



Machine learning and meta-ethnography:

Seven steps to synthesising 578 evaluations into four themes

NOVEMBER 2021

Stephen Porter, Harsh Anuj, Yingjia Liu, Kristin Strohecker

The Independent Evaluation Group, The World Bank Group

About

This paper is part of the BetterEvaluation Innovation Working Paper series. The series focuses on using innovative methods, processes, and approaches to overcome challenges in monitoring and evaluation.

This paper can be accessed at www.betterevaluation.org/resources/machine-learning-and-meta-ethnography

This paper was written by Stephen Porter, Harsh Anuj, Yingjia Liu, Kristin Strohecker (The Independent Evaluation Group, The World Bank Group).

This paper benefited from comments from Alice Macfarlan (BetterEvaluation).

Recommended citation:

Porter, S., Anuj, H., Liu, Y., Strohecker, K. (2021). 'Machine learning and meta-ethnography: Seven steps to synthesising 578 evaluations into four themes'. BetterEvaluation Innovation Working Paper series.

Introduction

What would you do if you were asked to rapidly identify practical insights on a development issue based on a potentially vast number of evaluation documents? An Independent Evaluation Group (IEG) team faced with reviewing 578 project evaluations combined evaluation and data science techniques to rapidly analyse and report on four themes related to improving interventions in contexts of migration. A worked example of developing a briefing on [the major factors improving project performance in the context of migration](#)

Evaluation often requires some form of synthesis based on documents. Doing this in the World Bank is a challenge, with around 450,000 electronic documents to draw upon. An evaluation of an individual project can require a review of 15 of these documents and be accomplished by a single person. Meanwhile, a focused review of the World Bank's COVID-19 response entails around 1,000 documents, while a more extensive review of lessons from projects in the context of crisis involves potentially over 10,000 documents. Even with a structured process, the multipliers mean that timely review is practically out-of-reach without the assistance of technology.

To solve this challenge, we experimented with combining data science and evaluation techniques, such as text mining, machine learning, and qualitative meta-ethnographic synthesis.

Text-mining

Text mining works to identify relevant sections of a document by automatically extracting parts of a manuscript, helping to discover new and previously unknown information ([Hearst, 2003](#)).

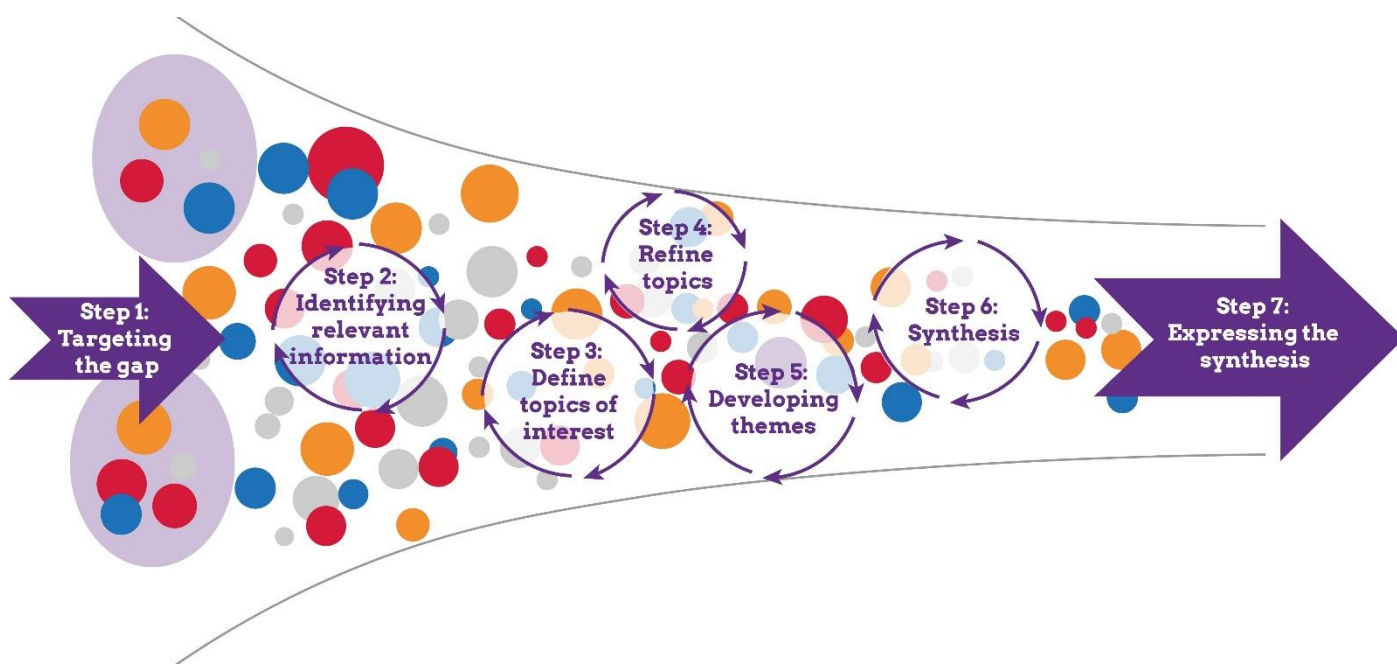
Machine learning

Machine learning simulates the characteristics of the human brain to structure pieces of text and learn from data, providing a process to rapidly map patterns across many documents ([Géron, 2019](#); [Rahman, 2020](#)).

Meta-ethnography

Meta-ethnography is a technique that has been developed in education and healthcare to produce knowledge through synthesising text from reports in a manner that maintains the structural aspects and inherent conflicts of the context ([Atkins et al, 2008](#); [France et al 2019](#); [Nobilt and Hare, 1988](#)).

Figure 1: Seven-step synthesis process



Box 1: Key terms used to search for migration-relevant evaluations

Asylum, border control, border crossing, border security, credential recognition, cross-border displacement, crossing border, diaspora bond, displace, emigrant, emigration, émigré, expatriate, forced return, foreign guest worker, foreign worker, guest worker, host communit*, human mobility, IDP, IDPs, illegal aliens, immigrant, immigration, international students, migrant, migration, multiple citizenship, non-resident deposit, overseas foreign worker, overseas guest worker, passport, refugee, remittance, residency permit, transhumance, visa.

Providing lessons on improving project performance in the context of migration

In late September 2020, the management of the Independent Evaluation Group (IEG) of the World Bank was asked to contribute to a technical briefing to the Executive Directors. Specifically, they were asked: **'Are there specific evaluations by IEG on projects related to migration?'**

One way to answer this question would be to poll IEG evaluators and provide a list of evaluations to the Board. However, this approach could easily have missed relevant pieces of work. More so, while the information gathered would have answered the question, it was not particularly useful information by itself.

Seven steps in a synthesis

To help undertake a valid synthesis, we informally adapted the meta-ethnographic process's seven steps (Figure 1) ([Atkins et al., 2008](#); [France et al., 2019](#); [Nobilt and Hare, 1988](#)). In each of the steps we sought to identify whether to combine data science and evaluator techniques. Below, we go into more detail about our approach in each step.

Step 1 – Targeting the information gap

The first step in this process involves defining the key question (or 'information gap') to be explored.

While our initial interpretation was to provide a list of evaluations that referenced migration, it became clear that it would be more helpful to provide insights on performance and factors that can help manage projects that operate in the context of migration, thereby incorporating other concerns of the Board.

To assess these issues requires opening the box of implementation to understand the mechanics of migration that affect performance positively or negatively in the contexts of sending, receiving and transit countries.

Considering the information gap in full, the overall question of the synthesis became:

What are the major factors improving World Bank project performance in the contexts of migration according to IEG evaluations?

Step 2 – Identifying relevant information

Based on this question, inclusion criteria consisting of keywords, document type and date ranges were developed. Getting these criteria to work well enough can be a lengthy and iterative process. However, this process is essential to identify relevant information from the mass of documents contained in the World Bank.

We selected an existing taxonomy of 37 keywords (box 1 below) developed by the [Global Knowledge Partnership on Migration and Development](#) or KNOMAD, which is implemented through a trust fund at the World Bank. Using this existing taxonomy sped up the process, as we could move directly to text extraction. Without an existing taxonomy, this step might include a literature review or expert consultation to define keywords.

Text mining is highly sensitive to the selection of keywords and the pre-processing of text. Failing to identify an important term can mean excluding very relevant information; having too many search terms means working through many unrelated text extracts. Getting the keywords working well involves some trial and error; for example, the keywords used for migration surfaced several examples related to migratory birds and other animals. We used our own code for the text

mining process of finding keywords and extracting sentences. Based on the types of search results we were getting, we refined our approach to include Boolean search terms¹ to help reduce selection errors. Lengthy processes of transitioning text into a form that our code could read (e.g. xml or csv) were not necessary as the reports we were using had already been prepared through previous efforts.

The selection of the inclusion criteria was also through an iterative process. We settled on including only IEG project evaluations after finding that including IEG reviews of World Bank project completion reports returned too much irrelevant text and jargon to resolve in a reasonable amount of time. Conversely, with the timeframe for inclusion, limiting the criteria to the past five-year period was too limiting, as there were too few project evaluations undertaken in the context of migration. Thus, we settled on a ten-year timeframe.

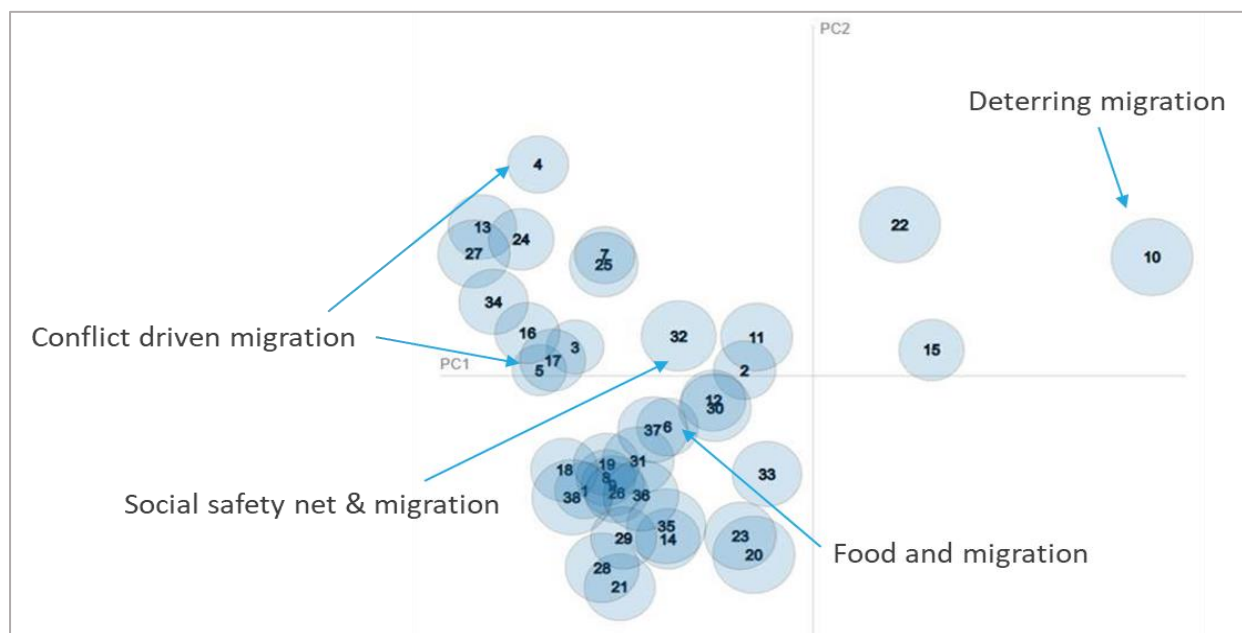
These limitations provided us with a sample of 578 project evaluations – an amount that would have been impossible for our team to manually review in the time available.

Table 1: Identified topics

| # | Topic |
|----|-------------------------------------|
| 1 | Animal and agricultural migration |
| 2 | Conflict driven migration |
| 3 | Deter displacement |
| 4 | Food and migration |
| 5 | Infrastructure and refugees |
| 6 | Mining and forestry displacement |
| 7 | Mitigation of migration |
| 8 | Monitoring Displacement/migration |
| 9 | Selecting migrants into the project |
| 10 | Social safety net & migration |

In applying the keywords, we reduced the total number to 123 project evaluations containing migration-related issues, which provided us with 1,197 sets of extracted text. The text extractions were a total of three sentences, with one sentence extracted before and after the sentence with the keyword(s) to assist with interpretation.

Figure 2: LDA mapping and evaluator defined topics



¹ A Boolean search combines or excludes keywords through operators (for example, AND, NOT, OR) to

help understand how they occur in different combination.

Step 3 –Defining topics

Having extracted the text, we combined machine learning and evaluation methods and processes to identify and refine topics of interest. Table 1 shows the list of topics and Figure 2 shows a mapping of those topics as represented by a machine learning algorithm.

We used a machine-learning algorithm specialised in topic discovery called Latent Dirichlet Allocation (LDA) to review the extracted text. LDA is an unsupervised machine-learning technique that automatically detects underlying topics in a body of text without using prior categorisation. LDA identifies patterns by using probabilities to define associations between text strings. In this way, the LDA provides mathematically correlated topics; however, this still requires the data scientist to tweak the algorithm and create a list of words to ignore. Without this, the LDA can produce topic lists that include administrative jargon and acronyms.

Evaluators then analysed the topics proposed by the LDA in more detail. The LDA provided a two-dimensional map of 38 topics (Figure 2). In addition, each topic (listed in Table 1) had 30 defining keywords and text extracts which were scored out of 100 for how well they fit within the topic's definition.

At the start of the review, we worked through the two-dimensional visual mapping to get a first look at the defining words of each topic and their coverage and frequency. To provide names to the topics, we examined the topics to review the associated words and read the text extracts. We counted a topic as relevant if we found a high proportion of coherent text extracts of interest to the question. In instances where only a small number of the extracts related to the question, we tagged the topic as not relevant. This reduced the number of topics, and by extension, evaluations considered important. Through the analysis, we identified 10 relevant topics that emerged within 49 project evaluations.

Box 2: Summary interpretation of Colombia social safety net project

The project context in regards to migration:

- the challenge in urban areas of a growing number of involuntarily displaced persons from extremely poor families with no access to basic social protection programs.
- The government contributed an additional US\$68 million to fund the expansion for the involuntarily displaced.
- Increasing numbers of people from conflict areas were displaced to urban areas creating new demands for the social safety net. Some 423 000 persons were displaced in 2002 alone and the accumulated number of displaced individuals from 2000 to 2005 reached 1.6 million. The displaced tended to be poor and moved mainly into medium and large municipalities.

The government executive branch not only had the pressure of humanitarian organisation and public opinion but a ruling from the Constitutional Court ordering humanitarian assistance to the displaced population as well.

Results of project were not disaggregated by displaced persons and included: Short-term education, nutrition, health and food intake indicators improved for program beneficiaries. Similar findings were reported for medium-term indicators of educational outcomes: high school completion rates increased by 4-8 percentage points for beneficiaries compared to non-beneficiaries.

Step 4 – Refining topics

To refine the topics, the evaluator developed a written interpretation of each topic and listed their main elements based upon an in-depth reading of the relevant text. In this step, the text reviewed was expanded from just the sentence extracts to entire paragraphs and sub-sections of the evaluation. In this project, the interpretation and write-up unfolded iteratively between multiple files and communications through three main activities described below.

First, to deepen the description of the topics, we sought to identify project evaluations likely to yield the richest insights on how World Bank project performance improved in the contexts of migration. To do this, we applied text mining to extract whole paragraphs from the remaining 49 evaluations using the migration keywords, providing 259 paragraphs. To focus the analysis, we ranked the evaluations by the number of extracted paragraphs. Reviewing these, we found that evaluations needed three or more paragraphs with keywords relating to migration to provide insight. Any less and migration tended to be a peripheral issue. This process led to the inclusion of 15 project evaluations. Should a larger number of evaluations have been included, we would have needed to modify our approach to coding to count the quantity of issues that arose, rather than in-depth qualitative analysis. Second, the paragraphs related to migration were read within the setting of the full evaluation report. We summarised the migration issue in each paragraph and read each evaluation's context, efficacy, and lessons sections. Targeting these sub-sections helped give an overall picture of the intervention, how it related to migration, what was achieved, and the challenges. The report formats had been kept consistent over the last 20 years, which made this review easier. An example summary interpretation of an evaluation is provided in Box 2.

Third, a colleague who had worked on evaluations on conflict and migration provided a sounding board for the emerging interpretations and advice on any non-project evaluations for review. This led to three additional thematic evaluations identified by our colleague also included in the review. Through reading, interpretation and exchange, rich descriptions emerged on the topics of conflict-driven migration, mitigating migration, food and migration and social safety nets and migration. The other topics, such as mining and forestry displacement, were more challenging to elaborate on as their issues were not well described across the evaluations.

Step 5 – Developing themes

A theme seeks to capture an issue common to several topics. To develop themes related to project performance in the context of migrations, the descriptions of topics were juxtaposed to identify shared issues. Then, the underlying text was analysed again to understand how the shared themes arose.

A limited range of issues that affected performance started to emerge that could be connected across the topics. The strength of the themes was reinforced when examples of both positive and negative performance arose within the text. For example, the targeting of groups affected by migration had been undertaken successfully and unsuccessfully in evaluations of social safety nets projects. Similarly, studies with content on both deterring displacement and food and migration found that a key issue in determining projects' performance was the development of the capacity of actors to engage in migration.

The extracted paragraphs were iteratively categorised to the emerging issues, with written interpretations of themes exchanged within the team. As the themes developed, the original studies were reread twice to ensure that the context of sending, receiving and transit countries were considered. The thematic evaluations that had been included based on expert advice had no extracted text to work with; therefore, these were reviewed through word and index searches. Consultation with our colleagues with expertise in

Box 3: Four themes that affect the performance of projects in the context of migration

1. Performance improves by targeting specific subgroups, locations, and time horizons.
2. Ongoing capacity development is needed to address migration.
3. Granular analysis serves as a basis for a coherent response.
4. The World Bank's support to coordination works well when based on comparative advantage.

migration helped to sense-check and suggested alternative interpretations. For example, while initially three themes were proposed, these were expanded to four by splitting capacity and coordination issues into their own themes. In this step, no data science techniques were applied, with the process drawing entirely upon the meta-ethnography approach.

The review identified that the performance of World Bank interventions operating in contexts of migration can be enhanced by paying attention to four themes. At this stage, the themes were quite rough. These were represented within a series of tables with assigned categories ground in the text;

Box 4: Description of targeting specific subgroups, locations, and time horizons

Subgroups – In contexts of migration, interventions should target differentiated groups based on categories such as migrants and nonmigrants, gender, ethnicity, and rural and urban spaces. For example, the Social Safety Net evaluation in Colombia identified that further efforts were needed to target interventions that "glue" displaced and local communities together to improve project performance (World Bank 2018c).

Location – Different evaluations highlighted that economic geographies need consideration in targeting migration. For instance, IEG's case studies from the rural nonfarm economy evaluation found that the World Bank often lacked a spatially orientated approach to the rural economy and needed to incorporate understanding of migration and remittance patterns in rural development interventions (World Bank 2017). Further, IEG's urban resilience evaluation highlights that identifying the spatial dimensions of migration helps identify risks and opportunities (World Bank 2019a).

Time Horizon – Targeting also has a time dimension. Separate paragraphs from evaluations in Azerbaijan, the Central African Republic, Haiti, and Lebanon reinforced that interventions often target either emergency or long-term responses to migration and that the failure to link the two can entrench poor economic outcomes (World Bank 2018a, 2018b, 2019b, and 2019d).

they still needed to be more finely shaped and potentially coalesced. These four themes are not comprehensive and represent the understanding that emerged from the reviewed evaluations.

Step 6 – Synthesis

To produce a synthesis, we further elaborated the themes and identified whether an overarching hypothesis about project performance in the context of migration could be articulated. The synthesis was conducted by the evaluators in the team, with ongoing exchanges with colleagues to draw out tacit knowledge. The output was reviewed by a series of managers who validated the synthesis.

In developing the synthesis, we described evidence from all 15 project evaluations and 3 thematic evaluations. The themes were articulated to present the evidence of factors that enhanced or decreased performance.

Although the meta-ethnographic process encourages attempts to combine texts to generate overarching hypotheses, this was out-of-reach in this case. The evidence did not describe how the different themes interacted, a limitation of the brief project evaluations that do not contain elaborated descriptions of observations or interviews. Consequently, the final synthesis was reciprocal; it contained mutually supportive evidence that identified issues that positively or negatively affected the performance of interventions.

At the end of this process, each of the four themes was fully described. Box 4 provides an example of these descriptions for the theme of "targeting".

Step 7 – Expressing the synthesis

We used an accessible briefing paper to present the elaborated themes to the management and executive directors of the World Bank. [It contained six pages of content, a table listing the evaluations, and a list of references.](#)

Providing a brief, rather than a presentation, panel debate or longer paper, was appropriate as it could be read and referenced as an accompaniment to the board briefing. It was potentially accessible to a broader audience following the briefing. This was the first of this kind of brief and was incorporated into the Board Brief and was reportedly used by World Bank management to review projects.

Reflection on this process

From this example, three lessons are worth noting that continue to inform our work. First, there are many opportunities for the interaction between evaluation and data science expertise throughout the steps. When we first started, we assumed that there would be a point where the data science ended and evaluation techniques were picked up, yet each step often requires the inter-play of expertise. For example, an evaluator can conduct a literature review, or a data scientist can suggest a keyword. Even expressing the synthesis can potentially be undertaken by machine learning using text summarisation techniques. Second, the representation of the seven-step process approximates the iteration, refinement and moving backwards and forwards in practice, especially between Steps 3, 4 and 5. Third, the synthesis requires both quantitative aggregation of text and qualitative interpretation. In the process, we used frequency counts, probabilistic machine learning and qualitative analysis to capture the rich interactions of the text. Working in this way marks a departure from meta-ethnography, which works solely through rich description.

We are still in the process of deepening our understanding of these seven steps. There's much to learn about how data science can augment evaluation skills throughout each part of the process, for example, by looking at additional techniques for identifying keywords or using supervised machine learning techniques to replicate the categorisation of existing datasets by evaluators.

Summary

The development of technologies that mine and analyse text can enable the rapid review and processing of a large amount of evidence. It can be used to augment qualitative review processes. The seven-step process outlined in this paper evolved from the meta-ethnographic process for synthesising reports. Based on this experience, we hope we've been able to show how data science and evaluation expertise can be used together to iteratively parse a large evidence base until a more focused and relevant evidence base emerges that can then be synthesised.

References

Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). 'Conducting a meta-ethnography of qualitative literature: Lessons learnt'. *BMC Medical*

France, E. F., Cunningham, M., Ring, N., Uny, I., Duncan, E. A. S., Jepson, R. G., . . . Noyes, J. (2019). 'Improving reporting of meta-ethnography: the eMERGe reporting guidance'. *BMC Medical Research Methodology*, 19(1), 25. Available from: <https://doi:10.1186/s12874-018-0600-0>

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies (Qualitative Research Methods Vol. 11)*. Thousand Oaks: SAGE. Available from: <https://dx.doi.org/10.4135/9781412985000>

World Bank (2017). *Growing the Rural Nonfarm Economy to Alleviate Poverty: An Evaluation of the Contribution of the World Bank Group*. Independent Evaluation Group. Washington, DC: World Bank.

World Bank (2018a). "Azerbaijan—Internally Displaced Persons Economic Development Project." Project Performance Assessment Report 128682, Independent Evaluation Group, World Bank, Washington, DC.

World Bank (2018b). "Central African Republic—Emergency Food Crisis Response and Agriculture Re-Launch Project." Project Performance Assessment Report 127422, Independent Evaluation Group, World Bank, Washington, DC.

World Bank (2018c). "Colombia—Peace and Development Project." Project Performance Assessment Report 127212, Independent Evaluation Group, World Bank, Washington, DC

World Bank (2019a). *Building Urban Resilience: An Evaluation of the World Bank Group's Evolving Experience (2007–17)*. Independent Evaluation Group. Washington, DC: World Bank.

World Bank (2019b). "Haiti—Port-au-Prince Neighborhood Housing Reconstruction." Project Performance Assessment Report 135543, Independent Evaluation Group, World Bank, Washington, DC.