



## **Effectiveness of Evaluation Evidence to Influence Change in Primary Education Sub-sector in Uganda**

**James Wokadala**

School of Statistics and Planning, Makerere University,  
P. O. Box 7062, Kampala-Uganda.

**Clare Kemigisha**

Uganda Martyrs University,  
P. O. Box 5498, Kampala-Uganda.

### **ABSTRACT**

**A well-designed impact evaluation study in the education sector is relevant in informing decision and policymaking based on high-quality evidence generated from the research. In this paper, we provide a chronology on how we developed and implemented an impact evaluation on school facilities grant (SFG) programme designed to support Government aided primary schools in Uganda – to construct classrooms, sanitation facilities as well as teacher houses. We describe the appropriate methodology, randomization designs implemented and the target groups. Results revealed that selecting a clear intervention with identified observable and unobservable characteristics is important for the robust results. Measuring attribution on outcomes may also require a balance on internal and external validity checks and analysis. The internal validity, in addition to statistical measures also requires stakeholders' engagement in the impact study process as a way of promoting social audits of public programmes. Finally, a well-planned and designed evaluation often leads to increased public trust in the findings and evidence uptake hence informing policy decision in management of programmes and projects.**

### **1. Background**

Conducting impact evaluations in the education sector is crucial in informing decisions and policy making based on high-quality evidence generated from both observation and experimentation research methods. Well-designed evaluations are relevant in determining whether a programme has had an impact on the target population by assessing how a project has delivered on its outputs in relation to intervention inputs and its cost-efficiency. In addition, researchers have been able to use evidence to make informed decisions, identify areas for improvement and hold policy makers accountable hence guiding proper implementation of development projects /programmes. Further, evidence can be used to determine the value for money in delivery of the education services (Khandker et al., 2010; World Bank, 2012).

Evidence was generated from impact evaluation of school facilities grant (SFG) on improving access, attendance and learning outcomes in primary education in Uganda. The SFG is a government programme introduced in 1998 under the poverty action fund. The SFG focuses on assisting the

populated and neediest schools acquire physical infrastructure, in response to the national and international (EFAs and MDGs) commitments of universal primary education. The SFG aims at: a) promoting equitable access to primary education; b) building capacity within districts and the local communities as well as supporting supervision on sites; and c) alleviating poverty. The evaluation sought to answer the following question: Do schools that receive SFG register better learning outcomes than those that do not receive SFG? According to Khandker et al. (2010), to attribute an outcome to a particular intervention, impact evaluation should determine the instruments through which beneficiaries are responding to the intervention. In his study, education programmes usually appear promising at their outset, but with time they do not generate the expected outcomes. Therefore, in such cases conducting an impact evaluation would enable policy makers understand what works, when, where and how. This requires use of appropriate research methods.

It is a universal fact that more education is associated with higher earnings with the global average rate of returns to schooling estimated at 10%. However, the positive benefits of education differs from country to another depending on factors related to gender, socio-economic status, geographical location, ethno-linguistic background among others (Montenegro and Patrinos, 2014; World Bank, 2012). The rigorous evaluation studies often isolate the factors and are able to measure attribution in the context. Well-conducted impact evaluations of programmes often lend empirical support to the various hypotheses on the education interventions needed by countries to adopt education reforms. According to Gertler et al. (2007); first, an appropriate model of behavior; second, detailed quantitative and qualitative data over an appropriate period of time that measures the response of beneficiaries to the intervention; third, an identification strategy that allows the measurement of a counterfactual. The counterfactuals are critical to understanding the effects of receiving the intervention relative to the control group. The valid comparison groups or counterfactual is important to determine the causal linkages and eliminates doubt of other attributes having an effect on the outcomes (Lewis and Patrinos, 2012). Evidence reveals that access to education enhances economic growth and poverty reduction as the education systems help expand knowledge and promote skills hence propelling labour productivity.

Rigorous impact evaluations also take into account the political economy (Grindle, 2004). There are stakeholders within the system with vested interests and understanding on how inputs, process and outcomes should be. Often finding the systematic ways to overcome these political interests and sentiments is challenging. Thus stakeholder engagement is key in this aspect, and effective communication is critical in paving the way for the reform and evidence-based research (Khemani, 2007; Majumdar et al., 2004).

The rigorous evaluations can be feasible and credible if the process is guided by the results based management framework. The United Nations define the Results Based Management (RBM) as an evaluation strategy by which all actors, contributing directly or indirectly to achieving a set of results, ensure that their inputs and processes contribute to the desired results to inform decision-making on the design, programme delivery, accountability as well as reporting (UNDG RBM Handbook, 2011). The evaluation framework hinges in three pillars: Monitoring, Evaluation and Planning; all which cascade into: setting the vision, defining the results map, planning the evaluation, implementing the design and measuring contribution to outcomes. Well- designed evaluations often help to answer other related questions , such as, are the goals intended at the policy formulation being achieved; and how can any policy impact be proved.

Thus the main objective of this study is to demonstrate the effectiveness of the robust evaluation to influence change in the primary education sub-sector in Uganda. In particular, the study attempts to answer the following research questions: First, how is the choice of the appropriate design of the

evaluation made? Second, what are the implementation fidelities and internal validity measures in robust impact evaluation? And third; how is evidence generated and communicated to the stakeholders?

The study demonstrates the evaluation design critical to the success of generating credible evidence. Besides, evidence can pass the internal and external validity tests if the intervention implementation arrangements are equally robust. The paper further reveals that, stakeholder involvement at every stage of the impact evaluation process is necessary as it increases ownership of results and up-take of evidence to inform policy in the primary education sub-sector in Uganda. In this paper, the remaining sections include: appropriate research design, data sources and types of data, sample size calculation and determination, timing and duration of the evaluation, designing an appropriate intervention, stakeholder engagement and evidence up-take, and finally conclusion.

## 2. Analytical Framework for Monitoring and Impact Assessment, and Theory of Change

### 2.1. Framework

An impact evaluation to be useful and robust, it depends on the evaluation framework and suitable estimation techniques employed. Besides, sufficient data must be collected to provide basis for sound analysis. Quantitative and qualitative should be collected using participatory methodologies. There are often key steps and questions in impact evaluation. The available literature offers a criteria and approaches, methods, tools and analytical techniques for the various kinds of analyzes (Baker, 2000 & Spath, 2004). The **Figure 1** provides a flow framework that is employed in the preparation, implementation, monitoring and measurement of the development interventions. The start off step is the review of the secondary baseline data and development of theory of change. Under the first phase, we discuss the evolution of the intervention and its logic, the initial problem analysis that form the hypothesis regarding impact.

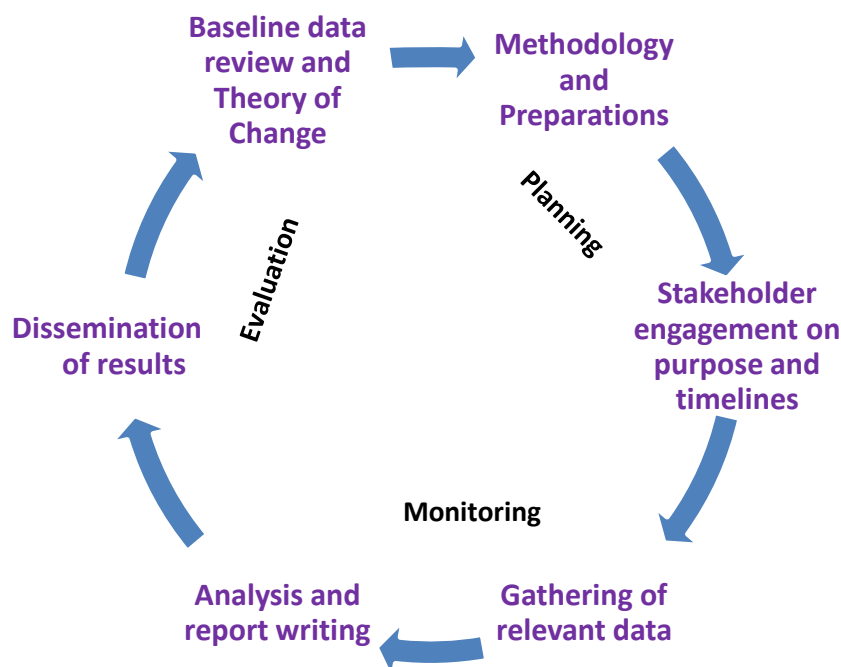


Figure 1: Framework for Monitoring and Impact Evaluation Assessment

The second phase entails the use of the logical framework to develop the methodology for evaluating the impact of the intervention as well as identifying the key evaluation questions to explore and test in the analysis. The methodology is often guided by certain critical assumptions that relate to internal and external validity. Phase three is the stakeholder engagement and sharing initial information with them particularly on purpose and timelines. The stakeholders should be aware of the methods to be used and how the results will be disseminated. Gathering of the relevant data is a critical stage in the evaluation framework. The extent of the collection of data and data types largely depends on the questions set and estimation techniques to be employed. Impact evaluations usually involve comparing the baseline and post-intervention data from both the control and treatment groups. The indicators to be compared depend on the type of the intervention being evaluated. Such indicators might include micro – individual or household level information; meso – institutional data; and macro – national level indicators, respectively. Data is collected through personal interviews, focus groups, key informants, and anecdotal information. The mapping tools are equally useful – not only basic, hand-drawn maps but computer tools that can overlay data to illustrate changes by the intervention (Spath, 2004). The analysis phase often traces the various activities implemented to determine the impact of the intervention on the target group. The evaluators assess the impact from the perspective of the beneficiaries themselves and identify factors that contributed to the outcome. The analysis is then transformation of information into the answer depending on the questions set – in form of a report. The final but relatively important stage is the dissemination stage that influences the level of uptake of the evaluation findings. Dissemination usually involves several methods and packages that suit particular audiences. Evaluations commissioned by the policy makers or technocrats in the public sector usually policy briefs are generated that inform their programme designs and implementation. The evaluation results may target the general public and disseminated through print media, television, CDs, PowerPoint as well as academic reference materials. The next section provides the appropriate designs suitable for education research.

## **2.2. Theory of Change**

The intervention logic or theory of change in education context depicts the causal links from inputs to impacts through processes, outputs and outcomes. The framework was developed through participatory approach i.e. workshop that attracted various stakeholders from the field of policy, media, politicians, funders, academicians and development partners. The evaluators synthesized the contributions made during the Theory of Change workshop to establish the key evaluation questions as well as a timeframe for undertaking this process evaluation activity.

The theory of change framework demonstrated that achieving the final outcome of UPE such as “*Better Learning Achievement*” is a process, based on a number of aspects such as program objectives and inputs holding certain assumptions true. This determines the activities that the programme stakeholders can implement. The implementation of the programme activities result into numerous outputs which lead to intermediate and final outcomes. It should be noted that the theory of change operates on a number of given assumptions within wider scope of the sector.

The ToC emphasizes SFGs and PTAs as important interventions to be evaluated. SFGs are supposedly meant to support school facilities construction and maintenance such as classrooms, teacher houses, water and sanitary facilities; supply of desks, chairs, school library, co-curricular facilities among others. The SFG funding was linked to UPE objective one to three – providing facilities to improve equitable access, thus making education affordable especially to the needy. The immediate outputs were improved enrolment, attendance, achieving gender parity, reduction in dropout rate thus improving completion rates and ultimately academic achievement.

### 3. Designs and Appropriateness of the Research Design

#### 3.1. Designs

There are two research methods employed to gather and analyze both quantitative and qualitative methods. In a programme evaluation, the classification is follows;

- (i) **Non-experimental (observation):** aimed at exploring the casual relationships but does not assign subjects to a treatment or control group to determine the effect of an intervention to a group (Hinton 2015). To attribute the changes to an intervention in a school two methods are considered. The methods include; first, **self-selection bias** that occurs when an individual, school or teacher participate in an intervention compared to those that do not participate; second, **non-random programme placement** which occurs by comparing the indigenous beneficiaries of an intervention with participants randomly selected to benefit from an intervention.
- (ii) **Quasi-experimental:** which focuses on assigning participants to either a treatment to receive an intervention or control group that does not receive an intervention. The analysis takes care of the potential bias that may result from non-randomization by applying a number of methods, namely;
  - **Randomized proportion or encouragement design:** especially used in situations where an intervention has less control over participant's compliance (Gertler et al., 2011). The research randomizes the encouragement group to receive treatment. Tracking of outcomes for those who receive and those that do not will obtain estimates of both the encouragement and intervention itself.
  - **Propensity score matching:** is applied to non-beneficiary characteristics to create treatment and control groups, however participants are not randomized. The focus is on observing the probability to participate characteristics (Krueger and Zhu, 2004). This method is limited by how to determine factors not observed by researchers, need for data from both participating and non-participating groups, a lot of data is usually collected hence making it hard to match all data collected.
  - **Differences-in-differences:** This method compares the baseline data and follow up data for treatment and comparison groups. It takes into consideration the mean difference between the after and before values of the outcome indicators for each of the treatment and comparison group. The advantage of this technique is that from time to time empirical techniques are used to differentiate any bias.
  - **Programme phase-in:** All schools and districts create an effective counterfactual, where all participants are legible to receive treatment at different timelines, starting with some participants while other join later (Marcus and Berman, 2013; Khandker et al., 2010). For this method to be effectively applied, there should be no any differences in characteristics or political influence that determine the participants at any given time.
- (iii) **Experimental designs:** this puts into consideration having an equal chance of participants being in a treatment and control group. This method randomly assigns participants to a treatment or control group to determine the effect of the programme. It ensures validity and eliminates other variables. Statistical methods are then applied to compare the observed outcomes of the two groups hence allowing for counterfactual (Hinton, 2015).

The advantage of applying randomized control trials is that it limits bias, spillover effects, although partial compliance and randomization bias may still occur. On the other hand randomized control trials are very costly, may lead to attrition bias where participants in the treatment group reduce, spillover effects may occur.

### 3.2. *Appropriateness of the Research Design*

The reliability and robustness of the results depend on the appropriateness of the design as well as the effectiveness of the implementation arrangements. There is no one-size fits-all evaluation method at addressing certain areas and issues in generating evidence in education research but rather, the expertise of the professional can be utilized to ensure that the evaluation meets all needs and expectations while producing rigorous evidence (Lewis and Patrinos, 2012). The research design is a framework in which a research is undertaken and monitored, and it employs one or more research techniques throughout the whole evaluation process. Often impact evaluation studies employ techniques to gain insights into the real world. Different designs are more or less suited to exploring the wider applicability of the research findings to a variety of study contexts.

The SFG evaluation study employed Randomized Phase in Design by randomly assigning surveyed schools into Phase I (treatment) and Phase II (control). The selection of the sample schools was based on a multi-stage stratified random sampling design. The evaluation was conducted at four stages with the first stage involving selection of UPE schools from 20 districts of the four regions of the country based on their dire need of school physical infrastructure, and high pupil-to-classroom ratio of 55:1 and above. At the second stage, 14 to 16 schools from the urban and rural parts of the districts constituted the study sample. A total of 301 needy (of infrastructure) schools from 20 districts were identified and surveyed. Within the study schools, a sufficient number of pupils (>20-24) both males and females from grade 3 (at baseline) and grade 5 (at endline) were also selected using random numbers generated by an App Random UX. At the baseline, 160 schools were randomly selected to receive the SFG grant in FY2016/17 and FY2017/18 (Phase I) and considered the remaining 141 schools to receive the grant after 2019 (Phase II). This part of the study used primary data to establish the short term impacts considering both the observable and unobservable characteristics. The randomization was participatory and transparent involving the evaluators and district/school officials.

Another entry point into impact analysis of SFG programme is the unit cost and effectiveness analysis. Cost-effectiveness ratio (CER) analysis is an approach to inform decision making and not a substitute for it. The ratio simply shows how much it costs per pupil to get so units of output in the treated relative to the control schools. Per unit cost and outcomes are computed and related as ratios. The costs are the facility grants disbursed to schools during the intervention period while per unit cost is the cost per pupil in the treatment schools. Outcome measures in this case are enrolments, attendance and test score gains. We measured or established the cost per student in the intervention schools by simply dividing the total SFG expenditure in the survey schools by the number of students. We then used the three phases of SFG implementation to assess the effectiveness of the programme. If we let P1 stand for pre-SFG period, and P2 (2014/15 to 2016/17), we can then identify possible scenarios and cases. The CER lead to clear decisions as to whether to adopt and use SFG intervention. In symbolic notation, the Cost- effectiveness measure can be expressed as:

$$\text{Cost – Effectiveness}_{BUPE} = \frac{(\text{cost} / \text{pupil})}{(\text{Effect size})} \dots\dots\dots (1)$$

Where cost is the SFG allocated to the treatment schools; effect size is the impact or the average gain in units per school relative to the control school. As noted above, cost-effectiveness is ideally an efficiency measure. The CER is an efficiency ratio for comparing two systems on the basis of a specific cost-effectiveness measure (cost per graduate/pupil). The decision rule is that if efficiency ratio = 1, then pre and post SFG intervention are equally efficient. If efficiency ratio >1, then UPE intervention is less efficient

### 3. Data sources and types of data

In evaluation research, data sources are described as belonging to one of two categories: quantitative data and qualitative data. Quantitative data are typically data that can be expressed numerically, while qualitative data typically involve categorizing and classifying information rather than using numerical values. Both qualitative and quantitative methods of data collection play equally important roles during an evaluation, although the mixed methods of data collection are commonly applied. Evaluations can be conducted at two stages of intervention, namely; retrospective, occurring after the conclusion of an intervention or prospective, designed before the implementation of an intervention. Conversely, retrospective evaluation is not cost effective but allows collection of original data.

Quantitative data emphasize objective measurement and numerical analysis of data to illustrate the development or correlational relationships. Using existing data is the most cost effective way of conducting an impact evaluation, but it is important to collect administrative data on an intervention that is also active in a particular school or districts so that control measures are put in place prior to an evaluation. Available data collected using school census on school drop outs, repetitions among others may also be used. Relatedly, administering questionnaires to a number of school staff would enable the research collect reliable information (Gertler et al., 2012). Qualitative data usually involve direct interaction hence enabling the research collect descriptive information which does not involve numerical values. It does not require a counterfactual to make unintended implication. It further allows collection of information on an intervention from those who are being impacted by an intervention or even those not receiving any benefits. Direct observation should be conducted by a person that is not involved in the evaluation to provide unbiased data on an intervention.

In the SFG evaluation, both primary and secondary data as well as quantitative and qualitative data were collected as guided by the estimation techniques employed in the study. The primary data was obtained from schools, pupils, teachers as well as school management committee. We administered a specific tool on each of these respondents.

**Primary Data Collection:** Data collection tools/instruments were developed in line with the defined respondents and sampling plans. The tools were developed for each category of stakeholders to be interviewed or consulted – Pupils, Teachers and Head teachers, SMCs, parents, DIS, DEOs, CAOs, District planner and Chief Finance Officers (CFOs).

**Quantitative Data:** The quantitative data was collected at school level and included aspects of school information, enrolments and attendance, school facilities (availability, adequacy and quality), staffing, pupil performance at PLE, school governance as well as parents participation in school activities, school facilities grant and related funds. More quantitative data was collected from teachers particularly on socio-economic characteristics, teacher training and deployment, teacher school attendance and perceptions on SFG related challenges. Moreover, quantitative data from pupils included information related to pupil characteristics and school attendance, learning environment and socio-economic status at their homes.

**Qualitative Data:** Key informant interviews were conducted to DHT/SMC/Senior teacher to schools that have ever and never received SFG. The information collected was on aspects of availability and condition of school facilities, the facilities needs of the schools, the funding sources available to construct these facilities. A checklist of questions was administered to DEO to find out how the schools were selected to benefit from SFG grant, how the SFG funds were distributed to the schools

and their adequacy, the activeness of the NGOs in providing the funds for facility construction, information on perception of SFG and the suggestions to make the SFG effective.

The school level questionnaire was used to collect data on enrolment, pupil/teacher attendance, staffing, infrastructure, whether a school provides mid-day meals or not, school location, whether a school has ever received support from NGO or not among other data; key information collected from the teacher and SMC questionnaires was on their demographic, experiences and perceptions on the relevance of school infrastructure on learning. Pupil questionnaire contained data on demographic characteristics, pupil living conditions at home and school, and their experiences on the learning and learning environment. Besides, the numeracy and literacy tests administered to the pupils were another key primary source of data. Other primary data and sources were the qualitative information gathered from key informants and FGDs as explained in the previous section.

The quantitative methods were used to measure impact of school infrastructure on access, attendance and ultimately test scores. The qualitative study complements the quantitative findings to explain the 'how', and also capture experiences and perceptions regarding the relevance of classrooms, toilet facilities as well as teacher accommodation to learning and learning outcomes.

Quantitatively, the study used the existing data in EMIS to establish the impact of SFG using Propensity Score Matching (PSM). The schools that had received SFG, for the period 2014/15 to 2016/17 were identified and compared with those schools that were similar in characteristics (based on both output and input variables prior to receiving SFG. Secondary annual school census data from 2006 to 2017 was collected from MoES, Office of the Prime Minister, Ministry of Finance as well as from Districts Local Governments and analyzed using Propensity Score Matching (PSM) Technique. Schools that received SFG were matched with those that did not receive the grants to estimate the impact of SFG from 2007.

#### **4. Sample size calculation and determination**

Appropriateness and adequacy of the sample size should aim at determining a sample large enough to have significant statistical power and to be a representative of the entire population. For instance, the cluster-based randomized control trial with a programme at the school level would require a unit of treatment of 40 to 50 schools with 40 to 60 students at each school to enable the survey measure an impact against their teachers and families (Gertler et al., 2007). This sample size allows for observation of differences in student achievement examination scores.

In the SFG evaluation, the total sample size required was determined by taking into consideration several factors, the three most important being: the degree of precision (reliability) desired for the survey estimates, the cost and operational limitations, and the efficiency of the design. The sample was designed to provide indicator estimates for all the UPE schools put together.

***At Baseline***, areas with UPE schools were stratified into 15 zones that included the following: Central 1, Central 2, Central 3, South West, Mid-west, Far West, North West, Far East, Near East, Mid East 1, Mid East 2, North East, Mid North 1, Mid North 2, West Nile. The selection of sample schools was based on a multi-stage stratified random sampling design. In the first stage, the selection of respondents was based on the regions. In each of the selected regions, schools were segmented into best performing versus worst performing and a random sample of schools was then selected systematically.

***At process evaluation Stage***, the districts were selected to represent the four regional blocks comprising the geographical setting of the country. A representative sample of needy schools was



selected from the respective sub-counties and parishes. As per the evaluation design requirement, a sufficient number of pupils (>20-24) both males and females from grade 3 (at baseline) and grade 5 (at end line) were selected from each sampled school.

With respect to the sampling and sample size determination, a sufficient number (i.e. >14-20) of needy schools were identified in each of the chosen 20 districts for this study. The choice of the study districts and schools was guided by the critical review and consideration of external validity items. All the districts and schools were eligible to participate in the RCT; the SFG intervention was implemented in all the study districts and particularly the treatment schools; there were no differences in key output and outcome indicators at the baseline and the study allowed comparison of baseline characteristics for both intervention and control group; the treatment was delivered as intended and the monitoring of the progress was regular; the SFG programme operates under clear policy framework and there was clear description of treatment alternatives or models (i.e. classroom, teacher houses and sanitation facilities). There was a clear literature search conducted to identify the knowledge gap and the evaluability of the SFG intervention. There were also clear sources of data and data management techniques equally defined beforehand; ultimately, the outcomes were measurable with clear valid choice of statistics. Other aspects are sufficient sample size particularly of pupils and schools as well as adequacy of the length of follow-up.

The power calculations of our evaluation largely followed Djimeu and Houndolo (2016), a manual prepared by 3ie Senior Evaluation Specialists. We adopted a two-stage sampling process. In the first stage, 301 clusters (i.e., primary schools) were randomly assigned in two (treatment and control) groups. In the second stage, students were selected from each grade. As is standard in the literature, we assumed the significance level to be 5% and the desired power of the test to be 0.8. We also estimated the intra-class correlation (ICC) and population average “within” standard deviation of percentage test scores to calculate Minimum Detectable Effect (MDE). We used statistics from the National Assessment for Progress in Education (NAPE) conducted for grades 3 and 6 students in 2014 to calculate these estimates. The within-school correlation for numeracy and literacy test scores are 0.43 for numeracy and 0.46 for literacy, respectively. We used the higher of the two. Similarly, the average standard deviation of test score is 20.1. For the power of the test to be 0.8, the MDE for the two tailed study is 6.5 percentage points. This resulted into the minimum pupil sample size of 6000 determined beforehand, of which 50% was treatment and the other half was in the control schools.

## **5. Timing and duration of the evaluation**

### ***5.1. Timing of the Evaluation***

The **timing** and duration of the evaluation is largely determined by what decisions the evaluation is intended to inform and when. The factors include but not limited to; the intervention itself, data sources and sampling units, and indicators being measured. For an evaluation based on indicators, the student success outcomes, should be measured before the intervention and after two complete school cycles, this would ensure ample time to have test scores adjust to a new programme.

Whereas for an intermediate quality of education outcomes, such as drop-outs, repetitions should be measured twice each year during which the intervention is implemented. One at the beginning of the school year and the other at the end of the school year. However, the attendance compiled by the school management should be measured consistently throughout the year.

Undertaking evaluation of the relevance of SFG is anchored on the various national development frameworks (e.g., National Development Plan I – II & Vision 2040). The national development agenda places school facilities as priority for the country's economic and social transformation. The Education Sector Strategic Plan (2016/17 – 2019/20) emphasizes school infrastructure development as the key priority for promoting quality teaching and learning. There is also information needed on the links between school facilities and learning outcomes. This evaluation observes that the SFG intervention has been running to date for a period of 18 years with no specific and comprehensive evaluation undertaken. Thus this evaluation was timely as the MoES was in the process of reviewing her strategic plan (i.e. ESSP 2016/17 – 2019/20). Usually investments in school infrastructure are heavy, costly and time consuming; hence it is right and just to use evidence to guide these policy decisions in the sector. This study further provided an understanding of the impact of school physical facilities on equitable access and learning outcomes in Ugandan government aided primary schools. The evaluation chiefly answered the question as to what would be the situation (in pupil access/attendance and learning outcomes) in case there was no SFG intervention (counterfactual), especially looking at the recipients of the grant.

### ***5.2. Duration of the evaluation***

The SFG impact evaluation was executed in four broad stages during the period 2015 – 2019. Stage one was the process evaluation conducted in November - December 2015. It involved the analysis of the situation to understand the then prevailing situation in aspects of teaching and learning in the larger sample of selected schools in Uganda.

Stage two was the baseline survey conducted in November 2016, and it involved the following activities: design of tools and mapping of the eligible and ineligible SFG schools for impact; field visits and data collection; preparation of field reports processing; cleaning and analysis of field data; preparation of draft baseline report, as well as holding of consultative/ dissemination workshop to validate the study findings.

Stage three focused on the midline survey activities undertaken in January 2018 and involved among others, review of the midline concept note; sensitization and mobilization of stakeholders at national and local governments (LGs) to implement the SFG evaluation programme effectively; collection, analysis and reporting on secondary data; development of tools for qualitative study and presentation to evaluation sub-committee; conducting primary data collection; analysis and preparation of draft midline report; holding the consultative/dissemination seminars with stakeholders as well as presentation of final report.

The final stage was the endline survey conducted in September 2018, and the key activities included: review of the secondary information and design of tools; recruitment and training of research assistants and supervisors; field visits and data collection; processing, cleaning and analysis of field data; preparation of draft evaluation report and policy briefs; consultative/dissemination workshops and meetings; presentation of final report to stakeholders and submission of final report and other deliverables.

## **6. Validity and Reliability Analysis**

### ***6.1. Validity analysis***

Kothari (2004) defines validity as the indication of the degree to which an instrument measures what it is supposed to measure. Findings of a study are considered as valid if the measurement instruments are reliable. Validity itself is the measure of trustworthiness or strength of the findings or conclusion;

the evaluators focused at arrangements or approaches that increased confidence on the originality and truthfulness of the results.

The sets of questionnaires (i.e. questionnaire for the households, and qualitative tools - KII guide and FGD guide) that were used to collect data for this study were created. To ensure having valid content the study allowed adequate coverage of the objectives under study. In addition, the evaluation team also ensured the existence of construct validity through developing research instruments (questionnaires) with sound base in theory and conforming to the theoretical body of knowledge. Furthermore, the research indicated a relationship among the variables of other empirically tested constructs in similar studies.

### **6.2. Reliability analysis**

Reliability is the extent to which the applied data collection techniques provide consistent findings (Saunders et al., 2009). The reliability scale ran first for each individual variable and for all variables in STATA version 12. The reliability (internal consistency) of the study measured using Cronbach's Alpha ( $\alpha$ ) approach (Cronbach, 1951). Miller et al. (2001) stated that for better reliability the Cronbach's alpha should at least be 0.50 but a higher score reflects better reliability. Therefore in this evaluation, all the constructs measures that had  $\alpha$  above 0.5 provided an indication that there was reliability and internal consistency since  $\alpha$  is within the acceptable range above the minimum recommended 0.5.

Data quality assurance was undertaken to discover any inconsistencies and anomalies in the data and this improved the data quality. The process was phased into three stages;

- Pre-test the tools/instruments, prior to carrying out the stakeholder consultation. The purpose was to test early enough the adequacy and responsiveness of the tools/instruments to collect the information and data necessary for compiling the report. Ambiguities identified in the tools i.e. lack of clarity of question posed to respondents were identified and rectified as part of this process;
- Carried out induction of the whole team on application of tools and interviewing techniques of key informants. To reinforce the natural friendly setting of the key participants, role plays were conducted as part of the induction processes.
- During the actual field work, consultations were constantly made to review the collected data and experiences and challenges were shared and this fostered continuous data quality improvement.

Throughout the data collection exercise, the evaluators observed some quality checks. For instance, interviewers were accompanied by the supervisors to ensure that the identification and selection of all the respondents as well as the interviews were being conducted according to protocol. The data management team provided all the necessary checks on the interviews conducted immediately after completion of interview with respondent. At minimum, the checks were done to ensure completeness and accuracy. Call back-checks were conducted by phone calls to confirm that the interview took place as per protocol. During the call, the back-checkers randomly picked questions on the questionnaire to confirm the responses. Research team also reviewed the daily updates template to assess field work progress. Moreover, the field teams trained on the importance of the details and contact information were reviewed by the interviewer and supervisor on a daily basis. Incomplete entries were rejected until completion.

## 7. Stakeholder engagement and evidence up-take

The government is the major stakeholder interested in education, also often the funder, main provider and regulator of education programmes. Therefore, early engagement through meetings enables the research solicit for buy-in of the evaluation. Stakeholder engagement meetings provide feedback on making the critical findings palatable by government hence enabling them monitor the evaluation outcomes transparently. The only sure way to conduct sound evaluations is to define the context at each aspect of the evaluation process. This should be implied at each stage of the evaluation from initial design, to definition of the treatment, identifying indicators and evaluation size.

The SFG evaluation applied a participatory approach in order to clearly articulate the program theory of change and its attribution. The evaluator sought the involvement of key program implementation stakeholders in identifying gaps, generating the evaluation questions as well as sharing the findings gathered from the desk review that had been conducted.

Achieving quality teaching and learning requires collective responsibility both on part of government and parents as well. Under the SFG Programme, parents were urged to support the children by providing books, uniforms and other basic and essentials like day meals. Parents' mobilization is usually done through PTAs.

For ownership and sensitization of the SFG study, District Education Officers were constantly consulted right from the process evaluation, baseline, midline and Endline phases. These were so helpful in the selection of schools that had and had not received SFG. Procedurally, the randomization exercise in each district was conducted at the district headquarters with a team of evaluators, representative from OPM and MoES and the District Officials i.e. the CAO, DEO, DIS, Senior Education Officer, District Sports Officer, Principal Education Officer, Chairperson Education Community Affairs, the Speaker to District Council, District Community Development Officer, LCV Chairperson, etc.

## 8. Conclusion

Robust evidence is essential in informing and guiding policy and programming decisions in social sector such as education. Credible research and evaluation results are a basis to inform judgments, develop options, and guided choices on how to effectively and efficiently use resources. For best performance in student learning outcomes, rigorous evaluations should be conducted to increase our understanding of the interventions that improve educational outcomes. In particular, since randomized studies require fewer assumptions and reduce biases, they allow researchers produce rigorous findings. The Government has continued to embrace improvement in educational outcomes and therefore important to conduct impact evaluations of different interventions to enable policy decision makers take a decision on allocation of resources based on rigorous evidence based decision.

## References:

- Baker, J. L. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Directions in Development. Washington, DC: World Bank.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334
- Djimeu, E. W., D.-G. Houndolo (2016) Power calculation for causal inference in social science: sample size and minimum detectable effect determination, *Journal of Development Effectiveness*, 8:4, 508-527, DOI: 10.1080/19439342.2016.1244555.
- Government of Uganda (2010). Nationak Development Plan I (2010/11 – 2014/15). Kampala, Uganda
- Government of Uganda (2012). Uganda Vision 2040, Kampala, Uganda
- Government of Uganda (2015). Nationak Development Plan II (2015/16 – 2019/20). Kampala, Uganda
- Ministry of Education and Sports (2015). The Education Sector Strategic Plan (2016/17 – 2019/20)

- Gertler, P., H. A. Patrinos and M. Rubio-Codina. 2007. *Impact Evaluation for School-Based Management Reform*. Washington, DC: World Bank.
- Gertler, P., H. A. Patrinos and M. Rubio-Codina. 2012. Empowering Parents to Improve Education: Evidence from Rural Mexico. *Journal of Development Economics* 99(1): 68-79.
- Gertler, P. J., S. Martinez, P. Premand, L. B. Rawlings and C. M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank.
- Grindle, M. 2004. *Despite the Odds: The Contentious Politics of Educational Reform*. Princeton, NJ: Princeton University Press.
- Hinton, R. 2015. *Assessing the Strength of Evidence in the Education Sector*. London, England: United Kingdom Department for International Development.
- Khandker, S., G. Koolwal and H. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, DC: World Bank.
- Khemani, S. 2007. Can Information Campaigns Overcome Political Obstacles to Serving the Poor? in S. Devarajan and I. Widlund, eds., *The Politics of Service Delivery in Democracies*. 20 References | Generating Evidence in the Education Sector
- Kothari (2004) Research Methodology Methods and Techniques. 2nd Edition, New Age International Publishers, New Delhi.
- Krueger, A. B. and Zhu, P. 2004. Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist* 47(5) 658-698.
- Lewis, L. and H. Patrinos. 2012. *Impact Evaluation of Private Sector Participation in Education*. Reading, Berkshire, England: CfBT and Washington, DC: World Bank, 2012.
- Majumdar, S., A. Mani and S. Mukand. 2004. Politics, Information and the Urban Bias. *Journal of Development Economics* 75(1): 137-165.
- Marcus, A. (ed.) and D. Berman. 2013. *Cambodia: Challenges in Scaling Up Preschools*. Washington, DC: World Bank Group.
- Miller, S.D., Duncan, B.L., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, 2(2), 91-100.
- Moffitt, R., J. Fitzgerald and P. Gottschalk. 1999. Sample Attrition in Panel Data: The Role of Selection on Observables. *Annale d'Economie et de Statistique* 55/56: 129-152.
- Montenegro, Claudio E. and H. A. Patrinos. 2014. *Comparable Estimates of Returns to Schooling Around the World*. Policy Research Working Paper 7020. Washington, DC: World Bank.
- Späth, B., 2004. "Current State of the Art in Impact Assessment: With A Special View on Small Enterprise Development". Swiss Agency for Development and Cooperation. Available at: [http://www.deza.ch/ressources/resource\\_en\\_24414.pdf](http://www.deza.ch/ressources/resource_en_24414.pdf)
- Saunders, M., Lewis, P. & Thornhill, A. (2009). *Research methods for business students*, 5th ed., Harlow, Pearson Education.
- UNDG (2011). *Results-Based Management Handbook: Harmonizing RBM concepts and approaches for improved development results at country level*. United Nations Development Group.
- World Bank. 2012. *World Development Report 2012: Gender Equality and Development*. Washington, DC: World Bank.