Identifying what works: a meta-evaluation of modern slavery evaluations

Conference paper prepared by Katharine Bryant, Research Manager, Walk Free Foundation, for Freedom from Slavery Conference, October 9th, 2015

Introduction
The purpose of this paper is to provide an overview, and update on work in progress on the “Promising Practices Toolkit”. The Toolkit is a database of 165 evaluations from the modern slavery sector. Within this database, each evaluation is categorised based on the type of modern slavery, sector, type of activity, and programme results. From this categorisation, we can begin to conduct a meta-evaluation, drawing conclusions about which practices have been proven to work, which look promising, and which are ineffective.

This paper sets out in summary the theory and purpose of the database; and overview of the process for developing the database; and begins to identify some early learnings and questions for discussion.


As noted in many different fora, one of the critical challenges of responding to modern slavery is knowing “what works”. For example, does providing training to local organisations about the legal framework on human trafficking result in the realisation of rights for that community? What works to reduce vulnerability of certain populations, like irregular migrant workers? Are victim support programmes meeting the needs of their client groups? While all these and more are critical questions, we have very few agreed answers.

While various studies have noted the lack of evaluations undertaken, it is also true that many programmes in the field have evaluation requirements attached to them as part of funding agreements. Starting from this observation, this study seeks to compile as many published and unpublished evaluations from the field as possible, classify and categorise them in ways that facilitate analysis, and then undertake a meta-analysis of learnings that can be observed from these evaluations. The theory is that we can learn from the evaluations already undertaken, even if the learning is “there is a lot we don’t know”.

2. Taxonomy and database

To facilitate the review, we created a classification taxonomy, reflecting what we considered to be key categories of analysis of the evaluation literature. These focus on the following, which are set out in more detail at Appendix 1:

1. Type of modern slavery
2. Sector of exploitation
3. Target population
4. Country/region

---

1 This number will fluctuate with the final removal and clean of evaluations over the coming weeks
5. Type of activities
6. Programme objectives
7. Evaluation author
8. Evaluation objectives
9. Evaluation methodology
10. Evaluation findings
11. Met objectives
12. Process, outcome or impact?
13. Impact on modern slavery
14. Promising?

We created a database within Endnote, to store and house the research results.

3. Searches
We then developed a standard set of search terms (Appendix 2) and did systematic searches to find evaluations. The team conducted internet searches, relevant international organisation websites and searches of academic databases to find examples of these evaluations. The simplified search terms are set out in Appendix 2.

4. Inclusion in the database
After identifying potentially relevant evaluations, members of the team reviewed the titles and/or abstract to determine if the evaluation should be included in the library. Using a standardised inclusion criteria (Appendix 3), evaluations were included if the title or abstract identified that the report, website or document was an evaluation of a programme or activity, that it had an explicit methodology of how the evaluation was conducted, and it referenced some form of modern slavery or related social issue (see Appendix 2 for list of related social issues).

5. Preliminary results: what is in the database?
The database following cleaning, contains 165 evaluations. Broken down by sector, we can see that while most evaluations do not specify a sector, sex work, domestic work and agriculture are the sectors that have been evaluated most frequently.

![Number of evaluations per sector](image)
By country and region, we can see that the majority of evaluations have been conducted in Sub Saharan Africa and the Asia Pacific region. At the country level, most evaluations have been conducted in the Philippines and India.
When looking at the evaluations by activity, we can see that raising awareness, service delivery and training have been evaluated most often.

Using the Maryland scale, the overwhelming majority of evaluations are conducted by a post assessment, often involving desk review of relevant literature, selected interviews with relevant participants, and a write up of results. Randomised Control Trials have been conducted, but for similar social issues such as Female Genital Mutilation/Cutting (FGM/C) and the impact of conditional and non-conditional cash transfers on child labour and education.
The majority of these evaluations are conducted independently. However, an independent body was defined broadly as any organisation or individual who was not involved in the implementation of the programme. It did not look into funding sources for this independent body.

Finally, although the searches were broadened to include other forms of social practice beyond modern slavery, the majority of evaluations were end of project reviews of trafficking programmes.
6. **Sample of key findings: what can these evaluations tell us about monitoring and evaluation in the anti-slavery sector?**

Most simply, we can learn that evaluations of anti-slavery and counter trafficking programmes have been conducted over the past 15-20 years, and thus we should be able to point to some lessons learnt in particular sectors, regions, or for specific types of slavery. We should be able to pull together some interesting points in terms of what does and doesn’t work in the garment sector, or forced marriage programming, for example. Although it could reveal limited access to evaluations, the database also points to areas where there are gaps in our understanding - an initial analysis would suggest that more work needs to be conducted assessing programming on debt bondage, or programmes looking at domestic servitude.

Traffic off programme deserves a special mention. What we can begin to deduce from the database is that nearly 50% of trafficking evaluations have no clear description of the sector where trafficking occurs, or a detailed description of the type of trafficking the programme was aiming to address. While this could be an indication of the limitations of monitoring and evaluation documents, such a significant proportion could suggest that trafficking programmes tend to focus on implementing a response that meets treaty requirements, rather than implementing a tailored response to a specific issue.

Moreover, the initial analysis reveals that there are issues in how we measure outcomes in anti-slavery and counter trafficking programmes. In several cases, evaluations identified that there were issues with the initial project logic, which linked activities to outcomes. Several evaluations identified that it was difficult to draw conclusions about if and how these outcomes had been achieved either because the outcomes were too ambitious or not clearly linked by any theory of change to activities implemented. This would suggest the need for more explicit descriptions of programme logic in order to facilitate later evaluation, or to build evaluation into initial programme design.

Finally, few evaluations attempt to measure the impact of the programme on the modern slavery itself, be it by decreasing the number of people enslaved, or changing behaviours that allow slavery to occur. The majority of evaluations evaluated the progress (achievement of activities or outputs) or the outcomes of the programme (achievement of project objectives or outcomes). While this
reveals the difficulties in designing an anti-slavery programmes that show direct impact, it also points to a gap in the field, and the need for more rigorous evaluations in order to test the effect of anti-slavery and counter trafficking programming.

7. Sample of key findings: content analysis of the database

Type of intervention: criminal justice responses
Traditionally counter trafficking programming has applied a criminal justice response to modern slavery, in line with treaty requirements as outlined in the UN Trafficking Protocol 2000. A criminal justice response is defined by three core elements:

- Reducing the vulnerability of victims and potential victims
- Reducing the motivation of offenders
- Increasing the oversight or capacity of guardians (police, government, community, etc.)

Despite the number of criminal justice programmes, and evaluations in the database (38 of 165), it is still difficult to draw concrete conclusions linking programmes which enforce the law to a reduction in the number of people enslaved. One project, IJM’s project lantern in Philippines, (Jones, Schlangen, & Bucoy, 2010) did provide anecdotal and some quantitative data that sex trafficking victims were ‘less available’ as a result of the project. Evaluators posed as customers and asked for a child, and identified that there was a reduction in the number of children available. There are suggestions, however, this reveals that child victims were more hidden as a result of the programme, rather than an overall reduction in numbers.

Evidence does exist, however, that these projects are effective in building the capacity of police and other law enforcement officers. A study of specialised training for police agencies in the United States (Broderick) found that statistically significant correlations between specialised training in human trafficking and the apprehension of human traffickers among police officers. Qualitatively, police perceived that they were more aware and competent to handle human trafficking situations as a result of the training they received. It points to promising results if training requirements for police are strengthened. Similar results were found in a counter trafficking project in Indonesia, where specialised police training led to increase in the number of trafficking cases investigated, although limited baseline data makes it difficult to trace this increase directly back to the intervention of the project (Budiharga & Arna, 2007).

Moreover, there is an example in the Philippines where, despite improvement in the capacity of police and increased knowledge of trafficking issues, there is no evidence that prevention, investigation or prosecution practices have improved (Williams, 2003). What is lacking is an understanding of why this occurred, and the difference between programmes that seem to follow a similar theory of change, but have not led to an increase in identification of victims and investigations.

One theory (see Mackie, forthcoming) is that enforcing the law will only work in contexts where there is already a common respect for the law. Programmes or projects operating in environments where respect for the law is limited due to high levels of corruption, or poor policing, should therefore take action to support the development of this ‘legal norm’, or seek alternative means to build respect for justice. Models which build capacity of communities to develop their own policing mechanisms, without the formal justice sector, could be one such approach.
Sector of slavery: Garments
Projects addressing worst forms of child labour and workers’ rights in the garment industry have been included under this section. Although these evaluations are not dealing directly with modern slavery, interesting lessons emerged during the review regarding outreach, capacity building, and monitoring.

An evaluation of a programme implemented by Global March Against Child Labour to contribute to the elimination of child labour and trafficking in the garment sector in India found that community outreach, raising awareness campaigns, and capacity building exercises did increase the understanding and exercise of worker’s rights in the workplace (Bhatt, 2013). Ninety-eight percent of factories that participated in the programme, for example, now have a child labour policy, while children sub-contracted to work in the home with their parents were able to enrol in school. Union membership increased, as did union support for workers.

However, the evaluation points to the difficulty of drawing concrete conclusions about a programme’s impact. Although all children identified in child labour in a baseline study were enrolled in school by the end of the programme, it is unclear whether these children attended school. Of 100 women identified in the baseline survey as engaged in home based garment manufacturing work on a piece meal basis, 25 took up the opportunity to attend training on sewing and embroidery. While those who attended claimed they could now earn more money as a result of the training, it is unclear how they will engage the remaining 75 women (Bhatt, 2013).

A 2013 evaluation of ILO’s programme ‘Better Factories Cambodia’ identified good practice in monitoring and reporting labour practices in the garment sector in Cambodia (Consortium, 2013). The independent evaluation points to the importance of transparency in monitoring visits and providing avenues for remediation when violations are identified. Evidence in support of these findings shows that despite the programme aiming to improve working conditions in garment factories, there is evidence that workers have suffered a drop in real wages, and are employed on temporary employment contracts, which may leave them vulnerable to abuse. Underpinning the causal link between the programme and declining working conditions is the assumption that since the programme’s monitoring reports are not always made public or available to workers, the abuse of worker’s right is able to continue unchecked (Consortium, 2013).

Type of activity: Raising awareness
Evaluations of raising awareness campaigns and activities have shown mixed results in terms of their impact when promoting identification mechanisms and safe migration. Various raising awareness techniques were used in a campaign in Israel, for example, to promote the use of a trafficking reporting hotline. These include distributing information, providing training, publicizing reports, using media advocacy, and giving lectures. There was, however, no marked difference in levels of knowledge of the hotline between pre- and post- test assessments, but also no negative impacts. It is unclear which elements of the campaign hindered improvements in understanding (Hashash, 2007).

The involvement of local communities in designing culturally appropriate information, education and communication (IEC) materials is shown to increase awareness of the concepts of safe and blind migration, and trafficking. After a road show on trafficking, for example, around 90 percent of participants could remember one key message, including definitions of trafficking. However, while general conceptual awareness has been established, it has not necessarily translated into the changed behaviours of participants, such as increased reporting of victims, or adoption of less risky behaviours (Sainsbury, 2006).
One component of an IOM project to improve protection of migrants in the Horn of Africa and Gulf of Aden regions included outreach to migrant workers to raise awareness of safe migration practices. An independent consultant assessed the outputs of the programme, and although no formal impact evaluation was conducted, multiple respondents to the independent consultant indicated that it was unclear what effect these outreach measures would have on migration patterns (Stewart, 2013). One of the recommendations of the evaluation was a broader assessment of the effectiveness of outreach and raising awareness activities.

**Type of slavery: Early marriage**

The success of community capacity building to enact social change is shown by the Tostan programme in Senegal. The Tostan programme was established to reduce occurrence of Harmful Traditional Practices (HTPs). It provided community training on adverse health effects of HTPs, democracy, human rights, and child development in order to mobilise these groups to publicly declare that they would abandon these practices.

The evaluation compared villages who had participated directly in the Tostan programme, those that had participated indirectly, and a control group of villages that had no experience of Tostan. It found, as a result of the programme, instances of female circumcision declined. The impact on early marriage was less clear cut; early marriage declined across all villages during the time period, including in the control villages. The decline of early marriage should be viewed in the context of increased education, media sensitization, greater freedom to contact other cultures, and fear of legal sanction. (Long-Term Evaluation Of The Tostan Programme In Senegal: Kolda, Thiès And Fatick Regions 2008). As the programme evolved from the late 1990s, principles of community ownership and mobilisation became increasingly important to ensure that the capacity building was both community-led and sustainable.

Education (both formal and non-formal), community consultations, use of mentors, and economic incentives have had an impact on early marriage in one kebele (a small administrative unit) in Ethiopia (Karei & Erulkar, 2010). Using pre- and post- testing with a control group and statistical controls, an assessment of the Berhane Hewan programme found that girls aged 10 to 14 who participated in the programme were 90 percent less likely to marry than the control group (Karei & Erulkar, 2010).

The programme was then scaled to 12 kebeles. However, without an analysis of the specific components that had the most impact on the age of marriage, and the removal of individual economic incentives, the assessment of the larger programme’s impact is limited. There also does not appear to have been an evaluation of the scaled up programme (Karei & Erulkar, 2010).

A 2012 study into the different strategies used in public declarations to respond to HTPs practices in Ethiopia found similar results (Lulit & Hailu, 2012). Successful strategies included:

- Utilising existing community structures to raise awareness of child marriage as a Harmful Traditional Practice (HTP),
- Sending girls to school, and
- Enforcing the law.

---

2 For the 0 to 9 age group in the statistical analysis, 15 percent of girls underwent circumcision in Tostan villages versus 47 percent in the control group. (Long-Term Evaluation Of The Tostan Programme In Senegal: Kolda, Thiès And Fatick Regions 2008)
Effective strategies include community ownership, the incorporation of religious teachings into trainings, and teaching the negative health effects of HTPs (Lulit & Hailu, 2012).

Other learnings to be explored:
The promising practices toolkit includes other sectors, types of programmes and activities that warrant further exploration. These include an assessment of programmes dealing with commercial sexual exploitation, criminal justice responses, technical support programmes for governments, and training programmes for police. These, and a more detailed analysis of what we know about monitoring and evaluation will continue to be added to the toolkit and this working paper over the coming months.

Bibliography


Stewart, R. (2013). Independent Evaluation of IOM’s Project “Horn/Gulf of Aden/Yemen: Improving Protection of Migrants – Phase III. Retrieved from If you are interested in one of the following evaluation reports, please contact OIG/Evaluation at Headquarters (eva@iom.int).
APPENDICES

Appendix 1: Definitions of term lists

WFF 1: Type of modern slavery

Type of slavery was added after the systematic searches had been completed. It was hoped to be able to tag evaluations solely by sector (WFF2), however, this led to the tagging of many evaluations as ‘other’ as the programmes they referred to aimed to combat or prevent ‘trafficking’ or ‘forced labour’ without specifying a sector. Type of modern slavery was therefore added to cover these programmes. The term lists come from Walk Free’s definition of modern slavery (see Appendix I), although evaluations were categorised as defined by the type of modern slavery identified in the evaluation. Wherever possible, evaluations were tagged with more than one type of modern slavery to ensure that all interpretations of these terms would be covered. For example, programmes which aimed to tackle trafficking for forced labour, were tagged ‘trafficking’ and ‘forced labour’.

1. Human Trafficking
2. Forced labour
3. Forced, servile or early marriage
4. Slavery
5. Worst forms of child labour
6. Domestic servitude
7. Debt bondage
8. Sale or exploitation of children
9. Use of child soldiers
10. Other- child labour
11. Other- migration
12. Other- sex work/ prostitution
13. Other- IDPs/ refugees
14. Other- FGM/C

WFF 2: Sector

Where possible, evaluations were tagged by the sector the programme or intervention targeted. This sector had to be made explicit- where the programme’s target populations were defined as belonging to a specific sector- rather than a more generic approach supporting governments to combat trafficking or forced labour. It could also include general descriptions of the problem the Where no sector was explicit, the evaluation was tagged as ‘none specified’. 77 evaluations were tagged thus.

- Agriculture
- Armed conflict/ military
- Begging
- Brick kilns
- Camel jockeys
- Carpets/ rugs
- Construction
- Domestic work

3 Other covers evaluations which related to a similar social problem to child labour, prostitution, refugees, internally displaced persons, female genital mutilation, safe migration and labour migration.
- Drugs
- Electronics
- Fishing
- Health
- Marriage
- Mining
- None specified
- Organ removal
- Quarries
- Religious sites
- Restaurants/ catering
- Rubbish/ garbage collection/ sorting
- Sex work
- Street vending
- Textiles/ garments
- Tourism

**WFF 3: Target population**

Defined as the direct beneficiaries of the programme- for example those who received training, or who were directly assisted by the programme. This category, therefore, does not always line up with victims of modern slavery. Those where the beneficiary was not defined were tagged ‘not specified’.

To improve searchability of the database, those evaluations where the programme targeted ‘children’, but did not specify a gender, were tagged ‘girls; boys’ and a special group was created that contained both these terms. The same rule applied for adults, with the use of ‘women’ and ‘men’.

- Men
- Women
- Girls
- Boys
- None specified

**WFF 4: Country/ region**

Defined as the country and region (where available) where the programme occurred.

**WFF 5: Type of activities**

Defined as the main activities taken under the programme or project. Examples include:

- Advocacy
- Audits
- Capacity building
- Certification
- Community mobilisation
- Direct pay
- Education
- Financial support
- Geospatial mapping
- Hotline
• Inspections
• Job placement
• Legislation
• Micro credit/ financing
• Other (please specify in summary)
• Raising awareness campaigns
• Regulation of recruitment agencies
• Research
• Service deliver for victims
• Technical support
• Training

WFF 6: Programme/ project objectives

Free text write up of the programme/ project objectives.

WFF 7: Evaluation author

Defined as whether the evaluation was conducted independently or internally.

WFF 8: Evaluation objectives

Free text write up of the evaluation objectives

WFF 9: Evaluation methodology

Defined using the Maryland Scale of Scientific methods.\(^4\)

WFF 10: Evaluation findings

Free text write up of the results

WFF 11: Met objectives?

Did the programme meet its objectives as outlined in the evaluation? Categorised as yes (all), yes (some), no, and inconclusive.

WFF 12: Process, outcome or impact?

Did the programme meet its objectives as outlined in the evaluation? Categorised as yes (all), yes (some), no, and inconclusive. Details of which objectives were or were not met can be found in researcher’s notes

WFF 13: Impact on modern slavery

If the evaluation does measure impact, what impact did the programme have? This is defined as decrease in number of people enslaved or change in attitudes or behaviours that facilitate modern slavery to occur

WFF 15: Promising?

If the evaluation claims outcome or impact success, and is able to demonstrate some evidence for these findings, the programme is tagged as ‘promising’. These promising practices are those which

\(^4\) This is currently being expanded to include other types beyond experimental and quasi experimental methodologies
then are subject to a more detailed content analysis; an example write up of these findings can be found in Section 7: Sample of Key findings: Content analysis of the database.
Appendix 2: Search terms
Evaluation (if necessary) OR Assessment (if necessary)

AND

- Modern slavery OR
- Human trafficking OR
- Worst forms of child labour OR
- Slavery OR
- Forced labour OR
- Domestic servitude OR
- Debt bondage OR
- Forced marriage OR
- Servile marriage OR
- Early marriage OR
- Child marriage OR
- Sale or exploitation of children OR
- Use of child soldiers OR
- Child labour OR
- Prostitution OR
- Refugees OR
- Internally displaced persons OR
- Female genital mutilation OR
- Safe migration OR
- Labour migration

A total of 1,787,748 sources were identified from these searches, largely due to the number of sources identified through internet searches. For internet searches that revealed a large number of evaluations, the team reviewed the first 10 pages for relevant evaluations. From all of our searches, a total of 410 evaluations were included in Endnote. These were then reviewed for duplications, leaving 344 evaluations in the library.

After further review, another 179 were then removed because they were descriptions of programmes, were lists of good practice determined by other organisations, were mid-term evaluations, were not in English, or did not include an explicit methodology of how the evaluation was conducted. Lists of good practice determined by other organisations were removed as these had inconsistent or contradictory criteria from the work being conducted of what constituted a good practice. Explicit methodologies were defined as the inclusion of a methodology section or a description of the actions taken by methodology team; those evaluations missing this were also excluded.
Appendix 3: Inclusion criteria

1) Is the report or website an evaluation of a programme or activity?

An evaluation measures progress towards outputs, or change in outcomes, or an assessment of impact, of a development programme, policy or intervention. An international development programme or project is a set of activities which deliver outputs and outcomes that promote social change and behaviour change related to modern slavery or related social issues.

An activity is a singular action taken to address modern slavery, which could be part of a programme, project or a standalone activity. Examples include giving training, conducting inspections, distributing leaflets, providing technical support to the government.

YES- Go to Q2

NO- Exclude from Endnote

2) Does the evaluation reference a form of modern slavery?

Modern slavery defined as:

- Human trafficking;
- Worst forms of child labour;
- Slavery;
- Forced labour;
- Domestic servitude;
- Debt bondage;
- Forced, servile or early marriage;
- Sale or exploitation of children; and
- Use of child soldiers.

YES- Include in Endnote

NO- Go to Q3

3) Does the evaluation reference a related social issue?

Similar social issue includes: child labour, prostitution, refugees, internally displaced persons, female genital mutilation, safe migration and labour migration. These are included as individuals who experience these issues may become vulnerable to being enslaved, thus successful programmes may become relevant as prevention activities.

YES- Include in Endnote

NO- Exclude from Endnote

---

5 Adjusted from 3ie definition of impact evaluation http://www.3ieimpact.org/en/evidence/impact-evaluations/