THE BUREAUCRACY LAB
A Joint Initiative of the Governance Global Practice
and the Development Impact Evaluation/Economics Research Group

HOW CAN WE MEASURE PUBLIC-SECTOR PRODUCTIVITY?
Our Mission

• To innovate bureaucracy through better data and empirics.
What is productivity?

The ratio of outputs ($) to inputs ($)

How efficiently are inputs (labor, capital, and raw materials) being used to produce the final good?
What are common measures of productivity in the private sector?

- Output per worker (number of shoes produced per employee)
- Output per expenditure on labor (number of shoes produced per dollar spent on labor)
- Output per expenditure on all inputs (number of shoes produced per dollar spent on labor, machines, and materials)
Administration-to-administration transactions (e.g. central government)

Task completion rates and processing times
- Rasul and Rogger (2018); Lewis-Faupel al (2016); Fenizia (2019)
  - Project completion rates (adjusted for complexity); road construction completion rates (quality-adjusted); process times (response times)

Individual-level indicators (employee performance)
- Ashraf et al (2014; 2016); Khan et al (2019); Deserranno (2019); Callen et al (2018)
  - Tax collection; performance evaluations; attendance

Subjective assessments of performance
- Bertrand et al (2018); Alonso & Lewis (2001); Brewer & Selden (2000)
  - Subjective assessments by multiple stakeholders; Employees’ subjective assessments of org. performance

Staff satisfaction and tenure
- Deserranno (2019); Janke et al (2019); Bright (2008)
  - Tenure/retention; Staff satisfaction; Job satisfaction and turnover intentions

Measures of corruption
- Hanna & Wang (2017); Meyer-Sahling et al (2016); Olken (2007); Ferraz
  - Lab game; Perceptions of other public officials; Stated expenditure v audits
<table>
<thead>
<tr>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget execution</td>
<td>Liebman and Mahoney (2017); European Commission</td>
</tr>
<tr>
<td></td>
<td>The rate of expenditure over the budget year (and the quality of expenditure)</td>
</tr>
<tr>
<td>Mystery shoppers (comparing to a hypothetical/ideal benchmark)</td>
<td>Das and Hammer (2007); Bertrand et al (2007)</td>
</tr>
<tr>
<td></td>
<td>Compare ‘ideal’ process to actual process experienced by independent users (e.g. comparison against checklist; subjective measures of experience)</td>
</tr>
<tr>
<td>The information, knowledge, or capacity of officials</td>
<td>Rogger &amp; Somani (2019); Hjort et al (19); Dal Bo et al (18); Das &amp; Hammer (05)</td>
</tr>
<tr>
<td></td>
<td>Objective v subjective assessments of core quantities; Mayor beliefs regarding impacts of policy; Manager guesses of treatment effects; Clinical competence</td>
</tr>
<tr>
<td></td>
<td>Prices and quality of road construction; Prices paid for standardized good categories</td>
</tr>
<tr>
<td>Experience satisfaction measures</td>
<td>Olken (2010); Boyne (2002); Andrews et al (2005)</td>
</tr>
<tr>
<td></td>
<td>Villager satisfaction, willingness to contribute, &amp; perceived benefits; User satisfaction [% satisfied with how dealt with/service/access]</td>
</tr>
<tr>
<td>Wages and promotion</td>
<td>Alonso and Lewis (2001); Meier and O’Toole (2002)</td>
</tr>
<tr>
<td></td>
<td>Performance rating and grade level; Wages over and above “normal determinants”</td>
</tr>
</tbody>
</table>
## Administration-to-user transactions (e.g. ‘frontline’ services)

<table>
<thead>
<tr>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task completion rates and processing times</td>
<td>Rasul and Rogger (2018); Lewis-Faupel et al (2016); Fenizia (2019)</td>
</tr>
<tr>
<td></td>
<td>Project completion rates (adjusted for complexity); road construction</td>
</tr>
<tr>
<td></td>
<td>completion rates (quality-adjusted); process times (response times)</td>
</tr>
<tr>
<td></td>
<td>Household visits; sales; tax revenues collected; household visits; attendance</td>
</tr>
<tr>
<td>Subjective assessments of performance</td>
<td>Bertrand et al (2018); Alonso &amp; Lewis (2001); Brewer &amp; Selden (2000)</td>
</tr>
<tr>
<td></td>
<td>Subjective assessments by multiple stakeholders; Performance appraisals;</td>
</tr>
<tr>
<td></td>
<td>Employees’ subjective assessments of org. performance</td>
</tr>
<tr>
<td>Staff satisfaction and tenure</td>
<td>Deserranno (2019); Janke et al (2019); Bright (2008)</td>
</tr>
<tr>
<td></td>
<td>Tenure/retention; Staff satisfaction; Job satisfaction and turnover intentions</td>
</tr>
<tr>
<td>Measures of corruption</td>
<td>Hanna &amp; Wang (2017); Meyer-Sahling et al (2016); Olken (2007); Ferraz</td>
</tr>
<tr>
<td></td>
<td>Lab game; Perceptions of other public officials; Stated expenditure v audits</td>
</tr>
<tr>
<td>Service delivery &amp; governance indicators (org. performance)</td>
<td>Propper et al (2007); Banerjee et al (2017); Atkinson (2005);</td>
</tr>
<tr>
<td></td>
<td>Health outcomes (eg. vaccines, mortality); Test scores; Drink-driving rates;</td>
</tr>
<tr>
<td></td>
<td>Service delivery indicators divided by budget/expenditure</td>
</tr>
</tbody>
</table>
Administration-to-user transactions (e.g. `frontline services’)

- **Budget execution**
  - Liebman and Mahoney (2017); European Commission
  - The rate of expenditure over the budget year (and the quality of expenditure)

- **Mystery shoppers (comparing to a hypothetical/ideal benchmark)**
  - Das and Hammer (2007); Bertrand et al (2007)
  - Doctor knowledge versus practice; Incentives for driving license applicants to acquire license faster (and verify driving skills)

- **The information, knowledge, or capacity of officials**
  - Roger & Somani (2019); Hjort et al (19); Dal Bo et al (18); Das & Hammer (05)
  - Objective v subjective assessments of core quantities; Mayor beliefs regarding impacts of policy; Manager guesses of treatment effects; Clinical competence

- **Procurement outcomes (quality adjusted prices)**
  - Prices and quality of road construction; Prices paid for standardized good categories

- **Citizen satisfaction measures**
  - Olken (2010); Boyne (2002); Andrews et al (2005)
  - Villager satisfaction, willingness to contribute, & perceived benefits; User satisfaction [% satisfied with how dealt with/service/access]

- **Wages and promotion**
  - Alonso and Lewis (2001); Meier and O’Toole (2002)
  - Performance rating and grade level; Wages over and above “normal determinants”
<table>
<thead>
<tr>
<th>Function</th>
<th>Per cent government spending in 2000</th>
<th>Date introduced</th>
<th>Main components</th>
<th>Devolved administrations</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>17.1</td>
<td>Introduced in the Blue Book 1998, with data from 1986</td>
<td>Pupil numbers – Quality adjustment of 0.25 per cent to primary and secondary schools</td>
<td>UK figure for pupil numbers in nurseries and primary and secondary schools</td>
<td>Economic Trends, October 1998</td>
</tr>
<tr>
<td>Administration of Social Security</td>
<td>2.7</td>
<td>Introduced in the Blue Book 1998, with data from 1986</td>
<td>Number of benefit claims for 12 largest benefits No allowance for collection of contributions</td>
<td>United Kingdom</td>
<td>Economic Trends, October 1998</td>
</tr>
<tr>
<td>Administration of Justice</td>
<td>3.0</td>
<td>Partially introduced in the Blue Book 2000, with full impact in the Blue Book 2001, with data back to 1994 Q1</td>
<td>Number of prisoners, legal aid cases, court cases, and probation cost-weighted activity index</td>
<td>Great Britain for prisons, otherwise England and Wales</td>
<td>Economic Trends, September 2000, for probation, November 2001</td>
</tr>
</tbody>
</table>
### Measures used in the UK

<table>
<thead>
<tr>
<th>Service</th>
<th>Score</th>
<th>Source</th>
<th>Measure Description</th>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire</td>
<td>1.1</td>
<td><em>Blue Book 2001</em>, with data from 1994 Q1</td>
<td>Number of fires, fire prevention and special services</td>
<td>England, Wales and Northern Ireland</td>
<td>November 2001</td>
</tr>
<tr>
<td>Personal Social Services</td>
<td>7.4</td>
<td><em>Blue Book 2001</em>, with data from 1994 Q1</td>
<td>Children and adults in care and provision of home helps</td>
<td>England</td>
<td>November 2001</td>
</tr>
<tr>
<td>Police</td>
<td>5.8</td>
<td>Experimental</td>
<td>Cleared-up crimes of different types</td>
<td>England, Wales and Northern Ireland</td>
<td>May 2002</td>
</tr>
</tbody>
</table>
Detailed Literature Review
Details: Service delivery indicators


_HOUSEHOLD VISITS; COMMUNITY MEETINGS; NUMBER OF WOMEN GIVING BIRTH AT HEALTH CENTRE; CHILD CARE; IMMUNIZATIONS; BREASTFEEDING RATES; DEWORMING TREATMENTS_


_NUMBER OF DAYS WORKED UNDER NREGS AND TOTAL NUMBER EMPLOYED IN NREGS_


_DRINK-DRIVING RATES_


_FOOD SUBSIDIES RECEIVED_
Details: Service delivery indicators


DEATH RATES FOLLOWING TREATMENT AFTER HEART ATTACKS


REVIEW OF VARIOUS DIFFERENT MEASURES USED (LISTED IN ARTICLE TABLES)


SEVEN-DIMENSION MEASURE OF ORGANIZATIONAL PERFORMANCE, INCLUDING SERVICE DELIVERY INDICATORS


TEST SCORES
Details: Service delivery indicators


**FINANCIAL PERFORMANCE AND QUALITY OF CLINICAL CARE AND STAFF SATISFACTION**


**TAX REVENUE COLLECTED (ORGANIZATION)**


**SCHOOL-LEVEL STUDENT TEST SCORES**

Details: Service delivery indicators


Examples of performance outputs at the individual/societal level include homicide rates, school test performance, and life expectancy of clients. Some individual/societal outcomes include changes in workforce quality and productivity, development of active citizenship among young people, and heightened public knowledge about health issues.
The following papers use one or more of these global indicators:

**Government Effectiveness Index; Worldwide Governance Indicators; International Country Risk Guide; Business and Environmental Risk Intelligence; Business International.**


Details: Project and Task Completion


PROJECT COMPLETION RATES AND COMPLEXITY


PROJECT COMPLETION RATES AND COMPLEXITY


PRICES PAID AND QUALITY OF ROAD CONSTRUCTION PROJECTS (CONSULTANCY AND CONSTRUCTION PACKAGES IN INDIA AND INDONESIA). QUALITY = TIME OVERRUN; AUDIT REPORTS OF CONSTRUCTION QUALITY)
Details: Task-processing times


CLAIMS PROCESSED, WEIGHTED BY COMPLEXITY AND ERROR RATE


NUMBER OF BENEFITS PROCESSED EACH MONTH
Details: Individual-level outputs

CONDOM SALES

HOUSEHOLD VISITS; COMMUNITY MEETINGS; NUMBER OF WOMEN GIVING BIRTH AT HEALTH CENTRE; CHILD CARE; IMMUNIZATIONS; BREASTFEEDING RATES; DEWORMING TREATMENTS

HOUSEHOLD VISITS AND TENURE
Details: Individual-level outputs


TAX COLLECTED


NUMBER OF PUBLICATIONS
Details: Attendance rates


**DOCTOR ATTENDANCE**


**TEACHER ATTENDANCE**


**ATTENDANCE OF TEACHERS AND HEALTH WORKERS**
Details: Subjective assessments


SUBJECTIVE ASSESSMENTS OF CIVIL SERVANTS BY MULTIPLE STAKEHOLDERS ("IAS OFFICERS, STATE CIVIL SERVANTS, ELECTED POLITICIANS, REPRESENTATIVES OF BUSINESS ASSOCIATIONS, LOCAL TV AND PRINT MEDIA, AND CIVIL SOCIETY ORGANISATIONS. FOR EACH IAS OFFICER THEY KNOW, WE ASK STAKEHOLDERS TO GRADE THEM ON A 1 (LOW) TO 5 (HIGH) SCALE FOR: EFFECTIVENESS, PROBITY, THE ABILITY TO WITHSTAND ILLEGITIMATE POLITICAL PRESSURES, PRO-POOR ORIENTATION AND OVERALL RATING")


ORGANIZATION PROVIDED PERFORMANCE APPRAISAL DOCUMENTS FOR INDIVIDUAL EMPLOYEES


SELF-ASSESSMENTS OF PERFORMANCE AND OWN QUALITY OF WORK
Details: Subjective assessments


SELF-ASSESSMENTS OF PERFORMANCE AND OWN QUALITY OF WORK


PERFORMANCE RATING (SCORE) AND GRADE LEVEL


EMPLOYEE ASSESSMENTS OF ORGANIZATIONAL PERFORMANCE
Details: Subjective assessments


EMPLOYEE ASSESSMENTS OF ORGANIZATIONAL PERFORMANCE


EMPLOYEE ASSESSMENTS OF ORGANIZATIONAL PERFORMANCE


CASE STUDIES INTERVIEWS OF PROVIDERS AND MANAGERS
Details: Subjective assessments: An example module


My organization has made good use of my knowledge and skills in looking for ways to become more efficient.

In the past 2 years, the productivity of my work unit has improved.

The work performed by my work unit provides the public a worthwhile return on their tax dollars.

In general, people of my race/national origin group are treated with respect in my organization.

Overall, how would you rate the quality of work performed by your current coworkers in your immediate work group

My organization provides fair and equitable treatment for employees and applicants in all aspects of personnel management without regard to their political affiliation, race, color religion, national origin, sex, marital status age, or handicapping condition.
Details: Staff satisfaction and retention rates


HOUSEHOLD VISITS AND TENURE


JOB SATISFACTION, PERCEIVED PERFORMANCE, QUALITY OF WORK, TURNOVER INTENTIONS


JOB SATISFACTION AND TURNOVER INTENTIONS
Details: Staff satisfaction and retention rates


**JOB SATISFACTION AND TURNOVER INTENTIONS**


**FINANCIAL PERFORMANCE AND QUALITY OF CLINICAL CARE AND STAFF SATISFACTION**
Details: Measures of corruption


**COMPARE RESULTS OF A CORRUPTION GAME AMONG PUBLIC AND PRIVATE SECTOR ASPIRANTS. HOW MUCH TO OVERREPORT AND TO UNDERPAY?**


**COMPARE THE STATED RESPONSE TO NUMBERS FROM DICE ROLLS TO A UNIFORM DISTRIBUTION.**


**PERCEPTIONS OF PUBLIC OFFICIALS ABOUT THEIR ORGANIZATION: "OVER THE LAST 12 MONTHS THERE HAVE BEEN RUMOURS OF KICKBACKS IN MY MINISTRY"**


**COMPARE STATED EXPENDITURES TO INDEPENDENTLY-ESTIMATED EXPENDITURES ON INFRASTRUCTURE PROJECTS**
Details: Measures of corruption


PERCEPTIONS OF CORRUPTION BY SERVICE USERS, EMPLOYEES AND BY EXPERTS (JOURNALISTS); COMPARE STATED EXPENDITURES TO INDEPENDENTLY-ESTIMATED EXPENDITURES ON INFRASTRUCTURE PROJECTS; SURVEY ESTIMATES OF BRIBES OR INFORMAL GIFTS; DIRECT OBSERVATIONS BASED ON PUBLIC RECORDS OR AUDITS OR BRIBE PAYERS; ESTIMATES OF GRAFT (MEASURE A QUANTITY BEFORE AND AFTER CORRUPTION IS SUPPOSED TO HAVE TAKEN PLACE, E.G. GRANT SENT V GRANT RECEIVED; COMPARE EXPORTS SENT VERSUS EXPORT RECEIVED; COMPARE STATED INPUTS V ACTUAL MEASURED INPUTS; ADMINISTRATIVE RECORDS OF GRANTS OR SUBSIDIES V HOUSEHOLD SURVEY); MARKET EQUILIBRIUM CONDITIONS (E.G. COST OF LOSING A POLITICAL CONNECTION; PUBLIC-SECTOR PAY RECEIVED VERSUS ASSETS OWNED; PROBABILITY OF RECEIVING PUBLIC PROGRAMME SUPPORT OR FUNDS BETWEEN PUBLIC AND PRIVATE EMPLOYEES, CONDITIONAL ON OBSERVABLES.
Details: Measures of corruption


Details: Budget execution and expenditure per unit


**ARE EXPENDITURE RATES FASTER AS BUDGET DEADLINES APPROACH? DOES THE QUALITY OF EXPENDITURE DECLINE?**


**PUBLIC-SECTOR PERFORMANCE MEASURES DIVIDED BY MEASURES OF "RELEVANT" PUBLIC EXPENDITURE. PERFORMANCE IS AN INDEX WHICH COMBINES SERVICE DELIVERY INDICATORS ACROSS EDUCATION, HEALTH AND PUBLIC INFRASTRUCTURE OUTCOMES AS WELL AS MEASURES OF ECONOMIC PERFORMANCE AND EQUALITY (INCOME DISTRIBUTION, STABILITY AND GROWTH) AND MEASURES OF CORRUPTION AND RED-TAPE, BASED ON MEASURES FROM THE WORLD ECONOMIC FORUM AND SCHNEIDER (2002) ON THE SIZE OF SHADOW ECONOMY.**
Details: Budget execution and expenditure per unit


**OUTPUTS DIVIDED BY INPUTS:** CHANGE IN LIFE EXPECTANCY DIVIDED BY HEALTH EXPENDITURE; CHANGE IN TEST SCORES DIVIDED BY NUMBER OF TEACHERS PER STUDENT; AVERAGE TRIAL LENGTH DIVIDED BY NUMBER OF JUDGES PER 1000 INHABITANTS; NUMBER OF CHILDREN IN DAY CARE DIVIDED BY EXPENDITURE ON CHILD CARE; TONS OF WASTE COLLECTED AND RECYCLED DIVIDED BY EXPENDITURE ON WASTE.


**SERVICE DELIVERY INDICATORS DIVIDED BY EXPENDITURE**


**OUTPUTS DIVIDED BY INPUTS (TEST SCORES DIVIDED BY NUMBER OF TEACHERS PER STUDENT); OUTCOMES DIVIDED BY EXPENDITURE (EXPENDITURE ON JOB-SEEKER SUPPORT AND CHANGES IN UNEMPLOYMENT RATES AND EARNINGS).**

A score/index is created which takes into account six dimensions of performance (quantity of outputs; quality; efficiency; outcomes; value for money; and consumer satisfaction scores. These are qualitatively assessed by a commission and each organization given a score.)


"Measures of inputs, throughputs (e.g. access to care metrics such as waiting times, which are important in a system where care is rationed), outputs (financial performance and measures of the quality of clinical care) and staff job satisfaction "... e.g. number of beds, number of admissions, lengths of stay, waiting times, cancelled operations, deaths, disease rates, readmission rates."

KILOMETERS OF BUS ROUTE PER EMPLOYEE; ACCIDENT RATES


TECHNICAL EFFICIENCY OF POWER PLANTS

**COMPARE TREATMENT OFFERED TO STANDARDIZED PATIENTS (MYSTERY SHOPPERS) TO VIGNETTES ABOUT HYPOTHETICAL CASES TO MEASURE KNOW-DO GAP (INCENTIVES V TRAINING).**


**VIGNETTES: MEDICAL PROVIDER PRESENTED WITH HYPOTHETICAL CASES AND COMPARE RESPONSES TO A CHECKLIST OF ESSENTIAL PROCEDURES. DIRECT OBSERVATIONS OF DOCTORS DURING A DOCTOR-PATEINT INTERACTION**


**COMPARE DOCTOR KNOWLEDGE (VIGNETTES) TO PRACTICE WITH COMPARABLE CASE MIXES OF DISEASE.**
Details: Mystery shoppers


**COMPARE DOCTOR RESPONSES TO HYPOTHETICAL CASE VIGNETTES TO ASSESS CLINICAL COMPETENCE**


**INDEPENDENT SURPRISE DRIVING TESTS TO ASSESS CAPABILITIES OF THOSE THAT ATTAINED A LICENSE. APPLICANTS INCENTIVISED TO ACQUIRE A LICENSE FASTER WERE MORE LIKELY TO GET A LICENSE AND FAIL THE INDEPENDENT TEST.**
Details: Information/knowledge of officials


COMPARE ESTIMATES PROVIDED BY PUBLIC OFFICIALS TO OBJECTIVE BENCHMARKS ON FUNDAMENTAL QUANTITIES REQUIRED TO COMPLETE TASKS


COMPARE ACTUAL TREATMENT EFFECTS TO SUPERVISOR ESTIMATES REGARDING A NEW MONITORING TECHNOLOGY AIMED AT IMPROVING THE PERFORMANCE OF AGRICULTURAL EXTENSION AGENTS (PERFORMANCE = SHARED OF ASSIGNED FARMERS VISITED IN A GIVEN WEEK).


MEASURE WILLINGNESS TO PAY FOR RESEARCH EVIDENCE AND EXTENT TO WHICH MAYORS UPDATE BELIEFS REGARDING POTENTIAL IMPACTS OF EARLY CHILDHOOD DEVELOPMENT PROGRAMMES. ALSO MEASURE PROBABILITY OF MUNICIPALITY IMPLEMENTING A POLICY AFTER A RANDOMIZED PRESENTATION OF THE EFFICACY OF TAXPAYER REMINDER LETTERS.
Details: Information/knowledge of officials


COMPARE TREATMENT OFFERED TO STANDARDIZED PATIENTS (MYSTERY SHOPPERS) TO VIGNETTES ABOUT HYPOTHETICAL CASES TO MEASURE KNOW-DO GAP (INCENTIVES V TRAINING).


VIGNETTES: MEDICAL PROVIDER PRESENTED WITH HYPOTHETICAL CASES AND COMPARE RESPONSES TO A CHECKLIST OF ESSENTIAL PROCEDURES. DIRECT OBSERVATIONS OF DOCTORS DURING A DOCTOR-PATIENT INTERACTION


COMPARE DOCTOR KNOWLEDGE (VIGNETTES) TO PRACTICE WITH COMPARABLE CASE MIXES OF DISEASE.


COMPARE DOCTOR RESPONSES TO HYPOTHETICAL CASE VIGNETTES TO ASSESS CLINICAL COMPETENCE
Details: Procurement outcomes


Details: Citizen satisfaction measures


VILLAGER SATISFACTION; WILLINGNESS TO CONTRIBUTE; PERCEIVED BENEFITS


EXAMPLES: PERCENTAGE OF BENEFIT CLAIMANTS SATISFIED WITH THE WAY THEY WERE DEALT WITH; PERCENTAGE OF LIBRARY USERS SATISFIED WITH STAFF AND OPENING HOURS; PERCENTAGE OF USERS SATISFIED WITH LOCAL BUS SERVICE; PERCENTAGE OF APPLICANTS SATISFIED WITH PLANNING SERVICE; PERCENTAGE OF CITIZENS SATISFIED WITH OVERALL SERVICE PROVIDED BY LOCAL AUTHORITY
Details: Citizen satisfaction measures


CITIZEN/CUSTOMER SATISFACTION MEASURES FROM USERS OF A LOCAL GOVERNMENT AUTHORITY


CASE STUDY INTERVIEWS OF SERVICE USERS AND STAKEHOLDERS (FAMILIES OF CARE-RECEIVERS)
Details: Wages/grades/promotions


PERFORMANCE RATING (SCORE) AND GRADE LEVEL


MANAGERIAL QUALITY MEASURED BY THE WAGES PAID TO SCHOOL PRINCIPALS OVER AND ABOVE "NORMAL DETERMINANTS OF SALARY"
The conceptual problem of measuring public-sector productivity (ch4)

Progress in what countries do to measure this in national accounts (Ch2/3/5/6)

Leading to the need to classify services into collective (hard to measure output) and individual (reasonable output measures available)
  - Collective => Task Completion
  - Individual => Quality-adjusted outputs over expenditure

Other classifications between services (Prendergast, 2007)
  - Those where organizational performance is better for individual “consumers” (social security, pensions, …)
  - Those where organizational performance is worse for individual “consumers” (taxes, crime/justice,..)

Importance of National Accounts => Macroeconomic policy decisions; => Measuring welfare
  - Measuring public-sector productivity may require different approaches for each of these purposes
    - What about just ignoring the public sector in the accounts?
    - Opportunity cost of resources used as inputs into public production? (Public as substitute)
    - Indirect effects of public production on private production (health/educ of labor)? (Public as complement)
    - Activities are often without market prices but are still welfare-enhancing such that they should be counted
      - Yet, increasing national accounts if often not the goal of major policy (peace, inequality, human rights)

There is a difference between output estimates (for national accounts, e.g.) and performance measures of public-service management
  - They use much of the same data.
    - But performance indicators need precise, transparent and simple measures that cannot be manipulated. They do not necessarily need to be stable over time and can be selective in their coverage
  - Output measures used for accounting purposes need to try and cover all activities and be stable over time

Outputs versus Outcomes
  - Outputs are useful performance targets (number of vaccinations per 1000 persons) and outcomes are a function of many external factors (GDP, diets/food/agriculture innovations, exercise habits, smoking/alcohol preferences,…). Assessing the performance of the NHS based on these is inaccurate
  - National accounts measures are not a substitute for performance measures and can generate perverse incentives.
The Atkinson Review 2005

- Government output is generally non-market. At the point of “purchase” there is no price. The lack of transaction underlies many of the problems in measuring output and productivity
- Government output is an important component of measuring growth. Comparing growth across countries requires comparable measures of government output
  - Blue book convention: value = value of total inputs
    - (Output= input convention). Wages/benefits of employees; procurement costs; capital rent costs.
    - Because it is difficult to measure the output of MoDefence => What is the exact output?
    - What is the value of a health service? What would be the value of the market transaction? Is that enough?
    - Output=input assumes that the productivity of each employee is constant over time
    - This limits productivity growth measures, especially for countries with large public sectors => seemingly unproductive!
      - Some countries (Germany) apply a productivity growth factor) while others don’t => comparison!
      - What about the delay in inputs=>outputs?
- Moving away from output=input has benefits & costs. It means that output/input is not always 1 anymore. But “no single number, however carefully constructed, can fully capture the performance of complex public services with multiple objectives.”
- Since 1998, the ONS has moved away from input=output. Direct estimates of government output now cover the majority of government services
The European System of Accounts and Eurostat Handbook…

Rejects outputs=inputs as it ignores changes in productivity
Rejects inputs+ assumed productivity change
Mandates that member states develop direct output measures

<table>
<thead>
<tr>
<th>Type of service</th>
<th>A/B/C methods</th>
</tr>
</thead>
</table>
| Individual services (such as Education, Health, Social Security, Recreation and Cultural Services) | A methods – output indicator approach where the indicators satisfy the following criteria:  
  a) they should cover all services provided;  
  b) they should be weighted by the cost of each type of output in the base year;  
  c) they should be as detailed as possible; and  
  d) they should be quality adjusted.  

B methods – output indicator approach where the criteria are not fully satisfied: eg the level of detail could be improved or the measure does not take into account changes in quality.  

C methods – if input, activity or outcome is used (unless outcome can be interpreted as quality-adjusted output) or if coverage of output method is not representative.  

| Collective services (such as General Public Administration, Defence, Police, and Research and Development). | Broadly the same as for individual services but:  
B methods – input methods are B methods, as are the use of volume indicators of activity.  
If input methods are used they should estimate the volume of each indicator separately, taking quality changes of inputs into account. Applying productivity or quality adjustments to the sum of the volume of inputs is not recommended.  

C methods – the use of a single input volume indicator is not a B method.  

The Atkinson Review 2005

- They are attempting to measure changes in output (performance). No attachment of $$$ to the output (in levels or changes).
- No attempt is made to assess how inputs ($$$) => outputs $$$
- They need to be continuously monitored to ensure that they capture changes in quality and that there are no adverse consequences in other dimensions. How to account for changes in labor productivity (e.g. through the increased use of tech)?

Expenditure at current prices on the inputs purchased by government to produce its outputs, usually referred to as General Government Final Consumption Expenditure at Current Prices: £140bn in 1995 and £230bn in 2003, an increase of 64 per cent;

Volume of government inputs, adjusted for the increases in the prices of inputs, using appropriate price indices: at 2001 prices, £166bn in 1995 and £214bn in 2003, an increase of 29 per cent. This implies that, on average, input prices had risen by some 35 per cent;


- Divide 1 by 3 to obtain cost per unit of government output (towards more classical measures of productivity)
- The output=input method, only 1 and 2 are used (inputs). Used for Defence, General Public Administration, Economic services, Environmental Protection, Recreation and Culture, Housing and Community Amenities

- Education inputs: labor (teachers, secretaries, caretakers, …) wages and national insurance/pensions. Good and services (books, pens, lighting, heating, transport…). Capital (buildings, equipment).
The Atkinson Report: Principles

- Adjust for quality (how does the service increasingly => outcomes)
- Adjust for the private sector
  - The value of the output of government services rises with the real value of private assets and incomes (e.g. fire/emergency services; justice system become more valuable when richer)
  - Similarly, as the private sector develops and demands higher skills, the returns to education increase, increasing the real value of “the number of children in primary school”
- Indicators should have “full coverage” (account for multi-dimensionality) as far as possible
- Indicators should be possible to transfer from region to region
- Inputs should be “comprehensive” (labor, capital, equipment)
  - Both “direct” (number of hours worked with weights for skills)
  - And “indirect” (deflation of pay/wage bills by labor-cost index [inflation])
- Indicators should have a transparent clear margin of error