning outcomes over the first of its kind, 12-week literacy radio education program in Haiti. This will primarily inform program management and help the team understand the efficacy of radio education for distance learning. These results will be used to understand how and if the radio program should be improved or expanded for subsequent iterations and in response to future school closures.

Measuring reach: The program was able to measure reach during weekly monitoring calls when they asked if a child has watched the program. However, they were unable to measure reach for those children and families for whom they do not have contact information.

Measuring engagement: Each week throughout the program, learners received a 2-3-minute monitoring call in which an assessor gathered information on whether the child listened to the week's episode and whether they could describe the main topic of that week's lessons. To collect these data, assessors called the caregivers' phone numbers at an agreed upon hour and asked to speak to the child.

Measuring outcomes: The pre- and post-tests were adapted from a formative assessment called ALiK (Ayiti Li Kreyòl), normally used by program supervisors and teachers during the school year. In this adapted ALiK assessment, pre- and post-test phone calls with learners lasted between 7-10 minutes and evaluated their ability to recognize letters/sounds, grade-appropriate vocabulary, and respond to comprehension questions after listening to two stories. To collect these data, assessors called the caregivers' phone numbers at an agreed upon hour and asked to speak to the child.

Equity and inclusion considerations and findings: Not all phone numbers that schools had on file were accurate, which raises challenges for measuring reach. Without being able to contact every family, there is no accurate way to know who wasn't reached by this program.

MONITORING ENGAGEMENT AND LITERACY OUTCOMES THROUGH PHONE CALL SURVEYS IN HAITI

Lessons learned: In an early iteration of this program, data were collected from parents via phone before materials were distributed. This led to frustration from parents, who felt they hadn't received enough from the program to warrant the amount of phone calls they were receiving. In light of this, the program now distributes materials before collecting any data.

In the future it will be important to train assessors to set up a regular time for connecting with a parent during the initial phone call to ensure they are not calling at inconvenient times.

References:
Strong Beginnings: Radio Haiti website

MEASURING REACH AND ENGAGEMENT THROUGH ANALYTICS ON A READING APP

Initiative: Keep Children Reading, BookSmart App/Worldreader App/Digital Library	**Implementers:** Worldreader	**Location/s:** Ghana, Kenya, South Africa, India, Jordan, and Peru (BookSmart app)
Distance learning modality(s): Mobile/online educational app	**Technology for collecting data:** Mobile phone, tablet	**Interface for collecting data:** Analytics

Program description: Worldreader released their BookSmart app in 2019. BookSmart comes in two forms, a preloaded tablet for schools and mobile (Android and web apps) for home use. The targeted readers are age 3-12. Books are always free to the end user and are currently available in five languages: Arabic, English, Hindi, Spanish, and Kiswahili. Worldreader also has a free digital library for older audiences used in 47+ countries. All books on the apps can be downloaded for offline reading.

With the onset of COVID-19, Worldreader implemented their Keep Children Reading initiative with the launch of its BookSmart app. Since March of 2020, more than 466,000 families have been reached and more than 109,000 are actively reading (as tracked by backend analytics). Worldreader also launched a 14-day "read a book a day" reading challenge in June 2020. The reading challenge noted that children read an average of 3 books and explored an additional 18 books on the platform within that 2-week period. Worldreader's established benchmark goal seeks to track readers reading 15 minutes per day, 3 times a week.

M&E approach: Worldreader uses analytics to assess location data, book views, book completions, time on page, and overall time spent reading per day. Tracking is programmed into both apps and the online version of the digital library to automatically collect data about reach and engagement.

Intended data use: BookSmart has a partner dashboard to inform teachers about student reading progress, popular materials, student reading levels, and reading time on task.

Measuring reach: Geographic location.

Measuring engagement: Time reading, frequency of reading, book engagement, books read, popular content by title, category and reading level.

Measuring outcomes: Worldreader's data science modeling allows them to track reading speeds along with the above engagement measurements as a proxy for learning, and Worldreader is building out a more sophisticated solution that will allow it to align more succinctly with EGRA.

Equity and inclusion considerations and findings: Readers need mobile phone or tablet access, so the app is not accessible to people without technology. However, statistics show that there are fewer and fewer communities without a semi-smart feature phone. In addition to Android, Worldreader supports affordable phone platforms like KaiOS to address digital inequity. The apps can function fully offline, further supporting digital equity considerations.

Challenges and other considerations: Tracking learning outcomes remains a goal, but has not yet been implemented. Data storage is also a challenge. Worldreader aggregates approximately a million rows of data per day, which is pooled into a database. However, data from phones and tablets differ slightly, and tablets are typically shared between multiple learners in a classroom, making it complicated to track individual learners.

Lessons learned: Building reading models based on machine learning requires a managing a huge amount of data and reliable data storage. Worldreader collects over 1M rows of reading data per day, which is stored in a complex custom-designed, cloud-hosted database system.

The content insights are highly reliant on well-structured and organized book metadata. In order to track books read by category and reading level, Worldreader's book metadata needs to be kept up to date and follow a clear schema. They recommend the THEMA classification system for anyone with a digital book collection.

Mobile phones can be leveraged to influence reading through push notifications, messaging apps and text-messaging reminders. During the 14-day reading challenge in June, those participating in the 'read a book a day" challenge read almost four times as many books as Worldreader's average reader as a result of in-app banners, messaging groups through WhatsApp, in-app push notifications, and radio and television advertisements.

References:
App Program and Monitoring Outline WR
Worldreader 2019 Annual Report
BookSmart website

MEASURING FORMATIVE NUMERACY OUTCOMES THROUGH PHONE SURVEYS

Initiative: "Low-Tech" Remote Learning Support	**Implementers:** Young 1ove	**Location/s:** Botswana
Distance learning modality(s): Mobile phone	**Technology for collecting data:** Mobile phone	**Interface for collecting data:** SMS, phone calls

Program Description: Prior to the COVID-19 pandemic, Young 1ove had been implementing an education program called "Teaching at the Right Level" in over 20 percent of primary schools in Botswana in partnership with the Ministry of Education in Botswana. As schools were about to close due to COVID-19, the Young 1ove team collected over 7,500 phone numbers in these schools to enable them to provide remote learning support during the school closure.

While schools are not in session during COVID-19, Young 1ove is using a combination of SMS messages to send weekly math problems to caregivers and phone calls to learners to provide math coaching and assessment. Caregivers also speak with the coaches to learn how to help their learners with the daily math problems. The intent is to prevent knowledge loss in numeracy skills as well as build knowledge of basic foundational skills.

The intervention was launched as a rapid randomized trial and is one of the first to produce experimental evidence on minimizing the fallout of the pandemic on learning. Two papers are available on lessons learned on measuring learning remotely[65] as well as the randomized trial results.[66]

M&E approach: Young 1ove is using mobile phone surveys (phone calls) for their M&E. They speak with the caregivers, then ask them to allow the learners to speak with a data collector (Young 1ove staff member/facilitator).

Intended data use: Helps the program development team tailor content and programming to levels and ensures the SMS interventions are using the right literacy levels. Informs the program management team on what is working and what is not working. Learning data inform students' learning levels and identifies any learning loss, retention, or gains.

Measuring reach: Who is getting the SMS and phone calls (by basic demographics).

Measuring engagement: Engagement on numeracy activities, number of hours the caregiver spends in educational activities with their child, and parent perceptions of their child's learning.

Measuring outcomes: Numeracy gains. The assessment is based on the Annual Status of Education Report (ASER) numeracy assessment. The 11 questions were administered over the phone to 4,500 youth who provided phone numbers. Using an experimental design, a third of participants were randomly assigned to a control group, a third received an SMS intervention with daily math problems, and a third received SMS + phone calls. The study discerned "53 reductions in innumeracy from phone calls and SMS test, and 34 percent reductions in innumeracy using SMS texts only"[66] among children grades 3-5. Young 1ove conducted reliability tests (comparing in-person reliability to testing reliability) and are currently conducting further tests. They time the tests and ask students to explain their work to see timing and to see if learners are getting help by others. The survey takes 20 minutes or less.

Equity and inclusion considerations and findings: The test is administered in English, but the explanations can be given in Setswana. Data collectors have basic demographic data (e.g., gender, age).

Challenges and other considerations: Numeracy might be easier to capture than literacy. They are trying to ensure that literacy levels do not confound the results on the numeracy.

Lessons learned: Young 1ove collected 7500 phone numbers in the week before schools closed. This made the pivot to distance learning easier for learners, caregivers, and data collectors. They have experimented with ways to encourage caregivers not to interfere with their children's testing by limiting the answering time frame to two minutes per question.

References:
Teaching at the Right Level - Young 1ove website
Practical Lessons for Phone-Based Assessments of Learning[67]
Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana[68]

MEASURING REACH, ENGAGEMENT, KNOWLEDGE, AND PRACTICES OF RADIO PROGRAMS THROUGH A MULTI-MODAL INTERFACE

Implementers: Farm Radio International (with radio stations and many partners).	**Location/s:** 11 countries in Africa (provide resources across the continent)	
Distance learning modality(s): Radio	**Technology for collecting data:** Phone	**Interface for collecting data:** IVR, SMS, Facebook, Phone calls, MMS

Program description: Farm Radio International partners with radio stations in rural communities across Africa to provide local radio stations with resources (training and support) and radio innovations (pioneer digital solutions and formats), as well as implement targeted radio projects. The intention of FRI is to "see the hard work of African farming families lead to prosperity and food security for themselves, their communities, and their countries"[69] using radio as a tool for building capacity and support. While not an education organization, FRI's work is educational and draws on social and behavior change communication (SBCC) and communication for development approaches. Two of their projects, Her Farm Radio in Ethiopia, Malawi, Tanzania and Uganda (targets female farmers) and Integrated Approach to Addressing the Issue of Youth Depression in Malawi and Tanzania were studied for this review.

M&E approach: FRI has an integrated M&E approach that is designed from the start of their projects and uses a number of representative sample approaches through two-stage cluster sampling.[70] As a general practice, FRI integrates listener interactions into all of their series. In each episode, listeners "flash" (call and hang up) a broadcast number, and an IVR system automatically calls the user back and walks them through a sequence of questions. These questions cover reach, engagement, knowledge (e.g., something covered in the program), and an opportunity to leave voice-recorded feedback. Callers' numbers are automatically recorded into a database. FRI uses a customized system that creates a physical map of where the FM radio signals should reach. It uses a combination of Geographical Information System data, census data, and formulas to estimate how far different radio stations reach. They then use the map and census data to capture who is actually accessing the program. In the past, they have used SMS but have moved away from this because of the cost and required reading literacy, but they still use SMS for announcing programs through reminders. FRI also uses GeoPoll Surveys for a few of their projects to capture reach, engagement, and outcome measures.

Intended data use: Data are used to inform program management, radio program and materials development, and to help ensure activities and capacity are transferred to partnering radio stations. They also use data to measure summative findings, or the outcomes the radio programs have on knowledge, attitudes, and practices (described as behaviors in this review).

Measuring reach: Time listened, station where episode heard, which episodes are heard.

Measuring engagement: How many programs are listened to, what parts were interesting, recall of programming or characters, how engaging they thought the programs were.

Measuring outcomes: Basic knowledge acquired, attitudes or beliefs (e.g., gendered norms), behaviors/practice (e.g., applied learned knowledge), and myths and experiences stories (through WhatsApp or Facebook depending).

Equity and inclusion considerations and findings: FRI intentionally targets rural, small-holder farmer, and women farmers. They ensure women are central characters and guests on programs, and that content is gender-responsive. FRI collects gender and geography data and measures registered mobile phone holders; they do not measure those without phones or other data. They created a special call-in line for women as they found that women could not spend long times waiting on the phone per call-ins.

Challenges and other considerations: Requires substantial financial resources. Requires IVR infrastructure (E1 or T1 lines) that countries do not always have, requires working with the telecoms and governments or finding a workaround.

Lessons learned: FRI has shifted interface as technology shifts. They have experimented with different kinds of technology and still collect in-person data to account for communities that cannot be reached through mobile phone surveys. They have found that behavior change can be shifted through radio. Findings also show that surveys need to be 20 minutes or less to maintain attention and avoid respondents being interrupted or the call being dropped.

References:

Farm Radio International website
Interactive Radio program report: An Integrated Approach to Addressing the Issue of Youth Depression in Malawi and Tanzania
Final Report: Her Farm Radio in Ethiopia, Malawi, Tanzania and Uganda
GeoPoll Technical Expert: Farm Radio International Uganda – Listenership of Bushenyi FM Radio in Bushenyi and Kasese

MONITORING LITERACY AND HEALTH THROUGH PHONE SURVEYS IN THE DRC

Initiative: Accelere! I	**Implementers:** Chemonics	**Location/s:** DRC
Distance learning modality(s): Radio	**Technology for collecting data:** Mobile phone	**Interface for collecting data:** Mobile phone, voice survey, SurveyCTO (survey software)

Program description: Accelere! I was developed to meet two main goals in the DRC's national education strategy: increasing children's access to education and ensuring that all children master fundamental reading skills by the end of Grade 2. The program was located in 1,578 formal and nonformal private structures, and was set to wrap-up in 2020, just as COVID-19 broke out. Over 112,00 formal and nonformal learners in grades 1 and 2 and nonformal equivalency programs were targeted by Accelere! I in person.

To pivot, Accelere! I moved to interactive radio instruction with content on literacy and health topics relevant to the COVID-19 pandemic (e.g., handwashing, social distancing). The radio programming is offered in Kiswahili and Lingala, with 24 30-minute-long episodes in each language.

M&E approach: Chemonics developed a rapid sampling cycle, with phone survey data being collected on Fridays, analyzed over the weekend, and shared back out to project management and staff the following Wednesday. Data were collected by reading mobilizers who were program contractors and had existing relationships with partner schools. Reading mobilizers called caregivers and asked to speak to learners to gather literacy and health knowledge data based on the week's radio programming.

Intended data use: Improving radio programming.

Measuring reach: Who listened in specific geographic areas (Kinshasa and Haut Katanga, North Kivu), who knows where (what frequency) and when (what time) to listen.

Measuring engagement: Listenership, why learners and caregivers did not listen.

Measuring outcomes: Basic phonemic, syllabic, and word level knowledge; Health knowledge around COVID-19 safety.

Equity and inclusion considerations and findings: Families needed radio access to listen to the programming and mobile phone access to participate in M&E. Over 90 percent of families indicated that they had cell phone access. In many cases, lack of electricity prevented participation.

Challenges and other considerations: A number of challenges arose with the pivot to distance learning and remote M&E. With the radio programming, there were interruptions for news bulletins and a broken transmitter which delayed programming. Lingalaphone speakers noted some difficulty understanding some vocabulary as there are linguistic differences between the spoken language and the standard version of the language of instruction. M&E activities were made more difficult by caregivers working outside the home and having the only mobile phone with them. Chemonics also noted that it was difficult to reach caregivers and learners after the programs because caregivers sometimes departed immediately or soon after the program to support their livelihoods.

Lessons learned: One way that learners and caregivers overcame the lack of radio/electricity was to send learners to a nonformal school to participate in small socially distanced listening groups of 10 learners or less, or to listen to programming via mobile phone.

Reading mobilizers and Chemonics had access to caregiver phone numbers from schools. These data were collected long before the COVID-19 pandemic and significantly aided M&E efforts. ACCELERE!I has also extended the data collection period to Saturday, allowing field agents to reach caregivers upon their return home in the early evening or early Saturday before they depart the home.

References:
Radio Program and Monitoring Outline Chemonics DRC
Accelerating Access and Learning in the Democratic Republic of the Congo

INTEGRATION OF SOFTWARE ON MONITORING PLUS VIDEO ACCESS IN PAKISTAN

Initiative: Teleschool and Taleem Ghar	**Implementers:** Government of Pakistan and Government of Punjab	**Location/s:** Pakistan
Distance learning modality(s): Video/television, mobile	**Technology for collecting data:** Mobile phone*	**Interface for collecting data:** SMS*

Program description: In Pakistan, the governments reacted quickly to the closure of Pakistani schools due to COVID-19. In less than two weeks, the national Government of Pakistan launched the Teleschool initiative, while the regional Government of Punjab implemented Taleem Ghar. These programs air animated education videos on cable television every day from 8 a.m. to 6 p.m., covering math, science, and English for grades 1-12. Learners can also access Taleem Ghar videos through computer or smartphone application.

The content for this programming was sourced through partnerships and drawn from the existing curriculum, with some videos having been previously developed for use in the classroom. Television access has been prioritized as television has higher penetration than computers or smartphones in Pakistan.

M&E approach: To increase engagement and to monitor learning, an interactive SMS feature to be incorporated into this television programming has been proposed. If implemented, learners would see a keyword and code appear on screen at the end of a video lesson. They would send the keyword to the code via SMS and receive a lesson summary, tips and tricks related to the lesson, or an assessment. For an assessment, learners would receive a question with response options. They would respond via SMS with an option number and receive another question, until the assessment was complete.
*This evaluation method has been proposed by Tabadlab.

Intended data use: The results of this assessment could be shared with the learner, the teacher, or the school.

Measuring reach: Proposed, but not implemented at time of review.

Measuring engagement: Proposed, but not implemented at time of review.

Measuring outcomes: Proposed, but not implemented at time of review.

Equity and inclusion considerations and findings: Television was chosen as the primary modality for this program after determining that it allowed for the greater access. This program was designed to reach urban, peri-urban, and rural learners.

Challenges and other considerations: While this program was designed to have the highest possible reach, television is still not accessible in some rural areas.

Lessons learned: This program was implemented quickly in response to schools closing and prioritized an accessible, cost-efficient, and scalable model. While this program is sustainable in the short-term, it will need to be further developed to include more interactive tools in order to be sustainable in the long-term and allow for robust monitoring.

References:
EdTechHub Brief on Pakistan COVID-19 Educational Response
Pakistan: Tacking COVID-19 in Education
Government of Pakistan, Teleschool
Taleem Ghar Website

CAPTURING REACH, ENGAGEMENT, AND OUTCOMES FOR RADIO, TELEVISION, AND EDUCATIONAL APPS: A MULTI-MODAL AND MIXED-METHODS APPROACH

Initiative: Ubongo	**Implementers:** Ubongo	**Location/s:** Works in 18 countries in Africa (started in Tanzania) with team in 8 countries
Distance learning modality(s): Radio, Television/video, mobile phone apps, paper	**Technology for collecting data:** Primarily phone	**Interface for collecting data:** Phone calls, SMS, digital analytics (plus in-person)

Program description: Ubongo is an education-entertainment program that produces radio programs, animated videos, and mobile applications that aim to improve school readiness and learning outcomes for children and promote social and behavioral change for viewers. They "leverage the power of entertainment, the reach of mass media, and the connectivity of mobile devices, to deliver effective, localized learning to African families at low cost and massive scale. As Africa's leading edutainment company, we create fun, localized and multi-platform educational media that reaches millions of families through accessible technologies. Our programs significantly improve school readiness and learning outcomes for kids, and also promote social and behavioral change for kids, caregivers and educators."

M&E approach: Ubongo collects data on different types of reach, engagement, and outcome data. They use a series of in-person methods for assessing engagement and outcomes: in-person interviews, FGDs, observations of children and caregivers watching programs, and in-person eye tracking studies where they record children viewing programs and then code these videos. They also use GeoPoll for weekly mobile survey data on viewership and listenership and have used a number of nationally representative surveys (i.e. IBSO, 60 Decibels) to capture more data on demographics of households in the viewership and listenership. A researcher conducted a cost-efficiency study with Ubongo in 2019.

Intended data use: The data Ubongo collects are used for formative and summative purposes. Internally, the observations, FGDs, and interviews helps gather formative data to guide program development and revision. They also use data for program management and deciding where to broadcast and what types of programming to create. In-person data are also used to measure change in knowledge, foundational skills (e.g., numeracy and literacy), and social behavior change.

Measuring reach: Ubongo uses data from GeoPoll surveys for analyzing viewership and listenership at a national-scale–they use GeoPoll in five countries: Tanzania, Kenya, Uganda, Rwanda, and Ghana. Other countries, such as Nigeria, have national companies that collect similar data. GeoPoll utilizes SMS-based surveys and provides Ubongo with a dashboard that shows how many people are watching each of their programs during each half hour segment. Ubongo also collects data on reach by conducting phone surveys with adults. Interviewers ask parents if their child has watched or listened to the program in the last week. These data can be cross-checked by surveying children on character recognition.

Measuring engagement: Ubongo has measured engagement through SMS surveys. At the end of the program, a phone number is shared on television screen or over the radio, and viewers/listeners are encouraged to sign-up by sending a message to the phone number. Those who sign up are sent additional educational content via SMS, and a sample of those who sign-up are selected for a survey. Survey participants are asked a number of questions about their frequency of viewing/listening, what characters they remember (which is correlated with outcomes), through which channels they access the programs, and whether caregivers are watching with children.

In addition to television and radio, Ubongo videos are also available on YouTube. Though only a small subset of Ubongo's audience views their content through YouTube, it provides an opportunity to collect rich engagement data. The analytics YouTube provides offers information on audience retention, including when viewers stop watching.

Measuring outcomes: Most of Ubongo's measurement of learning outcomes has occurred in-person through 45 minute in-person interviews. Remotely, they have asked caregivers in SMS surveys if they think their children have learned from Ubongo programs, but this is just a perceived learning outcome. They are currently exploring other methods of measuring learning outcomes remotely. They have worked with research partners including the University of Maryland School of Public Health and Uwezo Tanzania for in-person data collection.

Equity and inclusion considerations and findings: In some countries, Ubongo has been able to collaborate with and add questions to nationally representative house-to-house surveys (such as those conducted by IPSOS). The data collected from these surveys help to provide a more accurate picture of who is not being reached by Ubongo programming. For example, they can disaggregate viewership and listenership by household income. Ubongo looks at gender data of viewers, as well as looks carefully at issues of gender and inclusion in their materials.

CAPTURING REACH, ENGAGEMENT, AND OUTCOMES FOR RADIO, TELEVISION, AND EDUCATIONAL APPS: A MULTI-MODAL AND MIXED-METHODS APPROACH

Challenges and other considerations: Like with other organizations using mobile phone surveys to measure reach, there here are challenges. For example, SMS polls are only polling adults; a child is watching or listening to a Ubongo program without an adult present is not captured. Additionally, Ubongo has found that data collected from caregivers often underestimate reach. In one example, 13 percent of caregivers said their child watched or listened to Ubongo programs, but 50 percent of children could name characters from the programs.

Ubongo has faced challenges with conducting longitudinal surveys on learning outcomes due to a high exposure rate to Ubongo programming among respondents, preventing a reliable comparison group.

There is the potential to collect meaningful analytics on engagement and learning outcomes from mobile applications. However, Ubongo has not been able to capture such analytics due to a number of challenges. First, most of Ubongo's target audience has basic smartphones, which do not allow for big apps with a complex backend. Additionally, to increase accessibility, Ubongo's apps are designed to work completely offline. While the apps can send data back when they are synched to the Internet, while offline they only collect data on how many times the app was opened and for how long. Finally, collecting specific data on learning outcomes from children via mobile app analytics requires that parents register through the app and consent to the data collection. Ubongo does require this for one of their apps and has found that it acts as a barrier to entry; this app has significantly fewer uses.

Lessons learned: Due to high exposure rates preventing reliable comparison groups, Ubongo is considering other measures beyond exposure. Ubongo has tried multi-modal M&E using a mixed-methods, multiple technologies, and different data sources (caregivers, learners, etc.).

References:
Ubongo Website
A Quasi-Experiment Examining the Impact of Educational Cartoons on Tanzanian Children[71]
The Impact of an Educational Media Intervention to Support Children's Early Learning in Rwanda[72]
Examining the Impact of Akili and Me's New Content on Tanzanian Children[73]

MONITORING AND SUPPORTING TEACHERS DURING DISTANCE LEARNING IN TANZANIA

Initiative: Tusome Pamoja	**Implementers:** RTI	**Location/s:** Tanzania
Distance learning modality(s): Mobile phone	**Technology for collecting data:** Mobile phone (connected to Cell-Ed's Learner Management System)	**Interface for collecting data:**

Program description: In Tanzania, RTI designed a proof of concept focused on improving SEL competencies for both teachers and students in order to create more positive school climates with reduced violence and improved learning outcomes. This proof of concept was designed to include an in-person co-creation workshop for 37 teachers on discrete classroom and teacher-based activities that they could do to support a positive school and classroom climates and SEL. However, when schools closed due to COVID-19, the team transitioned to an extended and virtual co-creation activity that more specifically addressed teacher concerns with the content (e.g., what will and will not work in their classrooms, teacher confidence in discussing and implementing certain activities with students).

At first, the team used WhatsApp to continue working with teachers virtually to address their concerns from co-creation and to continue refining activities. Then the team partnered with Cell-Ed, a mobile learning platform that has a data warehouse called a Learner Management System (LMS). Using Cell-Ed's platform, the team was able to construct streamlined, high-quality, audio text-based content in the form of weekly lessons over a 10-week period. This content always contained questions to check for teacher understanding and open-ended feedback where teachers could flag specific issues and concerns. All teacher responses were stored directly in the LMS for further data analysis. Cell-Ed's LMS also tracked teacher participation, which allowed the team to see which teachers were progressing and to reach out to, on an individual basis, the teachers who had fallen behind with the content.

When schools re-opened on June 29th, teachers were in the middle of the virtual Cell-Ed-delivered content. Having already established this system of communicating with teachers, the Tanzania team was able to continue working with teachers virtually as these teachers started classes again. As such, teachers could test out the content they learned during lockdown and flag issues as they occurred in the classroom. The team was able to respond to these issues via the LMS and support on-the-ground.

The Tusome Pamoja team also held WhatsApp conference calls with teachers to share feedback on the teachers' responses to the lesson questions.

M&E approach: Transitioning to a remote, mobile program provided a unique opportunity for monitoring not generally available for in-person training; the team was able to continually revise and enhance the content of this virtual co-creation based on teacher feedback. Additionally, the Cell-Ed LMS allows real-time monitoring of teacher participation and produces reports on engagement and lesson question responses.

Measuring reach: Not necessary with small pilot.

Measuring engagement: Participation in Cell-Ed's WhatsApp content post training.

Measuring outcomes: Now that in-person school has resumed in Tanzania, the team at RTI is collecting self-reported data on teacher attitudinal and behavioral change in the classroom via short questions sent to teachers through Cell-Ed's WhatsApp. Preliminary data seem to show that positive changes are occurring.

Lessons learned: This experience reflects RTI's broader focus on teacher well-being, teacher support, and teacher training within distance learning environments. They emphasize that, in addition to monitoring student outcomes, it is critical to monitor teacher outcomes, such as socioemotional well-being, access to necessary technology, skill in virtual instruction, and frequency of interaction with students.

Technology such as Tangerine, RTI's mobile assessment and coaching application, can be used to facilitate teacher monitoring and ensure teachers are engaged in distance learning. Tangerine allows for offline data collection on low-cost Android tablets. RTI has already used this application for teacher training in Malawi, where training content originally delivered through IVR was uploaded into Tangerine for teachers to revisit as needed. Tangerine's continuous assessment component, Tangerine:Teach, can be used to help teachers monitor student well-being and family outreach. As teachers record their interaction with a student or family in Tangerine, the application can alert teachers to a family or child they haven't spoken to in a while. Tangerine:Teach can also assist teachers in collecting, analyzing, and using results from continuous curriculum-based assessments, allowing teachers to remain involved in their students' learning while school is remote and preparing teachers to reenter the classroom when in-person school is reopened.

References:

Looking Beyond School Closures: Considering Teacher Well-being and Support
Virtual Assessment and Making the Right Technology Choices
Co-Creation of Teaching Activities During COVID-19

ANNEX E: KEY TERMS

Terms around distance learning are in constant negotiation. The definitions below pertain to this review but may be defined differently in other circumstances.

Asynchronous distance teaching and learning	Occurs at different times AND in different places (e.g., recording lectures and having learners respond with questions and comments on a discussion board on their own time). (Great Schools Partnership 2014a).
Basic phone	"A wireless handheld device that allows users to make and receive calls...[also] capable of sending and receiving text messages. As these devices evolved, they became smaller and more features were added, such as multimedia messaging service (MMS), which allowed users to send and receive images."[74]
Computer-Assisted Telephone Interviewing	"In the CATI research mode, interviewers make calls themselves, and they are the ones noting down the respondents' answers and reading the next question from a script. CATI interviewers use specialized software to dial phone numbers, record the answers they are given, and read the next question based on any skip logic that is included."[75]
Cookies tracking	"A computer 'cookie' is...a term for a packet of data that a computer receives and then sends back without changing or altering it...When you visit a website, the website sends the cookie to your computer. Your computer stores it in a file located inside your web browser."[76] Cookies allow for automatic tracking and data collection.
Distance learning (or distance education)	Teaching and learning where educators and learners are in separate physical spaces. Distance learning can occur through one of four modalities: audio/radio, video/television, mobile phone, and/or online learning platforms. Printed texts (e.g., teachers' guides and student materials) often accompany these modalities and could be a fifth modality in cases where technology is not (or cannot be) used for teaching and learning (e.g., correspondence learning). Distance learning can be synchronous or asynchronous.
Feature phone	"A feature phone is a type of mobile phone that has more features than a standard cellphone but is not equivalent to a smartphone...Typically, a feature

	phone has the basic characteristics of a mobile phone and has capabilities such as a portable media player, digital camera, personal organizer and Internet access.[77]"
Inclusive education	"Having one system of education for all students, at all levels (early childhood, primary, secondary, and post-secondary), with the provision of supports to meet the individual needs of students. Inclusive education focuses on the full and effective participation, accessibility, attendance, and achievement of all students, especially those who, for different reasons, are excluded or at risk of being marginalized."[78]
Interactive audio/radio instruction	Interactive audio instruction (IAI) is a distance education approach that uses interactive pedagogies to engage listeners in active and quality teaching and learning through pre-recorded audio programs. Audio lessons guide educators and learners "through activities, games, and exercises that teach carefully organized knowledge and skills."[79] Programs draw on songs, stories, and culturally based knowledge and content. Audio programs can either be digitized and listened to on an audio device (IAI) or broadcast through radio (IRI). IAI can be used with learners and educators in a range formal and nonformal settings to "improve educational quality and teaching practices."[80]
Interactive Voice Response (IVR)	"Interactive Voice Response (IVR) is an automated telephony system technology that interacts with the callers, gathers the required information, and routes the calls to the particular appropriate recipient...Conversations are either pre-recorded or generated audio which assist, direct, or route calls automatically without a live operator. Within these interactions, clients can communicate by using either the touch-tone keypad selection or voice telephone input. The responses take the form of voice, call-back or any other related media."[81]
(User) Interface	"User interface (UI) is a broad term for any system, either physical or software based, that allows a user to connect with a given technology."[82]
Learning management system	"Learning management systems (LMS) are software platforms for instructors to manage and organize educational courses online and provide students a single location for all course material...LMSs are

	composed of a document management component and communication capability, allowing teachers to upload course files such as rubrics, assignments, calendars, and gradebooks, as well as communicate with students via chat rooms or forums. Many LMSs may also offer online assessment functionality, such as quizzes and tests, or offer features that integrate multimedia components such as videos and photos. Students can use LMSs to submit assignments, and many platforms allow instructors to grade assignments within the platform."[83]
Learning outcomes	"Learning outcomes are statements that describe the knowledge or skills students should acquire by the end of a particular assignment, class, course, or program, and help students understand why that knowledge and those skills will be useful to them. They focus on the context and potential applications of knowledge and skills, help students connect learning in various contexts, and help guide assessment and evaluation."[84]
Log file (customized for education app)	"A log file is a file that keeps a registry of events, processes, messages, and communication between various communicating software applications and the operating system. Log files are present in executable software, operating systems, and programs whereby all the messages and process details are recorded. Every executable file produces a log file where all activities are noted."[85]
Messaging services: SMS (Short Message Service) MMS (Multimedia Messaging Service)	Text and multimedia messaging technologies. "SMS stands for Short Message Service...it is one of the oldest texting technologies. It is also the most widespread and frequently used. MMS stands for Multimedia Messaging Service. It was built using the same technology as SMS to allow SMS users to send multimedia content. It's most popularly used to send pictures, but can also be used to send audio, phone contacts, and video files."[86]
Massive Open Online Courses (MOOCS)	An online distance education mechanism (platform) where teaching and learning is global (e.g., a general class on monitoring and evaluation). Some MOOC providers (e.g., Coursera, Udemy) offer a certificate or credit for a cost. They can be taught asynchronously, with active teacher monitoring, or be completely automated.

Nonformal education	"Non-formal education takes place both within and outside educational institutions and caters to people of all ages. It does not always lead to certification. Non-formal education programmes are characterized by their variety, flexibility, and ability to respond quickly to new educational needs of children or adults. They are often designed for specific groups of learners such as those who are too old for their grade level, those who do not attend formal school, and adults. Curricula may be based on formal education or on new approaches. Examples include accelerated 'catch-up' learning, after-school programmes and literacy- and numeracy-focused programmes. Non-formal education may lead to late entry into formal education programmes, in which case it is sometimes called 'second chance' education."[87]
Online education	A distance learning modality that refers to teaching and learning that occurs via the Internet. Online education (or online learning) can be used to supplement in-person education (e.g., learners follow along on tablets during a guided reading exercise) or be the primary mode of delivery in distance learning settings (also called online distance education). Online learning can be asynchronous (where learners control time and pace) or synchronous (where teaching and learning happens simultaneously in real time either in an online space or through a concurrent broadcast).
Smart phone	"A smartphone is a mobile phone with highly advanced features. A typical smartphone has a high-resolution touch screen display, WiFi connectivity, Web browsing capabilities, and the ability to accept sophisticated applications."[88]
Survey software	"An application used to collect feedback from a targeted sample through a computer-assisted method, which comes in different ways. At its core survey software solutions help you design, send, and analyze surveys, usually via the Internet and using drag-and-drop tools and automated functionality."[89] For example, Tangerine, KoBo Toolbox, Survey to Go, Survey Monkey, and Qualtrics.
Synchronous distance teaching and learning	Occurs simultaneously, but not in the same physical space. It often refers to online learning that happens in real time via digital, video, audio, or online forums

	(e.g., class discussion on Zoom). (Great Schools Partnership 2014b)
Universal Design for Learning (UDL)	"Universal design for learning (UDL) is a set of principles for designing curriculum that provides all individuals with equal opportunities to learn. UDL is designed to serve all learners, regardless of ability, disability, age, gender, or cultural and linguistic background. UDL provides a blueprint for designing goals, methods, materials, and assessments to reach all students including those with diverse needs. UDL is an approach to instruction that prioritizes meeting the needs of learners with disabilities."[90]

ENDNOTES

[1] Programming in this review refers to the audio, video, phone, or online content, plus any accompanying materials per the design.
[2] The scope of this review was pre-primary and primary school level (or equivalency) learners. While the guidance can be applied to secondary and tertiary level learners, there is notably more research on mobile phone and online learning modalities at the secondary and tertiary levels than is covered in this review.
[3] USAID's Education Policy (2018) defines Inclusive Education as having one system of education for all learners, at all levels (early childhood, primary, secondary, and tertiary), with the provision of supports to meet the individual needs of learners. Inclusive education focuses on the full and effective access, attendance, participation, and achievement of all learners, especially those who, for different reasons, are excluded or at risk of being marginalized.
[4] Outcomes include learning outcomes, such as acquisition of knowledge and skills, as well as changes in attitudes, behaviors, and practices.
[5] Hempel, Kevin and Nathan Fiala. *Measuring Success of Youth Livelihoods Interventions: A Practical Guide Monitoring and Evaluation.* The World Bank, 2012. http://documents.worldbank.org/curated/en/204611468320058537/Measuring-success-of-youth-livelihood-interventions-a-practical-guide-to-monitoring-and-evaluation
[6] See Education Reporting Toolkit and Guidance For USAID Education Sector Implementing Partners: Monitoring, Evaluation, And Learning During The COVID-19 Pandemic for more information
[7] Though not discussed in this report, cost metrics are also important to measure alongside program planning and development and are thus included in Table 1.
[8] According to the USAID, the term "social and emotional skills" is used in the context of formal and non-formal education programming and the term "soft skills" is used in the context of workforce development programs and higher education.
[9] Users encompass all learners, educators, and others who use the programming and materials. Intended users are those designated as beneficiaries and target audiences of the programming, while unintended users are those who access and use the programming but were not identified as direct beneficiaries in the design.
[10] Simpson, Robert. "Can Interactive Radio Instruction turn post-conflict educational challenges into opportunities? A case study of the 'Speak Up!" English language programme, South Sudan." Liverpool School of Tropical Medicine, 2013.
[11] Education Development Center (EDC). *Final report of the Somali Interactive Radio Instruction Program.* Prepared for the United States Agency for International Development. Education Development Center. n.d. https://pdf.usaid.gov/pdf_docs/pdact951.pdf
[12] Education Development Center (EDC). *Final report of the Somali Interactive Radio Instruction Program.*
[13] Anzalone, Stephen, and Andrea Bosch. "Improving education quality through Interactive Radio Instruction." Africa Region Human Development Working Paper Series No. 52. Washington, DC: The World Bank, 2006.
[14] Education Development Center (EDC). *Final report of the Somali Interactive Radio Instruction Program.*
[15] Education Development Center (EDC). *Radio Instruction to Strengthen Education (RISE) Tanzania. Final Report. Prepared* for the United States Agency for International Development. Dar es Salaam, Tanzania: Education Development Center. 2010.
[16] Leigh, Stuart, and Andrew Epstein. *South Sudan Interactive Radio Instruction performance evaluation report.* Prepared for the United States Agency for International Development. Washington, DC: Management Systems International, 2012. https://www.eccnetwork.net/resources/south-sudan-interactive-radio-instruction-performance-evaluation-report
[17] Socio-Economic Data Centre Limited. *Extent of listenership to Southern Sudan interactive radio instruction programs.* Prepared for the United States Agency for International Development. Education Development Center, 2011. https://pdf.usaid.gov/pdf_docs/PNAEC248.pdf
[18] See *Best Practices in Generating Data on Learners with Disabilities.*

[19] As described in USAID's *Distance Learning Toolkit*, a national or regional needs assessment of radio, television, mobile phone, and Internet coverage should be conducted before determining which distance learning modalities to implement.
[20] See USAID's Delivering Distance Learning in Emergencies: A Review of Evidence and Best Practice for more details.
[21] Dahya, Negin. *Education in conflict and crisis: how can technology make a difference? A landscape review.* Bonn, Germany: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2016. https://www.eccnetwork.net/resources/education-conflict-and-crisis
[22] Anzalone, Stephen, and Andrea Bosch. "Improving education quality through Interactive Radio Instruction." Africa Region Human Development Working Paper Series No. 52. Washington, DC: The World Bank, 2006.
[23] Burns, Mary. *Distance education for teacher training: modes, models and methods.* Education Development Center, 2011. http://idd.edc.org/sites/idd.edc.org/files/Distance%20Education%20for%20Teacher%20Training%20by%20Mary%20Burns%20EDC.pdf
[24] Christina, Rachel and Nathalie Louge. *Expanding access to early childhood development using interactive audio instruction.* World Bank Group and Education Development Center, 2015. http://idd.edc.org/sites/idd.edc.org/files/Expanding%20Access%20to%20ECD%20Using%20IAI%20-%20Summary.pdf
[25] See UNESCO's *Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak* for additional examples of age-appropriate programming.
[26] Anzalone, Stephen, and Andrea Bosch. "Improving education quality through Interactive Radio Instruction." Africa Region Human Development Working Paper Series No. 52. Washington, DC: The World Bank, 2006.
[27] Borzekowski, Dina L. G. "A quasi-experiment examining the impact of educational cartoons on Tanzanian children." Journal of Applied Developmental Psychology, 54 (2018): 53-59.
[28] Borzekowski, Dina L. G., L. E. Kauffman, E. A. Dura, M. Chale, and C. B. Bauer. *Examining the impact of Akili and Me's new content on Tanzanian children.* Prepared for UBONGO and the Human Development Innovation Fund. College Park, MD: University of Maryland, 2020.
[29] Inclusive Development Partners (IDP). *Formative Evaluation of the Soma Nami Radio Program on Inclusive Education.* Internal Program Document for Tusome Pamoja, 2020.
[30] Interestingly, Young 1ove found that caregivers struggled to accurately identify learners' numeracy skills and level.
[31] While this review is focused on evidence for preschool and primary school-age learners and interventions, language is intentionally inclusive of nonformal out-of-school youth and adult learners.
[32] DeJaeghere, Joan, Emily Morris, and Richard Bamattre. "Moving beyond employment and earnings: reframing how youth livelihoods and wellbeing are evaluated in East Africa." Journal of Youth Studies 23, no. 5 (2019): 667-685. https://doi.org/10.1080/13676261.2019.1636013
[33] Xie, Xinyan, Qi Xue, and Yu Zhou. "Mental health status among children in home confinement during the coronavirus disease 2019 outbreak in Hubei Province, China." JAMA Pediatri. 174, no. 9 (2020): 898-900.
[34] Bruce, Giles. "Youth suicides were rising before COVID-19. How should we support kids now?" Center for Health Journalism, May 14, 2020. https://www.centerforhealthjournalism.org/2020/05/13/youth-suicides-were-rising-covid-19-how-should-we-support-kids-now
[35] Mitra, Esha. "India didn't prioritize mental health before COVID-19. Now it's paying the price." CNN, September 7, 2020. https://www.cnn.com/2020/09/06/india/india-mental-health-dst-intl-hnk/index.html
[36] Adigwe, Joseph. C. "Ethnicity, test anxiety and science achievement in Nigerian students." International Journal of Science Education, 19, no. 7 (1997): 773-780.
[37] Heissel, Jennifer A., Emma K. Adam, Jennifer L. Doleac, David N. Figlio, and Jonathan Meer. "Testing, stress, and performance: How students respond physiologically to high-stakes testing." Education Finance and Policy, (2018) 1-50.
[38] Morris, Emily. "Performing Graduates, Dropouts, and Pushouts: The Gendered Scripts and Aspirations of Secondary School Students in Zanzibar." PhD diss., University of Minnesota, Twin Cities, 2018.
[39] Sedere, Upali M. *Towards a Stress Free Education: International Perspective.* Online Submission. 2010.

[40] Watson, Joe. "The relationship between educational television and mathematics capabilities in Tanzania." PhD diss., University of Cambridge, 2019. https://www.repository.cam.ac.uk/bitstream/handle/1810/307624/Watson_Thesis_PhD.pdf?sequence=4&isAllowed=y

[41] Essay grading software exists but is costly and can have serious technical and ethical flaws (Smith 2018).

[42] Note that this was an in-person study of youth well-being, but the main delivery of content was through distance learning interventions.

[43] Danjuma, Malti. "Girls Documenting Covid-19 Experiences." YouthhubAfrica, July 23, 2020.

[44] Vota, Wayan. "Stop Giving Away Smartphones!" ICTworks (blog), September 15, 2016. https://www.ictworks.org/stop-giving-away-smartphones/#.X2uJWmhKiUl

[45] Alvi, Muzna, Shweta Gupta, Ruth Meinzen-Dick, and Claudia Ringler. "Phone surveys to understand gendered impacts of COVID-19: a cautionary note" (blog). CGIAR Research Program on Policies, Institutions, and Markets. IFPRI, July 14, 2020. https://pim.cgiar.org/2020/07/14/phone-surveys-to-understand-gendered-impacts-of-covid-19-a-cautionary-note/

[46] World Bank. *World Development Report 2016: Digital Dividends. Overview booklet.* Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO, 2016. http://documents1.worldbank.org/curated/en/961621467994698644/pdf/102724-WDR-WDR2016Overview-ENGLISH-WebResBox-394840B-OUO-9.pdf

[47] If collecting data during times of emergency, such as COVID-19, it is important to follow USAID and country-level guidance for how to do this safely (see Guidance for USAID Education Sector Implementing Partners: Monitoring, Evaluation, and Learning (MEL) During the COVID-19 Pandemic).

[48] YouthHubAfrica referenced in Step 2 used photograph diaries collected by adolescent girls in Nigeria during Covid-19, and Gender and Adolescence Global Evidence (GAGE) documented photo and audio diaries in their *GAGE Virtual Research Toolkit: Qualitative Research with Young People on their Covid-19 Experiences* (2020). The GAGE Toolkit offers guidance and best practices for collecting a range of qualitative data remotely with youth.

[49] Ruiz, Richard. "Orientations in language planning." NABE Journal 8, no. 2 (1984): 15-34.

[50] People in Need. *People in Need Czech Republic.* n.d. https://www.clovekvtisni.cz/en/what-we-do/humanitarian-aid-and-development/nepal/social-inclusion-and-protection

[51] Rhalmi, Mohammed. "Definition of testing, assessment, and evaluation." My English Papers (blog) November 26, 2019. https://www.myenglishpages.com/blog/definition-of-testing-assessment-and-evaluation/#:~:text=Simply%20put%2C%20a%20test%20refers,such%20as%20validity%20and%20reliability

[52] Alliance for Catholic Education's Read Haiti used phone calls, paper-based data entry, and photo messages to collect and return information on a 12-week interactive radio program. Data collectors received paper survey materials from a supervisor wearing PPE, called learners and caregivers to conduct the survey, and took pictures of the completed surveys to text to the central office.

[53] Elliott, Roxana. "Interactive Voice Response vs. Computer Assisted Telephone Interviewing for survey research" (blog). Geopoll, October 3, 2019. https://www.geopoll.com/blog/interactive-voice-response-vs-computer-assisted-telephone-interviewing-research/

[54] As FRI described, not all countries have this E1 or T1 infrastructure. If IVR systems do not have sufficient call-in capacity, they are only accessible to people who own their mobile phone and can wait long periods on the line (as described earlier).

[55] The Worldreader BookSmartApp and Rumie tablets capture reach and engagement data using analytics. The apps are typically pre-loaded on a smartphone or tablet or accessed on a computer. They log users' locations (and age and gender for registered Worldreader users) in addition to the content they engage with and the frequency of engagement. Rumie tablets have been used in 30 countries to provide educational videos, activities, games, and books (Moon et al. 2017, Rumie n.d).

[56] Gaëlle Simon, personal communication, September 22, 2020

[57] See Trochim for more details on sampling https://conjointly.com/kb/sampling-statistical-terms/

[58] Population Association of America. "Data collection during the pandemic: challenges and opportunities." Webinar, August 18, 2020.

[59] GlobalEdTechHub. "Over the last 2 months, we've been working with @BRACworld & @SAK_Sweden to test a free telephone helpline connecting students and teachers in rural #Afghanistan☎ Here are 6 things we've learnt from our #MVP - testing the concept with 50 teachers + their existing students." Twitter, July 23, 2020. https://twitter.com/GlobalEdTechHub/status/1286282087348174849

[60] Walls, Elena, Caitlin Tulloch, and Christine Harris-Van Keuren. *Cost analysis guidance for USAID-funded education activities.* Washington, DC: USAID, 2020. https://www.edu-links.org/sites/default/files/media/file/USAID%20Cost%20Analysis%20Guidance_Final%20Feb20_0.pdf

[61] Trochim, W. M. K. "Introduction to evaluation." Conjointly. n.d. https://conjointly.com/kb/introduction-to-evaluation/

[62] GeoPoll. *GeoPoll Audience Measurement (GAM) Methodology and Interface Overview.* GeoPoll, 2019. https://cdn2.hubspot.net/hubfs/325431/GAM_Methodology_and_Interface_Overview_August2019.pdf?utm_medium=email&_hsmi=76620038&_hsenc=p2ANqtz--qyA4znGUt0_weF3wWTScU4DdqSKWSesm0RE5F4omQmIDrhwobaF728-RkgUdS_FVr0HX6FN3GCkc_wHP4YetMhCPyFw&utm_content=76620038&utm_source=hs_automation n

[63] Education Development Center (EDC). "Interactive Audio Instruction (IAI)." EDC, 2019. http://idd.edc.org/our_work/interactive-radio-instruction-iri

[64] Education Development Center (EDC). "Audio Now!: responding to COVID-19" (blog). EDC, April 1, 2020. https://www.edc.org/audio-now

[65] Angirst, Noam, Peter Bergman, David K. Evans, Susannah Hares, Matthew C. H. Jukes, and Thato Letsomo. "Practical lessons for phone-based assessments of learning." BMJ Global Health, 2020. 5:e003030. doi:10.1136/bmjgh-2020-003030

[66] Angrist, Noam, Peter Bergman, Caton Brewster, and Moitshepi Matsheng. "Stemming learning loss during the pandemic: a rapid randomized trial of a low-tech intervention in Botswana." Centre for the Study of African Economics Working Paper WPS/2020-13, 2020.

[67] Angirst, Noam, Peter Bergman, David K. Evans, Susannah Hares, Matthew C. H. Jukes, and Thato Letsomo. "Practical lessons for phone-based assessments of learning." BMJ Global Health, 2020. 5:e003030. doi:10.1136/bmjgh-2020-003030

[68] Angrist, Noam, Peter Bergman, Caton Brewster, and Moitshepi Matsheng. "Stemming learning loss during the pandemic: a rapid randomized trial of a low-tech intervention in Botswana." Centre for the Study of African Economics Working Paper WPS/2020-13, 2020.

[69] Farm Radio International. n.d. https://farmradio.org/

[70] Hudson, Heather. E., Mark Leclair, Bernard Pelletier, and Bartholomew Sullivan. "Using radio and interactive ICTs to improve food security among smallholder farmers in Sub-Saharan Africa." Telecommunications Policy, 41 no. 7-8 (August 2017): 670-84. http://dx.doi.org/10.1016/j.telpol.2017.05.010

[71] Borzekowski, Dina L. G. "A quasi-experiment examining the impact of educational cartoons on Tanzanian children." Journal of Applied Developmental Psychology, 54 (2018): 53-59.

[72] Borzekowski, Dina L. G., Agnes Lucy Lando, Sara H. Olsen, and Lauren Giffen. "The impact of an educational media intervention to support children's early learning in Rwanda." International Journal of Early Childhood, 51 (2019): 109-126.

[73] Borzekowski, Dina L. G., L. E. Kauffman, E. A. Dura, M. Chale, and C. B. Bauer. *Examining the impact of Akili and Me's new content on Tanzanian children.* Prepared for UBONGO and the Human Development Innovation Fund. College Park, MD: University of Maryland, 2020.

[74] Techopedia. "Mobile Phone." Techopedia August 20, 2020. https://www.techopedia.com/definition/2955/mobile-phone

[75] Elliott, Roxana. "Interactive Voice Response vs. Computer Assisted Telephone Interviewing for survey research" (blog). Geopoll, October 3, 2019. https://www.geopoll.com/blog/interactive-voice-response-vs-computer-assisted-telephone-interviewing-research/

[76] Norton. "What are cookies?" Norton. n.d. https://us.norton.com/internetsecurity-privacy-what-are-cookies.html

[77] Techopedia. "Feature Phone." Techopedia, February 5, 2016a. https://www.techopedia.com/definition/26221/feature-phone

[78] USAID. *USAID Education Policy.* Washington, DC: USAID, 2018b. https://www.usaid.gov/documents/1865/2018-usaid-education-policy
[79] Ho, Jennifer and Hetal Thukral. *Tuned in to student success: assessing the impact of Interactive Radio Instruction for the hardest to reach.* Washington, DC: Education Development Center, 2009.
[80] Anzalone, Stephen, and Andrea Bosch. "Improving education quality through Interactive Radio Instruction." Africa Region Human Development Working Paper Series No. 52. Washington, DC: The World Bank, 2006.
[81] Mitranescu, Miruna. "IVR Definition and Benefits." Aircall (blog), August 25, 2016. https://aircall.io/blog/call-center/interactive-voice-response/
[82] Techopedia. "User Interface (UI)." Techopedia ,November 11, 2016b. https://www.techopedia.com/definition/4685/user-interface-ui
[83] G2. "Best Learning Management System." G2. n.d. https://www.g2.com/categories/learning-management-system-lms
[84] University of Toronto. "What are learning outcomes?" University of Toronto, 2020. https://teaching.utoronto.ca/teaching-support/course-design/developing-learning-outcomes/what-are-learning-outcomes/
[85] Techopedia. "Log File." Techopedia, January 11, 2017. https://www.techopedia.com/definition/5445/log-file
[86] Twilio. "What are SMS and MMS and How do They Differ?" Twilio. n.d. https://www.twilio.com/learn/messaging/what-are-sms-and-mms
[87] Human Rights Council. *Report of the Special Rapporteur on the right to education: realizing the right to education through non-formal education.* Human Rights Council, June 23, 2017. https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session35/Documents/A_HRC_35_24_AUV.docx
[88] Techopedia. "Smartphone." Techopedia, February 25, 2019. https://www.techopedia.com/definition/2977/smartphone
[89] Robinson, Christopher. "What is survey software? A comprehensive guide to benefits, features, types, pricing and more." Finances Online, 2020. https://financesonline.com/survey-software-comprehensive-guide-benefits-features-types-pricing/
[90] Teaching Excellence in Adult Literacy (TEAL). *Universal Design for Learning.* American Institutes for Research, 2010. https://lincs.ed.gov/sites/default/files/2_TEAL_UDL.pdf