The Evaluation Metro Map

Sebastian Lemire
Abt Global

Background: The practice and profession of evaluation continue to evolve. Over the years, evaluation scholars and practitioners have developed a broad—and still broadening—range of evaluation approaches. These evaluation approaches are often combined with various study designs in real-world evaluations. To provide an overview of these approaches and designs, and inspired by my son's love of metro trains, I created an evaluation metro map.

Purpose: The purpose of this article is to describe the background, purpose, and benefits and limitations of the evaluation metro map.

Setting: Not applicable

Intervention: Not applicable

Research Design: Not applicable

Data Collection and Analysis: Not applicable

Findings: Not applicable

Keywords: *evaluation theory; evaluation theory tree; evaluation metro map; visualization.*

Background and Motivation

I think I have always been a visual thinker. As a kid I would draw maps, mazes, and buildings—always from a helicopter perspective. I loved grid paper. One of my favorite children's books was called *The Yellow House* (*Det Gule Hus* in Danish), and true to its title the book was shaped like a big yellow apartment building. The book opened up and had layers; for each apartment, you could open the front door and see who lived inside, what they were doing, and their stories unfolding. It was magical.

My gravitation toward visual thinking has shaped my professional work in several ways. As an evaluation practitioner, I have always been interested in theories of change and theory-based evaluation—that it, visualizing the underlying logic of programs. I have conducted countless theory of change workshops and love the collaborative aspect of conceptualizing and visualizing how programs are intended to work. I recently published a method guide on different ways of visualizing program theories (Lemire et al., 2023), developed a poster of experimental designs (Lemire et al., 2022), and have used network maps to visualize evaluation capacity building communities (Nielsen et al., 2023). Mapping different aspects of the evaluation landscape is a sustained interest of mine.

I created the evaluation metro map within my first year of working at Abt Associates. Soon after starting in my new job, I had been asked to coauthor an article on trends in evaluation (Lemire, et al., 2020) and to give a webinar on that same topic for some of my colleagues at Abt. For the article, we had used an earlier version of Christie and Alkin's (2013) evaluation theory tree that we had adapted to focus on evaluation approaches instead of evaluation theorists. However, the adapted evaluation theory tree in the article did not really provide an overview of the many different evaluation approaches and study designs I had come across during my career. It was more selective and tailored to the approaches discussed in the article. I initially tried to develop a new tree with all the different approaches and designs, adding them one by one to the different branches. However, I quickly realized that I could not fit all the approaches and designs on the tree. There were leaves everywhere. As such, I was looking for a different visual for the webinar.

At the time, my family and I had just moved from Los Angeles to Washington, DC. Benjamin, my five-year-old son, just loved riding the DC metro. Most Saturdays we would get up early, head to the nearest metro station, and spend most of the day riding around on the metro and talking about trains. It was on one of those Saturday mornings that I got the idea to use a metro map as my visual for the webinar. I don't remember which station we were at.

The first iteration of the map, shown in Figure 1, was created a couple of weeks later. On a Sunday morning, coffee in hand, I was working on the opening slide for my webinar. I first played around on a piece of paper and slowly outlined the basic structure of the map, settling on three main metro lines: one for study designs, one for theory, and one for use—all of which are central to evaluation. I then transferred my doodles to a PowerPoint slide and added stations for all the evaluation approaches and study designs I could think of. These included approaches I had used in my own practice, read about in evaluation journals and books, or simply heard about at conferences.

I did not think much of the metro map until I used it in the webinar for my colleagues. They loved the metro map, and some even wanted to use it in client presentations and proposals. I am not sure they ever did, but I decided to post the map on LinkedIn a few months later. The initial post generated a lot of interest, with thousands of views and numerous reactions and comments. I have used various iterations of the metro map in several presentations since then—it always generates a lot of exchanges on evaluation. I continue to modify it.

Figure 1. The Evaluation Metro Map (Early Version)
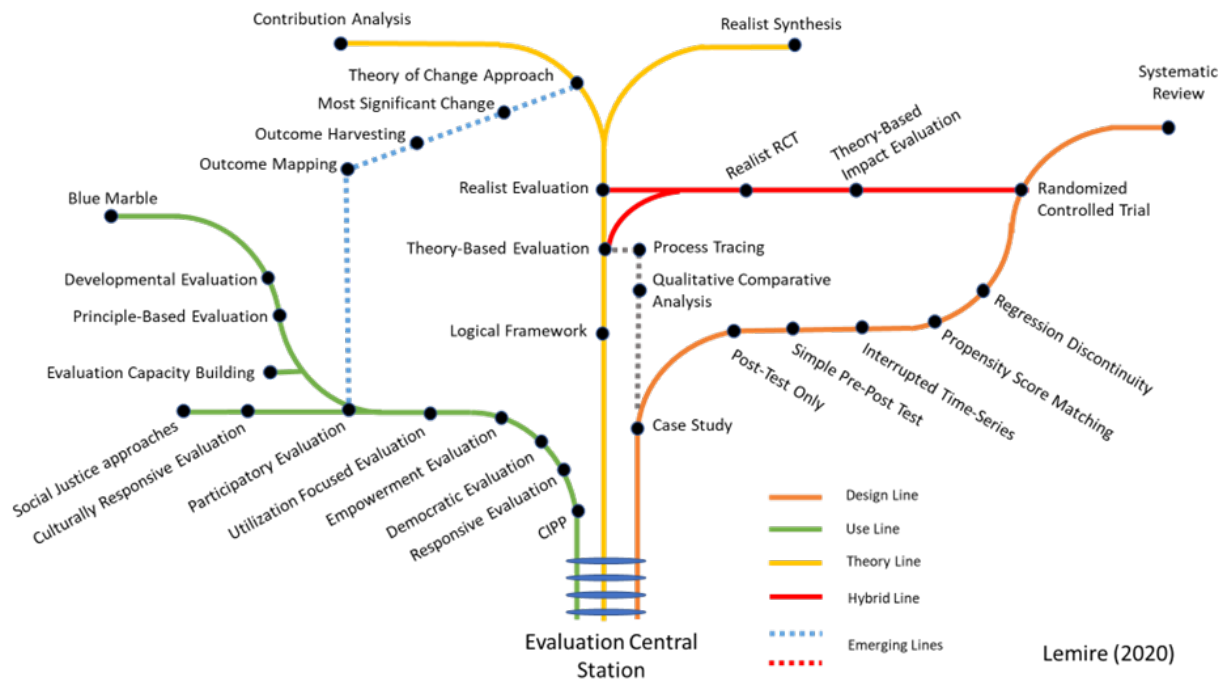
Structure and Content

Over the years, the evaluation metro map has evolved in many different ways. However, and as shown in Figure 2, the map is still structured around the initial three main lines: a design line, a theory line, and a use line. On each of these lines, which I will describe below, stations are labeled according to a specific approach or design. To the extent possible, related approaches are located on the same line and in proximity. For example, on the use line I have positioned principle-based and developmental evaluation in close proximity, because they are closely related (yet distinct) evaluation approaches.

The design line depicts a broad range of study designs, ranging from case study designs and non-experimental designs to quasi- and experimental designs. I generally don't think of these study designs as evaluation approaches and thus wanted them to be presented on a separate line from the evaluation approaches on the map. As discussed in the introduction to this special issue, study designs are distinct from evaluation approaches. I have sometimes been criticized for including the design line on the map. However, in the context of federal government evaluation, where I do most of my current work, I find that evaluations are commonly referred to and identified by their study designs, as opposed to the evaluation approaches they use. In fact, the latter is rarely mentioned.

The yellow theory line depicts evaluation approaches that center the evaluation around the development and testing of a program theory (or a variant thereof). Broadly referred to as theory-based evaluations, these approaches include realist evaluation and contribution analysis. The approaches on this line differ in important ways but share the common goal of developing and testing a program theory. I decided to split the line at the top of the map to allow it to travel in two different directions. The split line travelling towards the left side of the map is moving toward contribution analysis—the end station (for now). On the way, the line connects with an emerging line, consisting of evaluation approaches that connect the theory and use lines. The split line traveling toward the right side of the map ends at the station for realist synthesis, an evidence review approach that follows realist principles. I wanted the realist synthesis station to be located closer to the design line, which includes a station on systematic evidence reviews. Most of my past evaluation practice in the Danish public sector and in international development was located on the theory line, and I have done research on various aspects of theory-based and realist evaluation.

The use line depicts the many evaluation approaches focusing on the utilization of

evaluation, most notably approaches centering on equity and social justice. The same line includes participatory and culturally relevant approaches, as well as transformative and equity-centered evaluation approaches. I allowed the use line to split into two separate lines and end stations. Over the past few years, there has been a renewed surge of interest and commitment to culturally responsive and equity-centered evaluation approaches in evaluation circles. As a result, there have been and continue to be a lot of new stations added to this line. I wanted these approaches to have a separate line to allow for future growth.

In addition to the main lines, the metro map also includes several emerging lines (though I am not sure the approaches located on these lines still count as emerging). One emerging line connects the theory and use lines, with stations for outcome harvesting, outcome mapping, and most significant change, among other approaches. Another emerging line connects the theory and design lines. Stations include qualitative comparative analysis

and process tracing, both of which are anchored in case study designs and often applied in evaluation circles in the context of theory-based evaluation.

Finally, I added a red hybrid line to the map. The hybrid line connects the theory line and the study design line. The idea for this line emerged from my research on theory-based evaluation and my growing interest in how evaluators combine theory-based evaluation approaches with experimental designs. As part of my review of theory-based and realist evaluations, I had come across realist trials, evaluations that integrate a realist evaluation approach with an experimental design. These hybrid designs are not without controversy (Nielsen et al., 2023). After all, the realist evaluation approach was initially developed as a counter to experimental designs. However, I have come across several realist trials and theory-based impact evaluations in my research on evaluation and wanted these hybrid evaluation designs to be reflected on my metro map.

Figure 2. The Evaluation Metro Map (Recent Version)
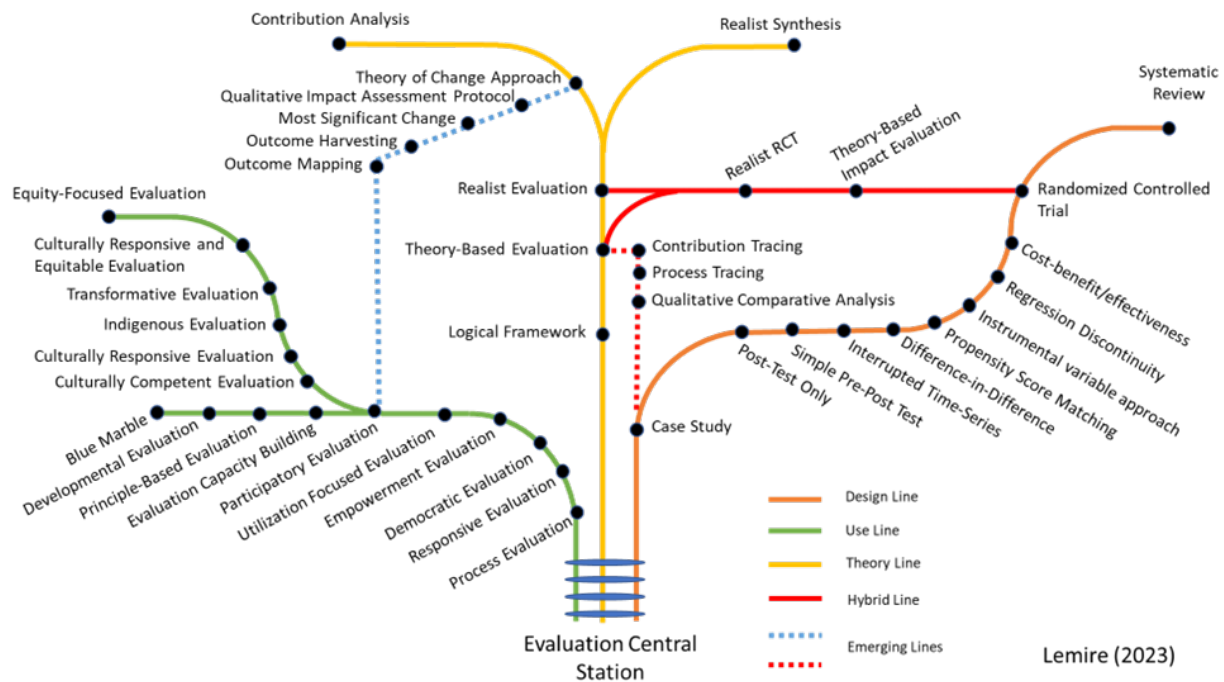
Lemire (2023)

Intended Purpose and Use

I initially wanted to use the map for webinars and conference presentations. The intended purpose was to provide an overview of the various evaluation approaches I have come across in my work. Over the years, I have used the metro map as a starting point for discussions on the difference between evaluation approaches and study designs, how various evaluation approaches are related and connected, and where new approaches are emerging. The intended users of the map are anyone and everyone with an interest in evaluation.

Benefits and Limitations

I think one benefit of the evaluation metro map is that it always generates conversation. Why is this evaluation approach next to this one? Why is this approach included or not included? If included, which line would it fit on? Is it a new line? I think these types of questions lead to new questions about what evaluation is or could be. I have sometimes been encouraged to add more stations to the lines. For example, in response to the early versions of the map, people asked me to include more culturally responsive and equity-oriented approaches to the use line of the map. Informed by conversations with Katrina Bledsoe, I have added several new stations on the use line, including culturally competent evaluation, culturally responsive evaluation, and culturally responsive and equitable evaluation. I try to use these discussions to further refine the metro map—it should continue to evolve.

Another benefit is that the map relies on a fairly common visual—the metro map—that many evaluators are familiar with. In this way, the map is both intuitive and accessible to many evaluators. I always encourage people to develop their own variant of the metro map or develop some other type of visual that speaks to them. I think different visualizations of evaluation theory—maps, rivers, trees, or flowers—reflect the experiences, perspectives, and preferences of those who created them. As noted in the introduction to this special issue, each of these visualizations provide *a* but not *the* lens through which to view and depict the practice and profession of evaluation. As such, and especially given the diversity of the evaluation community, we should encourage and celebrate a diverse range of visualizations.

The subjective nature of these visualizations also speaks to an important limitation of the metro map; namely, that the map covers the evaluation approaches and study designs that I am aware of and have come across in my work. One example of this limitation is that my design line leans heavily quantitative and focuses on impact designs with comparison groups. This likely reflects my current work on evidence reviews for the What Works Clearinghouse, where the focus of our reviews is exclusively on these types of impact designs. Although I continue to revise and update the map, there are surely many approaches and designs that are not included. I am sure I missed a lot. As mentioned earlier, I always encourage people to develop their own maps. After all, could there possibly be a better way to start a Sunday morning? I think not.

Another limitation that speaks less to the content and more to the structural part of the metro map is that real-world evaluations often combine stations from more than one line. As just one example, a utilization-focused evaluation (located on the use line) might involve a theory of change (located on the theory line) and case studies (located on the method line). Most of the impact evaluations I come across in my current work are perhaps best described as theory-based impact evaluation in that they involve quasi-experimental and experimental study designs combined with theory-based implementation studies. In this way, the map does not really reflect how designs and approaches are integrated in real-world applications of evaluation.

Closing Thoughts

I think the metro map can be viewed as a way of telling my personal story of evaluation—my perspective on the evaluation landscape. In this way, the metro map provides a narrative—a way of making sense—of the evaluation theories that I have come across in my evaluation practice. As such, the metro map also reflects my personal background and experiences in evaluation. In this way, it provides an example of how I as a practitioner am making sense of evaluation approaches. Reaching beyond my individual experiences, I would love for other evaluators to map the evaluation landscape, reflecting their evaluation experiences and practices. I believe we should have many different maps of this thing called evaluation. Collectively, this can help all of us make sense of the growing landscape of evaluation.

References

Christie, C. A., & Alkin, M. C. (Eds.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences.* Sage.

Lemire, S., Peck, L. R., & Porowski, A. (2020). The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation. *Policy Studies Journal, 48*(S1), S47–S70. https://doi.org/10.1111/psj.12387

Lemire, S., Peck, L. R., Porowski, A., & Kaur, P. (2022). *Experimental designs* [Poster]. https://www.abtassociates.com/experimental-designs

Lemire, S., Porowski, A., & Mumma, K. (2023). *How we model matters—visualizing program theories. Abt method guide.* Abt Associates.

Nielsen, S. B., Jaspers, S. Ø., & Lemire, S. (2023). The curious case of the realist trial: Methodological oxymoron or unicorn? *Evaluation, 30*(1), 120–137. https://doi.org/10.1177/13563890231200291

Nielsen, S. B., Lemire, S., Bourgeois, I., & Fiero, L. (2023). Mapping the evaluation capacity building landscape: A bibliometric analysis of scholarly communities and themes. *Evaluation and Program Planning, 99*(2), n.p. https://doi.org/10.1016/j.evalprogplan.2023.102318