U N I V E R S I T Y   O F   C O P E N H A G E N

Reflections on vocabularies in research on evaluation

Dahler-Larsen, Peter

Reflections on vocabularies in research on evaluation
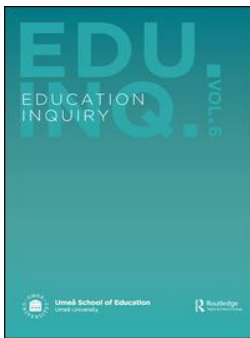
Peter Dahler-Larsen

Published online: 01 Sep 2016.

Submit your article to this journal

Article views: 100

View related articles

View Crossmark data

Reflections on vocabularies in research on evaluation

Peter Dahler-Larsen[†]*

The contributors to this issue (Vol. 7, No. 3, 2016) deserve praise for taking on the important research project "Consequences of evaluation for school practice: governance, accountability and school development" and for making their results available in this collection of well-written articles. Their insights into the role of evaluation in schools in Sweden are many. Based on multiple methods and various approaches, we obtain a comprehensive picture of how key actors engage in evaluation at various levels of school governance.

I thank the editor for the invitation to reflect upon this interesting collection of articles. I accept it fully aware that I cannot be equally attentive to all contributions. Somewhat unfair to the contributors as a collective, I shall talk not so much about what they have already accomplished, but instead focus on what lies ahead of us as researchers and practitioners who wish to further the understanding of this strange phenomenon called evaluation.

Let me focus on reflections about our vocabularies, our terminologies, our conceptual frames of reference. As students of evaluation, we struggle to describe a moving target. Yet, our insights depend on our ability to conceptually catch up with a reality that is moving too fast.

Therefore, my first reflection has to do with the very definition of evaluation. Most of the articles (including Hanberger 2016) subscribe, for good reasons, to a broad definition of evaluation which includes inspection, quality assurance, ranking and a variety of other practices that document teaching and its results. Sometimes the broad definition of evaluation also includes as one of its elements ... evaluation!

My purpose is here not to engage in a hair-splitting discussion of definitional formalities, but to ask the question: What does it give us to define the overall 'monster' of documentation, ranking, outcome-orientation etc. under the rubric of evaluation?

[†]Correspondence to: Peter Dahler-Larsen, Department of Political Science, University of Copenhagen, Copenhagen, Denmark. Email: pdl@ifs.ku.dk
*Department of Political Science, University of Copenhagen, Copenhagen, Denmark.

373

What is the 'baggage' that comes along with the term evaluation? Can we discuss the pragmatics of our definitional strategy?

I think it has to do with cultivation of a particular set of socio-historical expectations. Evaluation is often connected with expectations about systematic, methodological inquiry. It carries with it a tradition for attention to value issues and sometimes value controversy. And it carries the expectation that systematic inquiry based on explicit values should be used for some identifiable purpose in society, more often than not some form of social improvement. All of these aspects of evaluation have evidently been challenged over the years and the basics continue to be under discussion (Schwandt 2015).

Evaluation has not lived up to its original grand and often rationalistic assumptions. However, without *any* set of normative expectations, research into the broad set of empirical phenomena we call "evaluation" would lack direction. While we should pay attention, of course, to the ever-changing social formation of the evaluation wave, the very term "evaluation" supplies that research with some degree of unity and common interests and also some normative expectations that would be lacking if we randomly chose another umbrella term for our research object. But perhaps we can be clearer about these foundations for our research on evaluation.

The same applies to our understanding of how multiple evaluative phenomena act together. Lindgren, Hanberger and Lundström (2016) direct our attention to the astonishing number of "approximately thirty evaluation systems" operating in the Swedish education system. It is an extremely important observation as it raises an interesting question about the political and democratic nature of this plethora of evaluative phenomena.

Conceptually, however, it can be debated whether the overall interaction between these diverse elements could, in some situations, also be regarded as a 'system' in itself. It seems like a "web of evaluation" is becoming the preferred term.

I have suggested that at least some of these phenomena can be conceptualised as "evaluation machines" (Dahler-Larsen 2012). This metaphor suggests inhumane overtones. It is usually not nice to interact with machines. But it is just a metaphor. The way we use metaphors as early and preliminary analytical terms points to the need for further theoretical specification. Indirectly, the various metaphors – and analytical terms we may develop in the future – bring with them connotations about how evaluative elements in society are connected, how smooth their operation is, the role of human judgement, who has responsibility for the interconnectivity, and whether some overall meaningful coherence can be expected. I would be eager to see the notion of "web of evaluation" further specified, metaphorically and analytically. Computers are connected into webs. Webs facilitate the movement and spread of information. Webs are also used by spiders to catch insects. When an insect is caught, the web moves and informs the spider that something interesting has happened that requires the spider to move, too.

In our studies of interconnectivities among forms of evaluation and people and their practices we are challenged to revise our fundamental assumptions and categories. The articles in this volume made me think of one of our key notions, that of levels of analysis. True enough, the effects of the evaluation wave manifest themselves at the local level in the form of a reduced space for professional discretion and autonomy for teachers as well as for school managers (Hult, Lundström and Edström 2016).

And true enough, this has to do with evaluation instruments being used to enhance uniformity and comparison, all in the interest of central control. It seems to me, however, that several of the articles also suggest more complex patterns of interconnectivity. Otherwise, it might be too easy to say that the real roots of the problem lie outside the frame of the analytical picture which keeps its focus on the *local* school and the immediate actors around it, who may, on a bad day, be reduced to 'victims' of evaluation. We can too easily regard the concerns for control as what the economists call an exogenous variable: All bad things come from the outside. I assume, in Sweden, more specifically: From Stockholm.

If we ask instead which of the actors under study help enhance evaluation in more active (though sometimes more or less sceptical) ways, it seems that both managers (see Hult et al. 2016) and teachers (perhaps especially younger ones) in fact keep evaluation alive through several practices, including the time-consuming production and transportation of many forms of data. And the parents, for whom evaluation plays an increasingly important role in their capacities as 'right-holders' and 'customers' in the 'market' for education, are these parents not in fact to some extent identical with the voters who have elected the policymakers who are responsible for the legal and institutional frameworks the effects of which are under study? So, some of the 'producers' are embedded in the analytical picture, not just comfortably relegated to 'the central level'.

What I am suggesting is that, in some situations, the notion of 'levels' helps keep the 'producers' and 'victims' of evaluation apart. In fact, it may be one of the social functions of evaluation to do exactly that, as evaluation is capable of holding someone accountable across time and space for achievement of goals created at far distances. But if there were no interconnectedness, evaluation would not work. If the central is not present in the local, it has no impact.

'Levels of analysis' are not physical entities (Latour 2005), they are imaginations of how the world is organised. In other words, we can alternatively imagine other ways of framing our studies so that interconnectivities become further highlighted and the 'production' and 'use' and 'maintenance' of evaluation appear as being constructed in one and the same analytical move and the actors are portrayed as (even more) deeply involved in all the ambiguities and tensions that follow. Some of the articles testify to such ambiguities, at least in my reading. As an example of evaluation that teachers themselves see as close to their practice, they mention "how they succeed in raising

student awareness of their own performance in relation to new grade criteria" (Hult and Edström 2016). I read this as a call for – at least – a third-order reflection (among teachers) over a second-order reflection (among students) over their performance (in the first order) in relation to new grade criteria. I assume the latter *were introduced from the national level*. So the 'levels' are incredibly intertwined. Maybe the several orders of evaluative reflection at the local level are themselves productive, or at least they multiply the effects of everything that came from Stockholm. In my reading, the new grade criteria *are* not close to the teachers, but they are *drawn* closer by means of evaluative practices where students and teachers engage in organised reflections. There is something about the *drawing* part of practices that is, in my mind, not sufficiently captured if our ontology is already shaped by assumptions of pre-existing 'levels of analysis'.

Perhaps we can even analytically capture or even re-create socio-political situations in which the tensions between different roles (producer, user, expert, professional, victim, customer etc.) in relation to evaluation are brought closer together and their tensions become more visible *within* our case study, maybe in the moderated form as reflections over reflections, or maybe even in the form of mini-publics where deliberation takes place live as we study it. Perhaps the tension is already there but just needs to be made (even) more manifest and explicit through our methodological strategies.

This brings me to a short methodological reflection on how to document the consequences and use of evaluation. Obviously, one approach is to ask actors in various roles and positions whether they find a particular form of evaluation valuable and useful, and whether they can exemplify forms of use. But this is '*user-heavy*' since it relies very much on the perspective of the persons in focus. It only brings forward what they can see and like to talk about.

Another approach (finely illustrated by Carlbaum 2016) is reading the 'technologies' of evaluation in terms of the agencies and roles they make possible (even when not acknowledged as such by the actors involved). In turn, this perspective might be '*theory-heavy*' as it highlights forms of use already expected within a particular theoretical framework.

I think both of these methodological strategies can contribute to our understanding of the consequences of evaluation. But they do not give the same type of insights, and insights based on these two strategies cannot be mixed at will. They are not commensurable. Consider, for example, the observation that evaluation data are being used for marketing and self-promotion purposes (mostly by those with good scores, not surprisingly). Actors themselves may find the data 'valuable and useful'. If the same observation is understood through an alternative, critical analytical lens, it shows the deep ambiguities of navigating in a strategic landscape defined by evaluation: It is tempting to use a bad indicator for good purposes for those people whose score happens to be fine. But that kind of behaviour will take place only in a

social universe in which the management of education is already defined as having to do with marketing and promotion. This logic appears, then, 'behind the back' of the actors; and their explicit interview statements cannot be trusted at face value. Our methodologies to capture the consequences of evaluation deserve further differentiation and specification.

To capture the consequences of evaluation, this collection of articles often refers to 'constitutive effects'. I have read with great interest and enthusiasm how the contributors have made the concept of 'constitutive effects' work for them. A big challenge related to this concept is how to keep it sharp enough to achieve a sense of unity in the findings and at the same time open enough to inspire a search for its many empirical manifestations.

Constitutive effects include, for example, the reduction of creativity in teachers' work (Hult and Edström 2016), the introduction of a more consumer-oriented set of roles in relation to education (Carlbaum 2016) and the gradual acceptance and taken-for-grantedness of ranking and performativity as key principles in education (several articles, e.g. Hanberger, Lindgren and Lundström 2016). Perhaps increasing levels of distrust can also be understood as a constitutive effect. If evaluation is introduced in order to control schools and teachers in a situation of deteriorating societal trust in education, it is difficult to imagine a theory that explains how evaluation brings back trust. Instead, every new step of evaluation only increases distrust, and it is not difficult to imagine a vicious circle of more distrust and more evaluation.

It appears to me that the concept of constitutive effects is both useful and productive in the material at hand. Let me combine the many observations of constitutive effects through the articles with a relevant observation made by Hanberger (2016). He observes that sometimes evaluation systems are introduced in situations where all requirements for classical accountability (clear principal-agent relationships) are not met. So, even if the (normative) requirements for accountability are not in place, evaluation systems succeed in – paradoxically – producing 'accountability effects' even if all preconditions are not there. Perhaps the 'tentative', 'indirect', 'discoursive', 'symbolic' and 'interpellating' aspects of constitutive effects are the keys to understanding how this is possible. Perhaps the constitutive power of evaluation manages to cut across imperfect accountability. Maybe the power of evaluation is so much stronger because it manages to make people do something they technically do not always have to do. Maybe this power even helps amplify constitutive effects and send them off in various productive directions which are difficult to map because they are non-linear and describe an 'overflow' or 'overproduction' of effects rather than just what would be exactly expected in a more deterministic model.

However, before the concept of constitutive effects flies off in all directions, it is also important to remind ourselves (including myself!) which kind of responsibility the concept requires. Since the concept does not use any 'original' political intentions

as a standard and benchmark, we are left with the question of how to make explicit the norms, values and expectations in contrast to which some constitutive effects are noteworthy. Further, the concept itself does not suggest which effects are positive/ negative – and they may be both, depending on which value framework is assumed – but if they are neither/nor, they may be of little interest. Constitutive effects are visible only because we have frameworks that allow us to see them. So the concept itself challenges us to be explicit about our frameworks.

Hanberger et al. (2016) are admirably clear in pointing out how the negative view of the customer-oriented approach to education stands out clearly when contrasted with the ideal of deliberative democracy. So it is good to be explicit about which theoretical frame allows us to see which constitutive effects. I would add that the relevance of our observations increases if we publish observations about constitutive effects that are potentially interesting in society. So we must make the 'actor-heavy' and the 'theory-heavy' ingredients in constitutive effects meet in the public arena or, as a minimum, explain how they can potentially meet (even if they are not automatically commensurable). In this way, research on constitutive effects sends researchers on evaluation directly back to report their findings in the contest ed public arena where the constitutive effects are themselves produced in the first place.

To succeed in increasing attention to constitutive effects in the public arena, evaluation researchers have to do argumentative work. In combining theory-directed and actor-directed perspectives, it is sometimes necessary to use lengthy argumentation to open people's eyes. Some perspectives have so limited representation in the public arena that it is difficult to make particular research results resonate with their concerns. One example is unborn generations who are, by definition, absent in the debate. Another example may be – may I provocatively say: Children and students! Even if the articles in this collection cover an admirably broad range of actors, I have not succeeded in finding an in-depth coverage of the constitutive effects of evaluation upon the daily life of students and their experience with education. I may be wrong, but with students 13–15 years old it should be methodologically possible and an important thing to do.

My own hunch is that evaluation regimes interact with a number of measures in educational policies (and with new images of competition in the labour market) to produce new generations with a more instrumental view of education – and with more anxiety among those who do not 'perform'. Only future research will show if this hunch is more or less correct, even if too simple, I am sure. May it suffice here to note that all observations of potential constitutive effects are by definition incomplete and depend and theoretical, analytical and normative choices.

I agree with Lindgren et al. (2016) that constitutive effects are too often ignored in policymaking and in the public arena at large. So we have a contribution to make in

bringing attention to these effects and in keeping the deliberative and democratic discussion about these effects alive.

As a counterpoint, I also partly disagree with myself. While we discuss, what should we not do? For how long a time should we discuss these things? And what should we do when that time is gone? Do we expect a moratorium on all evaluative activity while the discussion is ongoing? Do we have the capacity to discuss all aspects of practice?

Can we theoretically accept the empirical fact that there are evaluative activities taking place which are practical but not subject to discussion?

Deliberative discussion is a good thing, but we should be careful not to define in advance what counts as democratic contributions (Vattimo 2005). Maybe we should begin to think of all these evaluative phenomena as socio-political interventions *sui generis*. Maybe they work all the time in their own way, at best parallel to some democratic deliberation but, perhaps frighteningly, they are not only loosely and partly subject to democratic ruling, they are not all totally transparent, their effects are elusive and confusing, and perhaps evaluative practices do not obey appeals to normative ideals. Perhaps they work according to new and remarkable logics that we have not yet fully understood. Perhaps evaluation is less dependent on democracy than we conventionally think.

To study these things requires a lot of courage and we have only just begun. I enjoyed this collection of articles a lot.

Peter Dahler-Larsen, PhD, is a professor at the Department of Political Science, University of Copenhagen, where he is the leader of CREME, Centre for Research on Evaluation, Measurement and Effects. His research interests focus on the social and organisational aspects of evaluation and evaluation systems. He has conducted qualitative studies in Estonia, Moldova, Transylvania, Namibia, Greenland and Denmark. He is past president of the European Evaluation Society. He is the author of "Kvalitetens Beskaffenhed" (University of Southern Denmark Press 2008) and "The Evaluation Society" (Stanford University Press 2012). Email: pdl@ifs.ku.dk

References

Carlbaum, S. (2016). Customers, partners and rights-holders: School evaluations on websites. *Education Inquiry*, 7(3), 29971, doi: http://dx.doi.org/10.3402/edui.v7.29971

Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.

Hanberger, A., Lindgren, L. & Lundström, U. (2016). Navigating the Evaluation Web: Evaluation in Swedish Local School Governance. *Education Inquiry*, 7(3), 29913, doi: http://dx.doi.org/10.3402/edui.v7.29913

Hanberger, P. (2016). Evaluation in Local School Governance: A Framework for Analysis. *Education Inquiry*, 7(3), 29914, doi: http://dx.doi.org/10.3402/edui.v7.29914

Hult, A. & Edström, C. (2016). Teacher ambivalence towards school evaluation: promoting and ruining teacher professionalism. *Education Inquiry*, 7(3), 30200, doi: http://dx.doi.org/10.3402/edui.v7.30200

Hult, A., Lundström, U. & Edström, C. (2016). Balancing managerial and professional demands: school principals as evaluation brokers. *Education Inquiry*, 7(3), 29960, doi: http://dx.doi.org/10.3402/edui.v7.29960

Latour, B. (2005). *Reassembling the social*. Oxford, UK: Oxford University Press.

Lindgren, L., Hanberger, A. & Lundström, U. (2016). Evaluation systems in a crowded policy space: Implications for local school governance. *Education Inquiry*, 7(3), 30202, doi: http://dx.doi.org/10.3402/edui.v7.30202

Schwandt, T. (2015). *Evaluation foundations revisited*. Stanford, CA: Stanford University Press.

Vattimo, G. (2005). *Nihilisme og Emancipation*. Århus, Aarhus Universitetsforlag.