Faith-based Metrics & Measurements

Date: November 12, 2015

Notes and Reflections on the meeting from Helena Manguerra, Dean Pallant and Jean Duff

**Meeting Participants, Organizations, and Contact information**

David Bronkema Eastern University dbronkem@eastern.edu

Mark Forshaw Global Scripture Impact mforshaw@americanbible.org  
Sneha Stephen Innovations for Poverty Action sstephen@poverty-action.org

Isabel Onate Innovations for Poverty Action ifalomir@poverty-action.org

Lincoln Lau International Care Ministries lincoln.lau@caremin.com

Jean Duff JLIF&LC JeanDuff@pfaithdev.org

Helena Manguerra JLIF&LC HelenaManguerra@pfaithdev.org

Dean Pallant The Salvation Army Dean\_Pallant@salvationarmy.org

Lily Talapessy UNFPA talapessy@unfpa.org  
Jill Olivier University of Cape Town Jill.Olivier@uct.ac.za

Bruce Wydick University of San Francisco wydick@usfca.edu

Invited:

Ellen Strohm Global Scripture Impact estrohm@americanbible.org  
David Sutherland International Care Ministries David.Sutherland@caremin.com

Azza Karam UNFPA karam@unfpa.org  
Scott Todd Compassion International stodd@us.ci.org

Alistair Sim Compassion International ASim@compassion.com.au  
Dean Karlan Innovations for Poverty Action dean.karlan@yale.edu

**Meeting Overview**

Many faith-based organizations are increasingly concerned that secular systems of measurement and evaluation do not fully reflect their organizations’ work or desired impact. Calls have been made for the development of “faith-based measurement and metrics”. This framework includes ‘spiritual’ outcomes valued by faith-based organizations but often not measured in mainstream impact evaluations, such as a person’s experience with trust, hope, and love.

A meeting was held in New York City at Salvation Army NYC Offices in-person and via conference dial-in to review current work related to “Faith-based Metrics and Measurement”.

Jean Duff, JLIF&LC, providing contextual framing for meeting by giving background on the [Faith-based Action Framework](http://jliflc.com/2015/09/joint-faith-based-action-framework-to-be-announced-on-september-24/), which was launched Sept 24th to provide a basis for joint faith-based collaboration in support of the SDGs. Three pillars of work around Evidence, Advocacy and Collaboration are being advanced. As part of the objectives and actions set out for the Evidence Working Group, a study of measurement and metrics for faith-based work has been requested. Jean also briefed participants on the Joint Learning Initiative on Faith and Local Communities, and the possibility, if there is interest as a next step, of setting up a JLI Learning Hub on this topic, or establishing an online platform for sharing of related information resources. She asked the group for discussion of what we mean by faith-based indicators, and how standards of measurement can be improved when it comes to distinctive aspects of faith based activities and contributions to development. In addition to those present, Jean mentioned interest expressed in the topic by others including Alastair Sim in NSW Australia.

Dean Pallant provided a conceptual framing for the discussion hoping for a landscaping of work in this area, towards strengthening of respective measurement activities, and an exploration of the value of a possible on-going conversation. In his view, the relational outcomes of, for example, hope, trust, and love distinguish faith-based interventions.

This report provides a brief summary of the discussion, as well as recommendations for possible next steps.

Brief Review of Current Work

|  |
| --- |
| **Dean Pallant, Salvation Army** The Salvation Army has long recognized the importance of measurement in improving the quality and effectiveness of their programs. However, the Salvation Army was unsatisfied with the existing measurement systems, which they believed did not capture the full outcomes of their work. One critique was that available measurement frameworks were not ‘person-centered’ and did not reflect the Salvation Army’s conception of a complete and fulfilled person.  The Salvation Army, working closely with Bridgespan Consulting Group, developed the Unified Measurement Framework to reflect their organization’s goals with measurement and learning with their conception of a person at the heart of the framework. The Unified Measurement Framework is to be primarily intended for use by program leaders to stimulate self-assessment and learning; however, it is also important that the evaluations help form a credible story to share outwardly.  The Unified Measurement Framework includes both contextual and relational indicators. Contextual indicators relate to a particular program and its context; relational indicators relate to the quality of relationships, such as trust and hope. The Salvation Army has drawn from the World Values Survey as a basis for its relational indicators. However, there has been disagreement within the Salvation Army on whether the World Value Survey’s indicators adequately measure the Salvation Army’s conception of relational indicators. In response, Pallant argues that though the measurement may be imperfect at the moment, it is better to measure than to not. He is interested in ways to improve SA evaluation. |
| **Mark Forshaw, GSI & David Bronkema, Eastern University**  Forshaw and Bronkema help evangelical Christian organizations assess impact by tracking “Spiritual Metrics”. Forshaw serves as the director of Global Scripture Impact, which is the research and development unit for American Bible Society. David Bronkema of Eastern University provides training on organizational effectiveness for Christian organizations through the Accord network. They work with organizations for which metrics might not be “front and center”—unlike World Vision, Samaritan’s Purse, or Compassion International, which frequently engage with the UN and other secular institutions.  Forshaw and Bronkema highlighted the importance of first clarifying an organization’s theory of change and purpose of the impact evaluation—to improve the program or to improve the effects?—as these have very different design processes. They use a mix of qualitative and quantitative methods, and have reviewed existing tools rather than attempt to create their own. Their goal is to provide local people with tools they can use to learn more about the impact of their work on local communities.  Forshaw and Bronkema are currently writing a book that will serve as a practical manual for organizations on the principles and challenges of working with spiritual metrics. They have also conducted workshops on measuring spiritual metrics for members of ACCORD. They agreed to share information from the workshops. |
| **Lincoln Lau, ICM**  Lau is a researcher at International Care Ministries, a faith-based organization focusing on the ultra-poor in the Philippines. ICM runs its program through local churches, in which pastors and local community members provide education, life skills, and business training. ICM collects a wide array of metrics (from debt reduction, health, sanitation, etc.) through its programs and conducts randomized controlled trials using that data.  In order to examine the role of the faith influence versus secular on their programs, ICM is conducting an RCT with four groups: 1) delivery of the program in a faith-based setting; 2) delivery of the program in a secular setting; 3) delivery of no program; and 4) delivery of only religious messages. ICM is in the second round of data collection.  Lau acknowledged that although RCTs are not easy to conduct, they do provide a cleaner study design and can manage confounders when measuring impact of the program. |
| **Sneha Stephen & Isabel Onate, Innovations for Poverty Action**  Researchers at the Innovations for Poverty Action, under the direction of Yale economist Dean Karlan, are also analyzing the data collected by ICM. IPA is examining broader well-being indicators, such as aspirations, satisfaction, and spirituality. With subjective indicators, Isabel emphasized the need to be as specific as possible in one’s questioning. IPA also includes questions that related to the well-being indicators, but are not meant to be directly asked. For example, as a proxy for hope, they may ask “Where do you see yourself in the future?” Onate and Lau have agreed to share operational definitions. IPA is now collecting 6 month follow up data, and will do a further follow up on the 4 groups at 12 months. They agreed to share information about measurement tools with the working group members. |

Jill Olivier, University of Cape Town, and Bruce Wydick, University of San Francisco offered comments.

|  |
| --- |
| **Jill Olivier, University of Cape Town**  Olivier noted that the purpose of developing the metrics should be clarified, as the purpose will determine what kind of metrics will be developed. She pointed to two different purposes in the meeting’s conversation: 1) Developing metrics to prove to international agencies that FBOs make a difference and provide an added value; 2) Developing metrics not to make the case for FBOs but to measure internal processes of hope and trust that standard measurements do not assess, and to evaluate the success in achieving those goals. |
| **Bruce Wydick, USF**  Wydick echoed Olivier’s comments that purpose should be clarified and also highlighted the importance of study design in addition to metrics. In order to prove impact, the study design needs to identify the ‘counterfactual’—the situation that would have happened had the program not been implemented. Wydick stressed the importance of measuring correctly in order to yield reliable data, as unreliable data could lead to incorrect implications about policy and program design.  Wydick said that when measuring indicators such as trust or hope, asking people what they believe results in data with very little variation. It is much more reliable to look at behaviors—asking for example, “Have you given to the poor?” Wydick explained the concept of a ‘pyramid of psychological data’, representing a hierarchy of data quality. At the base of the pyramid is the least reliable data, which includes data on hypotheticals (questions asking a person what he thought he would do in a situation). The next highest level included data taken through direct questioning about reported characteristics, such as “Are you a hopeful person?” The next level was reported actual behaviors, and the highest level was observed behaviors by researchers in non-experimental and experimental settings. |
| **Lily Talapessy, UNFPA**  As a Senior Advisor at UNFPA, Talapessy is involved in the development of Global Indicators for the Sustainable Development Goals. Talapessy provided an overview of the process of development for Global Indicators. She encouraged the working group to continue to develop these “perception indicators”, as they were in large part not addressed in the development of the Global Indicators.  As part of the development of the Global Indicators, a ‘political forum’ will be created to discuss the lessons learned and the hurdles in implementing the SDGs. Talapessy suggested the outputs from the working group could feed into this ‘political forum’, which could serve as a platform to highlight what is not being measured in the current set of indicators. |

Meeting Note Reflections

Different perspectives on faith-based indicators

Meeting participants in their academic and professional careers have approached and engaged with faith-based metrics from different perspectives. A review of the meeting suggested a few differences among participants that could be clarified to help move the working group forward:

1. **Goal of evaluation and type of approach**

Differences in understanding shared terms around evaluation could be clarified, particularly when understanding each other’s purpose in engaging with faith-based metrics.

Impact Evaluation

The term “impact evaluation” was referred to often throughout the meeting’s conversation, but can be understood differently depending on the context. Among M&E professionals, economists, and academics, “Impact Evaluation” is atechnical term that is part of outcome evaluation.[[1]](#footnote-1) It is defined in this context as below:

*An impact evaluation assesses changes in the well-being of individuals, households, communities or firms that can be* ***attributed*** *to a particular project, program, or policy. The central impact evaluation question is what would have happened to those receiving the intervention if they had not in fact received the program. Since we cannot observe this group both with and without the intervention,* ***the key challenge is to develop a counterfactual*** *– that is, a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention.* ***This comparison allows for the establishment of definitive causality*** *– attributing observed changes in welfare to the program, while removing confounding factors.[[2]](#footnote-2)*

When used in this context, an impact evaluation is a measurement tool that quantifies the changes the program made upon a community through proof of causality—that the observed changes in the community are caused by and can be attributed to the program, as opposed to other external factors. The gold standard of impact evaluation are Randomized Controlled Trials (RCTs), but other study designs exist that can show (to a lesser degree of scientific rigor) the amount of change attributed to the program.

However, a broader community, including program managers, may be interested in assessing how effective its programs are at reaching their intending goals. They may term this as evaluating their programs’ impact—while not referring specifically to impact evaluation, as understood in the definition above. **“Impact evaluation”**—as understood by economists and M&E professionals—may be useful to evaluate some of its programs and gain an overall understanding of their impact; however, the use of **“impact evaluation”** to provide assessments of *all* of programs of a large institution is likely unrealistic. Impact evaluations require substantial time, resources, training, and information to implement and analyze; more realistically, impact evaluations could be a tool used in combination with other evaluation strategies to assess all an institution’s programs.

One question program managers and economists/academics might want to discuss from their respective perspectives: Can useful information about impact on outcomes be gained from an evaluation that is not specifically conducted as an “impact evaluation”?

1. **Focus on relational or spiritual metrics?**

Is the interest to measure more the outcomes of faith based interventions, or faith and faith-related indicators themselves? Pallant said that the Salvation Army was interested in an evaluation approach that looks at the “fruits of faith” in terms of relational outcomes, seeing this as an approach that also is useful for communication to the secular world. Bronkema and Forshaw work more on measurements of spirituality of faith itself, noting that spiritual metrics are front and center for Evangelical Christian organizations where growth in faith is a key outcome of the combination of mission and ministry programming (“integral mission”). Bronkema noted that secular studies of causality are pointing to faith as a key variable. We have a common interest in showing that “faith causally effects movement into virtues” / relational characteristics; from his perspective, the mechanism of causal transmission is spiritual change resulting from an encounter with Christ or Christian people.

1. **Is there a common purpose to ground a Working Group?**

Whether to understand organizations’ individual interests or tease out the joint working group’s purpose, the questions below could be helpful to clarify the primary purpose of evaluations, which in term can help determine the kind of metric and study design to develop and use:   
  
- Who are the audience and primary users of the evaluation?

- What level of rigor do they require? What level of accessibility?

Tensions exist between the rigor and accessibility of evaluation results. When determining the metrics and study design to use in an evaluation, understanding the audience and user of the evaluation can help achieve a balance between rigor and accessibility. Some audiences need a certain level of proof through rigorous data to make decisions; other audiences are constrained by the accessibility and usability of the data collected from the evaluations.

For example, policy makers and some funders often require a minimum level of rigor and proof to make decisions from an evaluation. Wydick described during the meeting and in a World Bank blogpost entitled “Measuring Hope”[[3]](#footnote-3) the concept of a “pyramid of psychosocial data”. Observed behaviors, and the proxies that are used to measure these observed behaviors, are the most reliable data and at the top of the pyramid. Proxies can serve as creative ways to measure seemingly “unmeasurable” indicators. In his blogpost, Wydick describes the example of measuring hope in children. Trained researchers were able to deduce levels of hope in children by interpreting pictures drawn by the children. The children were asked to draw themselves in the rain, and those with higher aspiration levels drew umbrellas or positive images of themselves; those with lower levels of aspiration drew themselves uncovered or unhappy. This data combined with psychological backing can be considered high quality on the pyramid of psychosocial data, as observed behavior.

This kind of data can be used to make a compelling case of the impact of a program on children’s levels of hope. Furthermore, this kind of study would allow for the possibility for impact of a program to be measured while collecting data from a relatively small sample of individuals.

While this data could have important implications for which programs should be prioritized and funded, this kind of data may not be appropriate for a program manager, or a pastor seeking to understand change in his congregation. The program manager or pastor may need the results from the evaluation to be more readily accessible and interpreted. A program manager might need information more quickly, or not have the resources to conduct this kind of evaluation. A pastor might seek data he could interpret himself and use as a benchmark with individuals who he counsels regularly.   
  
- Is measuring the outcome or “virtue” itself important?

The necessity of measuring the virtue itself or measuring via proxy was brought up during the meeting. This question is linked again to the clarifying the purpose of the evaluation. To use hope as an example, one could ask: Is the purpose to show that the program delivers hope to individuals? If so, the use of a proxy may be appropriate. Or, is the purpose to see if an individual or community’s levels of hope are changing over time? The organization may consider capacity for hope to be a central in assessing a person’s wellbeing, and wish to base their organization’s learning off of those kinds of metrics. In that case, it may be worth creating a ‘hope metric’ that can assess this change.

**Standardized measurement tools mentioned in the meeting:**

* World Values Survey
* Living Health Measurement Survey: World Bank
* Baumeister scale measures self0control, focus on long term goals

**Possible benefits of further work**

* Field based faith based practitioners have tools to assess impact and own the results
* Possibility of comparative measures with standardized tools
* Basis for exchange and partnership between faith and secular worlds

Possible next steps:

1. Share Theory of Change paper developed collaboratively by global faith based networks as a framing for their work to strengthen the capacities of local faith communities in meeting the needs of their people
2. Participants in the meeting to exchange further information as agreed at the meeting
3. Develop agenda for further conversation to explore working group
4. Establish common platform on JLI F&LC site to support sharing of information and ongoing work
5. Identify additional researchers and practitioners working in this area to exchange information
6. Consider collaborative literature review/scoping study

7. Explore Creation of JLIF&LC Learning Hub on Faith-based Metrics & Measurement

* Involve more interested partners into this topic
* Draft a scoping paper on current work and gaps

1. Presentation of progress reports:

* “Partners for Change” – Meeting in Berlin, Feb 17/18
* Donor-UN-FBO meetings
* Moral Imperative ongoing process, and Annual Meeting in September 2016
* Feed into SDG indicator Political Forum process

1. When outcome evaluation and impact evaluation are seen as separate types of evaluation, impact evaluation is seen as an evaluation of the longer-term effects of the program, compared to outcome evaluation, which evaluates short and medium-term effects. [↑](#footnote-ref-1)
2. <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,menuPK:384339~pagePK:162100~piPK:159310~theSitePK:384329,00.html> [↑](#footnote-ref-2)
3. http://blogs.worldbank.org/impactevaluations/measuring-hope-guest-post-by-bruce-wydick [↑](#footnote-ref-3)