Performance Management: Economics and Policy

Jeffrey Smith
University of Michigan
econjeff@umich.edu

Institute for Research on Poverty Seminar
Madison
February 23, 2012
Introduction

Thanks for the invite!

Always good to be in Madison

Focus on education, relate to active labor market programs but very broadly applicable due to GPRA (ex: NSF)
What are performance measures?

Performance measures typically use short-term outcomes to proxy for long-term impacts

Ex: test scores in grade 11

Ex: fraction of Workforce Investment Act (WIA) participants employed 13 weeks after leaving the program
Literary antecedents

James Q. Wilson’s (1989) *Bureaucracy*

Personnel Economics e.g. Lazear and Gibbs (2009) *Personnel Economics in Practice*

Heckman, Heinrich, Courty, Marschke and Smith, eds. (2011) *The Performance of Performance Standards*


Is the education literature re-learning things we already knew from the Active Labor Market Policy (ALMP) literature?
Big picture

There is no perfect way to organize the education industry (or the job training industry)

Key questions:

How large a role should performance measures play?

What performance measures to use?

What other institutions complement performance measures?

What to do about strategic behavior?
Features of the education industry
Principal agent problems (parent-school, parent-government)
Information asymmetries
Poorly understood production function
Long-delayed outcomes
Hard to infer impacts
Bottom line: This is not the market for pencils
ALMPs have some of these issues as well
Principal-agent problem

Uses of performance measures

1. Preference alignment
2. Effort elicitation
3. Systemic learning ("working smart") elicitation
4. Local knowledge elicitation (the Hayekian argument)
5. Personnel management - related to effort elicitation
6. Inform customer choice

US: 5 is most important (rewarding and punishing teachers)
UK: 6 is most important (school league tables in quasi-market)
Performance management without flexibility

How much control do the performers whose performance is being managed (e.g. principals, teachers or districts) have over …

Hiring teachers

Firing teachers (see recent media attention in NYC and LA)

Curriculum

Teaching methods

Ex: the unemployment rate as a performance measure for ALMP
Performance management without knowledge: curriculum

It is stunning how little we know about a lot of basic questions

US: a good start with IES and the What Works Clearinghouse

Aside: Evaluating “Teaching American History”

Aside: Literature survey on teaching history

Reduced form: abject failure of an entire discipline to produce useful output
Performance management without knowledge: teachers

Teachers matter! See e.g. Rivkin, Hanushek and Kain (2005)

Other than experience $< 3$, we know very little about how to measure the quality of individual teachers ex ante

Principals do know a little bit, so can fire ex post (if the institutions allow them to do so, as they usually do not)

This is a task (in part) for the educational psychologists

Ongoing measurement projects in the US
Performance measures: heterogeneity

Different dimensions: social, athletic, cultural, religious, political

Where are these other dimensions in the league tables?

Academics: is there but a single factor, quality, in the education production function?

Put differently, is there just one quality dimension as suggested by existing league tables and the statistical models that underlie them?
Performance measures: heterogeneity (continued)

What about complementarities between academics and other factors which vary among students? See Smith (2008).

What about heterogeneous academic styles, each of which is best for some students? Ex: controlled versus open environment, visual versus textual, etc. See e.g. Dearden, Micklewright and Vignoles (2011) *Fiscal Studies* on education and recent work on QTEs.

Link back to systemic knowledge generation and use of local knowledge (and to the ALMP context)

Performance measures: strategic responses

Clearly more of an obsession (among researchers) in the US than in the UK (but lots of UK evidence from the NHS)

Education: cheating by teachers, gaming of special education status, gaming of attendance, carb loading on exam day, states mucking around with their assessments and proficiency thresholds; see David Figlio’s talk.

ALMP: enrollment (e.g. cream-skimming), enrollment timing, termination timing, outcome measurement etc.

Responses to non-linearities in reward functions

Want agents to have just the right kind of flexibility
Performance measures: statistics

Outcome levels versus impacts (i.e. value added)

Bad examples: US Job Training Partnership Act (JTPA), Workforce Investment Act (WIA) and No Child Left Behind (NCLB), Job Corps


Better example: UK contextual value-added measures

Do the assumptions of underlying value-added models hold empirically? See e.g. Kane and Staiger (2008), Koedel and Betts (2009), Rothstein (2010)
Performance measures: curse of micronumerosity

The sample size issue is fundamental

Classrooms are not that large in most developed countries
Most extant systems test once and in only a few grades

Ex: Kane and Staiger (2002)

Ex: Leckie and Goldstein (2009, 2011)
Performance measures: consumption by parents

Know a bit about how parents use performance information but could know a lot more

Lottery papers and MTO

Giustinelli and survey measures of household bargaining processes

Transparency and understanding, standard errors, trust in experts

Example: JTPA and regression adjustment (the Westat study)
Example: What newspapers report in the UK
Complements to performance management

Performance management clearly cannot do it all

Teacher training / professional norms / intrinsic motivation
  Example: drill team
  Example: tenured professors and teaching

Choice (performance measures aimed at consumers)

Centralized knowledge creation
Bottom lines

This is hard!

What exactly are current performance schemes trying to accomplish?

Is effort a relevant margin in developed countries? If not, what is?

The perfect and the good: friends or enemies (e.g. the UK reaction)

Lots yet to learn on many dimensions of the problem