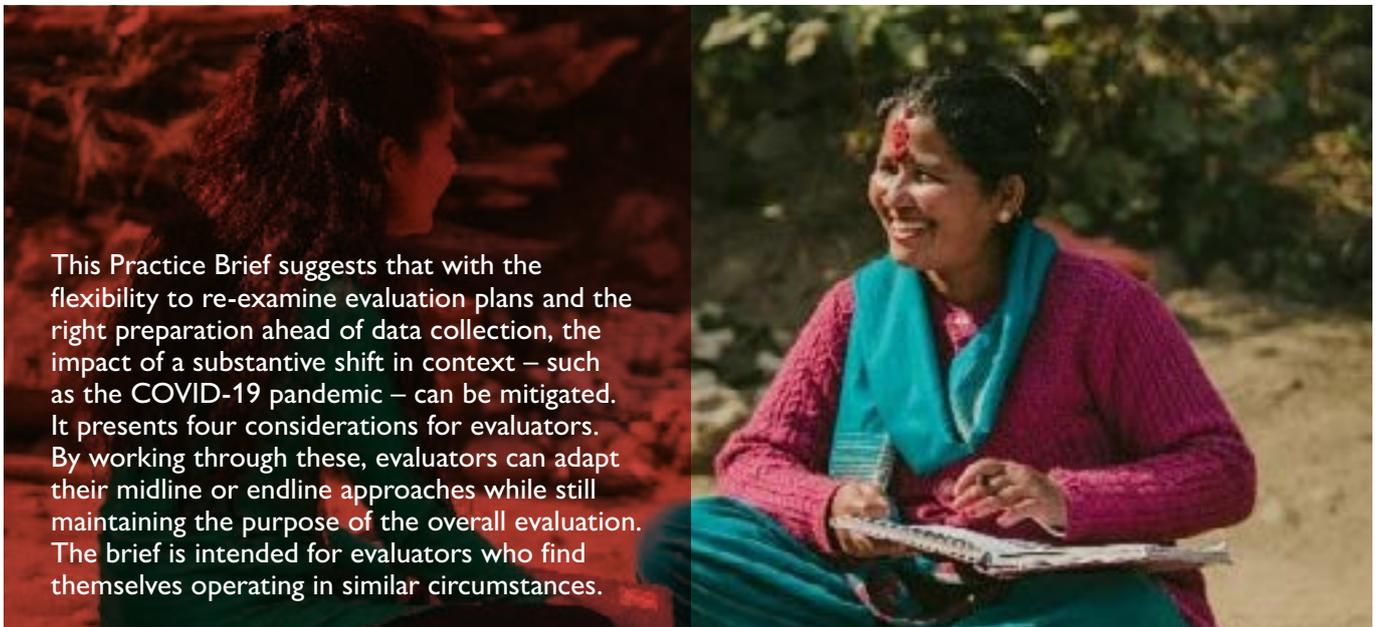


Pivoting evaluation designs during (and after) COVID-19

Four considerations for successfully adapting an ongoing evaluation in the midst of a pandemic



© VSO Nepal

This Practice Brief suggests that with the flexibility to re-examine evaluation plans and the right preparation ahead of data collection, the impact of a substantive shift in context – such as the COVID-19 pandemic – can be mitigated. It presents four considerations for evaluators. By working through these, evaluators can adapt their midline or endline approaches while still maintaining the purpose of the overall evaluation. The brief is intended for evaluators who find themselves operating in similar circumstances.

For Girls' Education Challenge (GEC) projects, evaluation designs were established at the start of the project, ahead of baseline data collections. The evaluation designs guided all evaluation points (baseline, midline and endline) by articulating the evaluation questions, establishing sample sizes, analytical models and requisite tools. This Practice Brief summarises adjustments made by evaluators to previously established evaluation designs in the midst of COVID-19 school closures – during which many assumptions that underly the original evaluation questions, sample sizes, analytical models and tools no longer held true.¹

When the World Health Organisation declared a worldwide pandemic on 11 March 2020, most of the 41 GEC projects had completed baselines and, in many cases, midlines. In the months that followed, several were due to complete midline and endline evaluations. In order to carry these out, evaluators had to revisit their original evaluation designs to accommodate emerging restrictions to data collections and programmatic changes to interventions. Restrictions and adaptations to interventions changed who can and should be included in the evaluation,

their experience with project interventions, the role of the project since the last evaluation point in girls' learning, transition and sustainability of activities, and outcomes. In other words, evaluators had to revisit the evaluation design to adapt upcoming evaluation points to remain relevant to current activities while also determining the extent to which comparability to previous evaluation points remained viable – particularly GEC's primary outcomes of learning, transition and sustainability.

Adapting designs is not a new venture on GEC projects. Even before COVID-19 school closures, evaluations were routinely adapted. The list below provides a few examples:

1. Adapting evaluation designs in response to findings: a high proportion of girls report lack of interest in a project activity, warranting a follow-up survey.
2. Adapting evaluation designs when availability of respondents shifts: to reach a particularly hard-to-reach population, evaluators need to develop a targeted set of tools and data collection visits during a subsequent evaluation point.

¹ See the following paper for a review on the challenges, opportunities and lessons learned to advocate for rigorous impact evaluations in fragile and humanitarian settings: <https://www.unicef-irc.org/publications/pdf/Impact-Evaluation-in-Settings-of-Fragility-and-Humanitarian-Emergency.pdf>

- Adapting evaluation designs when the context changes: if a new policy mandates that all students transition from one grade level to another then an alternative approach to assess readiness of transition among participants is needed.
- Adapting evaluation designs when there are delays to interventions: an intervention is extended from nine to 12 months requiring tools to capture additional information regarding the additional skills, knowledge or behaviours acquired by participants.

However, the COVID-19 school closures triggered a seismic shift in GEC evaluations. While previous adaptations were made on a project-by-project basis – tweaking items on a survey, adding respondent groups to the sample, or adjusting the indicator reporting methodology – March 2020 represented a pivotal point in **all** GEC evaluations. Across the portfolio, evaluation designs had to be re-examined so that upcoming evaluation points adequately reflected the on-the-ground realities of learners, project interventions and/or the context in which girls were now learning.

This brief describes four areas of consideration to guide mid-course corrections to evaluation designs. It draws from the experiences of re-scoping four endline evaluations between July 2020 and June 2021.^{2,3} The four areas of consideration are intended to support evaluators to make decisions about upcoming evaluation points that strike a balance between maintaining the original design and adapting to reflect the current reality of the intervention. As a result of these considerations, an evaluator may still find that an upcoming evaluation point remains relatively similar to the original design; on the other hand, they may make decisions that result in notably different tools, samples, analysis plans and/or evaluation questions.

GEC Evaluations Prior to COVID-19

Since the GEC started in 2012, evaluation designs featured a quasi-experimental, difference-in-difference (DID) approach⁴ to assess the impact of activities on girls involved in project activities. As a result, evaluations were conducted by external evaluators and were required to have a tracked, longitudinal sample. Furthermore, the evaluations used tools that measured outcomes in learning, transition and sustainability, as well as intermediate outcomes such as attendance, life skills, quality of instruction, attitudes towards girls' education and safeguarding practices.

Quantitative and, in equal measure, qualitative approaches were used to collect a range of evidence against key evaluation questions:

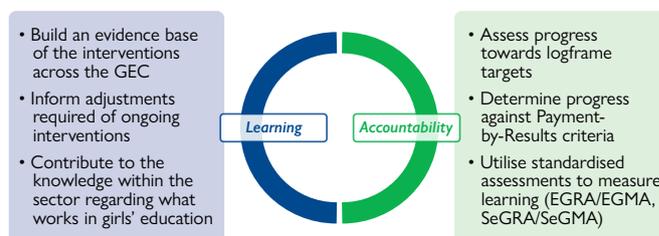
- What impact did GEC funding have on learning and transition of participating girls?
- What works to facilitate the transition of marginalised girls through education stages and to increase their learning?
- How sustainable were activities funded by the GEC?

Findings had to be generalisable to the populations involved in interventions, so quantitative samples had to be large enough to satisfy power requirements while also including all sub-groups in the population. Qualitative samples were drawn to ensure depth of inquiry and specifically to include hard-to-reach populations. All tools were intended to draw conclusions about the effectiveness

of programming and, to the extent possible, measure constructs consistently across projects. For instance, to assess gender perceptions among teachers, evaluators adapted items from a GEC teacher survey template of existing items from surveys.

Broadly, GEC evaluations were designed to serve two high-level purposes, illustrated in Figure 1:

Figure 1: Learning and accountability purposes for GEC evaluations prior to COVID-19

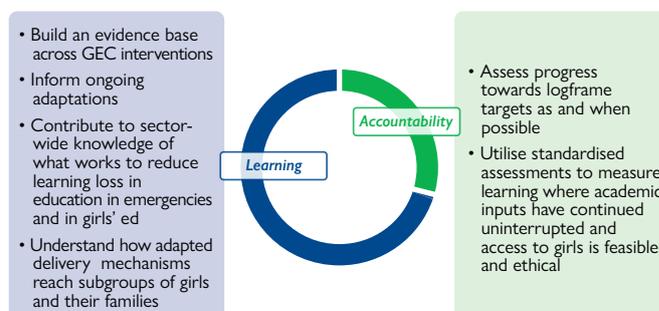


GEC evaluations in response to COVID-19

As a result of the disruptions to learning during the COVID-19 pandemic, the ability to uphold the accountability purposes dwindled. Logframe targets were either no longer valid due to lost intervention time or did not reflect adaptations in interventions. Payment-by-Results was already under revision across the portfolio prior to COVID⁵ and the ability to track girls across time became significantly more challenging, as school buildings or community-based centres were no longer the primary point of contact with girls.

At the same time, the demand for learning increased. Projects pivoted into uncharted territory in **how** they delivered inputs in addition to **what** those inputs were. Project teams partnered with **community-based structures** in place of school-based stakeholders to forge new pathways for delivering learning and wellbeing supports to girls. Projects expanded the types of activities they engaged and provided a wider variety of mental health and physical wellbeing supports to girls, their families and communities in response to the pandemic. All the while, projects kept learning interventions on the radar, albeit in varying levels of intensity depending on the needs and accessibility of girls. With the shift in needs, restrictions and programmatic changes on projects, the balance between learning and accountability tipped, as illustrated in Figure 2:

Figure 2: Shift in GEC evaluations during COVID-19



[link to thematic review]

² Discovery Project implemented by ImpactEd and endline evaluation completed by Oxford Policy Management; Improving Girls' Access through Transforming Education Transition (IGATE-T) implemented by World Vision UK and endline evaluation completed by Limestone Analytics; GEARR-ing Up for Success After School implemented by PEAS and endline evaluation completed by Jigsaw Consult; and Girls Access to Education (GATE-GEC) implemented by Plan UK and endline evaluation completed by NFER.

³ The Practice Brief is based on endline reports, inception reports, and the endline terms of references for four GEC projects that conducted endline evaluations between July 2020 and June 2021.

⁴ Difference-in-difference designs compare the difference from one timepoint to the next (first difference) between a treatment and a comparison sample (second difference).

⁵ Not all projects were eligible for PbR as not all project evaluations were able to establish a comparison group either for ethical reasons or no reasonable comparison group was possible.

The balance of learning and accountability varied from project to project. Evaluators and project teams, as well as FCDO through the Fund Manager team, regrouped to reconsider what they wanted to achieve in upcoming evaluation points. In some cases, this exercise triggered a complete overhaul of tools, sampling approach, analyses and guiding evaluation questions. In other cases, indicators, tools and/or methodologies were tweaked or underwent slight revisions. Many projects fell somewhere in between.

Although the focus of the GEC remained on learning, transition and sustainability, and remained at the core of GEC evaluations, the project designs and COVID-19 restrictions dictated how these could be measured. For instance, evaluations carried out during 2020 and early 2021 included learning assessments where feasible but were complemented by national exam scores or project monitoring data to triangulate evidence and draw conclusions about learning during school closures. Similarly, the types of intermediate outcomes and how they were measured shifted in accordance with intervention redesigns and delivery mechanisms. For example, understanding the home or remote-based learning experience, as opposed to the classroom-based learning experience, required survey and observation protocols that were unique to each project’s context and approach and included parents, caregivers, and community members rather than trained teachers or mentors. Evaluations were also adapted by embracing qualitative data collection approaches as they afforded greater flexibility to respond to changes in restrictions and interventions and allowed the exploration of unanticipated outcomes. The constraints to data collection also influenced tools and methodological decisions more

fundamentally – determining the respondents that could be reached, the length of time available to conduct assessments and surveys, and the nature of questions that it was appropriate and feasible to ask.

What resulted was a range of evaluation designs that adapted evaluation questions, methodology, tools, sampling approach and data collection protocols to varying degrees to achieve a successful evaluation. From these experiences emerged four key considerations when adapting an ongoing design in the midst of the pandemic.

Four considerations when adapting evaluations

Within the shift in balance between learning and accountability, GEC evaluations were able to make the necessary mid-course adjustments to their design appropriate to the project and the context in which they operated. The experience of adaptations from GEC evaluations during the COVID-19 pandemic offers four main considerations for evaluators to deliberate in the future:

1. [Adapt the evaluation scope and timeline to match the pivot in project design during COVID-19](#)
2. [Adjust evaluation tools to capture modifications to project design while remaining logistically feasible](#)
3. [Make methodological choices in light of on-the-ground data collection and logistical constraints](#)
4. [Make sampling decisions based on what is likely to be true and not just what is needed](#)

Each of these considerations is described in the sections below,

	Case study 1	Case study 2	Case study 3	Case study 4
1 Adapt the evaluation scope and timeline to match the pivot in project design during COVID-19		✓	✓	✓
2 Adjust evaluation tools to capture modifications to project design while remaining logistically feasible	✓	✓	✓	✓
3 Make methodological choices in light of on-the-ground data collection and logistical constraints	✓			
4 Make sampling decisions based on what is likely to be true and not just what is needed	✓		✓	✓

1. Adapt the evaluation scope and timeline to match the pivot in project design during COVID-19

Why do the evaluation scope and timeline need to change?

Evaluation designs rely on assumptions about the project and its interventions, the population being studied and the context. For example, who the project intends to serve dictates the sampling frame and the Theory of Change dictates the nature and types of changes the evaluation seeks to measure. When these assumptions change, an evaluation that retains the same approach may compromise the quality, rigor and appropriateness of the resulting findings. For example, when projects shifted to remote learning, the evaluation questions had to be revised so that the role of parents, caregivers and others who supported girls' learning in place of a trained classroom teacher or mentor, could be assessed to interpret girls' learning outcomes. This shift in how the project intends to operate indicates a shift in its underlying Theory of Change.

In addition, operating project activities during a pandemic may lead to **unintended outcomes** and/or alternative activities that were not captured or anticipated in the original Theory of Change and, by extension, in the original evaluation questions. Without adjusting the evaluation, important pivots in the project could be left unexamined and findings may be disconnected from the adapted project design.

What strategies could be used?

A. Conduct evaluation design workshops. Such workshops with project teams will help the evaluators to understand any new or adapted interventions and changes to the project's Theory of Change, to consider the nature and extent of changes to the assumptions behind the original evaluation design and to revisit the evaluation questions.

How to conduct evaluation design workshops with international and local project staff and GEC's Fund Manager technical advisors during the inception phase of the endline evaluations

B. Better 'm' for better 'e'. First, take advantage of data available from monitoring exercises to improve evaluation points. Learning assessment data collected at isolated evaluation points can be triangulated and interpreted in the context of learning data from monitoring exercises and/or available national exam data. By considering all available evidence on student learning, evaluators can fill gaps in conclusions drawn from assessments conducted at evaluation points. For example, monitoring data and national exam data can fill gaps in topic areas or about specific populations that could not be assessed during the evaluation point. Often, national exam data is discarded because it does not directly align with topic areas or populations of focus in a project. However, relationships between evaluation learning data and secondary national data can provide important counterpoints to project-collected data, not to mention data available across a population (versus a sample at evaluation points).

Second, during the design phase and before beginning any primary data collection, map the available monitoring data to the evaluation questions to determine whether these data can be used to:

- 1) answer, even to some extent, the evaluation questions
- 2) refine and target evaluation tools, especially removing redundant items
- 3) identify specific subgroups of beneficiary for whom limited data is available and/or who have higher attrition rates from the project. By identifying sub-groups who have been strongly affected by the pandemic, alternative data collection methods, tools and methodologies may be required to capture their voices.

How to use monitoring data collected by the project to inform sample selection and identify 'missing voices'

C. Explore alternate methodologies. Outcomes Harvesting⁶ is particularly well-suited to projects operating in dynamically changing settings, where the predicted outcomes and Theory of Change do not capture the project's whole impact. Outcomes Harvesting can shed light on emergent outcomes that are not included in the original Theory of Change, and outcomes whose contribution pathways are not well understood through other credible sources of evidence. Contribution Analysis⁷, which relies on the relationship between inputs and outcomes as expressed in the Theory of Change, has also been adopted by some GEC evaluations during the pandemic. Contribution Analysis can generate a narrative analysis of the contribution of project inputs to observed outcomes without relying on a counterfactual. GEC projects often faced additional challenges accessing a counterfactual group during the pandemic for ethical and practical reasons.

How to use alternate methodologies including Outcomes Harvesting and Contribution Analysis to assess impact

⁶ https://www.betterevaluation.org/en/plan/approach/outcome_harvesting

⁷ https://www.betterevaluation.org/en/resources/guides/contribution_analysis/ilac_brief

2. Adjust evaluation tools to capture modifications to project design while remaining logistically feasible

Why do the tools need to change?

Changes to project design, assumptions of the evaluation, or changes in the target population may render items in existing tools insufficient or no longer appropriate. Without adjustments to items to bring them into alignment with the evaluation questions at hand, the reliability of the measures can be compromised. For example, to examine changes over time in intermediate outcomes such as parent's attitudes towards girls' education or instructional practice, reliability of the tools is paramount. Contextual changes can also affect response trends in less obvious ways. As context changes so too can the factors which affect **how** respondents will answer questions. For example, even if an item on a survey is deemed relevant during the pandemic (e.g. it asks about parents' perceptions of the quality of a learning packet being used in the same way as before school closures) it may not capture the change in parents attitudes about the ease with which they were able to obtain the packet from their local community centre.

Additionally, the logistics of the evaluation can influence the scope of the tools. For example, remotely led evaluations may be necessary where international or national travel is challenging (or unsafe or banned) and presents new challenges to enumerator training, tools development and validation, data collection and data preparation steps, such as translation, transcription and data management. School closures and physical distancing measures may render school visits or face-to-face research impossible. In turn, an evaluation design may need to adjust to the logistic constraints of data collection including adapting the tools for remote data collection, shortening tools to fit within the timeframe available for training and/or field visits, and accommodating the skillsets of available enumerator teams.

When multiple logistical challenges are present, such as remote training, remote data collection and/or changing field restrictions, the compounding effect can render data collection and preparation steps even more difficult. Instead, narrow and focused tools can yield targeted and reliable evidence that can be easier to manage and use.

What strategies to consider instead?

A. Use an iterative process. Sequencing field activities allows earlier data collections to inform subsequent ones. Sequencing can inform both tools development and sampling.

For example, a first phase of quantitative surveys with beneficiaries can be used to inform areas to explore further during a second phase of key informant interviews. In doing so, evaluators can strategically target subgroups and topic areas during subsequent data collections while maintaining a manageable, higher quality, evidence base. Second, earlier data collections can help refine response options and/or items to use in a later data collection tool, particularly when evaluators need to ask respondents about topics not previously measured. Third, quantitative surveys (conducted with a sample) can be followed-up with qualitative interviews conducted with a targeted subgroup to gather additional details from specific respondents (such as those with the strongest opinion about girls' rights to education). The additional details provided by the subgroup of respondents can be used by the evaluator to determine whether the quantitative results from the larger

group were reliable (e.g., if those with the strongest opinion about girls' right to education mention concerns about safety to and from school, then the responses on the quantitative survey likely are measuring attitudes about girls' education **and** safety).

An iterative process can also inform sample selection, where the initial sample provides contact information for other respondents of interest. This approach can be particularly useful when contact information for respondents is missing or incomplete, or when the respondent group – such as shadow audiences to a distance learning programme – are difficult to reach by the project teams alone. In this case, girls and their families participating in the project activities can identify other community members who may also be following the distance learning programme.

How to use an iterative process to inform tools and sampling

B. Consider adaptability when making methodological decisions.

Consider data collection methods that answer the evaluation questions while also being adaptable in real time. For example, consider replacing items with pre-determined response options with open response options. By using a blended survey (including closed and open questions), emerging issues can be captured in real time. Enumerator training becomes more important to the quality of the evidence collected while reducing the importance of piloting item and wording of response options. Particularly when the situation is ever evolving – such as establishment home-based learning centres on a rolling basis – tools have to be flexible so that enumerators can probe for more information depending on the girls' experience with the centres. Blended questions can also capture contextual information that may be missed in closed questions. Finally, consider replacing focus group discussions with key informant interviews. If face-to-face data collection is not feasible, congregating a focus group remotely can be quite challenging. Key informant interviews can be done either face-to-face or remotely by phone.

- How to replace closed-ended survey questions in favour of open-ended survey and interview questions
- How to conduct more interviews in place of large-sample surveys
- How to reach more respondents by designing tools that are appropriate for face-to-face or phone-based data collection
- How to conduct online surveys with government officials and project staff and use participatory research methods

3. Make methodological choices in light of on-the-ground data collection and logistical constraints

Why do we need to accommodate logistical constraints?

Typically, evaluation questions determine the tools and methodologies that are most appropriate to answer the questions. This linear approach assumes that resources and access are available to support the field activities required. During the pandemic, and when resources and/or access are constrained, field realities influence the types of evaluation questions and associated methodologies evaluators can pursue.

For example, data collection activities typically have a narrow window of time to be completed because of budget, availability of enumerators, access to respondents, among other constraints. When the most efficient access to respondent groups is compromised – such as the closing of school buildings – delaying data collection until reopening is often not an option. Instead, more resource-intensive and challenging household-level data collection may be required, in turn influencing the number of tools and sample size feasible within the resources. Ultimately, the evaluation questions that can be realistically answered by the evaluator are limited.

Without allowing the evaluation decisions to be iterative and non-linear, evaluators can find themselves in situations where the reality on the ground up-ends their best laid plans. For example, even where a project is able to measure learning outcomes among girls in project and counterfactual communities, the ability to assess girls in counterfactual communities may be compromised when enumerators reach the field and are unable to obtain consent from girls and their families in the absence of project-established connections with them. Instead, evaluators can assess the pros and cons with project staff, determine options to secure a counterfactual group, and consider a methodology that does not rely on a counterfactual group to answer alternative evaluation questions. By exploring multiple aspects of the evaluation design simultaneously – evaluation purpose, ethics, enumerator availability, number of days in field, to name a few – the evaluator can make an informed decision regarding scope of the evaluation, questions and tools.

Although the iterative process can feel disorganised and messy, it opens space for unanticipated constraints and affordances to come into the decision process. For example, establishing an evaluation question that relies on well-trained enumerator teams in qualitative approaches can be compromised if the evaluator has not simultaneously assessed the availability of enumerators during the pandemic and their ability to move around the community in light of restrictions.

What strategies to consider instead?

A. Prioritise ethical data collection practices that are feasible.

Re-visit established protocols for data collection and reconsider **what is feasible, ethical and of interest to the stakeholders**. This may include revising how you obtain consent, the administration protocols for instruments and/or how you store and use data. The ethics of remote or phone-based data collection are different from face-to-face methods and can influence the tools used and evaluation questions that can be answered. Also, consider how to balance the ethical considerations for data collection with the practical. Some types of information may not need to be collected by independent enumerators and

instead can be gathered by project staff or community-based contacts; other data may still need to be collected by an external party if it is particularly sensitive or central to the Theory of Change or core evaluation questions.

How to prioritise ethical and feasible data collection practices, including revised consent statements for phone-based surveys

B. Obtain updated contact information as early as possible to inform methodological decisions. Review existing contact information of respondents from previous data collection points and utilise this information to inform sampling decisions, data collection methods and even tool design. Challenges with how to reach respondents when accessibility is constrained can shape much of the evaluation design. To maximise coverage across respondent groups, **determine primary and secondary modes of contact** for the targeted population– for example, obtaining phone numbers for girls and their families via their teachers or mentors, community leaders, community organisations, or project staff within the community. With respondent groups for whom you do not have complete contact information, consider using face-to-face data collection methods in order to prevent building in bias to the sample which may be linked with why the contact information is incomplete (e.g., households who do not own a phone or respondents who are less connected to the project). In this case, reserving limited resources for face-to-face field activities with a targeted subgroup of respondents can be an important consideration when forming evaluation questions and making tools and sampling decisions.

•How to capture easier to obtain contact information early to inform methodological decisions
•How to determine primary and secondary modes of contact by reaching out to students via teachers, head teachers, and requesting teachers support where network coverage is especially challenging

4. Make sampling decisions based on what is likely to be true and not just what is needed

Why does the sampling design need to change?

Shifts in context in the midst of an evaluation typically undermine assumptions underlying the sampling frame, or the parameters about the population from which the sample is to be drawn. For example, evaluation designs that rely on a comparison between a treatment and a control group may be compromised if groups that were comparable before the crisis differ during the crisis in meaningful ways. For example, girls in a counterfactual district might have previously only differed from girls in the project district in their participation in interventions. However, during the crisis, the counterfactual district also faced comparably greater challenges in accessing water and sanitary resources. A comparison between these two districts on measure of wellbeing can no longer hold across time as the differences between districts have change.

Furthermore, sampling approaches may involve a trade-off between desired level of power and generalisability of findings. For example, a representative sample may no longer be feasible when the sampling frame is limited by the girls and families who own phones. Instead, the evaluator may gain depth and breadth of insights by targeting subgroups of respondents strategically but knowing that the overall sample mean the findings can be generalised. The trade off in this case may be warranted and welcome, since a representative sample from girls with phones would lead to biased findings as they exclude the voices of girls who do not own phones but have participated in project activities. By excluding girls without phones, the findings cannot be generalised because they inadvertently exclude girls from other subgroups. For example, girls in hard-to-reach areas (where phone signals do not reach) or girls from poorer households.

What strategies to consider instead?

A. Use an iterative process to improve upon subsequent sampling decisions. Deploy sequenced data collection so that earlier data collections can inform subsequent ones. For example, administer surveys with head teachers and teachers first so that they can use their networks to identify students in remote or heavily impacted areas. Teachers and head teachers can also help verify or fill in missing contact information or reach out to girls and their families to confirm their availability.

- How to target interviews with girls from the midline sample who were not already being tracked through project monitoring data
- How to leverage school staff to help identify and reach girls

B. Use what you know to select a sampling strategy, especially to mitigate known sampling challenges. Use existing data and/or a sequenced design to determine the nature and extent of changes in a longitudinal sample. This also needs to take into account rates of attrition, extent of project participation or differences between treatment and comparison communities as a result of COVID-19. Project monitoring data and interviews with project staff can shed light on the types of changes that may have impacted the sample,

including higher levels of attrition in some communities than others and the inability of the project to reach certain beneficiaries (and as such, unlikely to be reachable for the evaluation).

Using a sequenced data collection strategy, survey data from teachers or head teachers can help verify a beneficiary sample for later data collection, assess the extent of their participation in activities and identify respondents who may not have participated at all. They may also help identify challenges with reaching targeted subgroups and even help find solutions. Once you know what those 'gaps' are, select methods (such as qualitative approaches) during later parts of your data collection activities to target the populations you are likely to miss out on.

- How to determine the extent of changes in the sample by using midline data and monitoring data,
- How to identify early in the process any changes to the sample that may compromise assumptions of comparability

C. Have a Plan B. When attrition rates are difficult to estimate, before heading into the field, determine the minimum sample size and response rate required for the design and embed contingency plans into field protocols. For example, prepare enumerators – and budget – to conduct additional site visits or additional days of data collection (remotely, if necessary), in order to reach the minimum number of cases. Protocols for reaching out to unresponsive stakeholders are also important to establish, particularly during the pandemic when migration, changes in contact information, and evolving restrictions can result in a biased response sample. When respondents cannot be reached, establish how and how often enumerators attempt making contact, if and how they are to identify a replacement, and the methods and criteria for establishing replacements.

How to determine the minimum sample size to avoid confusion in the field regarding sample 'top-up'

D. Use techniques to check assumptions regarding the comparability of groups. For purposively selected groups (such as schools), there may be systematic differences that can compromise findings. Such differences may include those present before the pandemic and others introduced as a result of the pandemic. Either way, use strategies like Coarsened Exact Matching (CEM)⁸ to assess the comparability of treatment and control groups both before and after data collection, especially when your goal is to complete a difference-in-difference analysis.

How to use Coarsened Exact Matching Difference-in-Difference approach to assess the comparability of treatment and control groups during inception phase

8 Blackwell, Matthew, et al. "CEM: Coarsened exact matching in Stata." *The Stata Journal* 9.4 (2009): 524-546.

Conclusions

The landscape of evaluations under the GEC evolved through the COVID-19 pandemic to yield a broad array of designs, methodologies and methods. Evaluators responded to the enforced and unpredictable realities of school closures and re-openings, and restrictions on movement and physical gatherings at national and local levels in each country. They also responded to the heightened need to consider the health and wellbeing of data collectors and subjects alike, as well as changes, often dramatic, in the implementation activities of the projects.

This technical brief draws on just four of the evaluations that have been adapted because of the COVID-19 pandemic. It offers a variety of tips and recommendations for other evaluations impacted by the COVID-19 crisis. The four evaluations from which these considerations were drawn were successful in completing their endline evaluations, including responding against relevant indicators, ensuring comparability where appropriate, and ensuring all stakeholders had a shared purpose for the evaluation.

These evaluations also struck a balance between learning and accountability purposes. At times they retained measures for outcomes and intermediate outcomes across all timepoints. Sometimes they introduced new measures that captured new approaches in implementation or activities. The ability of projects and their evaluators to genuinely make mid-course corrections was afforded by the shift in requirements from FCDO. In the ever-evolving school opening and closing situations during the COVID-19 pandemic, the GEC Fund Manager and FCDO sought, and will continue to seek, learnings alongside holding projects accountable – and not instead of it. The lessons that emerge as a result of pivots in evaluations during the pandemic can guide upcoming GEC evaluations and the sector at large.

Importantly, the lessons learned may indicate opportunities for 'building back better' evaluations even when the effects of the current crisis come to an end. For example, iterative approaches to evaluation designs can lead to evaluators proactively aligning evaluation questions with ground realities during the design phase rather than as a consequence of data collection constraints. Additionally, techniques such as Coarsened Exact Matching can help check assumptions about comparability of treatment and comparison groups regardless of the context. Coarsened Exact Matching can be a routine part of evaluators' approaches to sampling. Finally, proactively preparing enumerators with alternative plans when respondents are not reachable can produce data with fewer biases and retain generalisability, and need not only be invoked during the pandemic. As in this technical brief, evaluators are encouraged to reflect on the assumptions that informed their original designs and assess how those have changed while also assessing the changes to data collection logistics and requirements. The net sum of that reflection may lead an evaluator down a different path entirely, or reinforce the decision to remain on the same path.

Girls'
Education
Challenge



*This brief was prepared by the Girls' Education
Challenge (GEC) Fund Manager Evaluation Team:
Hetal Thukral, Mehroz Alvi and Aimee Reeves.*

For more information, contact: learningteam@girlseducationchallenge.org | www.girlseducationchallenge.org

The Girls' Education Challenge is a project funded by the UK's Foreign, Commonwealth and Development Office ("FCDO"), formerly the Department for International Development ("DFID"), and is led and administered by PricewaterhouseCoopers LLP and Mott MacDonald (trading as Cambridge Education), working with organisations including Nathan Associates London Ltd. and Social Development Direct Ltd. This publication has been prepared for general guidance on matters of interest only and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, PricewaterhouseCoopers LLP and the other entities managing the Girls' Education Challenge (as listed above) do not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.