

Policy evaluation in the context of reflexive governance

Per Mickwitz

Finnish Environment Institute (SYKE)

February 5th 2008

ARVIOINTI TÄNÄÄN JA HUOMENNA

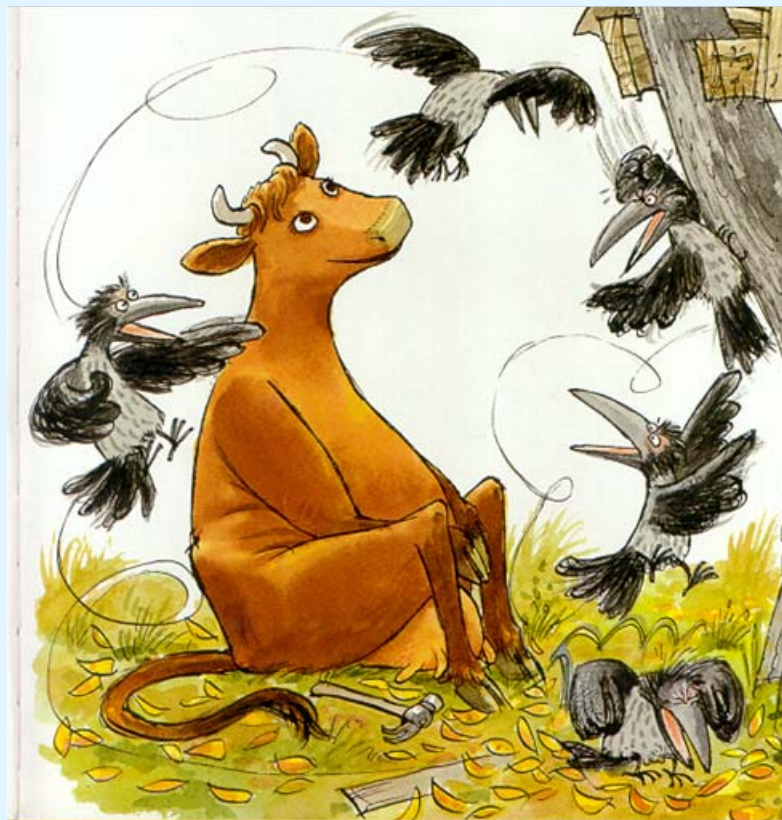
– Ammatillisen koulutuksen arvioinnin kehittämisseminaari

Structure of the presentation

➤ **Some key aspects of evaluation**

- **The changing context of evaluation**
 - **Globalisation**
 - **Governance**
- **Two evaluation trends**
 - **Effectiveness evaluation**
 - **Evaluations for Evidence Based Policy**
- **Evaluation as part of reflexive governance**

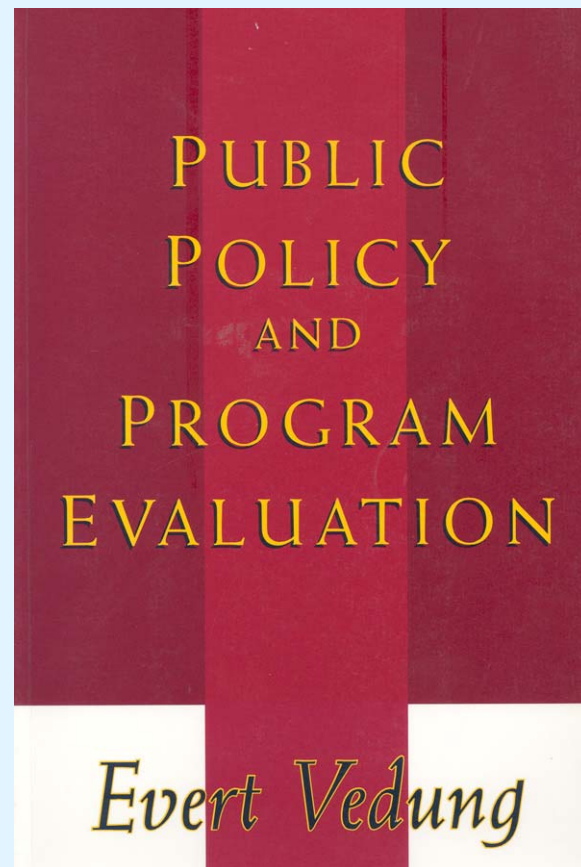
My perspectives



What do I mean by evaluation?

**Evert Vedung (1997)
defines evaluation as:**

*"~~careful retrospective~~
assessment of the merit,
worth and value of
administration, output and
outcome of government
interventions, which is
intended to play a role in
future, practical action
situations."*



What do I mean by evaluation?

**Evert Vedung (1997)
defines evaluation as:**

*"~~careful retrospective~~
assessment of the merit,
worth and value of
administration, output and
outcome of government
interventions, which is
intended to play a role in
future, practical action
situations."*

**The rationales for
evaluation:**

- Learning & Development
- Accountability
- [knowledge production]

Evaluation is normative!

Evert Vedung (1997) defines evaluation as:

"careful retrospective assessment of the merit, worth and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations."

Judgment has to be based some value criteria.

i.e. criteria for merit, worth and value.

The choice of criteria is normative.

What do I mean by evaluation?

**Evert Vedung (1997)
defines evaluation as:**

"careful retrospective assessment of the merit, worth and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations."

The focus of evaluations can vary:

- Programmes
- Policy instruments
- Institutions
- People, e.g. teachers
- Policy principles
- ...

What do I mean by evaluation?

**Evert Vedung (1997)
defines evaluation as:**

"careful retrospective assessment of the merit, worth and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations."

Evaluations are intended to be used.

Evaluation is part of policymaking and also political.

The changing context of public policy and thus evaluation.

- **Globalisation**
- **Governance**
- **The role of policy is different**
- **However, so is also the role of evaluation**
- **This is both the background for and the context of effectiveness evaluation & evidence based policy**

Usually when we think about globalisation we think about:

- **The closing of factories:**
 - Summa
 - Kemijärvi
- **Laying of employees**
 - Eriksson
 - Nokia Siemens Networks
 - ...
- **Imported goods from China and India**

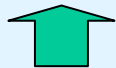
Increased connections between regions globally



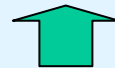
Global level



Policies
&
Regulation



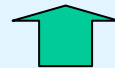
The
Economy



Technology



Information
&
Science



The
Environment



Local level

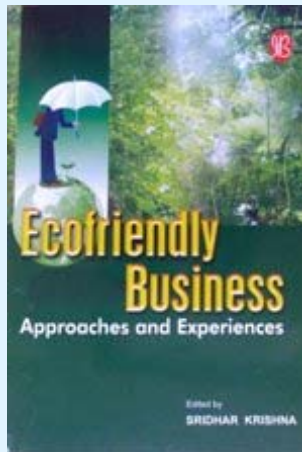


Has huge impacts on the substance of education, the means of education, the demands for education ...

Some personal reflexions on globalisation and education



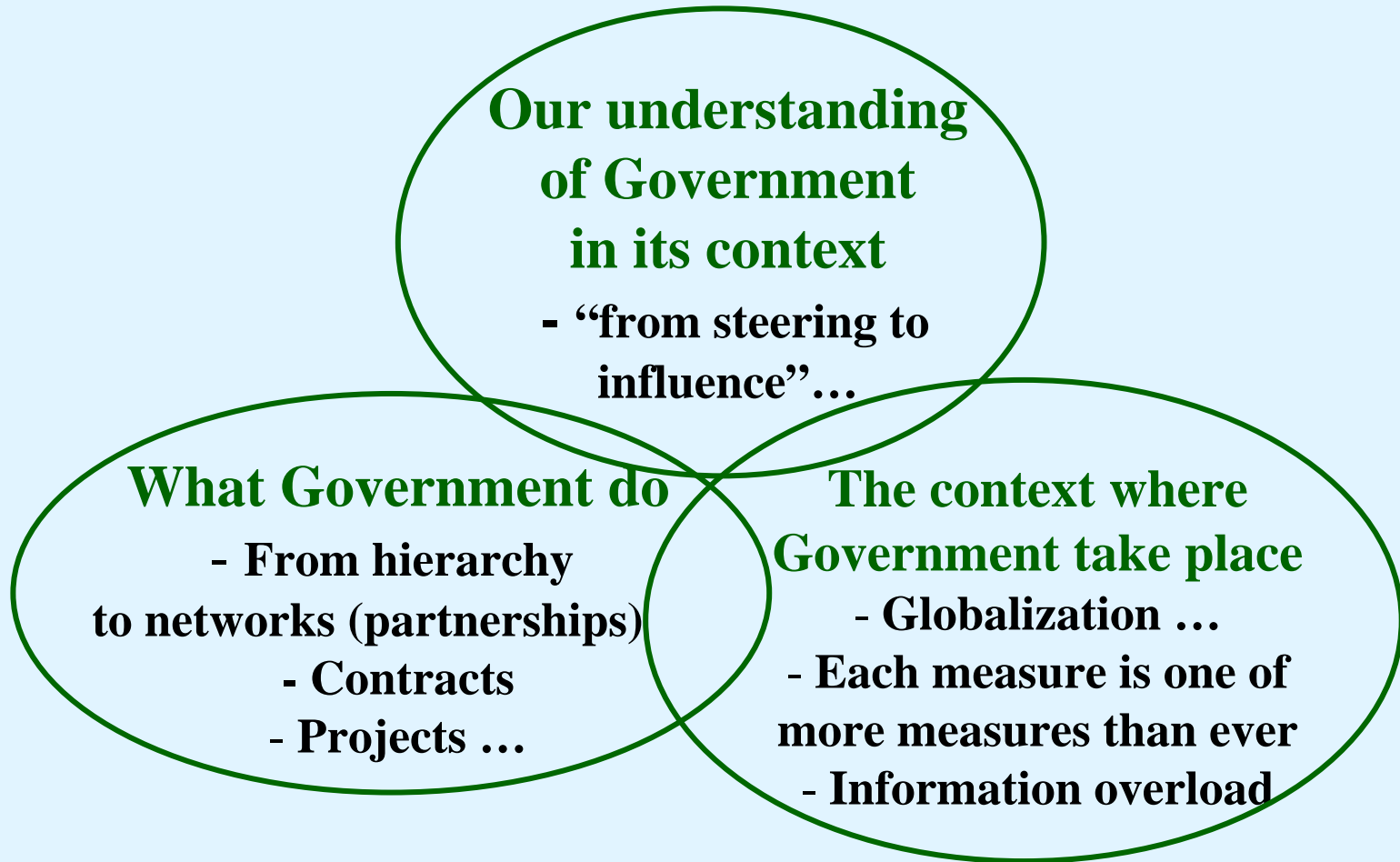
The International Institute for Industrial Environmental Economics
at Lund University



- For more and more positions increased amount of applicants with education abroad
 - Paradox: in environmental social science we have better university education in Finland than ever before, but e.g. SYKE more often employ people who have studied abroad, e.g. London, Wales, Edinburgh, St. Petersburg...

Evaluation within Governance

"From Government to Governance"

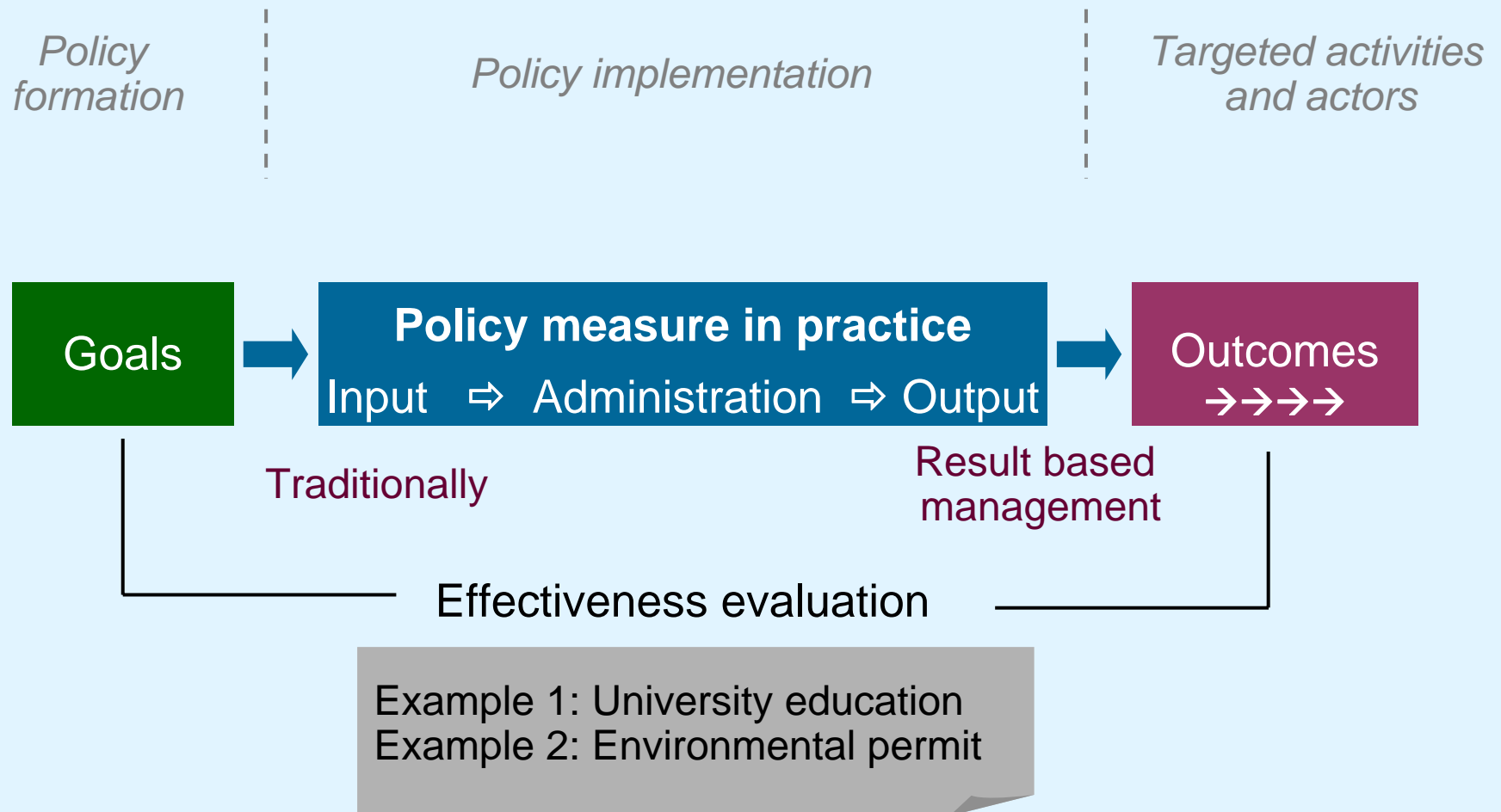


The Challenges of evaluation within governance in a globalising world



- **How to handle the dynamic relationships and more complex interactions?**
- **How to utilise the evaluations in networks**
 - Who should change and what?
- **The brevity of policies as well as targeted actions**
 - Projectification, i.e. project proliferation – long term effects

The simple idea of effectiveness evaluation



Basic challenges in effectiveness evaluation

- **Deciding the evaluand**
 - The effectiveness of what should be evaluated?
- **Determining the objectives**
- **Finding the effects of a policy**
- **Combining effects with objectives**

Finding the effects of a policy

Challenges involved

- **to find the relevant change in the target area**
and
- **to determine to which degree these effects are due to the policy and not caused by other factors, (how much are the Finnish pupils excellent reading results due to teaching, subtitles in television, characteristics of the Finnish language...).**

= **ATTRIBUTION**

Also the background to evidence-based policy



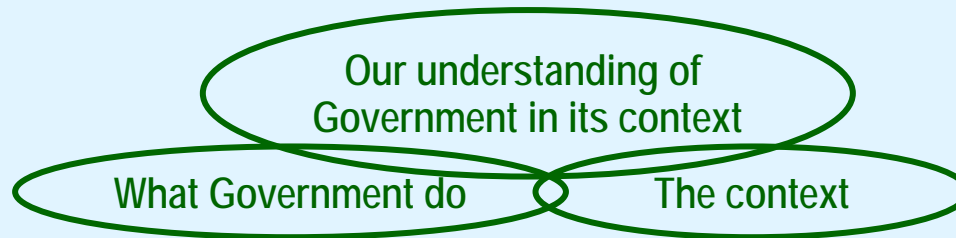
Addressing the impact problem

- **No single clear and universal solution**
- **Often several methods instead of just a single one.**
 - Include document analyses, statistical analyses and interviews.
 - Bartlett (1994, p. 183): *"Clearly desirable are multiple evaluation, done with a keen appreciation of the strengths and limitations of each approach and a frank recognition of the advantages of others."*

Often the term “triangulation” is used

- **Four types of triangulations can be distinguished:**
 - **multiple methods;**
 - **multiple sources within one method;**
 - **multiple analysts; and**
 - **multiple theories.**

Evaluation within governance in a globalising world



- **The increased interactions between policies, policy levels and actors make side-effects even more likely** → Evaluations can reveal these, but not if they are limited to effectiveness
- **The increased participation/co-operation in governance needs to be reflected in evaluation as well**
- **Evaluation, as an activity, may be “permanent” when evaluated measures and activities are constantly changing**
- **Evaluation may be a way to produced a new understanding of what government does in its new context.**

Evidence based policy – Background

- **The intent to base policy making on knowledge about what types of efforts work and with which means** (U.K. Cabinet office 1999)
- **Better use of evidence and research in policy making and better focus on policies that will deliver long term goals** (U.K. Cabinet office 1999)
- **Evidence based policy & effectiveness**
- **However, evidence take different forms and is valued differently in:**
 - The legal system
 - Natural science
 - Medicine
 - Humanities

Evidence-based policy in Europe

Sectors				
Health				
Social service				
Education				
Criminology				
Environment				

Based on Hansen & Rieper 2006

Evidence-based policy in Europe

Sectors				
Health	Cochrane Collaboration			
Social service	Campbell Collaboration			
Education				
Criminology				
Environment	Biodiversity e.g. ConservationEvidence.com			

Based on Hansen & Rieper 2006

Evidence-based policy in Europe

Countries	The U.K.	Denmark	Sweden	Finland
Sectors	Labour			
Health	Cochrane Collaboration			STAKES
Social service	Campbell Collaboration			STAKES
Education				
Criminology				
Environment	Biodiversity e.g. ConservationEvidence.com			

Based on Hansen & Rieper 2006

The U.S. discussion about evidence-based policy

- **The hottest issue in the U.S. debate about evaluation and research for policy**
- **The issue is discussed in every policy area, but especially in the context of education**
- **U.S. Department of Education position paper 2003**
 - **“proposed priority” to experimental design studies**
 - **Best for determining effectiveness**
 - **Only ED studies will be used to determine project funding**



What Works Clearinghouse

The “supreme court” of credible evidence



Policies based on systematic review & a “hierarchy of knowledge”

- Randomized controlled trials
- Quasi-experiments studies
- Before and after comparisons
- Cross-sectional, random sample studies
- Process evaluation, formative studies and action research
- Qualitative case studies and ethnographic research
- Descriptive guides and examples of good practice
- Professional and expert opinion
- User opinion (Pawson 2006)

Key phases of systematic review, i.e. evaluations for evidence-based policy, and some comments

- 1. Formulation of review question**
 - 2. Systematic search**
 - 3. Criteria for exclusion – inclusion (hierarchy of knowledge)**
 - 4. Summary of results**
- **Most studies show “no evidence or insufficient evidence” that interventions have effects.**
 - **Few studies of use of systematic review, but indications of barriers to use and actual use of alternative sources of information.**
 - **Evidence in law is often a better model for evidence for policy than evidence in medicine**

Evidence in law, according to Wikipedia and some interviews

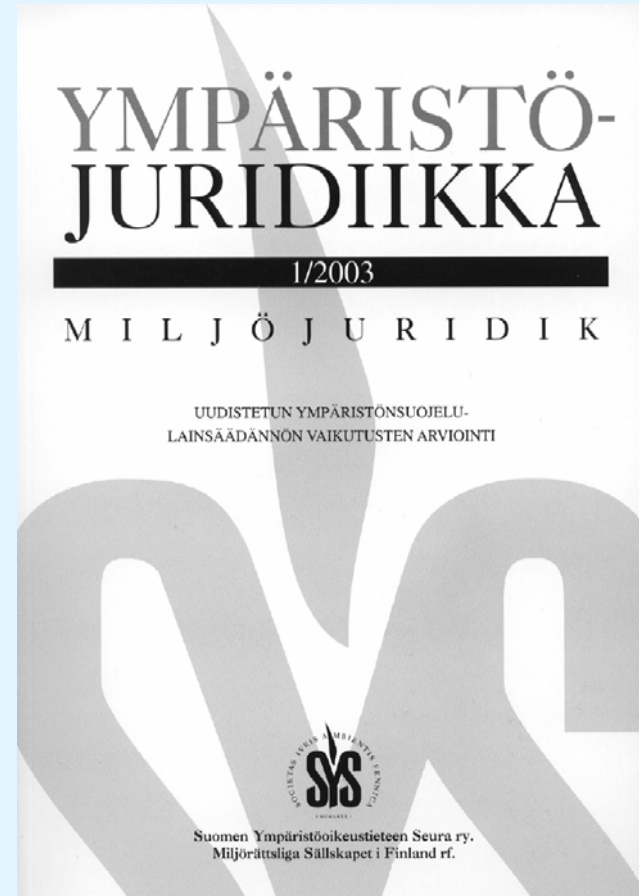
- *"The law of evidence governs the use of **testimony** (e.g. oral or written statements, such as an affidavit) and **exhibits** (e.g. physical objects) **or other documentary material** which is admissible (i.e., allowed to be considered by the trier of fact, such as jury) in a judicial or administrative proceeding (e.g. a court of law)."*
 - Relevance & Authentication
- The "Hierarchy" of the evidence depends on the case
- Burden of proof depends on type of case (criminal – civil law)
 - *"Beyond reasonable doubt" vs. "preponderance of the evidence"*
- *All the evidence is judged together, i.e. many different sources that would not be enough proof alone may jointly be considered enough – **triangulation in evaluation**.*
- *Narrative synthesis, very different from Cochrane Collaboration.*

What are the practical situations where the evidence is intended to play a role?

- ***“Intended use by intended users”***. Patton (1997)
- ⇒ **What can be considered evidence depends on by whom the evaluation is intended to be used and in what context.**
- **What can be considered evidence, or what degree of evidence varies from the EU to the local level.**
- **What evidence evaluations should produce depends on the window of opportunity for change and ongoing policy discussions.**

Example: the evaluation of the Finnish Environmental Protection Act

- **9 peer-reviewed articles in a Finnish law review**
- **Credibility as evidence among Finnish lawyers**
- **Much interest among those involved in implementation of the act**
- **The EU level has been a problem**
- **Manuscripts + additional report in advance to committee**



Take home message

- **A fixed “hierarchy of knowledge” approach to evidence is too narrow.**
- **The relevance of different types of evidence depends on the practical situations where it is intended to be used.**

The U.S. discussion about evidence-based policy

- **The hottest issue in the U.S. debate about evaluation and research for policy**
- **The issue is discussed in every policy area, but especially in the context of education**
- **U.S. Department of Education position paper 2003**
 - **“proposed priority” to experimental design studies**
 - **Best for determining effectiveness**
 - **Only ED studies will be used to determine project funding**



Intense debate, statements, panels, articles, books ...

- **AEA statement – NOT AEA statement**
- **Donaldson & Christie 2005. The 2004 Claremont Debate: Lipsey vs. Scriven - Determining Causality in Program Evaluation and Applied Research: Should Experimental Evidence Be the Gold Standard? *Journal of MultiDisciplinary Evaluation*, 3, 60-77.**
- **Informing Federal Policies on Evaluation Methodology: Building the Evidence Base for Method Choice in Government Sponsored Evaluation, *New Directions for Evaluation*, 113, Julnes & Rog (eds). 2007.**
- **Book: Donaldson, S.I., Christie, C.A. & Mark, M. (forthcoming). *What Counts as Credible Evidence in Evaluation and Evidence-based Practice?* Thousand Oaks, CA: Sage.**

AEA Statement vs. NOT AEA statement

AEA: “we believe the **proposed priority manifests fundamental misunderstandings** about
(1) the types of studies capable of determining causality,
(2) the methods capable of achieving scientific rigor, and
(3) the types of studies that support policy and program decisions.
We would like to help avoid the **political, ethical, and financial disaster** that could well attend implementation of the proposed priority.”

Not AEA: “This interpretation of what constitutes scientifically based evaluation strategies for assessing program effects **is consistent with the presentations in the major textbooks in evaluation** and with widely recognized methodological standards in the social and medical sciences.

Randomized controlled trials have been **essential to understanding what works, what does not work**, and what is harmful among interventions in many other areas of public policy including health and medicine, mental health, criminal justice, employment, and welfare.

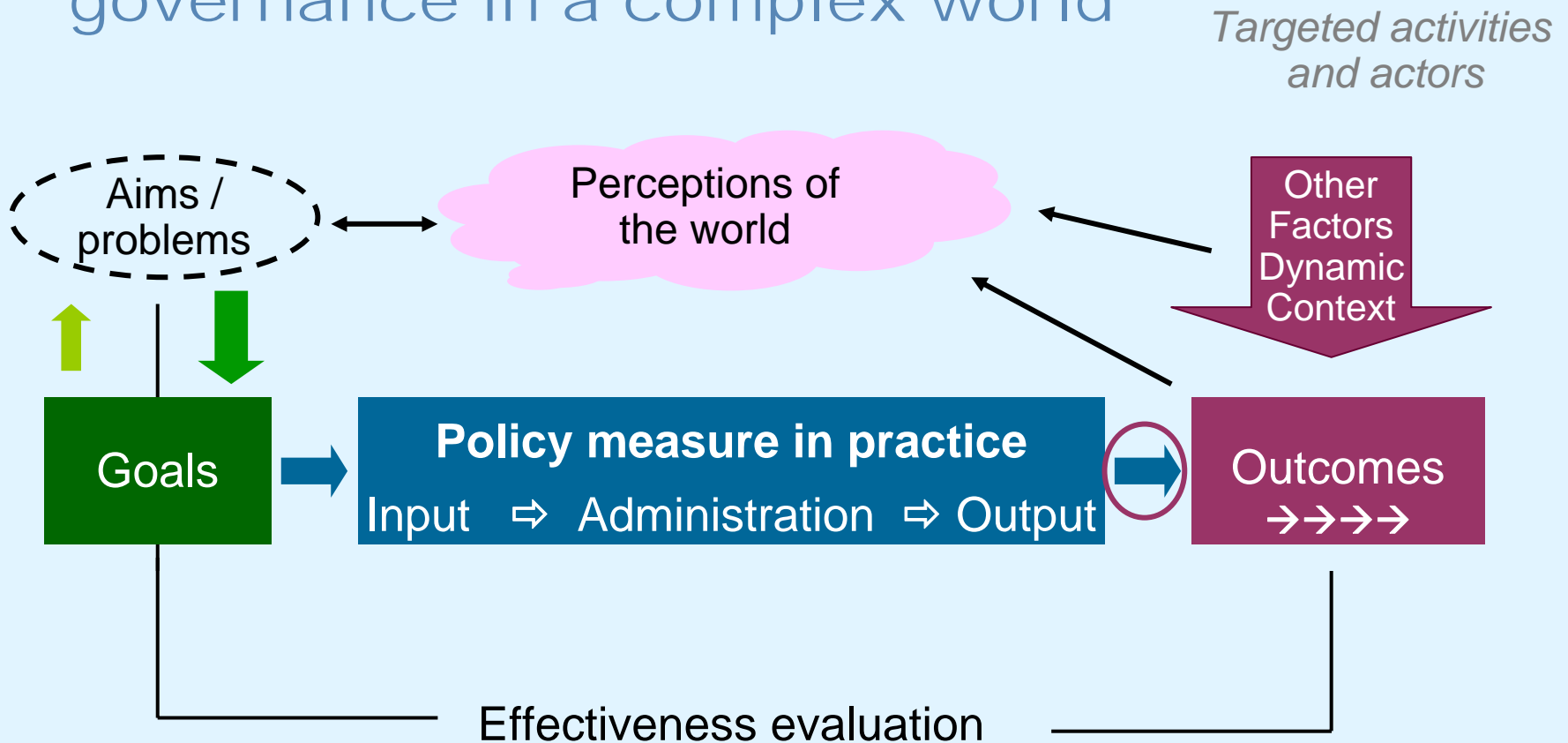
Furthermore, attempts to draw **conclusions about intervention effects based on nonrandomized trials have often led to misleading results** in these fields and there is no reason to expect this to be untrue in the social and education fields.”

Eleanor Chelimsky (2007)

with 14 years experience from the U.S.
Government Accountability Office (GAO)

- “All evaluation questions, however, are not cause-and-effect questions.”
- “experiments can be a good way to address certain questions under certain circumstances, it is important that agencies try to conduct at least some randomized studies of their programs.”
- “an a priori judgment about methods without a serious study of the context and specifics of a question is both unsuitable and imprudent in relation to likely evaluation success.”

Evaluation as part of reflexive governance in a complex world



Targeted activities and actors

Effectiveness evaluation not as goal achievement, but as evaluation of the **ability to address underlying aims**.
Learning & Development → Aims & Goals
Accountability → Goals & Aims

Analogy to Side-effect evaluation

- The “side-effects evaluation” model **extends the scope of effects, but maintains the special role of anticipated effects within the target area.**
- Similarly, effectiveness evaluation as ability to address underlying aims should **broaden the view of aims (in time as well as among actors), but maintain a special role for goals decided by elected bodies or administrations accountable to them.**

“if evaluation is abandoned, how will we be able to check whether public programmes are working?

Just by causal impressions?

Or by the statements of politicians, most/all of whom have some axe to grind?

Or by the claims of senior managers, who will usually bend over backwards to put the best possible face on whatever happened (defending their own institution is, after all, part of their job).” (Christopher Pollitt 2003)

Remember, evaluation is not to find the final truth, it is *careful normative assessment, which is intended to play a role in future, practical action situations.*