The Future of Evaluations, Crime Policy and Penal Law Reforms: Some (largely) Dutch Experiences

FRANS LEEUW

Table of Contents

Summary

Evaluations of crime policies have a long history and are in many (Western) countries booming. Data bases are used as are interdisciplinary approaches and different research designs, underlying policy and intervention theories are studied, attention is paid to data visualization and knowledge transfer and (sometimes) [some] evaluation results are used by policy makers, public prosecutors, police, other law enforcement agents and civil society representatives.

However, the situation is no 'paradise on earth'. Sometimes adequate data are lacking (e.g. on cyber crime), sometimes intervention theories are no theories, but only arrows, circles and a few words; sometimes inadequate designs are used, sometimes there are (error) costs of evaluations etc.

And, there are important new challenges: how to integrate the exploding field of brain, cognition and social neurosciences studies into crime, violence and aggression? How to link up with Big Data and evaluations in the field of (criminal) law? How to evaluate new technology that

is used in prisons and probation (like serious games)? And: what will the future of Quantified Self bring to the world of evaluation?

1. Evaluation of Law Enforcement

Evaluations in the field of law enforcement have a fairly long history. In the Netherlands it started in the early 20s, but really took off in the 1970s. In the USA Stuart Chapin and others did experimental evaluations in the 20s and 30s, and later became a, if not the leading country in doing penal law and crime and justice evaluations. Everybody knows the Sherman report which was one of the very first large scale review studies of hundreds of evaluations of law enforcement actions, programs and policies, published mid 90s. Special journals like on experimental criminology and law, behavior and policy have existed since a couple of decades.

While in the beginning money was very tight for this work, during the Tony Blair period in the UK, and during similar periods in other countries including the Netherlands, serious money became available for designing and evaluation crime prevention and deterrence policies.

So, how booming or blossoming is the law enforcement evaluation field in western, industrialized countries actually nowadays?

I follow the approach that Rist et al (2002) have used for first edition of the International Atlas Of Evaluation, which was recently updated in an article by Jacob et al (2015). Rist et al. used these indicators:

I. evaluation takes place in many policy domains;
II. there should be a supply of evaluators specializing in different disciplines;
III. discussions and debates fuel a national discourse regarding evaluation;
IV. a national evaluation society exists;
V. institutional arrangements in the government for conducting evaluations and disseminating their results exist;
VI. institutional arrangements in Parliament for conducting and disseminating evaluations exists;
VII. pluralism exists within each policy domain;
VIII. evaluation activities occur within the supreme audit institution; and
IX. evaluations do not just focus on inputs/outputs, but also on outcomes.

In the Netherlands there is. We not only have specialized institutions but also data spaces and data warehouses combining data from the police, prosecutor, judges, Statistics Netherlands, Child Custoday Board etc.

206

The UK has the Justice Data Lab which is a somewhat similar example, and also specialized (research) institutions for issues regarding the police, terrorism and others, including 'evidence-based policy making'.

Another aspect of the infrastructure is that with new laws and legislation, EU countries more and more are active in ex ante (impact) evaluations ánd in evaluating compulsory after the existence of laws and regulations for a period of 3 to 5 years.

Switzerland is also active in this field, a fact that has become obvious during this annual conference. The contributions of Bussmann and other colleagues provide further proofs.

Regarding the development and implementation of judiciary behavioral modification programs, with which prisons and other organizations in the penal law world operate, an important indicator of an infrastructure is the existence of so called certification and accreditation boards. In the Netherlands such a board exists, as is the case in several other countries.

And the final aspect of the evaluation – infrastructure deals with R and D/ innovations taking place in the world of crime policies. Is R and D taking place and are these activities evaluated? The answer for the Netherlands is affirmative. And that is true for a number of other countries.

Let me mention two examples from the Netherlands: One regards the use of insights from neurosciences, including neurobiology and neuropsychology dealing with finding out what root causes are of violent behavior but also and in particular how to improve the sanctions/interventions/programs to prevent or reduce antisocial behavior, aggression etc. That is done in collaboration with brain and cognition research, which is booming, partly due to new methods of data collection. In the Netherlands a 5 year national program has recently run which stimulated this kind of research and evaluations, with a total budget of 7 million Euro. Some 15 research projects were carried out ranging from diagnosing delinquents to find out who is indeed 'fit' for interventions like Aggression Regulation Therapy and Enhanced Thinking Skills; how new methods from neuro-criminology can be used diagnosing and assessing delinquents and how virtual reality can help in socializing delinquents in a more effective way. This includes interventions with the cuddle hormone (oxytocin) and noninvasive neurofeedback, like TCDS.[1]

---

[1] See for more information: https://www.hersenenencognitie.nl/contents/1115? locale=en

The second is an initiative on cyber security, Big Data and the Law, where legal predictions is the center piece: can one try to predict the acceptance of new legislation and regulation by using Big Data, from social media, Google search practices, and Legal Big Data. In the USA Daniel Katz et al (2013; 2014) is active in this field.

Indicator 1: What about the supply of evaluators for the law enforcement field?

As far as I monitor the field, the supply of researchers and evaluators for the legal field and law enforcement in particular, is sufficient. Over the years new master courses have been established for the field of criminology, and according to some there actually is an oversupply in some countries (like the Netherlands). In most countries attention is also paid to police research and policy evaluations; the Global Policing Database http://whatworks.college.police.uk/About/News/Pages/Database.aspx gives interesting insights, numbers and examples. Slowly but steadily empirical evidence on cybercrime and Big Data is increasing too.
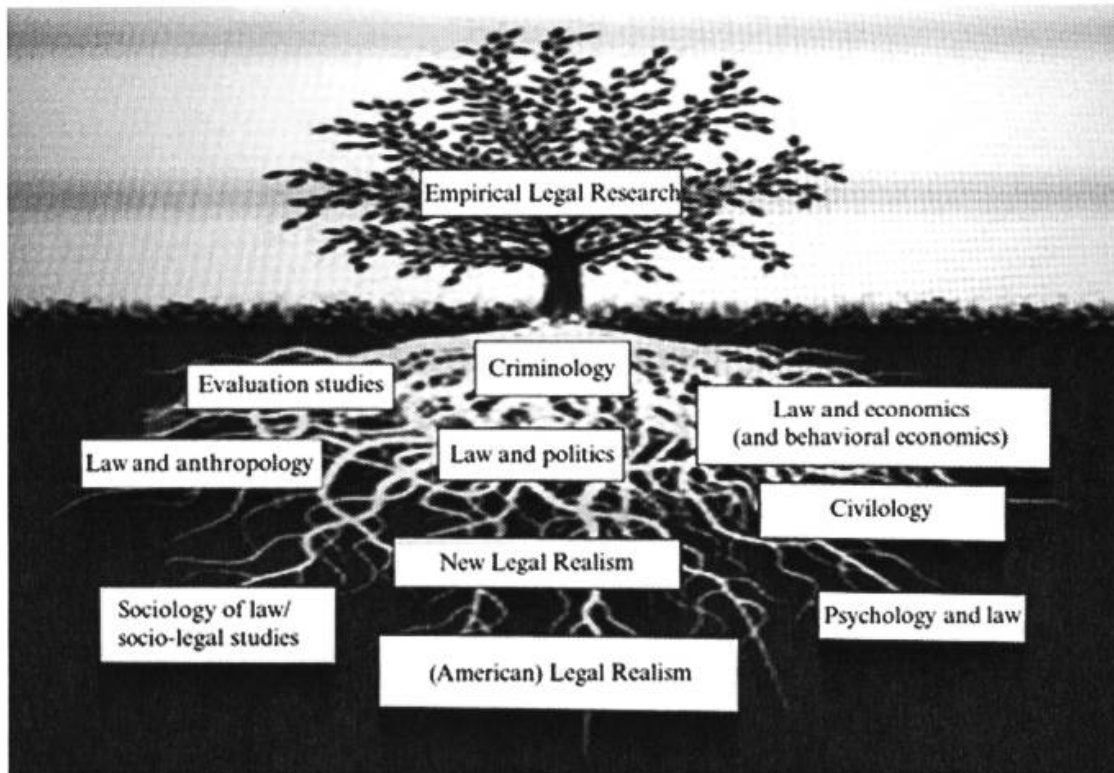
Figure 2.1   The roots of empirical legal research

What is also blossoming in many western countries is what can be called Empirical Legal Research: this field studies the very old idea from Roscoe Pound (1910, Harvard Law Review) that 'law in the books is not equal to law in action'. ELR combines different specialties like Law and economics, legal evaluations, civilology (private law and behavioral research), but also sociology of law and legal psychology.

Indicator 2:   Evaluations do not only look into inputs and throughputs of policies and programs, but also on outcomes, impact and effects.

Focusing on the Netherlands, the answer is largely positive. Next to quite a number of rather small evaluations done in the law enforcement field, also big ones are addressed. Just two examples:

(1) the reorganization of the police. Several years ago the police law was fundamentally changed, focusing now on a national instead of a largely regional police. The policy theory behind the reorganization addresses more police efficiency, better law enforcement results, more value

to the people, less complaints and many other aspects. A national committee has been installed to guide the evaluation which started in 2014 and will produce its final report in somewhere between 2017- 2019. The underlying theory of change has been reconstructed and is one of the central points of attention, longitudinal data collection is part of the study while at the same time attention is paid to more qualitative aspects like the police culture.

(2) evaluating the Counter Terrorism Policy: it is a study that looks into interventions, surveillance, mechanisms and impact of counter terrorism policy in the Netherlands. It is organized around 5 program theories that underlie this overarching policy.

To some of you all this might look like a paradise on earth for the evaluator-researcher. Well, that is not the case. First, there is the continuous problem of utilization of research and evaluation results. Still, there are not many superb results to be reported. Lum (2009; 2012) recently published several studies on this phenomenon for the police and was not at all positive; I did, for the Dutch policy the same and I too found a serious distance between street cops, management cops and policy cops on the one hand and evaluation on the other hand (Leeuw, 2015; Leeuw, 2016). Others studied how research influences judges and had to report findings like that 'prior orientation, peer knowledge, traditions are more important than results from research and evaluation'. One of my colleagues studied the extent to which ex-ante evaluations are referred back to when ex-post evaluations of regulations, laws, enforcement are carried out. In less than 10% this was happening (Klein Haarhuis, 2015).

The second issue is that sometimes inadequate research designs are used: designs that do not fit the research problem nor the type of knowledge that has to be produced. So, if you want to find out to what extent the graduated response policy on curbing digital piracy is working, designs that only post hoc ask people whether they have changed their downloading behavior, are totally inadequate for addressing this impact question. You are definitely all familiar with the classic Logan study (1971). Logan reviewed 100 studies of programmes designed to reduce criminal behaviour; 58 of these had designs with no control groups – that is, they featured only people receiving the programme, and not anyone receiving any alternative treatment or no treatment at all. The remainder had control groups, 23 with non-randomized control groups and 19 with randomized control groups. He showed that where there were no control groups the researchers were very likely to judge the programme as a

success, and that where there were randomized control groups they were much more likely to judge the programme as a failure.

Although it is better now, due to the MSM-Scale and many other activities, LQD's are still easy to be find. Low Quality Designs or studies that continue paradigmatic wars by saying that thinking in terms of designs, methods, explanatory theories and testing are 'positivistic old-school-studies'.

2   The Crime Drop

The 'crime drop' was first heralded in the United States where total violent crime fell over seventy percent between 1993 and 2011. The sharp drops in Canada's homicide and other crime has generally been remarkably similar to that in the US while New Zealand experienced rather similar declines in property crime (burglary, motor vehicle theft, and other theft) but not all personal crimes. Also in the mid-1990s, the UK began to experience serious declines across a wide range of property, personal and violent crimes, many falling by half or more (see Farrell et al, 2015; Clarke, 2015; Van Dijk et al, 2012).

For the Netherlands, WODC has data shown that there appears to be a crime drop, though not since the early 90s but starting in the early/ mid 2000s'. The conclusion is based on a combination of statistics/registers and counting the number of felonies experienced by samples of Dutch people annually in the Crime Victim Survey. A visualization of the crime

Total recorded crimes per 100k 1950 – 2014 NETHERLANDS
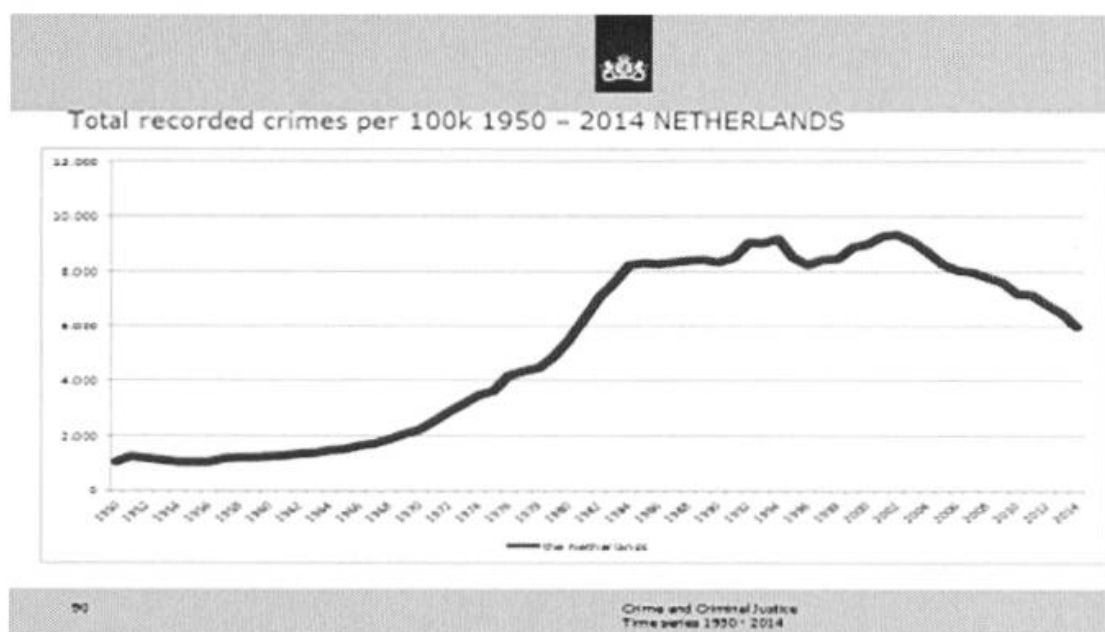
drop is available at WODC through sending me an email (f.leeuw @minvenj.nl). Based only on police data, this sheet also is illuminating.

What about a few other European countries? Paul Smits of WODC pooled data and made them as comparative as possible. For Switzerland I refer to Kilias and Lanfranconi (2012) and to recent data from the Swiss Federal Statistical Office. What can be concluded, is that it indeed looks like a crime drop. However, one has to be careful in yeah-yeah saying. Sir Karl Raimund Popper taught us that even 99 white swans do not imply that all swans are white. Moreover, crime is to some extent a 'dark number-phenomenon'. So, is there evidence that challenges the nowadays belief that we live in a Crime Drop world? Yes, there is to some extent.

A first challenge has to do with the issue whether or not, and to some extent, a measurement failure causes the crime drop. This possible failure has to do with the difficulties to measure cyber crimes in victim surveys and by the policy and the judiciary. One of the difficulties is that people do not always know if something is a cyber crime. Another is that valid data on cyber crime attacks on companies, institutions, organizations, that have passed the scrutiny of statistical offices and research organizations,

Total recorded crimes per 100k 1950 – 2014 OTHER EUROPEAN COUNTRIES
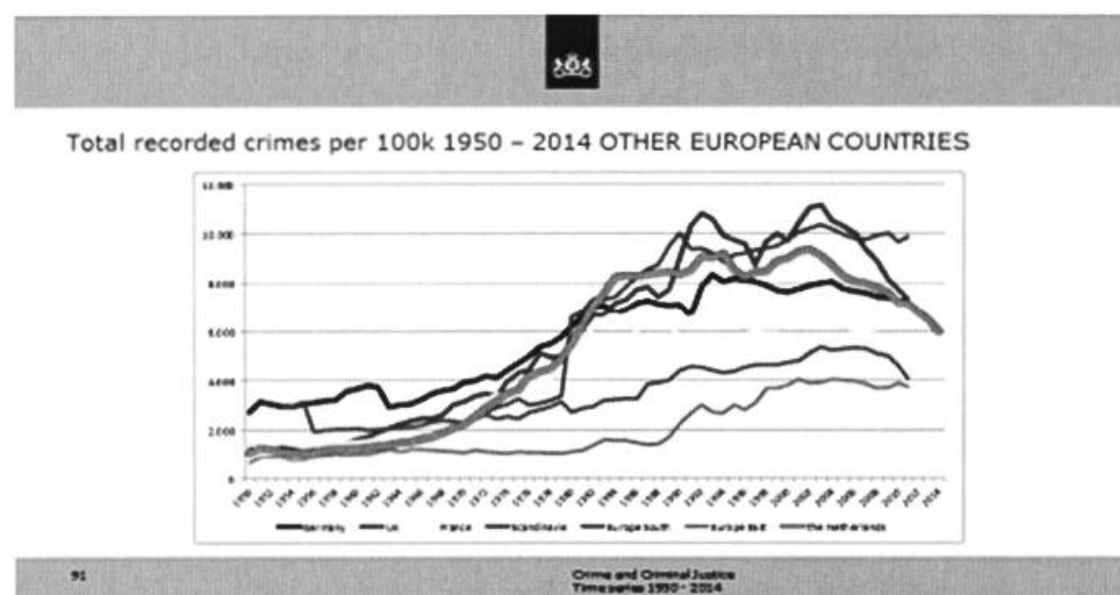
are very difficult to find. Of course ISP's and vendors of software like Avans or MacAfee will tell us that are trillions of attacks, but is it true what they say about daddy?

Now, the good news is that a recent study by the Office of National Statistics (ONS) in the UK showed that, with a new type of questioning in the victim survey for Wales and England, more cyber crimes came to the

212

fore than before. Instead of focusing on the method by which crimes occur during the interview, the crime itself was focal point but added to that (as part of the field test ONS carried out) was ONE question to record whether there was any cyber element to the offence. The findings were that "the field trial estimated there were 2.5 million incidents of crime falling under the Computer Misuse Act, the most common incident where the victim's computer or other internet enabled device was infected by a virus; it also included incidents where the respondent's email or social media accounts had been hacked" (http://www.ons.gov.uk/ons/rel/crime-stats/crime-statistics/year-ending-june-2015/sty-fraud.html).

In the most recently published Youth Crime Monitor in the Netherlands (2016), we measured that 3 out of 10 juveniles between 10 and 22 were, last year, engaged, in cyber crimes ranging from sexting, identity fraud, starting of DDos attacks (less than 2 %, though) and similar ones. To tackle the problem of cyber attacks on organisations and institutions, WODC together with the National Cyber Security Center is active in developing the cyber dashboard for institutions, companies, public sector organizations: how many times DDos attacks, ransom ware, and other cyber stuff they have encountered? So, how certain are we with regard to the crime drop? Not 100%, definitely not. It may be that depending on the way cyber crimes are measured, they may even partly counterbalance the drop in 'classic', brick & mortar crimes'.

There is also a second challenge, coming from the USA where a debate has started on the Ferguson effect, as it was coined by Mac Donald: the phenomenon of police officers backing off of proactive policing and thereby emboldering criminals. Her Wall Street Journal article from earlier 2015 starts as follows. "The nation's two-decades-long crime decline may be over. Gun violence in particular is spiraling upward in cities across America. A little later the Brennan Center for Justice published a report that indeed crime goes up, but, according to MacDonald, this statistic was hailed and 'explained away'(see Friedman et al, 2015). Although the empirical validity of this 'Ferguson effect' is discussed, and not undisputable, the debate raises issues on when a crime drop is coming to an end. As the crime drop in the USA was reported much earlier than in –for example—the Netherlands or Switzerland, this is also an interesting issue for policy makers in the (near?) future.

3.   Penal Law Change

In the Netherlands the Ministry of Security and Justice is involved in a rather fundamental process of change of the Penal Law Code. It started a couple of years ago and has to be finalized early 2017. Let me discuss only one aspect: the focus is on the penal law and law enforcement CHAIN. The current Penal Code prescribes things to do and not to do for different actors but is less clear when handling chain processes, interdependencies, legal logistics, rework etc. The first characteristic of the reform is that much more than in the existing Code attention is paid to implementation issues. Instead of assuming that when the law in the books, to quote Roscoe Pound is "okay", law in action will also be okay, the perspective is changed towards implementation aspects of the behavior of officials within and between the different chain actors. To handle that in an effective way and to produce high quality Legal Data on the execution, a so called "traffic control tower" for the penal law chain and its execution is developed. Such a 'tower' is a node taking care of preparing legal cases for the Public Prosecutor and the Judges and where, in a digital world, all information on when the court will convene, on the basis of which documents, including background information, and including checks on the completeness of the dossiers are carried out. The focus is higher efficiency, a better time management ('justice delayed is justice denied') and higher quality.

The complexity of this chain can be illustrated through the sheet on the next page. It describes, in Dutch, most of all actors and routes within the penal law chain. Such a 'traffic control tower' has to be organized in such a way that the huge amounts of data (on cases, persons, locations, times) are readily available in a digital space. WODC therefore has been working on realizing that through the development of dataspaces and datawarehouses, including the new interrelated research field: legal logistics (LL).

Legal Logistics can be defined as a framework to collect, analyse and integrate all relevant data to gain descriptive, predictive or comparative insight into the functioning of legal systems (Netten et al, in preparation). The focus is on:
– Judicial System Architecture
– Monitoring Tools
– Forecasting Tools
– Dissemination/Disclosure and "Open Data"

214

– Interplay of Man and Machine with(in) Legal Systems (e.g. exception management)
– Decision-support systems.

The relevance of this new 'legal engineering'-type of work is immanent from these three reasons:
– Lack of knowledge about efficiency of 'the Legal World' (How much rework is done when cases and persons are transported through the system?)
– Elapsed time problems
– Decision-making problems (how to handle huge amounts of data within even individual cases)

Of course, after working for some time with the new Chain philosophy regarding Penal Law, with its Traffic Control Tower and with Legal Logistics, the effects and possible side effects have to be evaluated.

References

Friedman, M., N. Fortier and J. Cullen (Brennan Center for Justice), Crime In 2015: A Preliminary Analysis. New York, 2015.

Chapin, F. Stuart, Experimental Design in Sociology: Limitations and Abuses, Social Forces, Vol. 29, No. 1 (Oct., 1950 – May, 1951), pp. 25-28.

Chapin, F. Stuart, Experimental Designs In Sociological Research, Harper & Brothers (1947).

Clarke, R.V. The Security Hypothesis and the Crime Drop, National Research Council Washington DC, May 18, 2015.

Farrell, G., G. Laycock and N. Tilley, Debuts and legacies: the crime drop and the role of adolescence-limited and persistent offending, in: Crime Sci (2015) 4:16.

Jacob, S., S. Speer and J. Furubo (2015), 'The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later', Evaluation, 21 (1), 6–31.

Katz, D. and M. Bommarito (2013), 'Measuring the Complexity of the Law: The United States Code' (1 August), http://ssrn.com/abstract= 2307352, accessed 4 July 2015.

Katz, D., M. Bommarito and J. Blackman (2014), 'Predicting the Behavior of the Supreme Court of the United States: A General Approach' (21 July), http://ssrn.com/abstract=2463244, accessed 2 July 2015.

Killias, M. and B. Lanfranconi, The Crime Drop Discourse – or the Illusion of Uniform Continental Trends: Switzerland as a Contrasting Case, in: Jan van Dijk et al (eds), The International Crime Drop. New Directions in Research 2012,p. pp 268-278.

Klein Haarhuis, C. (2015), Ex-ante onderzoek in metaperspectief. Aard, aantallen en gebruik van ex-ante analyses door de rijksoverheid, WODC, Den Haag.

Leeuw, Frans L. Lecture on utilization of police research and evaluations by the police, Police Academy, May 21, 2015.

Leeuw, Frans L. with Hans Schmeets, Empirical Legal Research: a guidance book for lawyers, legislators and regulators, EE Publishers, Cheltenham, 2016: chapter 9.

Logan, C. (1972), 'Evaluation research in crime and delinquency: a reappraisal', Journal of Criminal Law, Criminology and Policy Studies, 3, 378–87.

Lum, C. (2009), Translating police research into practice. Ideas in American Policing, Washington, DC: Police Foundation.

Lum, C., C. Telep, C. Koper and J. Grieco (2012), 'Receptivity to research in policing', Justice Research and Policy, 14 (1), 61–95.

Mac Donald, H. , The New Nationwide Crime Wave. The consequences of the 'Ferguson effect' are already appearing. The main victims of growing violence will be the inner-city poor, The Atlantic, May 29, 2015.

Rist, Ray et al (eds), The International Atlas of Evaluation, Transaction Publishers, Rutgers/London, 2002.