Rezension zu:

Patton, Michael Q.: Essentials of Utilization-focused Evaluation. Los Angeles: Sage 2012. 488 Seiten, 58,00 US $, ISBN 978-1-4129-7741-8

Frans L. Leeuw[1]

In the preface to this book, Patton informs the reader that one of the reasons to write *Essentials of Utilization-focused Evaluation* (UF-E) was that the original book on this topic doubled its number of pages from 303 in 1978 to 667 in 2008, which may have caused unintended side effects (reducing the use and impact of it). One was that the 667 pages of the most recent edition were used as a 'substitute pillow' during a trip in an overnight train in India rather than as a study book, while unexpected costs to have the book 'lifted' in and out of a house of a reader was (jokingly) another unintended and negative side effect of the almost 700 pages thick volume. Needless to say that these experiences for Patton, a man who is – and rightly so – keen on making a difference, i.e. realizing an impact with his scholarly work, triggered (single loop) learning. The result of this learning has been this new book *Essentials of UF-E* (2012), published by Sage and 488 pages thin (or long).

The structure of the Essentials-book is rather straightforward. Apart from sections covering the introduction, conclusions and summary, the book consists of seventeen steps described in the same number of chapters. The landscape of the book is articulated nicely and is designed to function as the central theme; it can be found in many places throughout the book. Patton calls this 'landscape' "the complex dynamic and adaptive systems graphic", picturing interactions among all 17 steps, but he also uses a simple 'stairway metaphor', which gets higher and higher as the book progresses covering more steps. I assume that the underlying assumption of this design is to make sure that no reader is left behind, gets lost or stops reading.

The 17 steps start with the inevitable assessment of the program and organizational readiness for UF-E and the assessment of the evaluator's readiness and competence to do the job. The book then moves on to the identification and engagement of primary users of evaluations, and to identifying primary intended uses. A little higher up in the stairway the question is addressed if fundamental areas for evaluations are being adequately addressed (like implementation, outcome and attribution questions). Step 9 moves to determining what the underlying intervention model or theory is that is evaluated, while subsequent steps discuss the process of negotiating which methods to generate findings are appropriate. From step 12 onwards the focus is fully dedicated towards questions like how to stimulate the use of evaluations, how to gather data with an ongoing attention to use, how to present data and findings to relevant audiences and how not to forget the "follow up with primary intended users to facilitate and enhance use". The final step is the meta-evaluation of the use: "be accountable, learn and improve" (p. 388 ff.).

It is a rich book, describing in detail how to handle an evaluation trying to stimulate utilization, but that also invites the reader to join short digressions. The first one is a section on 'evaluation anxiety', aka as the "pre-evaluation stress syndrome" (p. 23), which is immediately linked to a 'what to do-activity', i.e. to create a positive atmosphere and vision of the evaluation, together with the expected future user(s) of the evalu-

---

1   University of Maastricht and National Justice Research Center WODC, Den Haag, The Netherlands

ation. Another digression (p. 83) has to do with the several temptations evaluators can be lured into. Patton lists 10 of them. One of them is that "evaluators make themselves the primary decision-makers and therefore the primary users" of 'their' evaluation. Another one is assuming that the funder of the evaluation is automatically the primary stakeholder. An interesting but disputable one is temptation no. 7, "taking a stance of staying above the fray of people and politics". Patton suggests that trying to do this leads to "irrelevance … evaluation is action-oriented and action is in the fray". This statement sheds light on the epistemological frame that characterizes this book (and that I'll come to a bit later). The interesting section on six alternative evaluation goals (p. 115) is also a must-read, although I almost totally disagree with what he is writing about 'developmental evaluations'. Sometimes, Patton is a true believer. No references can be found to unintended side effects of the M(onitoring) & E(valuation) industry (à la lowering policy ambitions for example) or to the many other side effects, when auditing marches in (teaching to the test, the performance paradox, protocollization etc.), although a little later Patton indeed shortly discusses the performance measurement mantra (p. 152).

A very informative chapter is the one of formulating evaluation questions. A typology of questions is produced of which the IT question (what is IT?, does IT work? etc.) I like best (p. 179). IT has nothing to do with ICT, but IT is 'The Intervention' or 'The Program'. Here, Patton holds a plea to open up the black box of what in many evaluations often is kept silent, as IT usually is only the label of the intervention, program, law or other type of policy measure that has to be evaluated. Unfortunately, Patton is not addressing methodological literature on how to distinguish types of (evaluation) problems, and the methodological quality of the way in which problems are formulated. What makes research problems ill or even erroneously formulated? These are problems that are formulated against a background that consists of at least one incorrect statement. The background "is constituted by the antecedent knowledge and, in particular, by the presuppositions of the problem. The presuppositions of the problem are the statements that are somehow involved but not questioned in the statement of the problem and in the inquiry prompted by it" (Bunge 1997: 194).

The chapter on the intervention theory/logic is, as are many other parts of the book, basic and again loaded with charts and graphs. Unfortunately, not much is said about the methodology how to find ('detect', 'reconstruct') underlying assumptions, i.e. the intervention or program theory. Over the last decade a number of methodological suggestions have been formulated, including making use of argument mapping software, that can increase the level of professionalism when doing this kind of work. Now the reader may either wonder how to 'find' this intervention theory or, which is more dangerous, just starts to 'write the theory/assumptions up'. Tilley (1992), more than a decade ago, warned already for the danger of sloppy reconstructions of theories.

The most well-known chapters in the book are the ones on how to 'sell' evaluation findings to your audiences. They focus on how to write the evaluation report in a way that it is almost unputdownable, how to get stakeholders engaged in dialogues focusing on learning from the evaluations, to study the way in which 'information is beautiful'-websites help you to transfer the sometimes voluminous and heavy sets of data to the audiences involved and even to watch Hans Rosling (Swedish statistician) movies/you tubes, who animates statistical findings and trends like they are the European Soccer Finale.

Numerous exhibits, menus and boxes are part of the book. A shortcut to characterize the book would therefore be to label it the ABC3EFM-toolbox of UF-E. ABC3EFM stands for *Arrows*, *Boxes*, *Checklists*, *Charts*, *Cartoons*, *Exhibits*, *Figures* & *Menus*. Many examples from a wide number of fields and studies are also part of the book showing what in the respective practice has been going on and what can be learned from them.

Although exhibits, boxes and checklists certainly give guidance to readers and in particular to newcomers in the field, they also leave an impression of UF-E as something mechanical, and even 'techno-ish': evaluators should possess x, y and z, should practice Q, have participants share or do this and that, engender commitment to … etc. The actions to be taken during the 17 steps include the descriptions of situations where and when to carry them out, often with an impression of what the consequences will be. This, together with the large number of (normative) suggestions and guiding principles, makes me to bring up this

critique. Moreover, and in line with a recent presentation by Gary Henry,[2] the question should be asked how much systematic evidence there is for all the guidelines and normative statements. Of course Patton's personal decades of experiences doing utilization-focused evaluations is an important source as is the literature list, but I doubt that systematic and robust evidence on what works best, when and how, and for whom when transferring and using results from evaluations by practitioners, policy makers, politicians, managers, teachers, journalists and the public is available to underpin all of Patton's claims, suggestions and recommendations. Recently, I searched for this type of evidence in the field of knowledge transfer, translation, utilization in repositories of social, behavioral and technological studies, but found amazingly little systematic evidence.[3] Case studies, narrative material and surveys seem to dominate (Brandon/Singh 2009).

Moreover, there are two other serious and epistemological points of critique on the book. The first concerns its clearly articulated one-sidedness. One of the central arguments in the book is that almost everything that is in the hands of the evaluation commissioner, the evaluators and the users of the evaluation has to be done to make sure that evaluations are designed and carried out in such a way that they indeed are used. Non-utilization looks like a total waste of resources. How perfectly clear this statement is, it misses some points. The first is that there are good and bad evaluations. Bad evaluations should not be carried out, but for sure not be utilized. They wrongly inform society, policy makers and others; the error costs can be high (Leeuw 2011). The second point is that as evaluation has turned into a booming business with its own evaluation machines (Dahler-Larsen 2000), this increases the likelihood that other incentives than 'speaking truth to power' and 'learning' are on the agenda when doing and using evaluation results (like shopping for future clients and research money, bureau politics, teaching to the test). Dahler-Larsen's (2000: 70) finding that almost never evaluations are used to terminate policies of Danish municipalities is an exciting though somewhat worrying one. "Out of 600 evaluations, less than 1% led to termination of the evaluated activities, whereas 78% of the evaluations led to adjustments of program activities". There is no guarantee that these 'adjustments' de facto and letztlich will lead to more effective policies; the opposite phenomenon that 'interventions breed interventions' is also possible, but often neglected.

The second problematic aspect of Patton's book is the attention paid to collaboration between evaluators and primary users. His definition of UF-E is that it is "a process for making decisions about (these) issues in collaboration with an identified group of primary users focusing on their intended use of evaluations" (p. 6). This process is "highly personal and situational" (p. 4).

Although Patton refers to standards for evaluation like accuracy and accountability, in my opinion he does not pay enough attention to the cognitive-behavioral consequences of the way in which 'his' UF-E should be carried out. Take this example: "if evaluations are ignored or misused, we have to look at where our own practices and processes may have been inadequate" (p. 5). The possibility that the very act of collaboration between evaluators and evaluands may have contributed or even have caused evaluation findings not to be used, because officials are no longer certain about the independence and credibility of evaluators as well as the methodological quality of their designs and data, is not part of Patton's cognitive frame of reference. In my own world, which crosses over several decades of evaluations in Europe and in fields ranging from crime and justice, education and development aid to environment and energy and public administration, I have experienced many times that it is exactly the 'beloved' collaboration between evaluators and evaluands, including stakeholders, funders and others, which causes non-utilization or misuse.

Finally, a few other points:

– I miss the role internet and, more broadly, 'living in the web' plays in and for UF-evaluations. A couple of times Patton refers to it, but a systematic analysis of the strengths and challenges that developments like 'Big Data' (logs, mobile trans-

2   European Evaluation Society Congress, Helsinki, Session: Theories in Evaluation, October 2012.
3   The keynote was given at the EIPPEE Conference 2012: Advancing the use of research in education across Europe, Den Haag: "Transferring research (results) to policy makers and practitioners: Experiences from the fields of security and justice and development aid".

actions of many kinds, online user-generated content, such as blog posts and Tweets, online searches, satellite images, GPS-travel data etc.; see UN Global Pulse, www.globalpulse.org), digital policies and interventions and new ways of data collection (involving clients) by using apps, sensors or the ethnocorder, remain untouched. The same is true for the competition to come between the wisdom (sometimes: the stupidity) of the crowd and the wisdom of a few collaborating evaluators. As a profession, we seriously have to think about that.

– Comparing evaluations with DNA (see the double helix picture on the front page) and, further on in the book, linking evaluation to Gordon Moore's Law (computer capacity will double every two years) is a bit of an exaggeration. DNA and Moore's law are of a different nature and work on a different level. I always teach my evaluation students to be strong, will powered and tough on tenacity, but also … showing a certain humbleness.

References

Brandon, Paul R./Singh, J. Malkeet (2009): The Strength of the Methodological Warrants for the Findings of Research on Program Evaluation Use. In: American Journal of Evaluation, 30(2), pp. 123-157.

Bunge, Mario (1997): Philosophy of Science: From Problem to Theory. New Brunswick: Transaction Publishers.

Dahler-Larsen, Peter (2000): Surviving the Routinization of Evaluation: The Administrative Use of Evaluations in Danish Municipalities. In: Administration and Society, 32, pp. 70-92.

Leeuw, Frans L. (2010): Benefits and Costs of Evaluation – An Essay. In: Zeitschrift für Evaluation, 9 (2), pp. 211-229.

Tilley, Nick (1999): Evaluation and Evidence-(mis)led Policy. In: Evaluation Journal of Australasia, 11, pp. 48-63.