Performance measurement and evaluation indicators

February 18, 2022

Tatsuya Ohno
Professor at Tottori University

Lecture content

1. What is performance measurement?
2. Measures taken by Japanese government at local and central levels
3. How to set evaluation indicators
4. How to use evaluation indicators

1. What is performance measurement?

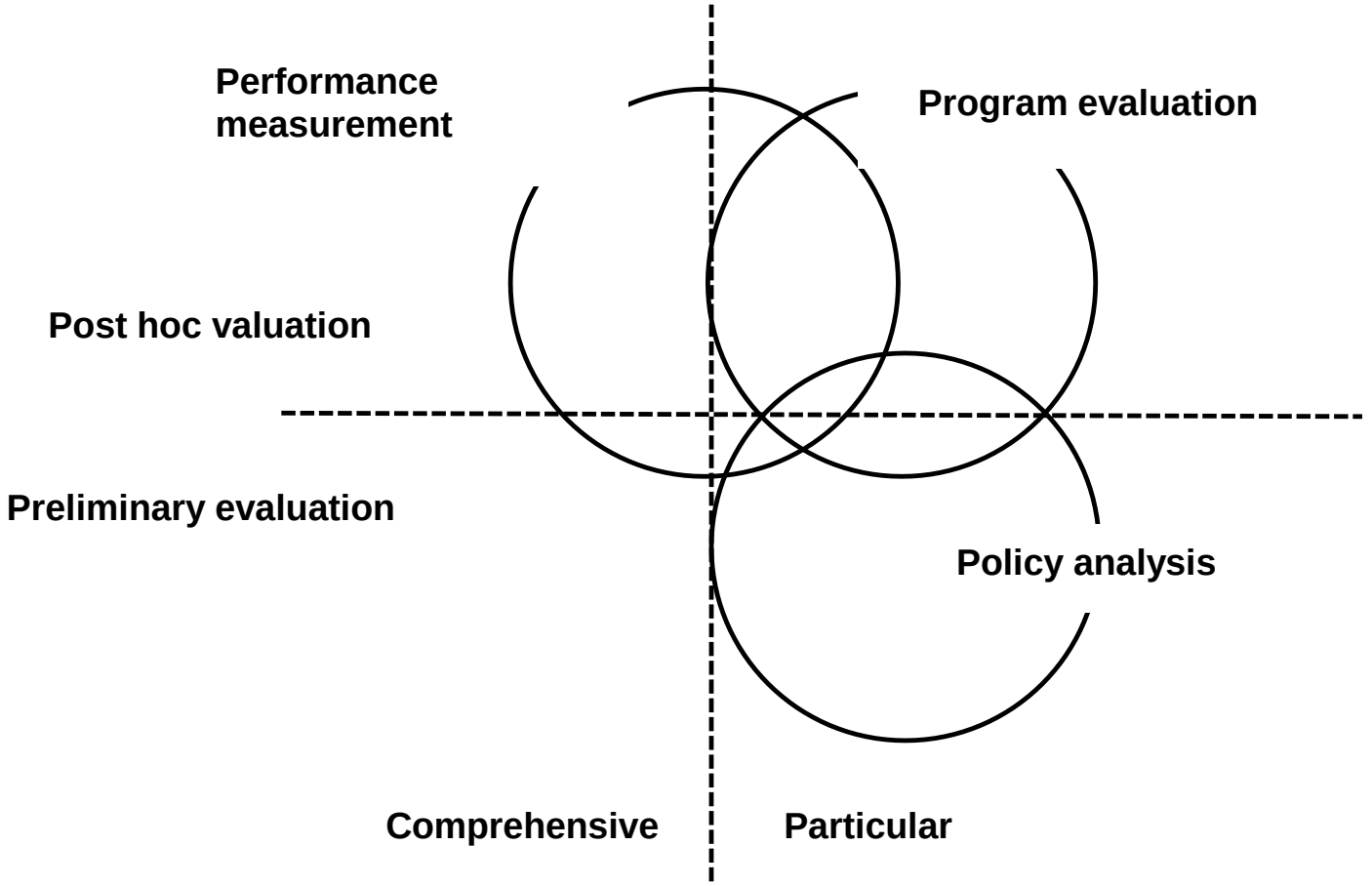
- "Regularly collect and report a range of data on 'input', 'action(s)', 'output', 'outcome', 'performance (performance)'" (U.S. General Accounting Office (1992). Program Performance Measures: Federal Agency Collection and Use of Performance Data Report GAO/GGD-92-65.)

- "Regularly measure the results and effectiveness of services and programs" (Hatry, H.P. (1999). Performance Measurement: Getting Results. Washington, D.C.: Urban Institute.)

- "Systematic and continuous monitoring of program performance, in particular the level of achievement of pre-established goals and criteria" (U.S. Government Accountability Office. (2012) Designing Evaluations 2012 Revision. GAO-12-208G.)

- “Regularly measuring performance, evaluating it, proposing activities for it, publishing it” (Ueno Hiroshi, Ueno Makiko (2007), “Measuring performance”, edited by Miyoshi Koichi “For students of evaluation theory”, Sekaishisosha)

Three systems in the theory and practice of evaluation



- The purpose of performance measurement evaluation is to manage the process of a program package based on the evaluation of individual programs and to manage the Deming cycle.

Characteristic features of evaluation by the performance measurement

1. The subject of the evaluation is a complex of ongoing administrative activities (post hoc evaluation).
2. Evaluation often involves setting goals in the preliminary stage.
3. Evaluation is carried out regularly in relation to a wide range of subjects of evaluation.
4. Often the emphasis is on the level of achievement of target figures.
5. Uniform measurement and evaluation criteria are used (so that programs can be compared and summarized)
6. The main focus is on the results of the work of the entire organization as a whole and the entire program.
7. It is necessary to conduct not so much an in-depth analysis of individual elements as a broad-based and superficial monitoring. Early notification and accountability play an important role.
8. An internal evaluation is usually carried out. But external verification is important.

2. Measures taken by Japanese government at central and local levels

○ The "New Public Administration" contributed to the establishment of the performance measurement system

- New Public Administration (English: NPM=New Public Management) is a theory and tool of a new system of public administration. It was developed mainly in the Anglo-Saxon states after the 80s of the last century, and then spread to many states, becoming the main direction in the field of reforming the public sector and administrative bodies.

- The characteristic features of this are the management method adopted in the private sector, the Deming cycle (PDCA cycle), targeted management with an emphasis on achieving results, accountability, the role and responsibility of the top manager and administrator, etc. The key concept is "effectiveness".

- To improve performance, it is necessary to manage the progress of implementation (that is, measure performance) based on quantitative data.

US Performance Measurement System

- In 1938, a book was published entitled "Measuring the Performance of Municipalities: a Study of recommended criteria for evaluating administration", written by Simon H.A and Ridley S.E.
- In the 90s of the last century, the number of activities undertaken by local governments increased dramatically. For example, the "Oregon State Target", etc.
- In 1992, a book was published entitled "The Formation of a New Type of Government", written by Osborne D and Gebler T.
- In 1993, the "Law on the Evaluation of Government Performance" (GPRA: Government Performance and Results Act) was adopted
- In 2011, the "Law on the Modernization of the Evaluation of Government Performance" (GPRAMA: Government Performance and Results Modernization Act) was adopted

Performance measurement system in Japan

< Local self-government bodies (LGSB)>

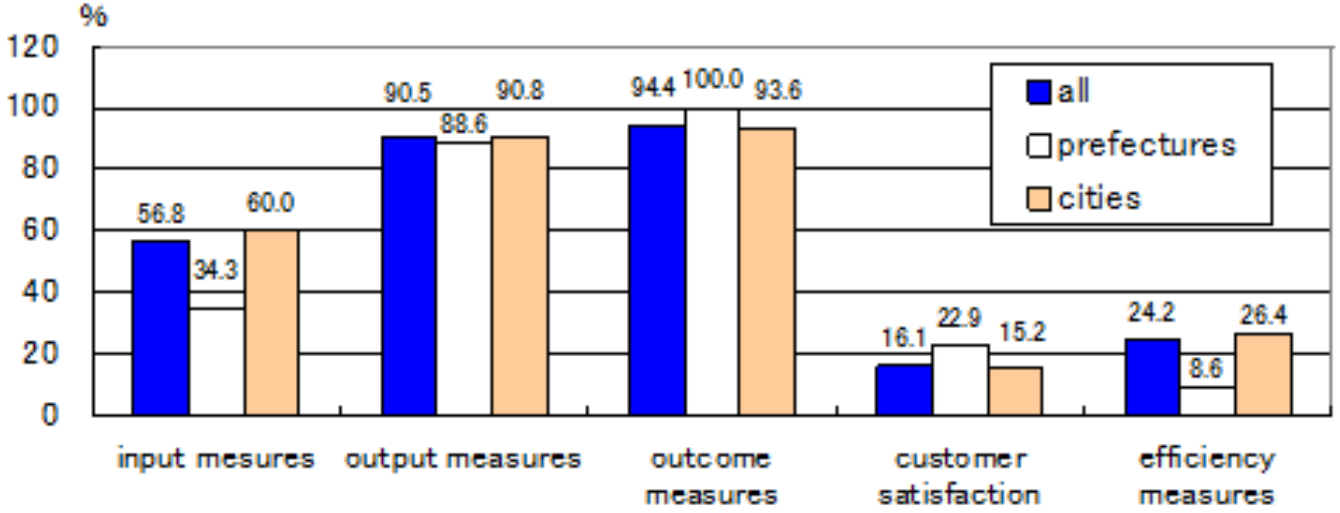
- In 1996, an "Assessment of administrative activities" was launched in the Mie Prefecture. It was the first full-fledged evaluation system in Japan.
- As of 2017, 60% of local governments have implemented the system. In prefectures and cities defined by Government decrees, it has been implemented by almost 100%.

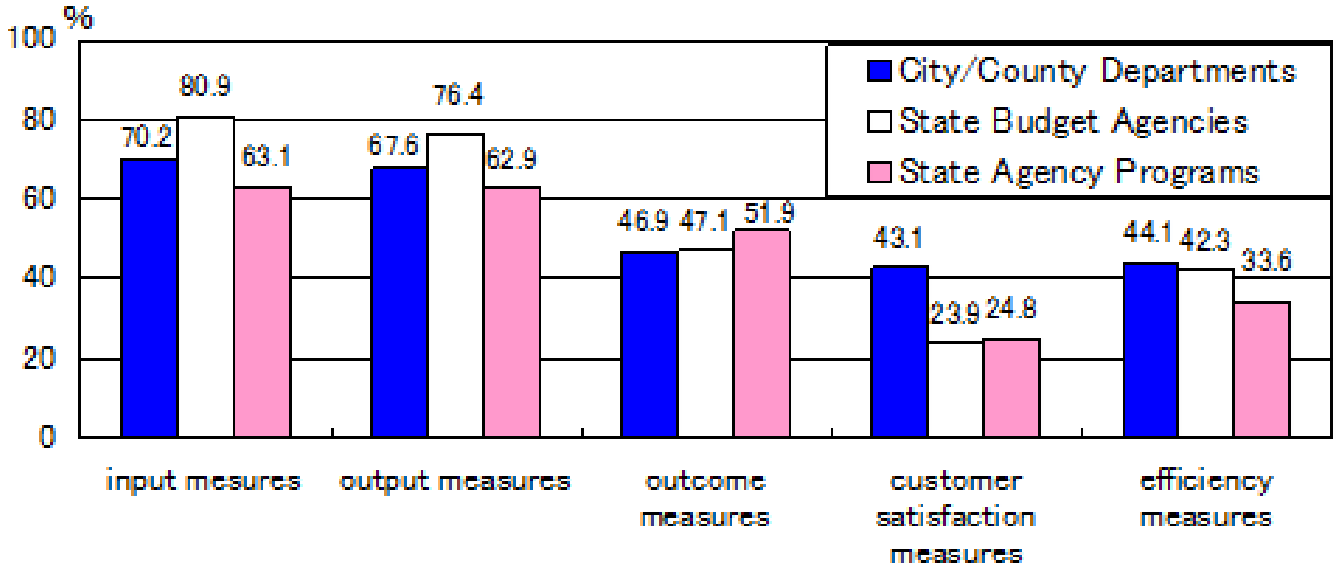
<Central government>

- In 2001, a "system for evaluating policy measures" was launched by all ministries and agencies in accordance with the law. In general, the subjects of evaluation in the Government are 500-600 policy measures.
- In 2010, the "audit of administrative activities" was launched. In general, 5000-6000 projects are subject to inspection in the Government.

Comparison of evaluation indicators between Japan and the USA

Table: Types of indicators used in Japan (top) and the USA





Source: Tanaka, H (2006), Current Status of Performance Measurement Systems in Japan: Are they functioning adequately? The article is presented at the annual EGPA Conference.

3. How to set evaluation indicators

- When setting indicators, “relevance” plays a crucial role.
- "Relevance" means that the items that should be measured are measured.
- At the stage of measuring evaluation indicators, "reliability" is important.

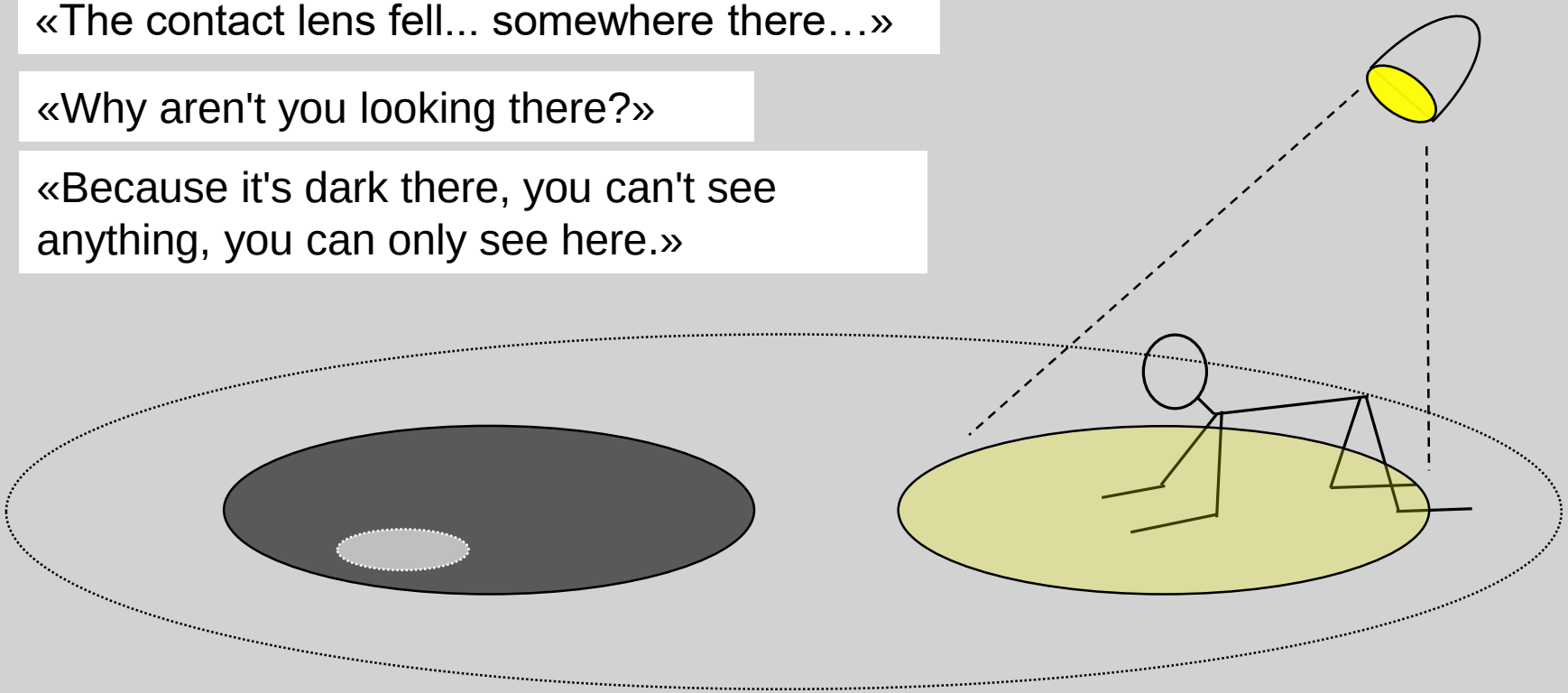
What is relevance? – A man is looking for something at night without moonlight under a street lamp.

«What are you looking for?»

«The contact lens fell... somewhere there…»

«Why aren't you looking there?»

«Because it's dark there, you can't see anything, you can only see here.»



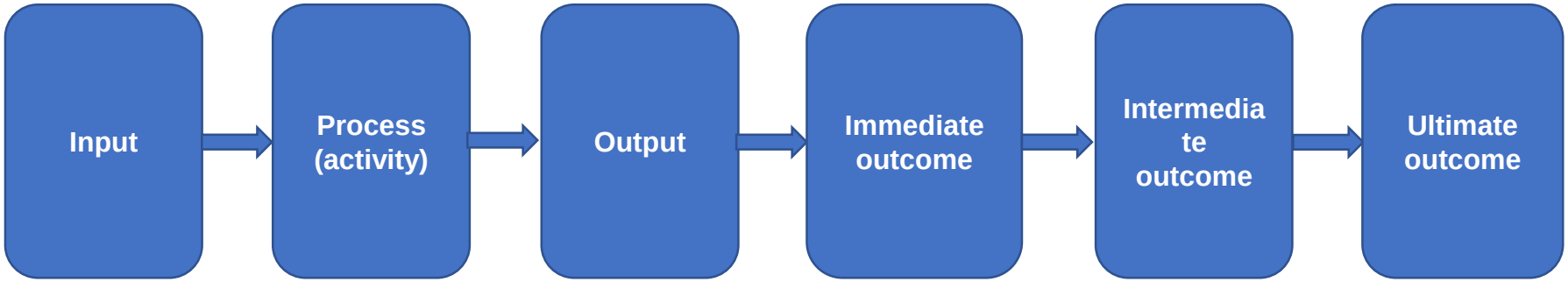
According to the parable of Swiss J.E. –'Performance Monitoring Systems'. in Ammons, D. N. (ed) *Accountability for Performance: Measurement and Monitoring in Local Government*. Washington, D.C.: International City/County Management Association, 1995

Policy Logic and Logical Model

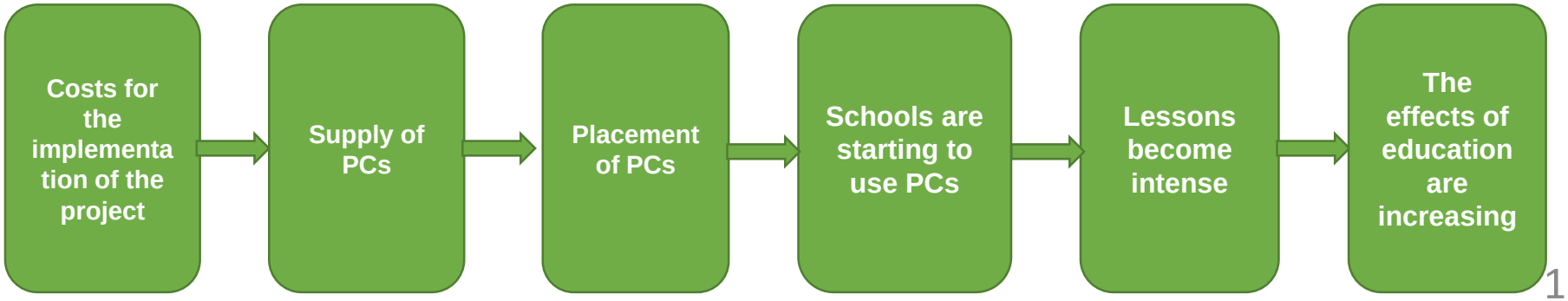
- Logic is a logically constructed process from the moment of investment of administrative resources to the moment of manifestation of policy results. It has almost the same meaning as a prorgam theory.

Input → Output → Outcome

- The logic that is expressed in the scheme is called a logical model.



- For example: A project to introduce PCs into schools



Setting indicators according to logic (Example 1)

< Typical example of a project to introduce PCs in schools >

Placement of PCs･･･ number of students per PC

↓

Using PCs in a lesson, etc. ･･･ degree of PC usage

↓

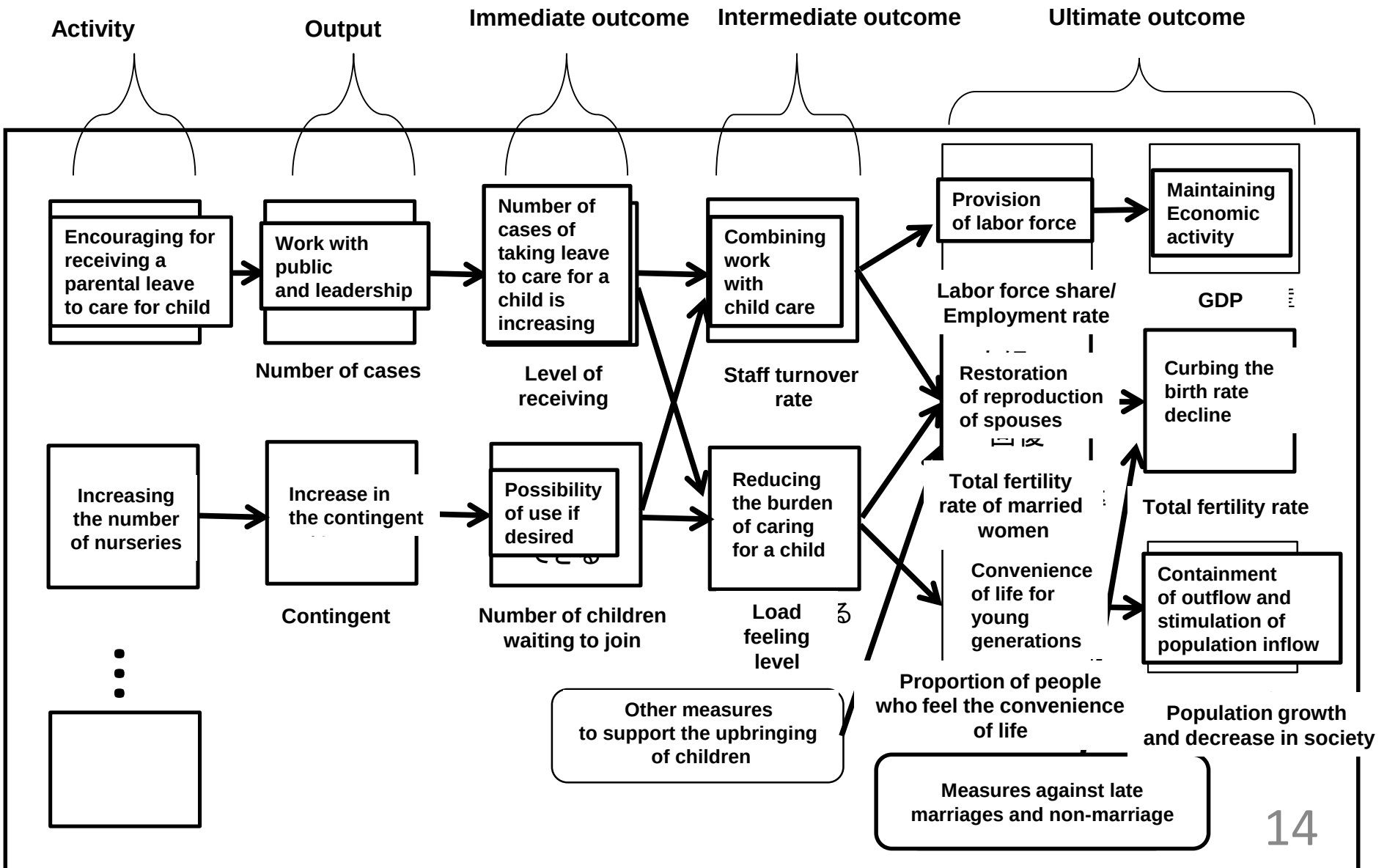
Intensity of the lesson･･･ assessment by schoolchildren, teacher's self-assessment

↓

Increasing the effect of education ･･･ testing for schoolchildren

Setting indicators according to logic (Example 2)

< A typical example of a project to support the upbringing of children >



Relevance of evaluation indicators

○ **What are the desired main indicators?**

1. The indicators should be close to the ultimate outcome. The following indicators should be included.
2. Such indicators are needed that allow, if necessary, to get acquainted with the progress of the output, the immediate outcome and the intermediate outcome.
3. We need such indicators that allow us to get acquainted with the own effects of this policy, that is, such indicators that are little influenced by external factors.
4. We need indicators that cover important processes leading to the ultimate outcome.

Another relevance

○ Two types of relevance with respect to evaluation indicators

(1) Relevance concerning logic – selection of logical model blocks

（2）Relevance regarding the definition of indicators – the definition of evaluation indicators reflecting the state of the blocks

< An example of a project for the introduction of PCs in schools >

- It was decided that the "Ratio of the number of lessons in which the PC was used to the total number of lessons" is used as the method for calculating the evaluation indicator "degree of PC use". Is it sensitive enough?

Reliability of measurement of the evaluation indicator

- Reliability is the point of view according to which it is possible to collect sufficient data in an impartial and comprehensive manner. This means that any person can get the same result from one phenomenon at any time.

< An example of the project to introduce PCs in schools >

- To measure the "ratio of the number of lessons in which the PC was used to the total number of lessons", it is difficult to collect data in all primary and secondary schools of the city during the year, so we came to the conclusion to record data in a school whose director is ready to cooperate with the city administration in this study for one month when the load is small.
- Is the data collected at this school reliable?

4. How to use evaluation indicators

○ **Comparison in two directions, which are required to measure performance**

A. Comparison of actual and target figures by evaluation indicators

B. The comparison between indicators and programs by the level of achievement of the goal is a comparison in quality. Generalization and summation of data in general

○ 4 questions that require a choice:

(1) Relationship between indicator digit measurement and time

(2) Growth models of the indicator value

(3) Difficulty in reaching the target figure

(4) Calculation method for determining the level of achievement of the goal

(1) Relationship between measurement of the indicator digit and time

Exercise 1. Indicator A rises and indicator B falls. Can we say that project A is going better than project B?

	5 years ago	This year
A. The share of houses and buildings with increased seismic resistance in the total number of houses and buildings in the city (%)	60	65
B. Level of participation in emergency exercises (%)	20	18

Stock and flow - "mixing is dangerous"

- "The share of houses and buildings with increased seismic resistance in the total number of houses and buildings in the city" is with a **stock indicator**.
- “The level of participation in emergency exercises” is an **indicator of the flow**.

- In general terms,

  Flow is a quantity that is measured in a given period of time.

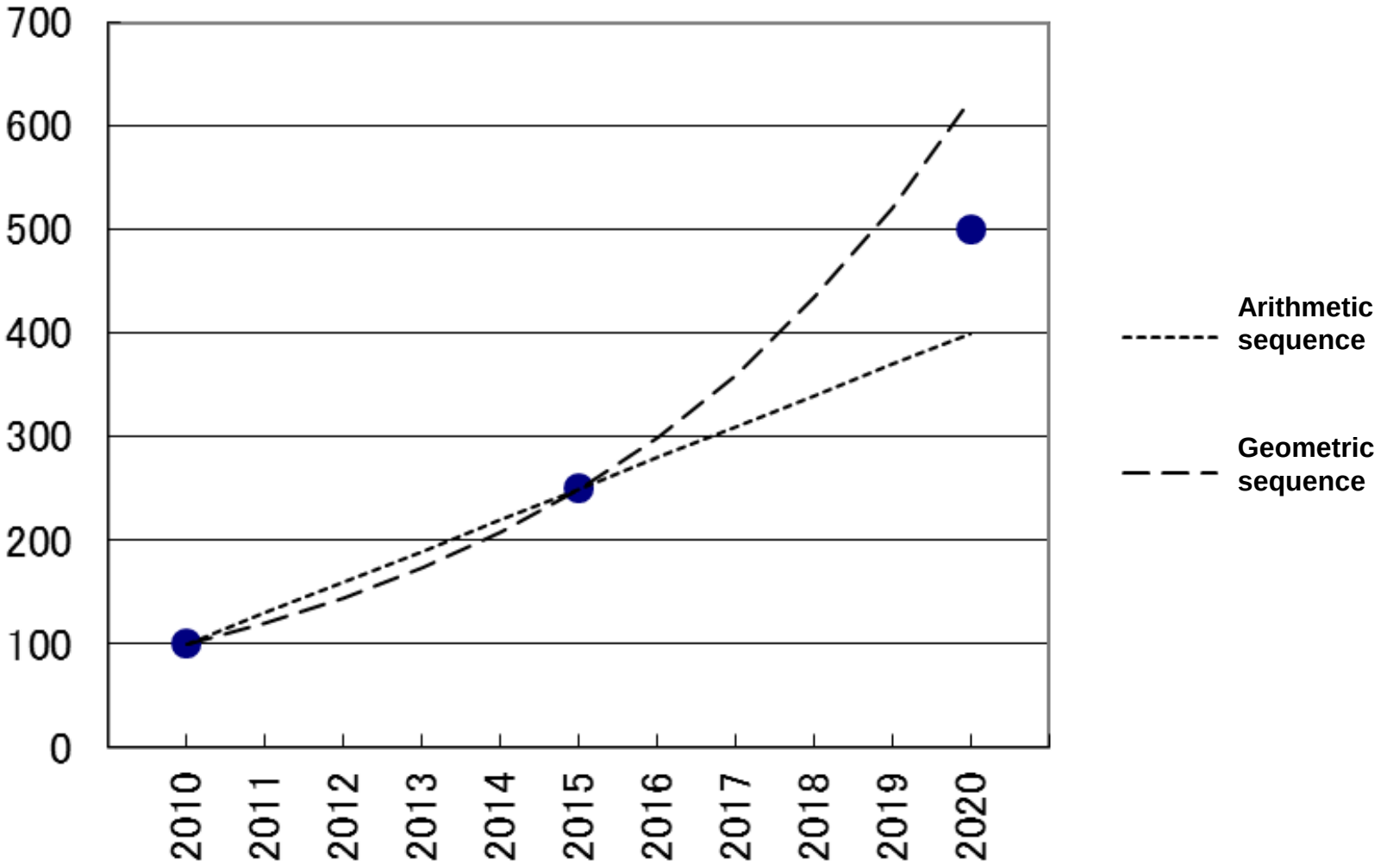
  Stock is the quantity accumulated at a given moment.

- To directly compare these two indicators, it is necessary to turn the stock indicator into a flow indicator by calculating the difference in the stock indicator (end of the period minus the beginning of the period).

(2) Growth models of the indicator value

Exercise 2 . In one project in FY 2010, the actual value was 100, and at the end of the same fiscal year, a target value of 500 was set by the end of 2020. If in FY 2015, the actual value was 250, can we assume that the project is going well?

Basically, there are 2 models for the growth of the indicator value



(3) Difficulty in reaching the target

○ Difficulty of reaching the target figure: an example of classification

A. Ideal level is a level that is very difficult to achieve, but you should strive to achieve it.

Б. Ambitious level is a level that requires the improvement of the tool and an increase in performance to achieve.

B. Realistic level is a level that the continuation of current measures is sufficient to achieve.

Г. Minimum level – requires improvement of a serious condition, etc.

○ Role of the target figure

- Is the target figure a promise or not?
- For one program, there can be two levels of the target digit (for example, B and C).
- For a new program, there may initially be a trial target figure.

(4) Calculation method for determining the level of achievement of the goal

Exercise 3 . The target figure and the actual figure of one indicator are shown below. How can I calculate the level of achievement of the goal?

	Start time	1st year	2nd year	3rd year
Actual figure	50	70	90	80
Target figure	—	60	80	100

- Goal achievement level (1) 80 : 100 = 80%
- Goal achievement level (2) 30 : 50 = 60%
- Goal achievement level (3) 240 : 240 = 100%

Summarizing data on the state of achievement of the goal as a whole

- Data were summarized on the state of achievement of the goal for 50 indicators at the time of the expiration of the 5-year period in the 10-year program.
- It became clear that 60% were grades B and above, which meant successful progress (the probability of reaching the final value is 50% and above). Can the program be considered successful overall?