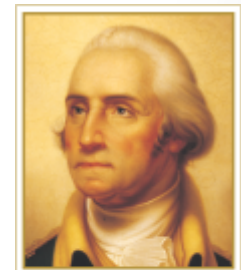


# Building-Blocks for the Knowledge Base: Recent U.S. Developments

**Michael Wiseman**

Ex Ante Evaluations - Preparing for EU 2014-2020  
4th International Evaluation Conference  
Budapest, 26-27 September, 2013



THE GEORGE  
WASHINGTON  
UNIVERSITY  
WASHINGTON, D.C.

# Background

- My presentation looks outside the EU context; I hope it will provide “thoughtful policy tourism”
- Ex Ante evaluation is large-scale activity; this talk is smaller-scale
- Prompted by issues raised at the Vilnius conference and other meetings:
  - Lack of attention to improving evidence base
  - The interpretation of “counterfactual” to include alternative strategies
- I will outline some small-scale RCTs

# Thinking about Building-Blocks

- Thinking small begins with “deconstruction”. Government programs are constructed from many elements, or building blocks.
- Focus of ex ante evaluation is often upon all parts of an initiative, including the logic of the intervention, the plan for process analysis, and the strategy for impact assessment
- Assessing the impact of the whole is often difficult, and the results are often at best suggestive of what might happen should a similar program be undertaken in another environment

- It is often possible to imbed within the general evaluation a more rigorous examination of alternatives for individual components—the building blocks
- Attention to blocks seems to be growing, for several reasons:
  - Uncertainty about external validity of results of large-scale experiments, even when rigorously evaluated
  - Emphasis on “Evidence-Based” policy
  - Growing appreciation of role of process analysis (and hence focus on inter-connection of program parts) in evaluation
  - Sense of greater external utility

- Literature from business, the UK, and elsewhere citing gains from experimentation with constituent elements of programs:
  - UK Cabinet Office *Test, Learn, Adapt: Developing Public Policy with Randomized Controlled Trials*
  - Jim Manzi, *Uncontrolled: The Surprising Pay-Off to Trial-and-Error for Business, Politics, and Society*
  - Sendhil Mullainathan and Eldar Shafir: *Scarcity: Why Having Too Little Means So Much*
- Each includes examples of small changes in program strategy or operation shown by rigorous (i.e. RCT) evaluation to have significant consequences for costs and impact

# Some Examples

- Turn now to examples of smaller-scale RCTs
- Overview is of necessity cursory
- Full citations for all work cited here appear at the end.

The first experiment involves income tax collection in Minnesota

# Example 1: Minnesota Tax Experiment

- General Issue: How does the state improve compliance with income tax?
- Traditional strategy was to conduct random, surprise audits, levy penalties
- Experiment involved *four* options, plus control:
  - 1) Increased examination and audit rate, with prior notice
  - 2) Enhanced customer service
  - 3) Redesigned tax form
  - 4) Letters reporting high compliance rate revealed by existing audits
  - 5) Control: Existing policy
- Letters motivated by survey evidence counteracted common perception of significant

# Outcomes

- Taxpayers randomly assigned; experiment involved only households without composition change over two tax years
- Impact assessed by comparing year-over-year reported taxable income for filing units
- Strategy (1) produced substantial gains, suggesting that incentive created by random audits left income unreported
- Strategies (2) and (3) had no significant impact
- Strategy (4) letter citing audit report produced surprising effect



# Replication in second year verified the strategy (4) effect

	Sample Size	Median increase in reported income, federal tax return	Median increase in state taxes collected
Treatment	8,537	\$1,411	\$86
Control	8,550	\$1,206	\$71
Significance Level		(.008)	(.0045)

*Source: Coleman (2007)*

\$15 (\$22.50 in 2013\$) gain in tax collection may seem small, but letters cost only about \$1 each.

# Important features

- Overall program—tax collection—is not being evaluated.; this study is about components—building blocks—of the collection strategy
- Utilized administrative data
- Random assignment enhances credibility of estimated impacts
- Done within the agency by responsible managers

## Example 2: Danish Childcare Centers

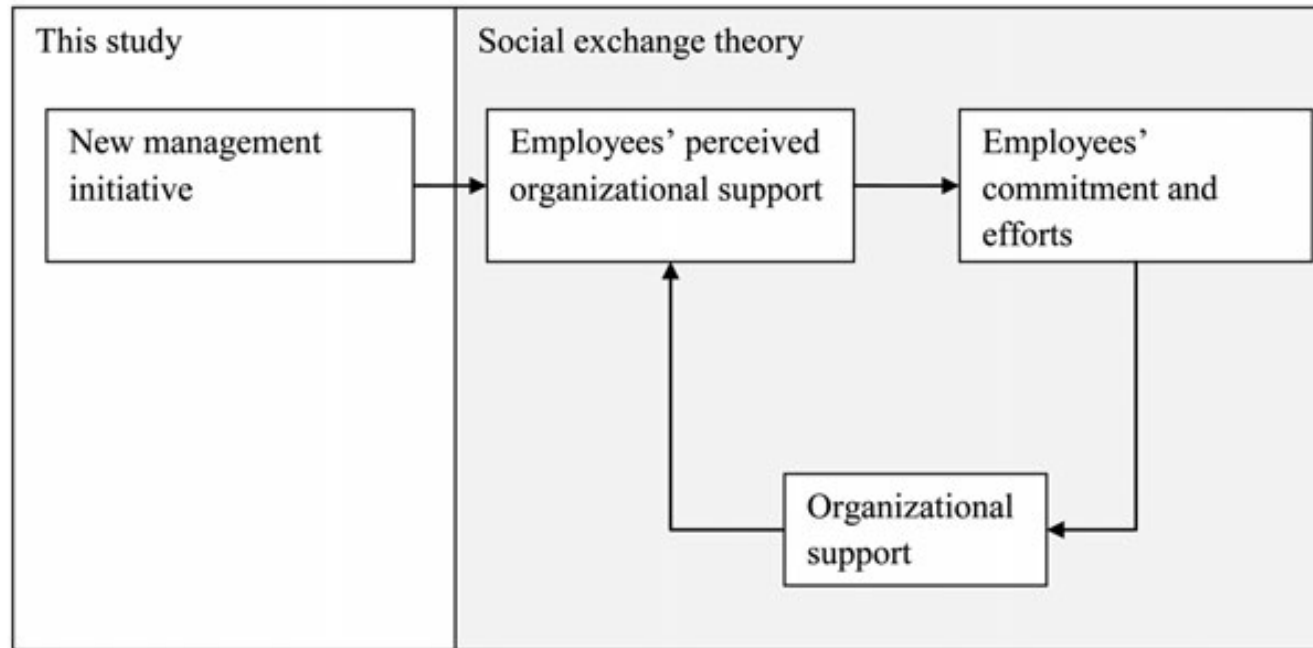
- Setting is bilingual childcare centers in Aarhus
- There are many, and in services such as this it is believed that much depends on the attitudes of and resources provided to center managers
- Theory is that employees' commitment and effort is a function of sense of administrative support
- Perception of Support (POS) refers to employees' beliefs about "the extent to which the organization values their contributions, cares about their well-being, and aids them so they can carry out their jobs effectively".
- Question: Can POS be enhanced?

# The Experiment

- Two treatments, plus control.
- Treatment one: Center manager offered a special language consultant whose function it was to visit the child-care centers individually and offer “decentralized, professional, and appreciatory support” to the employees. Both mode and extent of utilization left to managers. NO reciprocation expected
- Treatment two: “Each bilingual child in these centers received a suitcase containing books, games, and a video tutorial.”
- Both funded by central administration

# Exogenous change in resources

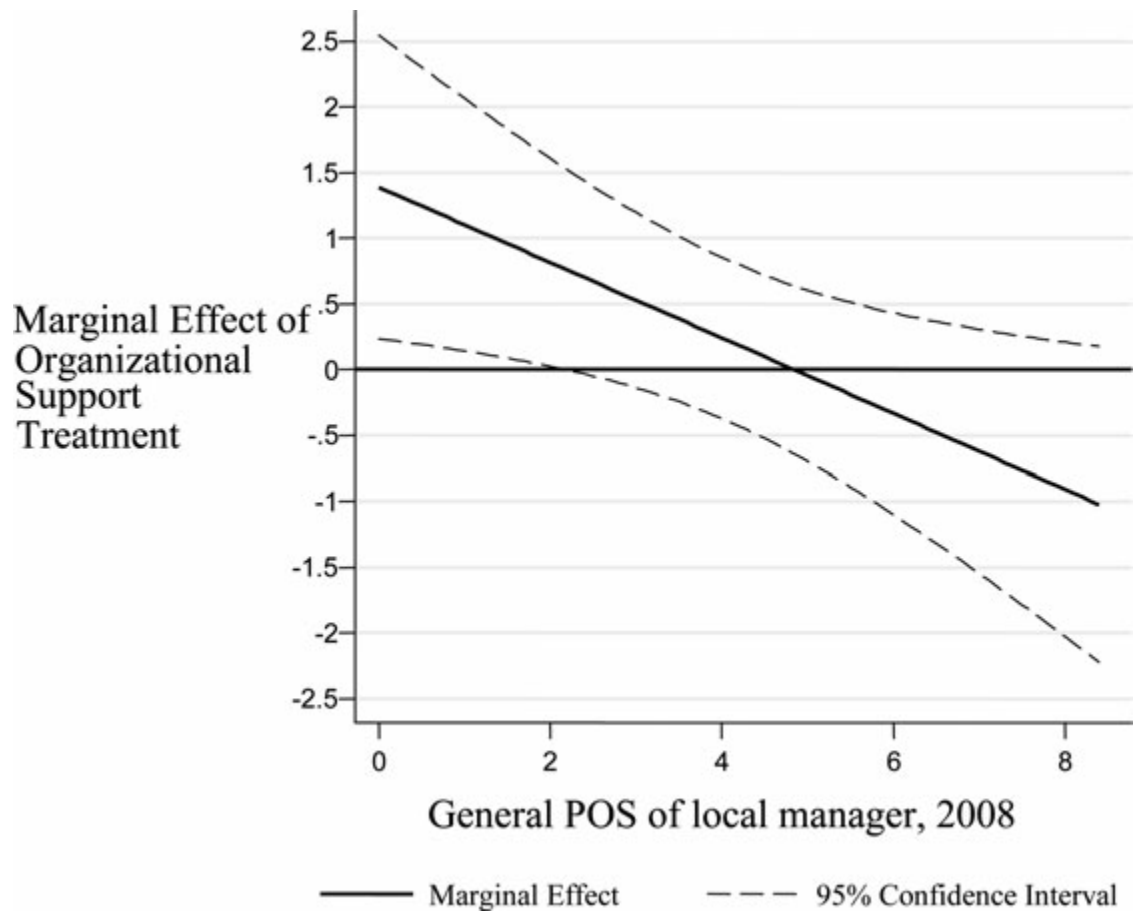
## The Organization–Employee Exchange Relation and New Management Initiatives



# Outcomes

- Cluster random assignment: Treatments are allocated to centers
- Impacts on POS measured by questionnaires administered before, after the intervention.
- NO average effect on POS measure of either intervention
- However, low POS managers showed gain.

# Treatment effect and initial manager POS



# Important Features

- Aimed at management/agency issues
- Comparative evaluation of two approaches to working with immigrant children
- Surely POS is a factor in service outcomes, but it is an intermediate outcome
- Example of collaborative work between government, academic investigators
- Ultimate objective is assessment of impact on children



## Example 3: The Cost of Delay

The last example involves administration of disability benefits.

- Like others, the US disability assistance system starts with disability assessment
- Central feature of process is determination of capacity for “substantial gainful employment”
- Done by examiners, and examiners have backlogs, so obtaining disability benefit can take a long time.
- Benefit payments are retroactive to date of application

# The Costs of Delay, continued

- An Important policy and management question concerns the consequences of delays in eligibility determination for the likelihood of return to employment.
- Candidates for disability benefit are randomly assigned to examiners
- Researchers employed variation in evaluator efficiency as instrument for time between application and determination.

# Outcome

- Delay not correlated with likelihood of certification
- Impact substantial regardless of outcome—employment likelihood inversely related to time to determination for both those certified, and those not
- A one standard deviation (2.4 month) increase in initial processing time reduces annual employment rates by 1 percentage point (3.2%) in years two, three and four post-decision.
- Implies the disability determination process as currently conducted directly reduces the post-application employment of denied applicants by approximately 3.6 percentage points (7%) and certified applicants by approximately 5.2 percentage points (33%).

# The Cost of Delay, continued

- Interesting because it's an “accidental” RCT
- Certification is one block in the disability assistance program
- Supports case for more diligent management oversight
- Product of interaction between Social Security Administration policy staff and external consultants
- In contrast to the other two examples, this involves “big data”

# The idea of a prize

- Development project under way In cooperation with the White House Office of Science and Technology policy
- Focus on plans, not outcomes, to get early response
- Steps involve
  - Letter of interest
  - Invitation to compete
  - Delivery of proposal
  - Awarding of prizes
  - Technical support, where necessary
  - Assistance in funding

# Evaluation Criteria

The criteria for prizes are being developed in consultation with administrators.

Example standards:

- Are project outcomes potentially useful beyond the experimenting government?
- Is the theory (logic model) sound?
- Will the proposed evaluation scheme reliably reveal both process and impact achieved?

# Blocks-work and ex-ante planning

- A case can be made that only work on components of policy can be portable and hence part of the evidence base
- Such experiments need to be considered in planning
- Goal is not only to add to the evidence base but also to build expertise and appreciation of benefits of rigorous evaluation

# Additional Resources

- Warsaw and Vilnius presentations, as well as links to evaluation documents can be found at: [MichaelWiseman.com](http://MichaelWiseman.com)
- Email: [WisemanM@gwu.edu](mailto:WisemanM@gwu.edu)





# References

- Autor, David, Nicole Maestas, Kathleen J. Mullen and Alexander Strand. *Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants*. Michigan Retirement Research Center Research Paper, No. WP 2011-258 (2011).
- Coleman, Stephen. *The Minnesota Income Tax Compliance Experiment State Tax Results*. Minnesota Department of Revenue, April 1996. <http://ssrn.com/abstract=1585242>
- Coleman, Stephen. *The Minnesota Income Tax Compliance Experiment: Replication of the Social Norms Experiment*. Minnesota Department of Revenue, 1 November 2007.
- Haynes, Laura, Owain Service, Ben Goldacre and David Torgerson. *Test, Learn, Adapt: Developing Public Policy with Randomized Controlled Trials*. U.K. Behavioural Insights Team, June 2012. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/62529/TLA-1906126.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf)
- Jakobsen, Morten and Simon Calmar Andersen. *Intensifying Social Exchange Relationships in Public Organizations: Evidence from a Randomized Field Experiment*. *Journal of Policy Analysis and Management*, Vol. 32, No. 1, 60–82 (2013).
- Manzi, Jim. *Uncontrolled: The Surprising Pay-Off to Trial-and-Error for Business, Politics, and Society*. New York: Basic Books, 2012.
- Mullainathan, Sendhil, and Eldar Shafir. *Scarcity: Why Having Too Little Means So Much*. New York: Times Books, 2013.