

3rd lecture on evaluation

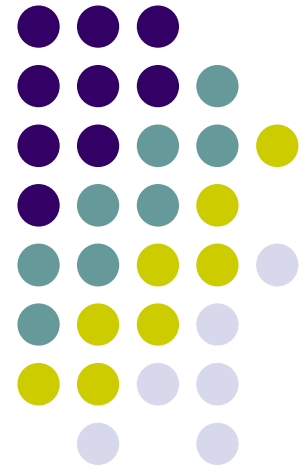
for members of the working groups of the National Development Council

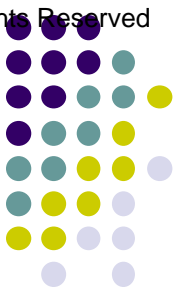
Program Evaluation Methods, Integration and Use of Evaluation Results (evidence), and their Reflection in the Policy Evaluation

---

Lecturer: Iwao OSHIMA

(Chairman of the Japan Evaluation Society / Vice-Rector and Professor of Tohoku Fukushi University)



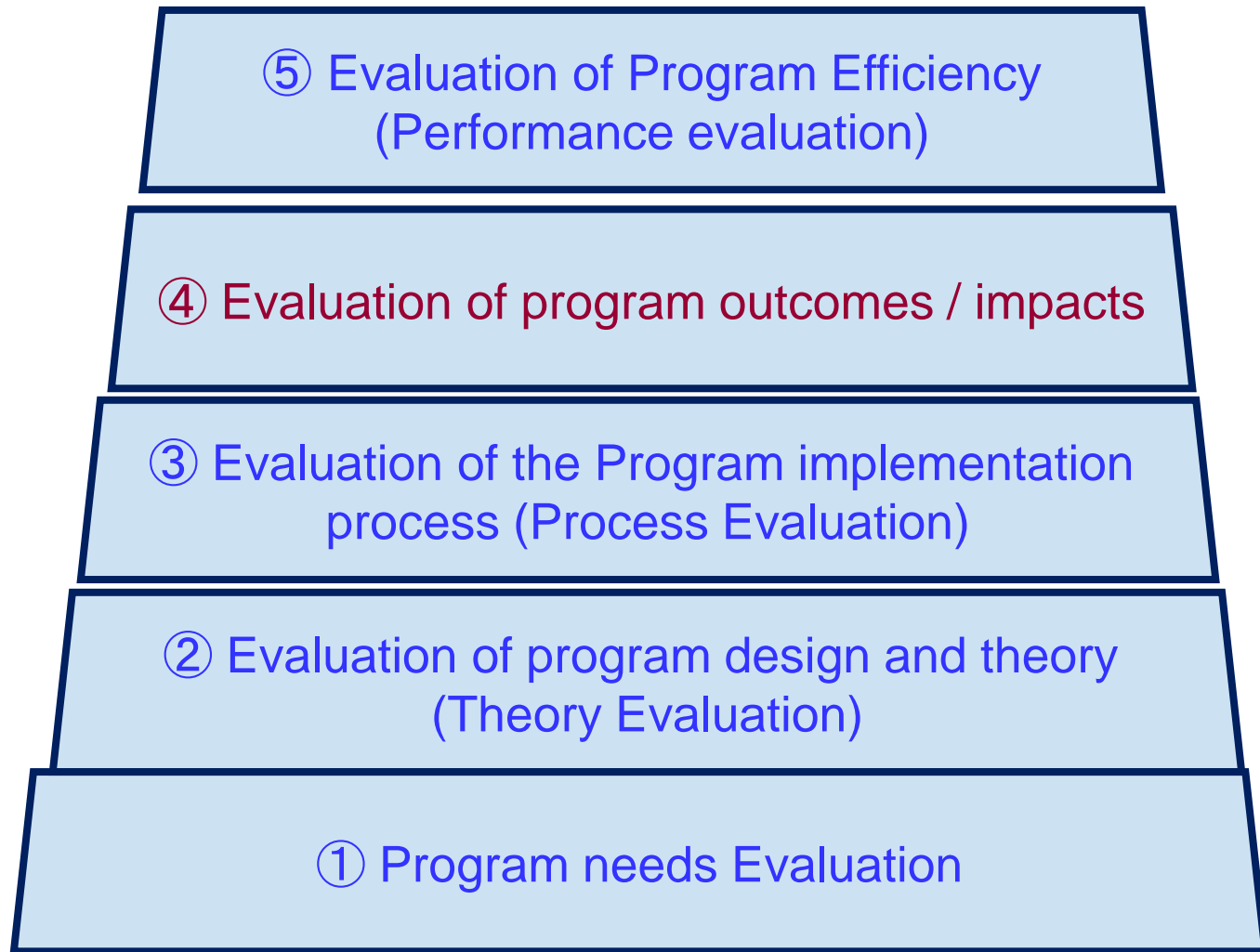


# Goals to be achieved during this lecture

- To understand ways (in general terms) to evaluate a program
  - ✦ To understand the positions and the mutual relations both of the "hierarchy of 5 levels of evaluation" and "Program Theory", which are two key theoretical foundations in the implementation of program evaluation.
- To understand the characteristic features of the program evaluation that lead to the acquisition of evidence-based knowledge and understand how they are reflected in the policy evaluation.
- To understand the importance of accumulation and integration of the evidence in order to carry out a scientifically based evaluation of the program.
- Evaluation planning (evaluation design) is essential for the proper implementation of a evidence-based program evaluation. To understand its general characteristics and points to pay attention of the evaluation planning, as well as to understand the position and mutual relationship of "5 levels of evaluation" and "Program theory".

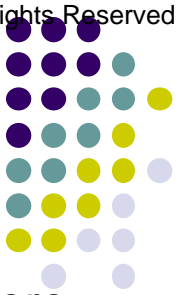
# What is program evaluation?

- Program Evaluation is an approach using the method of sociological research in order to systematically consider the effectiveness of the social program.
- This method is designed to adapt the program to the political and organizational environment and to provide useful information for social actions that improve the social standing (Rossi et al. 2004:16)
- Systematic review of performance: It is important to carry out a systematic, comprehensive and evidence-based evaluation, consisting of 5 aspects, forming evaluation levels (#).
  - ✦ (#)① adaptability to needs (needs assessment), ② relevance of program design and concept (evaluation of program theory), ③ expediency of the intervention process (process evaluation), ④ program effects (outcome assessment/impact assessment), and ⑤ performance (performance evaluation)



## 5 levels of program evaluation

Source: Rossi et al. (2004:80) partially updated



# 5 Levels of Program Evaluation

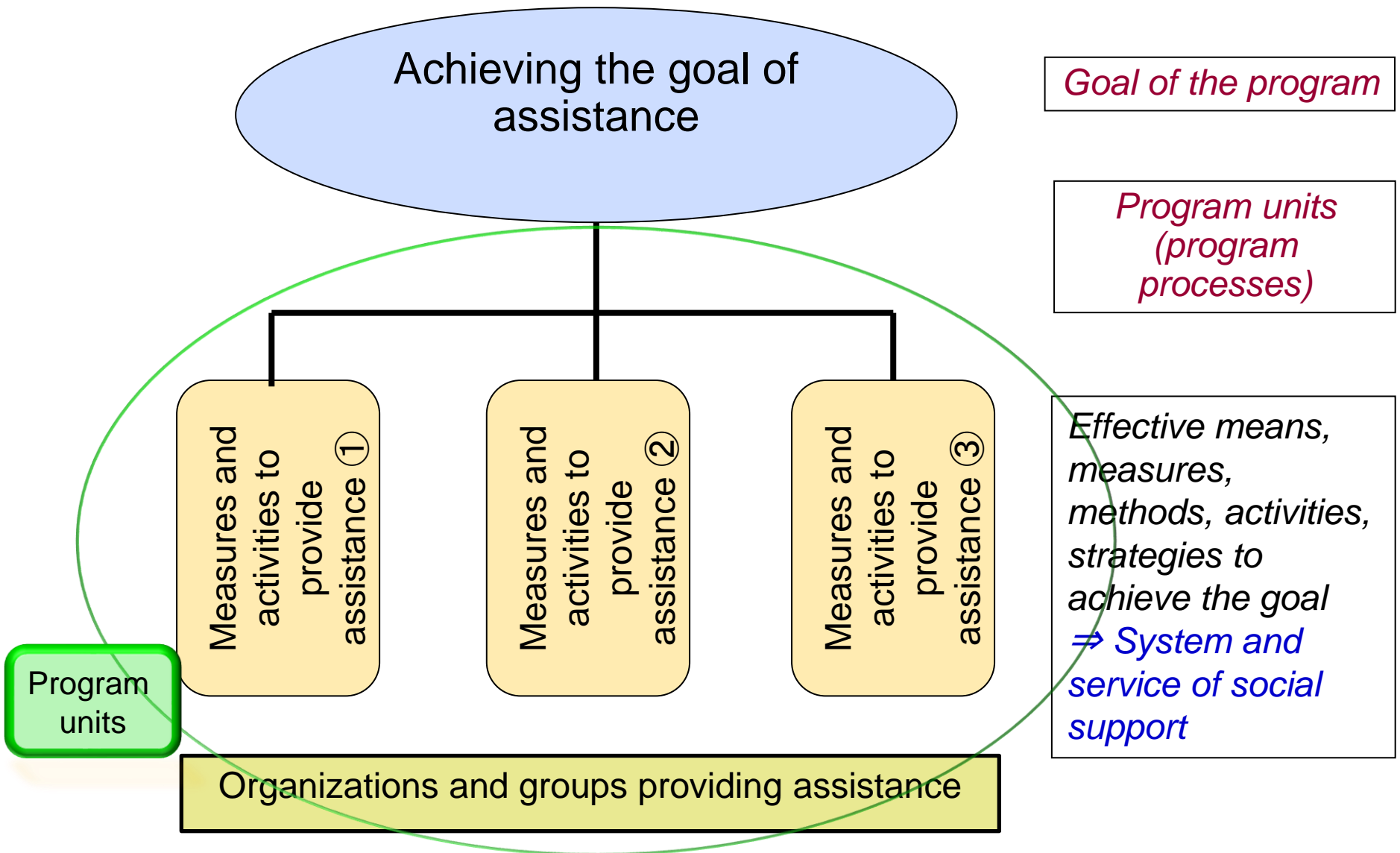
- Assessment of needs for the program: an assessment in terms of whether there are program needs, and if so, which program is most appropriate to meet those needs.
- Assessment of program design and theory (analysis of program theory): evaluation of program design.
- Assessment of the program process and implementation (process evaluation): assessment of the relevance of the program implementation process and implementation
- Assessment of program outcome/impact (evaluation of the program outcomes / impacts): evaluation of the outcomes/impacts that have emerged during the provision of services under the program.
- Assessment of program costs and efficiency (performance evaluation)

⇒⇒ *It must be taken into account that when evaluating a program, importance is attached to “outputs and outcomes”, but the evaluation is not limited to this.*

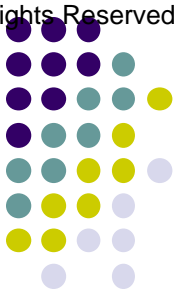
- ✦ In particular, when evaluating at the design and development stages of the program, as well as at of the formation and improvement stages of the program, it is important to properly ask “evaluation questions” according to their levels.

# What is the social program that is the subject of evaluation?

- Designed to solve social problems and improve the social environment, organized, planned and usually lengthy events.
- The purpose of the program and the units of the program (Oshima et al., 2019:3)
  - ✦ Purpose of the program: social problems and social environment to be solved and improved
  - ✦ Program units (program processes) : units of effective, well organized and planned processes (structure / functions / processes) in order to achieve the goal
- Program theory explores what social outcomes should be sought and how programs should be implemented by transforming program structure into a means-ends relationship. It provides an “theoretical chart” of assessment questions for each stage.



## About “program units” in social programs



# Program Evaluation and Evidence-Based Practices (EBP) Program

- Evidence Based Practices (EBP)
  - ✦ Evidence-based consistent practice program that improves the effect of assistance to its users ← formed as part of the program evaluation
- Trend towards policy evaluation with a focus on evidence-based program evaluation (EBPM) ← There is a global trend towards the development and use of effective program models, including evidence-based practice programs as a means of implementing policy.
- Activities for integrating and using the results of the assessment, designed taking into account the “Evidence-Based Policy Development” (EBPM: Evidence-Based Policy Making)
  - ✦ Cochrane Collaboration
  - ✦ Cambell Collaboration
  - ✦ National Institute for Health and Care Excellence (NICE) (United Kingdom)
  - ✦ The “What works?” Movement (UK), etc.

# Level of Evidence in Evidence-Based Practice (EBP)

*“The higher the level (the smaller the figure), the higher the expediency of the results”*

level	Source of information
I a	Meta-analysis of the results of several randomized controlled trials (RCTs)
I b	At least one randomized controlled trial (RCT)
II a	At least one well-designed randomized controlled trial (RCT)
II b	At least one well-designed quasi-experimental study of a different type
III	A well-designed non-experimental study, such as a comparative study, a correlation study, a case-control study, etc.
IV	Reports and opinions of the expert commission or the clinical experience of reputable specialists

(Agency for Health Care Policy and Research : AHCPR  
[Agency for Health Research and Quality : AHRQ])

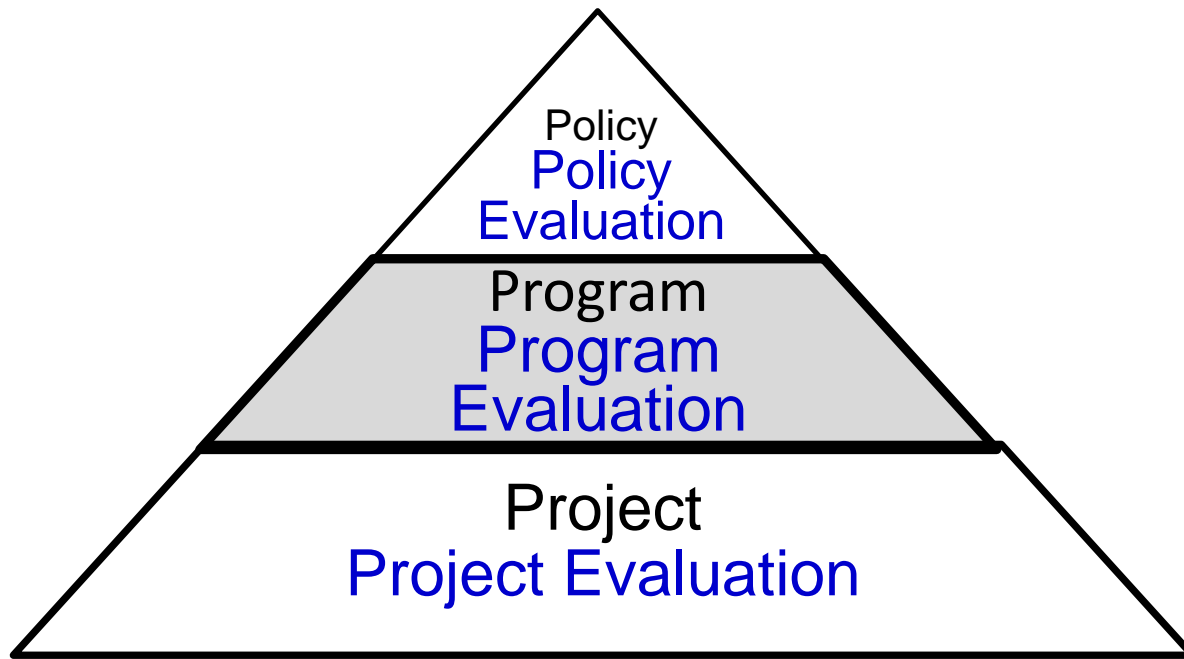


Figure 2: Policy system / policy evaluation system

(Yamaya, supplement to 2020: 3)

Table 1: Examples of the policy system and approaches to evaluation

Policy	Programs	Projects
Children's Welfare Act (1947, amended 2017) Convention on the Rights of the Child (ratified in 1994)	<ul style="list-style-type: none"> <li>① Child Abuse Early Detection Program</li> <li>② Program for the normalization of relations between parents and child</li> <li>③ Program for solution of the problem of child poverty</li> </ul>	<ul style="list-style-type: none"> <li>* Nationwide telephone number of the Child Welfare Advice Center</li> <li>* District Councils for Children in Need of Protection (District Child Protection Networks)</li> <li>* City, township and village centers for helping children and families</li> <li>* «Family Home»</li> </ul>

[See Professor Yamaya's paper \(October 8\)](#)

(Yamaya, 2020: 7)

# Program evaluation and policy evaluation

- Policy evaluation is carried out on the basis of ②' policy design which is consisted of ① policy objective and ② policy tools.
- ② The policy measures (from a scientific point of view) is evaluated on the basis of "several programs that make up the policy."
- Policy evaluation is carried out ② "on several programs", which are policy measures, based on the results of (evidence-based) evaluation of programs, including evaluation of the effectiveness, efficiency and relevance of implementation.
- Policy evaluation is carried out through a comprehensive assessment and reasoning of the ②policy measures, ②' policy design, effectiveness, efficiency and fairness. (Yamaya, 2012)
- To the extent possible, evaluation is carried out in a scientifically based way, in particular, by the method of sociological research in the process of program evaluation.
- Whereas public values are explicitly reflected in policy evaluation processes. For example, the final judgment on the results of policy evaluation is made by the population and, sometimes it depends on the election results, a change of power.

# Why is a program evaluated? :

## *The point of view of the formative assessment necessary for the policy evaluation*

- ① Distinguish between a significant programs and an ineffective programs (comprehensive evaluation)
  - ② Start a new program to achieve the desired effect and improve an existing program (formative evaluation)
- ✦ *When evaluating a program in social welfare, education and health, where personal services are provided, formative evaluation carried out to improve the existing program or to develop new effective programs are of no small importance  
⇒⇒ **to bring to the level of the program on the principle of evidence-based practice***
  - ✦ *Formative evaluation is carried out by systematically using several levels from the hierarchy of evaluation levels.*

# Diagram Schematic diagram of the development of "effect models"

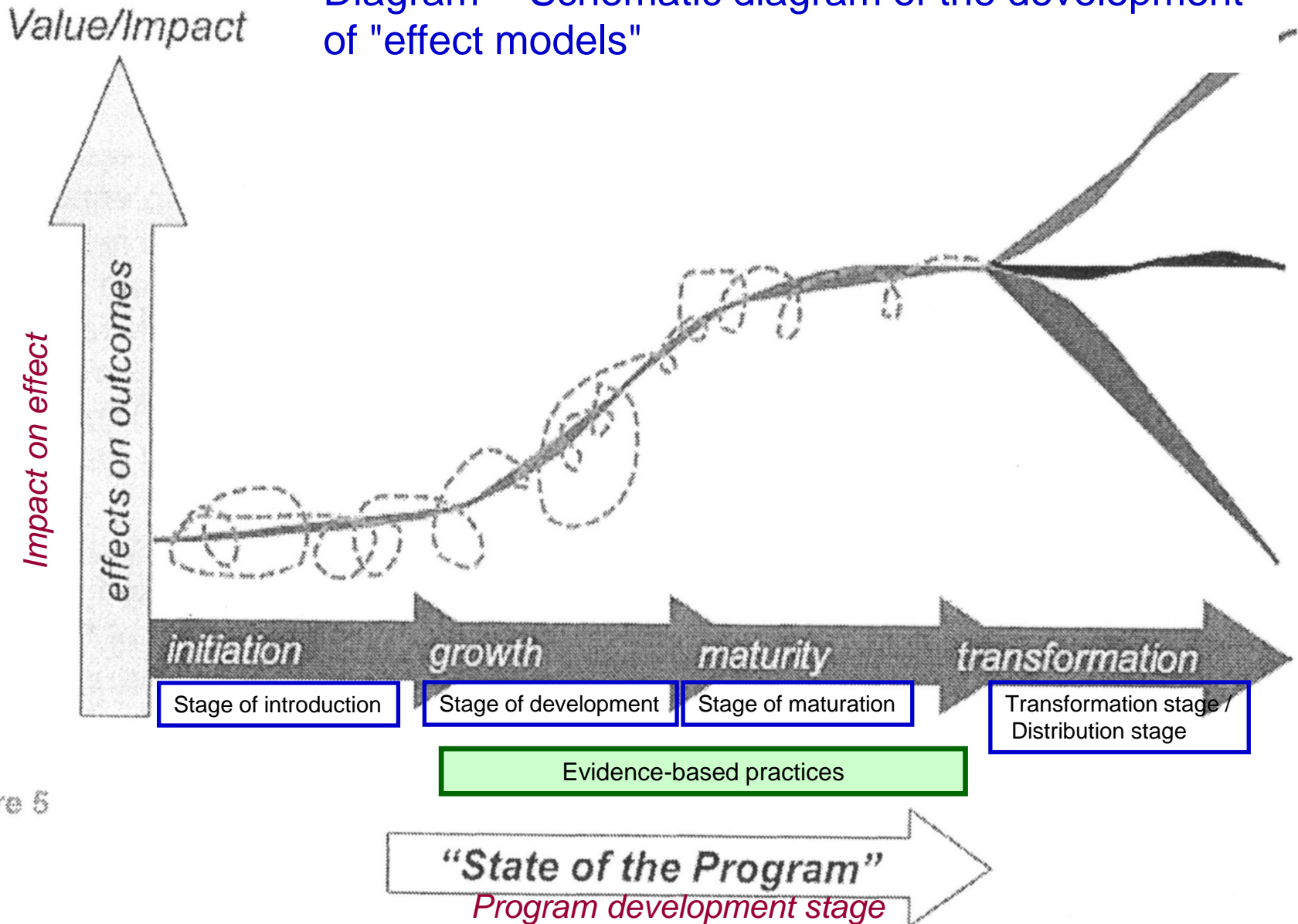
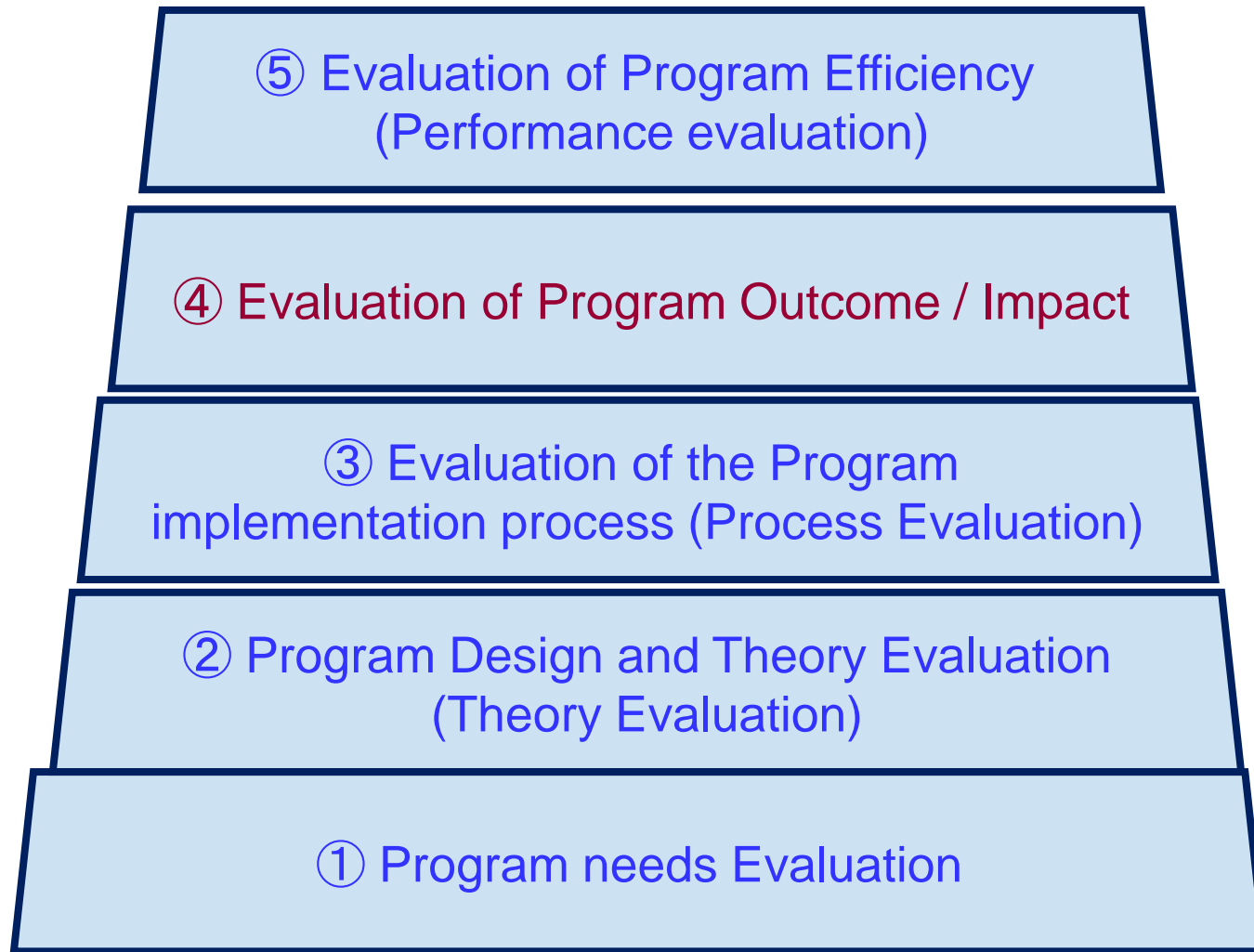


Figure 5

# Classification of social programs in terms of the level of evidence

- Evidence Based Practices Program (EBP)
  - ✦ Effective program based on sufficient evidence
- Best Practices Program (Best Practices)
  - ✦ Less scientifically sound than "evidence-based practice (EBP)", but grounded in practice.
- Program for which expert consensus has been reached (Expert Consensus)
  - ✦ Recommended by many professional experts
- Unreasonable program
  - ✦ In the course of practice, the validity of the effectiveness of the program is gradually achieved
  - ✦ Program that lacks the validity of effectiveness



## 5 levels of program evaluation (repetition)



# Types of program evaluation ①: needs assessment

- Identify social issues to be addressed
- Establish the scope of these problems ~when, where, what is the value?
- Identify the target groups of the program and identify them. Whose needs?
- Describe the target groups. Gender, age, region, diseases, social origin, etc.
- Describe the nature of the needs for services. Analyze the relevant services that are provided to target groups.
- Set program goals.
- Outline assumed stakeholders in the program.
  - ◆ ⇒⇒ *one of the assessments of "relevance"*

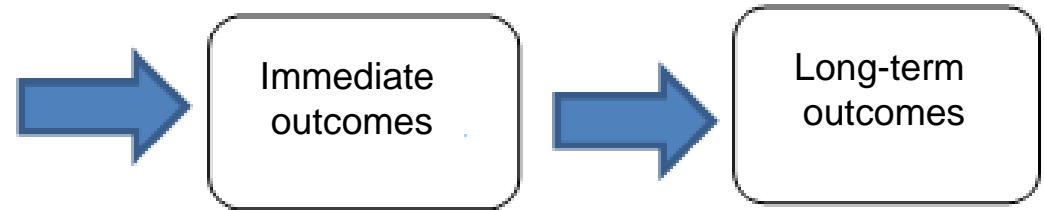
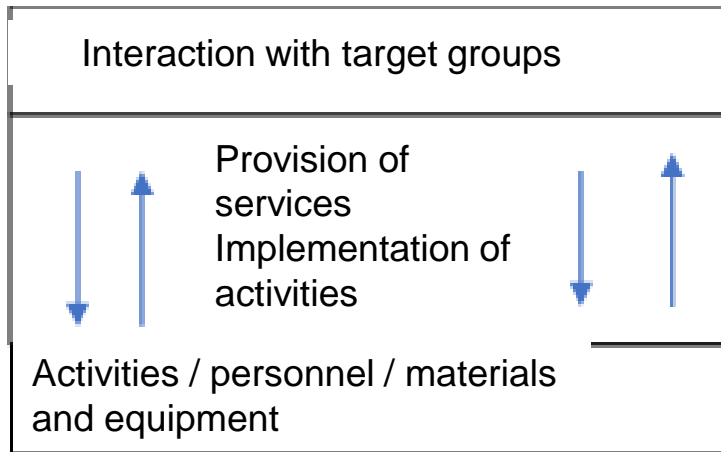
## Types of program evaluation ② : Theory Evaluation

- Program theory is a group of hypotheses regarding cause-and-effect relationships and program elements that give a clear perspective of what effects social programs can give, and what elements can affect the effects.
- Simply put, it is a "technical drawing of effects in relation to the structure, functions and results of a social program".
- The program theory consists of the impact theory, which shows cause-and-effect relationships with respect to the effects of the program and the process theory with respect to the elements of the program.
- It is expected to create an outstanding program that allows you to give the best results by carefully considering the program theory of individual programs.

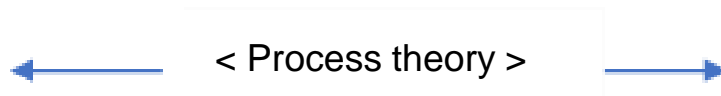
◆ ⇒⇒ *one of the assessments of "relevance"*

# Program Theory by Rossi et al. (2004) with a focus on the interaction between personal service providing programs and users

## Service usage plan



## Program organization plan



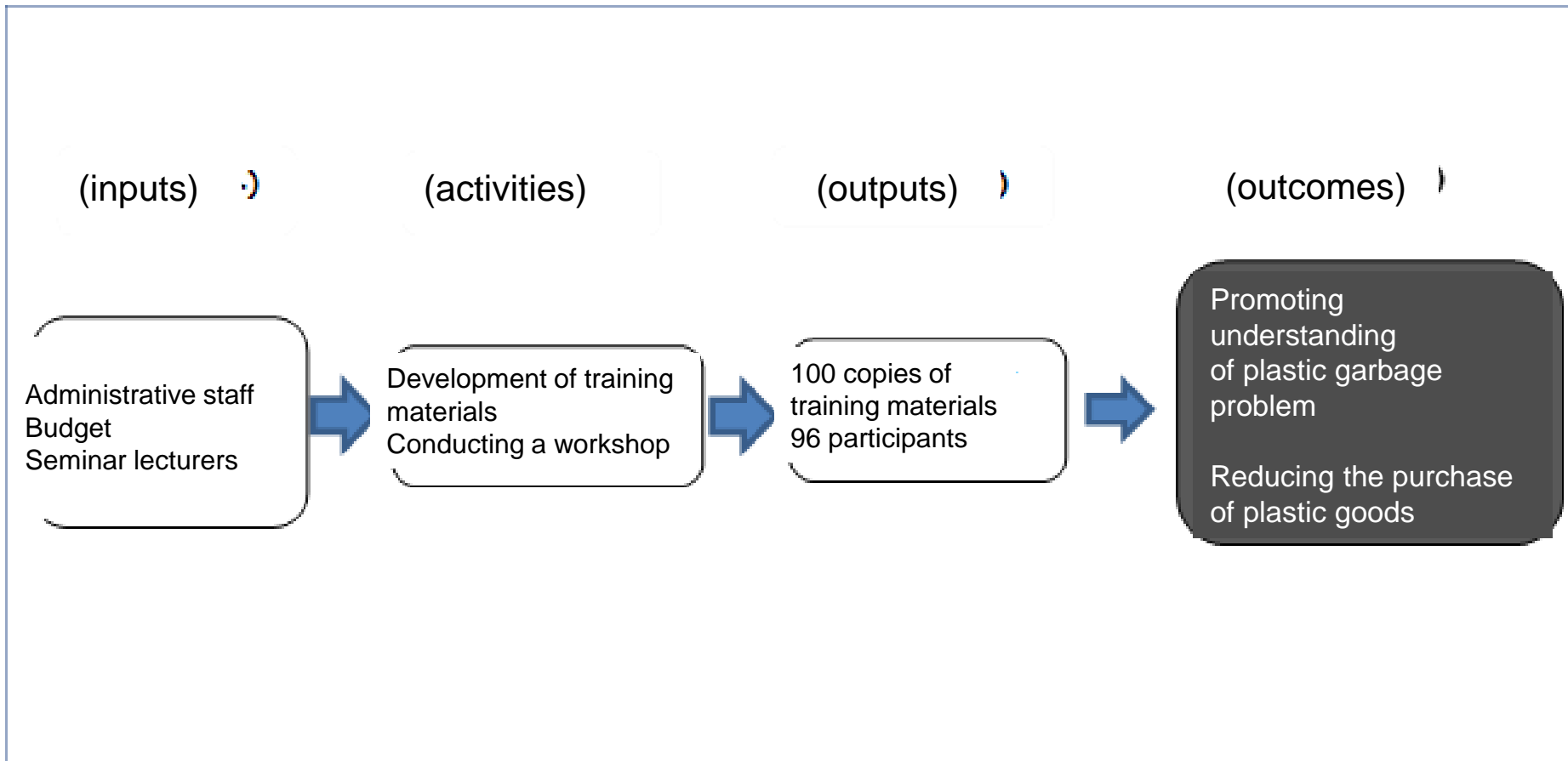


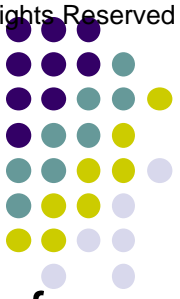
Diagram: Example of a simple logical model  
(Holding workshop on plastic garbage problem)

## Types of program evaluation ③: process evaluation (1)

### 2 types of process evaluation

- 1. Process Evaluation Evaluation  
 is an independent process by the evaluator not integrated into the day-to-day work carried out.
- < 2 roles of Process Evaluation >
  - (1) Evaluation on the effectiveness of program management and service provision. It is important for the process of a new program developing into stable one. ⇒ Study aimed at program process. Quality evaluation analysis is important.
  - (2) Evaluation to find out the effective elements of the program. Analysis of relationships with analysis of results and analysis of effective elements of care is carried out. ⇒ Study aimed at development of a fidelity rating scale (see next slide). Quality evaluation analysis is important.
- 2. Process monitoring used as a continuous program management system
  - ✚ MIS: Management Information Systems
  - ✚ Medical history, as an electronic medical record, etc..
  - ✚ Evaluation of Performance measurement, etc.

## Types of program evaluation ③: process evaluation(2) «Fidelity evaluation»



- It is analyzed how individual programs meet the requirements of effective evidence-based program (EBP) models (Bond et al. 2000)
  - ✚ It is carried out by identifying effective program elements in relation to the program goal (McGrew et al. 1994)
  - ✚ This is part of the evaluation process and aims to assess the degree of compliance and fidelity to effective program models.
- Since a program implemented in accordance with effective evidence-based program models is expected to produce good results, it has recently been given great importance.



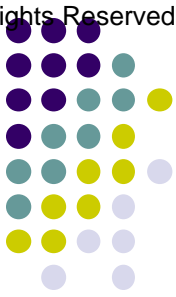
## Typical Fidelity scale: Dartmouth Fidelity Scale of Persistent Treatment in the Community (DACTS: Dartmouth Assertive Community Treatment Scale)

- This scale was developed as a scale for assessing the fidelity of "in the community" models by researcher Mr. Teague et al. (1998) and is widely used globally.
- 5-level rating scale, consisting of 28 criteria in 3 areas, such as "staff placement", "organizational structure", "features of services".
- The evaluator, as a third party unrelated to the program, receives information from the following various sources and summarizes it for evaluation.
- Sources of information: monitoring the assistance provided during home visits, monitoring the progress of the meeting, receiving interviews from the team leader, service coordinator, service user, clarifying records.



**Table 4-2 ACT: Criteria for fidelity evaluation**

<p><b>Human resources: structure and composition</b></p>	<ol style="list-style-type: none"> <li>3. Full responsibility for medication services</li> <li>4. Responsibility for emergency medical services</li> <li>5. Responsibility for inpatient care</li> <li>6. Responsibility for the hospital discharge plan</li> <li>7. Termless provision of services (termination rate)</li> </ol>
<ol style="list-style-type: none"> <li>1. Small number of patients who visit doctor</li> <li>2. Group approach</li> <li>3. Frequency of program meetings</li> <li>4. The leader of the group also carries out practical activities</li> <li>5. Personnel turnover</li> <li>6. No vacancies in the group</li> <li>7. Availability of a psychiatrist within the staff</li> <li>8. Availability of a nurse within the staff</li> <li>9. Availability of a specialist in drug addiction and alcoholism within the staff</li> <li>10. Availability of a professional specialist within the staff</li> <li>11. Program size (sufficient number of employees)</li> </ol>	<p><b>Features of the services</b></p> <ol style="list-style-type: none"> <li>1. Services provided at the local level</li> <li>2. Policy to ensure that no one drops out of treatment</li> <li>3. Mechanism for active participation</li> <li>4. Intensity of services (time of services alone per week - long)</li> <li>5. Contact frequency</li> <li>6. Treatment along with an individual support system</li> <li>7. Individual treatment of drug addiction</li> <li>8. Group treatment by the method of double diagnosis</li> <li>9. Using dual diagnosis models</li> <li>10. Role of consumer in treatment group</li> </ol>
<p><b>Organizational structure</b></p>	
<ol style="list-style-type: none"> <li>1. Clear requirements for entry</li> <li>2. New entries are supported at a low level</li> </ol>	



## Correlation between the fidelity scale and the decrease in the number of days spent in the hospital

Dartmouth Scale, sub-items	Rate
<b>Generally</b>	<b>0.60</b>
<b>Sub-scale for human resources</b>	<b>0.54</b>
Ratio of patients and employees	<b>0.19</b>
Group size	<b>0.35</b>
Availability of a psychiatrist in the group	<b>0.28</b>
Availability of a nurse in the group	<b>0.49</b>
<b>Sub-scale for organizational structure</b>	<b>0.56</b>
Patient load distribution	<b>0.65</b>
Daily group meetings	<b>0.49</b>
Direct provision of services by the manager	<b>0.46</b>
24 hour services	<b>0.55</b>
<b>Sub-scale for services</b>	<b>0.33</b>
Number of home visits	<b>0.31</b>
Total number of contacts	<b>0.59</b>

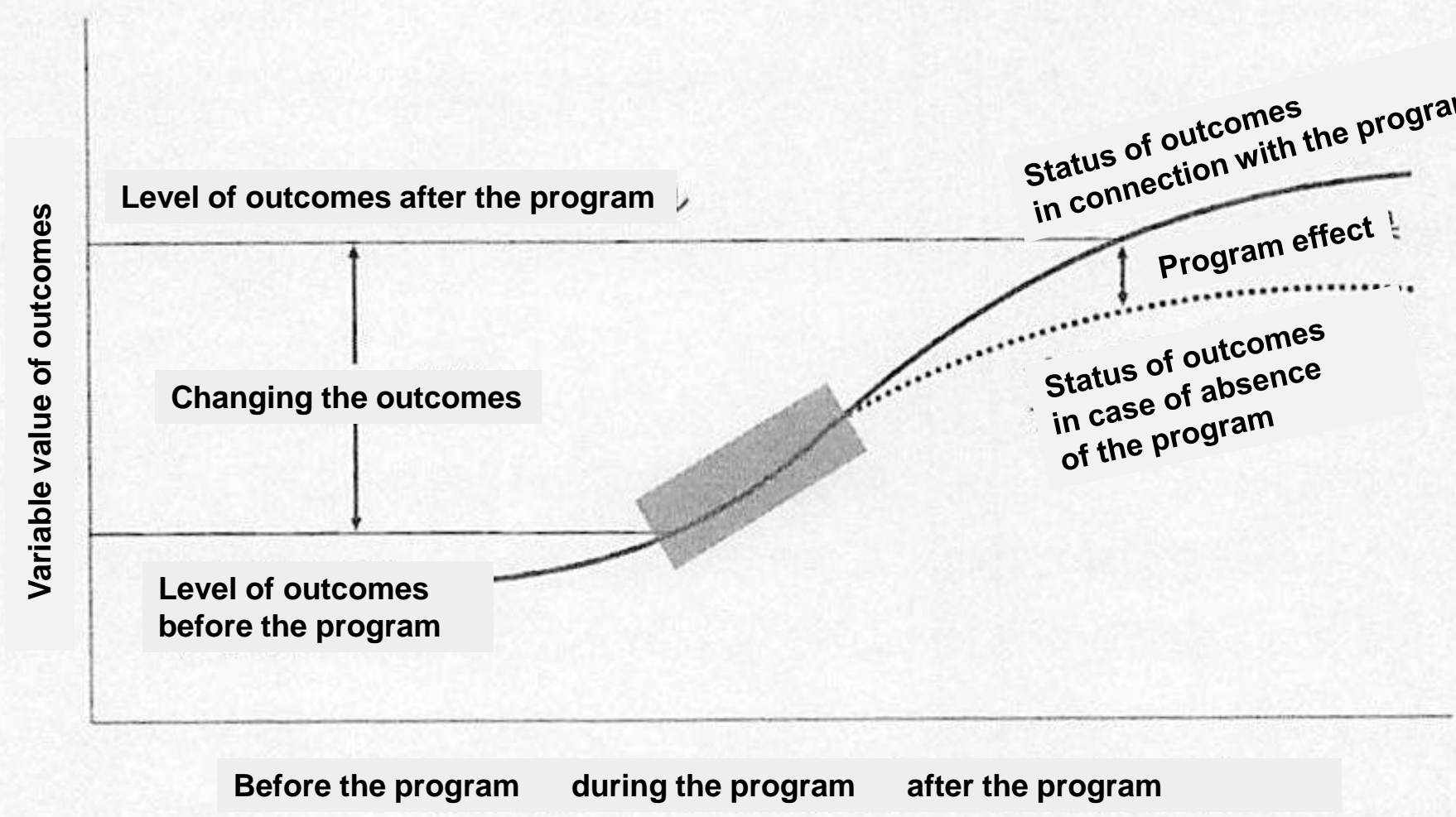
Note: n=18, Regional Mental Health Center (McGrew et al. 1994)

## Types of program evaluation ④: Evaluation of outcome / impact

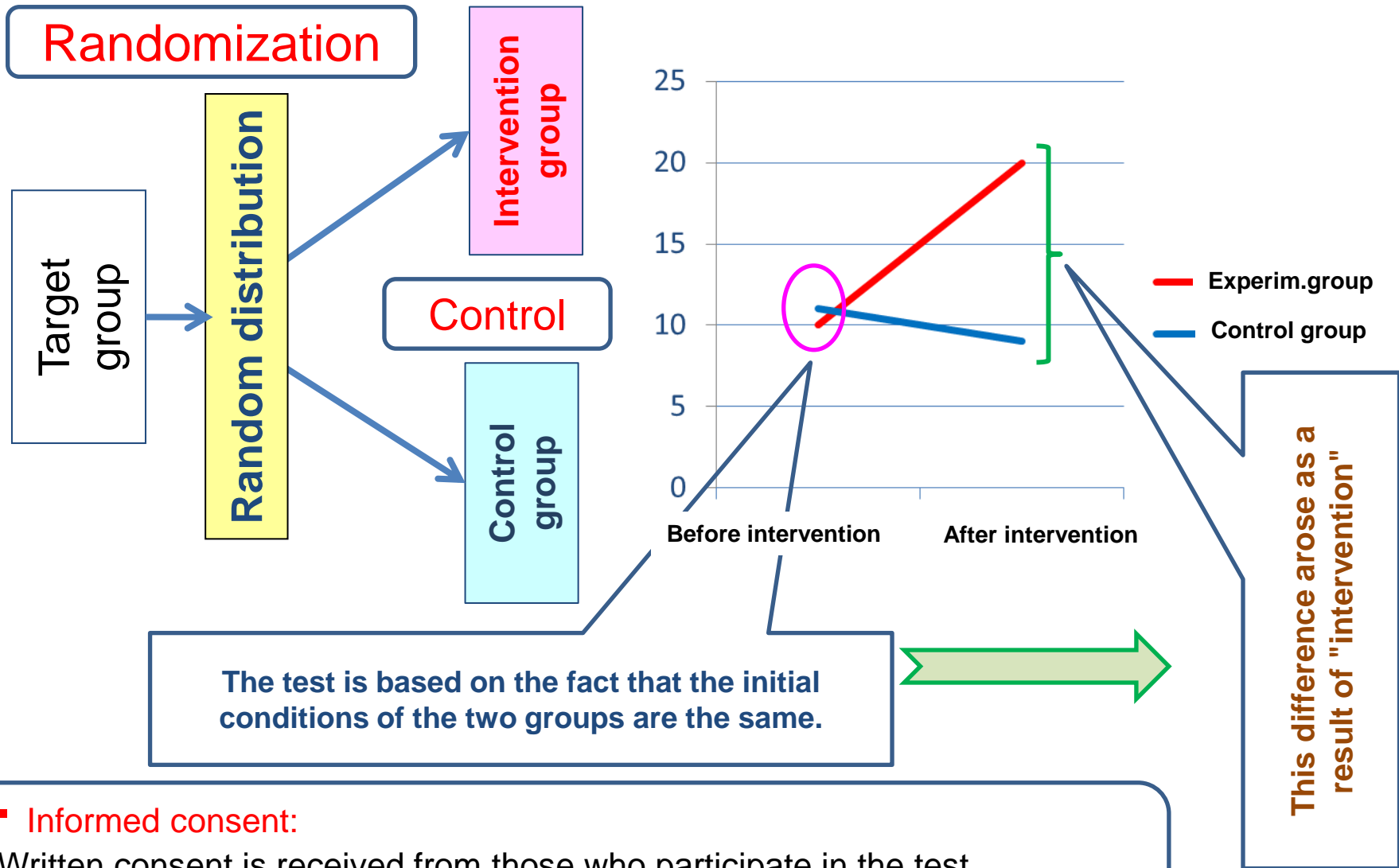
- program implementation and program outcomes are linked  
 ⇒⇒⇒ «performance and impact evaluation»
  - ✦ Determination the impacts of the program is finding out that this program itself is a factor of special change and influence, that is, establishing cause-and-effect relationships.
  - ✦ Target group undergoes changes due to influence factors other than the impact of the program.
  
- **Impact evaluation: control over influencing factors by selecting equivalent target groups:**
  - ✦ Experimental models, randomized controlled trial (RCT):  
 An ideal tool for establishing cause-and-effect relationships. The control group, which has no intervention, and the experimental group, which has an intervention, are compared, while in these two groups, apart from the intervention, the conditions are the same.
  
- **What kind of impact assessment is used?**
  - ✦ Sometimes there are cases that, in addition to the efforts made to implement the RCT, time and money constraints, as well as moral constraints, force appraisers to compromise. The severity of the impact assessment is determined depending on the social context. The appraiser, observing the principle of "good enough", determines the appropriate method of evaluation.
  
- **Gross effect (change in outcomes) and net effect (change as a result of the program=impact)**  
 ⇒⇒ *see the next slide*

# EXHIBIT 7-A

## Level of outcomes, change of outcomes, program effect



# Randomized Controlled Trial (RCT)



The test is based on the fact that the initial conditions of the two groups are the same.

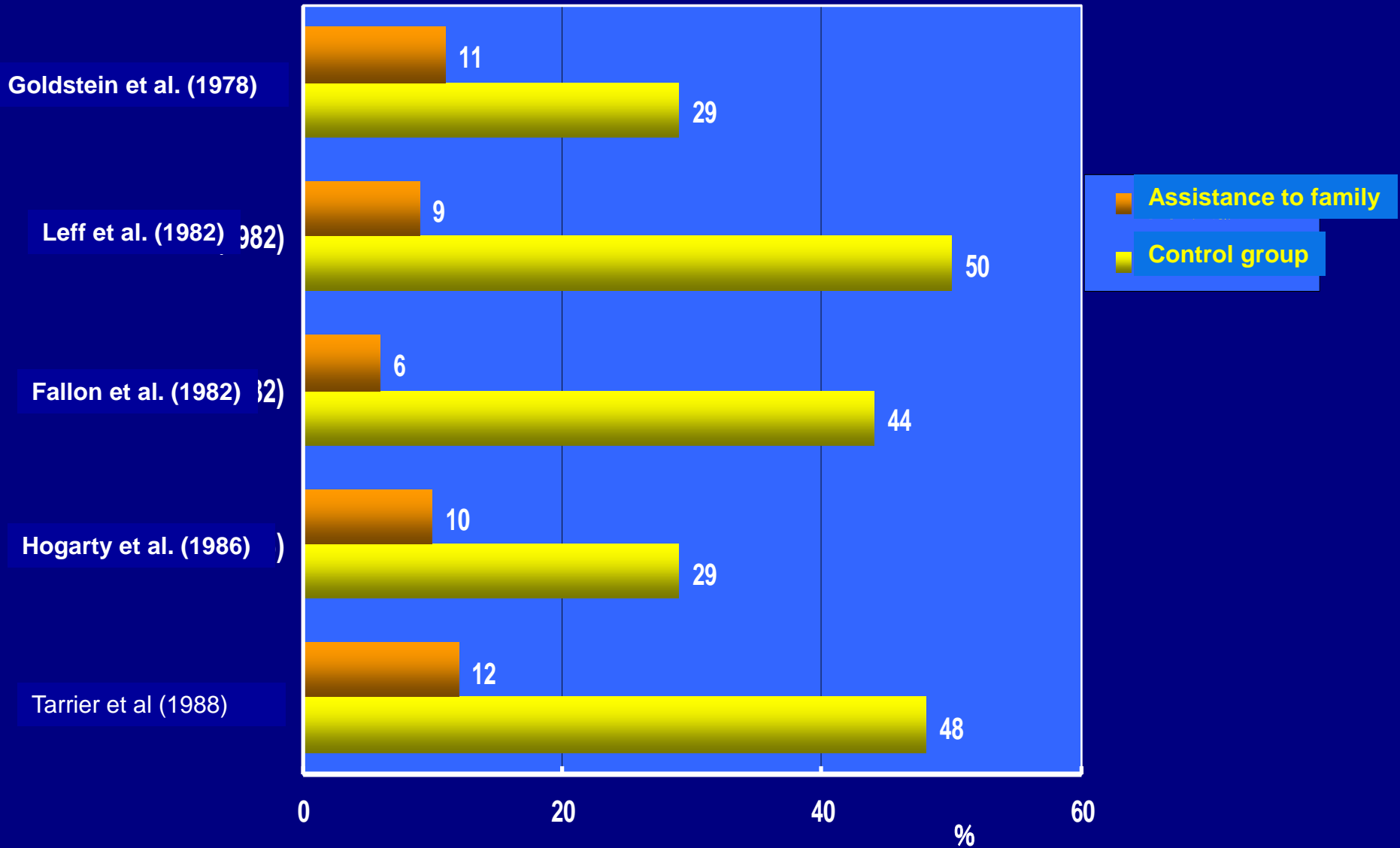
This difference arose as a result of "intervention"

## ▪ Informed consent:

Written consent is received from those who participate in the test.

✘ **Assessment of the difference between the experimental and control groups using statistical inference**

**Table: Comparison of relapse rates during the implementation of the family assistance program (forecast in 9 months)**



# Effectiveness of the family assistance program

## ■ Meta-analysis results :

Comparison of odds ratios in 6 studies (n=350 people) that meet the requirements for reducing the number of relapses:

6 months after the intervention	0.30(0.06–0.71)
9 months after the intervention	0.22(0.10–0.37)
2 months after the intervention	0.17(0.10–0.35)

## ■ Effects other than number of relapses:

Effect of improving the expressed emotions of families

Effect of reducing the burden of families

Effect of increasing the social functions of people with disabilities themselves

## Types of program evaluation ⑤: Efficiency evaluation

- Whether the benefits gained from the project can justify the costs spent for its implementation, is very important question from a social point of view, although the “focus” of program evaluation is the question of how sufficient the net effect (impact) of the program is.
- The measurement of costs and results (impacts) and the assessment of their relationship is called an assessment of efficiency. ⇒ judgment of "performance"
- Comparing the costs and the benefits is one of the most important issues when deciding whether to expand, continue, or cancel the social program,
- The analytical method by which only monetary costs are estimated and compared to the effects are called "cost-effectiveness analysis", and the analytical method by which the benefits of the program are also monetarily estimated is called "cost-benefit analysis".
- It is generally accepted that, just like impact assessment, efficiency evaluation is applied only to a sufficiently structured, matured and stable program.

Source: Rossi et al. (2004: 331-343)

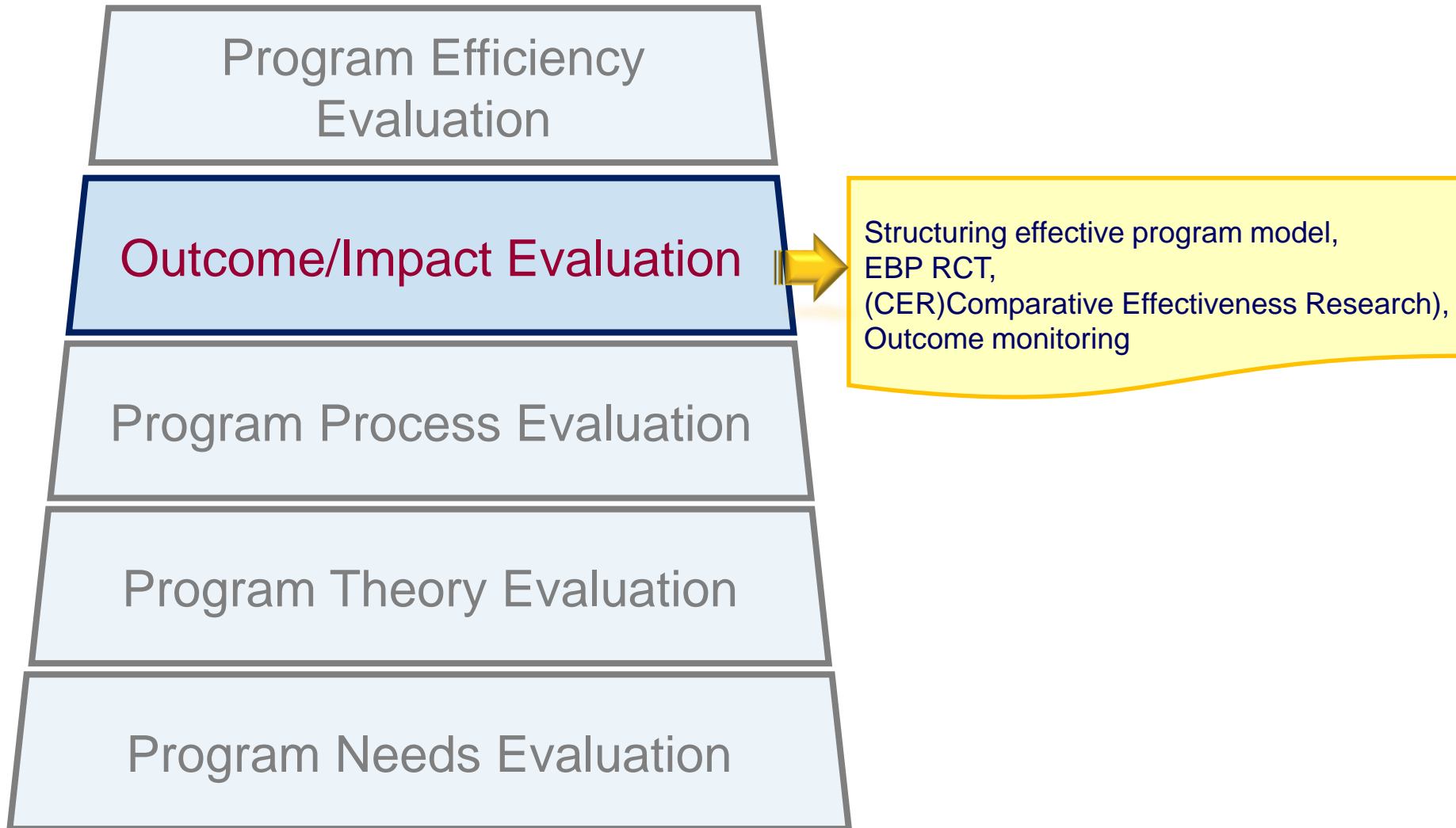
# Program Evaluation Planning ~ What is evaluation design?

- Program evaluation planning refers to general planning that covers all the details, such as the selection of target groups in a (scientifically based) sociological study, determining the number of subgroups in groups (individual target subgroups) for the study, the method of selecting these subgroups, the data collection interval, the method of comparing the results of the study, etc. (Weiss, 1998:87)
  - ✚ Evaluation questions ⇒ What to find out during the evaluation? Evaluation content
  - ✚ **Evaluation design** ⇒ As regards the design of the study for evaluation
- Program evaluation is a highly organized socio-political activity that aims to improve a difficult social situation. Therefore, for its proper implementation, it is necessary to design a well-considered and considered evaluation plan in advance. (Rossi et al. 2004:32-34)
  - ✚ Evaluation design from a scientific standpoint is important, but
  - ✚ It is necessary to take into account the reality that the evaluation design of social reality is developed on the principle of "good enough", taking into account various resources, the situation of stakeholders and the political situation.

# 5 evaluation levels and evaluation design

- Evaluation design from the viewpoint of “summative evaluation” and “formative evaluation”
  - ✦ ⇒⇒ There are different ways to use the evaluation of each level of the 5 evaluation levels.
- When implementing a “summative evaluation”, a proper evaluation design regarding the outcomes and impact is required.
  - ✦ Whenever possible, an evaluation design with high internal validity, such as RCT (randomized controlled trial), should be adopted. At the same time, it is required to increase the external validity.
  - ✦ The subject of evaluation can only be such program models, which, using the same program theory, have obtained certain outcomes (a certain level of compliance with the models has been established) in the course of the fidelity evaluation.
- When implementing a “formative evaluation” from the point of view of effectiveness, “needs assessment”, “theory evaluation” and “process evaluation” are organically and systematically used, representing the lower levels and supporting the level of “results/impact assessment” in the hierarchical structure of the evaluation.
  - ✦ **Evaluation at the design and development stages:** “needs assessment” and “theory evaluation” are used comprehensively.
  - ✦ **Assessment at the stages of formation and improvement:** “Process evaluation” and “Theory evaluation” are carried out, combined with “outcome/impact evaluation”

## Summative evaluation of social programs

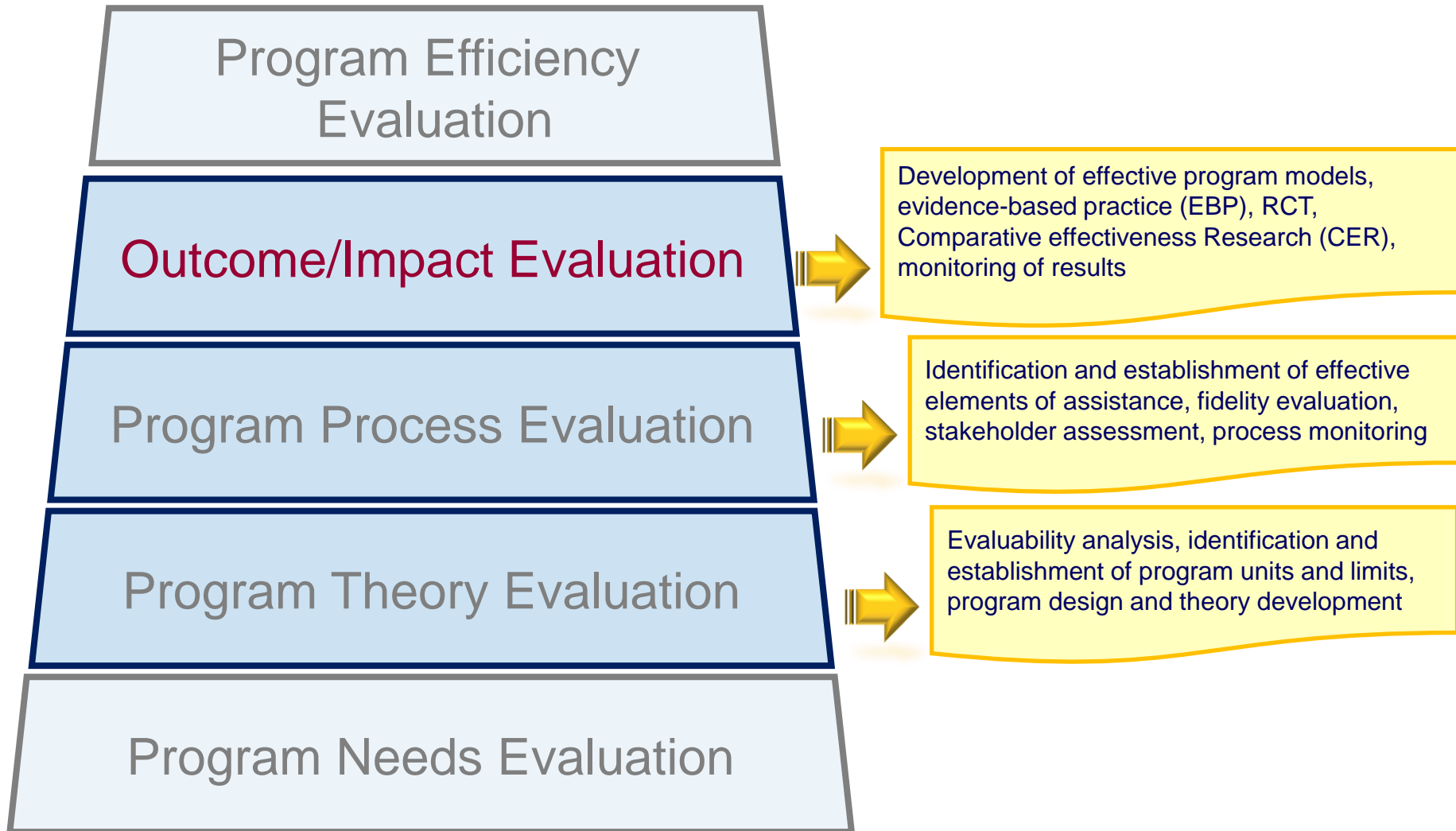


Summative evaluation of social programs  
**Outcome/Impact Evaluation**

# Design of summative evaluation

- To the extent possible, an evaluation design with high internal validity, such as a randomized controlled trial, should be used. At the same time, we need to try to increase the external validity.
- Impact evaluation design with high internal validity
  - ✦ ① Trial Design: Randomized Controlled Trial (RCT) Design
  - ✦ ② Quasi-Trial Design: Regression Discontinuity Design (RDD)
  - ✦ ③ Quasi-Trial Design: Statistical control design
  - ✦ ④ Quasi-Trial Design: Compliance Design
  - ✦ ⑤ Single group design (failure to perform intergroup comparison)
- Impact evaluation design with high external validity
  - ✦ ① Design of a randomized controlled trial being developed by several institutions as part of a joint study
  - ✦ ② Meta-analysis based on the accumulated results of evaluations carried out by the method of randomized controlled trial
- Time series comparison
  - ✦ ① Time series analysis with an increased number of measurement moments
  - ✦ ② Comparing two moments in a row      ③ Cross-group comparison only in case of evaluation at the end of the program

## Formative evaluation of social programs



Evaluation of social programs at the stages of formation and improvement and the hierarchy of evaluation

# Formative evaluation design

- From the point of view of effectiveness, "needs assessment", "theory evaluation" and "process evaluation" are organically and systematically used, representing the lower levels and supporting the level of "outcomes/impact assessment" in the hierarchical structure of the evaluation.
- When analyzing and considering the ratios of evaluation levels, as well as when developing criteria for individual levels, research in the form of a qualitative assessment is useful (see the next slide).
- With regard to "understanding context" and "understanding dynamic information", qualitative assessment plays an important role in understanding the relationship between the outcomes and the process of the program, as well as in a systematic understanding of the implementation of the program.
- In addition, qualitative assessment has a high adaptability to program theory. Consequently, it is well known that the use of qualitative data makes it possible to make the logic of the theory of the implemented program more understandable (Weiss=2014:352-353)
- Organizing a workshop (working group meeting) is also important, and contributes to reaching consensus on the program among stakeholders and increasing coherence on this issue.
- One of the ways to clarify the relationship between the outcomes and the process of the program by the method of qualitative assessment is the use of the fidelity evaluation scale.

# Conclusion

- For the implementation of evidence-based policy evaluation, the methodology of evidence-based program evaluation is important, which creates evidence-based practice (EBP) programs.
- There is a need for a deep understanding of the features of program evaluation that lead to scientific evidence, and due consideration should be given to how these are reflected in policy evaluation.
- In order to evaluate the program, it is necessary to clearly understand how to evaluate each level of the hierarchy of 5 levels.
- For the implementation of Evidence-Based Policy Making (EBPM), proper implementation of an evidence-based program evaluation is essential. For this, an appropriate assessment design must be used.
- In order to advance the development of evidence-based policy and its practical application, it is important to carry out program evaluation (formative evaluation), efforts and the establishment of a mechanism for integrating and combining scientific evidence obtained from evaluations.

Thank you for your attention !

