

Impact Measurement of Social Impact Bonds: How to Promote Social Innovation and Transparency

Hideaki BABA*, Takayuki YOSHIOKA**

Evidence-based policies have been introduced in several countries to improve public service innovation and reduce government spending. Impact measurement has also been attracting attention as a tool to evaluate outcomes under evidence-based policies. Unlike conventional numerical evaluation, impact measurement focuses on the economic or social impact produced by a program, and the impact must be verified by evidence. Social Impact Bond (SIB) is well-known for its unique funding scheme, and its main structure is an outcome-based payment using impact measurement. We conducted interviews with multiple SIB stakeholders in the United Kingdom, and found that performance is measured not only as impact, but also as outcome and output. It is important to conduct program or policy evaluation more comprehensively in order to measure the impact accurately.

Key words: evidence-based policy, outcome-based contract, program evaluation, policy evaluation

1. Introduction

Governments in many countries are required to reduce expenditure in severe financial situations. However, thoughtless budget cuts might cause a deterioration of public service quality. Thus, outcome-based contracts are introduced in some countries such as the U.K. and the U.S. to provide an incentive to achieve the expected level of outcome.

Outcome-based contracts are called Payment by Results (PbR) in the U.K. and

*Professor, Faculty of Business and Commerce, Kansai University

** Associate Professor, Institute of Global Human Resource Development, Okayama University

Pay for Success (PFS) in the U.S.; however, PbR and PFS do not simply mean outcome-based payments. They are schemes of public procurement under which the most suitable organization, whether it is a public or private entity, provides public services. In addition, attention to so-called “impact measurement” has been increasing due to outcome-based contracts. A standardized method of impact measurement has not been established yet, but many methods have been in development.

2. Impact Measurement under Evidence-based Policy

Performance of social services is usually measured as output, outcome or impact. In general, output means the amount of intervention provided by a program, and outcome means an objective achievement accomplished by such intervention. In contrast, impact refers to net long-term outcome, although the concept of impact is often not clearly defined, as mentioned in The Guild (2010). In Social Impact Bonds (SIBs), impact is frequently measured using proxy indicators. Figure 1 shows an example of impact measurement methods with the net outcome concept.

Net outcome is measured as a change in outcome achieved by a certain program. As shown in Figure 1, when we compare two programs to reduce jail days, Program B seems to be preferable because future jail days are fewer. However, when we consider the impacts of the programs, Program A is better because it achieved a greater change in outcome.

Evidence-based policy is one of the significant driving forces to introduce impact measurement. In the U.K., evidence-based policy was adopted in various fields under the Blair and Cameron administrations such as medical care, education, and job training. In the U.S., evidence-based policy and evidence-based

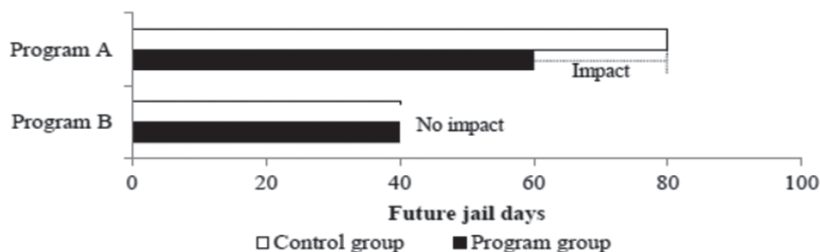


Figure 1: Method of Impact Measurement

Source: MDRC (2013, p.18)

evaluation were also introduced under the Obama administration to select effective and efficient programs using data collected from rigorous research.

Under such circumstances, impact measurement is becoming one of the well-known tools to estimate outcomes of public services provided under the evidence-based policy. The characteristics of impact measurement applied to public services are as follows:

- (1) Impact is not always substantiated in a short time, but sometimes takes several years to appear.
- (2) Impact is required to generate some budget cut effects.
- (3) Measured impact is sometimes used as criteria for payments for public services under outcome-based contracts.

The evidence-based evaluation is the most outstanding characteristic of impact measurement. To define impacts and outcomes, The Social Impact Investment Taskforce (2014, p.23) describes generally accepted indicators as “impact language.” The Taskforce also claims that supporting data infrastructure is essential for impact measurement. Big Society Capital produces a Social Outcome Matrix concerning nine social service areas.

In the case of a youth employment support project, the output may be the number of youths who have completed job training programs, and the outcome may be the number of people who have obtained jobs after completion of the programs. However, as for impact, it requires evaluation of broader and longer outcomes because if the programs have succeeded, welfare payments will be reduced and income tax revenues will be increased. To improve performance-oriented social service provision, the evidence-based impact measurement has been implemented in some countries.

To collect data for the evidence-based evaluation, The Social Impact Investment Taskforce (2014, pp.19–20) requires five data qualifications such as materiality, reliability, comparability, additionality, and universality. In addition, New Philanthropy Capital (2014) developed a “four-pillar approach” to determine the level of evidence required and to select data sources for evaluation.

3. Impact Measurement of Social Impact Bonds

SIBs are well-known for their unique funding scheme, and have already been implemented in several countries. According to Social Finance (2010), an

outcome-based payment contract is the main structure of SIB funding schemes, and simultaneously regulates methods of impact measurement of social services.

However, the impact measurement of SIBs is partly different from common impact measurement methods such as Social Return on Investment (SROI), because the impact measurement of SIBs often focuses on outcomes related to the reduction of government spending. In addition, several SIBs have recently been introduced into social service programs to improve public welfare rather than to considerably reduce government spending.

To explore these unique characteristics of the SIB impact measurement, we conducted interviews with government executives, intermediaries, service providers, and social investors in the U.K. These interview results are shown in Table 1.

The performance of social services is generally measured by output and outcome. Output is usually evaluated by measuring the quantity of provided services, and outcome is often evaluated by measuring the quality of achievement. However, both output and outcome are evaluated by measuring quantitative indicators in SIBs. In this study, we define output as indicators which do not directly determine long-term program objectives but affect short-term objectives, and outcome as those which directly relate to long-term program objectives but are not evaluated by comparing a target group with a control group. In addition, we define impact as the change of outcome which is actually induced by the SIB's intervention.

The Peterborough SIB project adopted a typical impact measurement method, namely measuring the reduction in the re-offending rate through a quasi-experimental matched control group (propensity score matching). The Essex SIB project can also be classified as impact measurement because the project is conducting a historical comparison of 650 cases.

In contrast, St Mungo's Broadway Street Impact SIB project did not conduct a comparison between the target group and the control group, and thus, the outcome evaluation method is not an impact measurement method. However, the target cohort of 416 rough sleepers was selected at the beginning of the SIB project, and thus, distinct outcomes would be measured by monitoring the people who do not require the intervention any longer. In addition, the output indicators were adopted in this SIB project. Reconnections to home countries, reduction in accident and emergency visits, temporary positions, and job training are not the main purposes of the SIB project, but are expected to improve individual well-being in the future.

Table 1: Impact Measurement of SIB projects

| Place | Interview | Term | Objective | Evaluation method | Performance indicators | Commissioner and external evaluator | Budget cut estimation |
|-----------------------------|---|---|--|--|--|--|---|
| Peterborough, England | The Young Foundation (2012.9.4) Social Finance (2012.9.5, 2015.9.16) | From 2010 to 2016 (Canceled in 2015 due to government policy change) | To provide a package of intensive support services for prisoners including housing assistance, drug and alcohol treatment, employment assistance, parenting assistance, and mental health support to reduce recidivism | Difference in differences: Comparison of control group and target group, 3,000 short-term male prisoners (sentences of less than 12 months, aged 18 and older, released from Peterborough Prison) | Impact: A reduction in the re-offending rate that the target group demonstrated, compared to the control group average in the 12 months following release from Peterborough Prison | (1) Commissioner: Ministry of Justice (2) External evaluator: National Institute of Economic and Social Research | Calculation of probability-weighted cost concerning police, court, and prison or community sentence |
| Greater Manchester, England | Teens and Toddlers (2016.9.19) | From 2012 to 2015 | To support disadvantaged young people to improve educational qualifications and secure employment | Cohort analysis: Up to 1,680 young people aged 14 to 17 years old who are in care or soon to be care, have a history of offending or at risk of offending, and other risky behaviors such as drugs or teen pregnancy | (1) Improvements at school: Outcome: Improved attitude towards school Output: Improved behavior at school Output: Reduction in persistent truancy (2) Qualifications: Outcome: Attainment of an accredited Qualifications and Credit Framework (QCF) Entry Level qualification Output: Attainment of basic skills by young people over the age of 16 Outcome: Achievement of National Qualifications Framework (NQF) Level 1-4 (3) Employment: Outcome: Entry into first employment including a training element Output: Entry into sustained employment | (1) Commissioner: Department for Work and Pensions (DWP) (2) External evaluators: National Centre for Social Research Insite Research and Consulting | Estimated potential benefit: Savings from preventing at-risk individuals from falling into long-term unemployment in the future |

| Place | Interview | Term | Objective | Evaluation method | Performance indicators | Commissioner and external evaluator | Budget cut estimation |
|-------------------------|---|-------------------|---|---|---|---|---|
| Greater London, England | St Mungo's Broadway (2014.10.30, 2015.9.14) | From 2012 to 2015 | To provide individual intervention plans to get people off the streets and into stable accommodation, thereby increasing prospects of employment or training and stabilizing health | Cohort analysis: Fixed cohort of 416 rough sleepers who were seen rough sleeping and/or staying in a rough sleeping hostel in the last 3 months, and seen rough sleeping at least 6 times over the last 2 years | (1) Outcome: Reduction of rough sleeping (2) Outcome: Accommodation for over 12 months (3) Output: Sustained reconnections to home countries (4) Output: Reduction in A&E visits (5) Outcome: Sustained volunteering, part-time or full-time work (6) Output: Job training | (1) Commissioners: Department for Communities and Local Government Greater London Authority (2) External evaluator: None (Greater London Authority monitors outcome indicators) | Estimation of the cost which would be borne by the local government if this SIB project was not adopted |
| Essex County, England | Essex County Council Action for Children (2013.10.29) | From 2013 to 2018 | To deliver family therapy in the home through highly qualified therapists over 3 to 5 months with the aim of keeping families together and avoiding out-of-home care | Cohort analysis: 380 youth allocated to 20 cohorts aged 11 to 16 years old at the edge of out-of-home care or custody, and their families | Impact: Comparison of 650 historical cases over a period of 30 months and the number of days spent in residential care for each family | (1) Commissioner: Essex County Council (2) External evaluator: Office for Public Management (OPM) | Comparison of the care cost estimation of the local government and that of this SIB project |
| Whole of England | It's All About Me (2015.9.14) | From 2013 to 2023 | To assist with the difficult adoption of children who are racial minorities, have disabilities, or have siblings | Quasi-experimental trial: 650 or more children in state care, ages 4 to 18, particularly waiting for an adoptive family for over a year | (1) Output: Child enters program registration (2) Output: Child placed with family (3) Outcome: 1st anniversary of placement (4) Outcome: 2nd anniversary of placement | (1) Commissioners: Local governments (2) External evaluators: Assessment of child's needs: South London and Maudsley Hospital, Centre for Children Impact on society, organization, and market: City, University of London, Cass Business School | Future costs, which are avoidable if the adoption succeeds, to preserve and raise children |
| Newcastle, England | Ways to Wellness (2015.9.11) | From 2015 to 2022 | To improve self-management for patients with Long-Term Conditions (LTCs) and reduce costs of secondary healthcare services | Difference in differences: Comparison between 8,500 patients with Long-Term Conditions (LTCs) in Newcastle West Clinical Commissioning Group (CCG) and control group | (1) Output: Improvement of patient's health and wellbeing in 8 areas that relate to the ability to live with LTCs by using patient self-reporting metrics, Wellbeing Star (2) Impact: Secondary care cost savings by comparing the control cohort in Newcastle North & East CCGs | (1) Commissioner: NHS Newcastle West Clinical Commissioning Group (CCG) (2) External evaluator: Newcastle University | Estimated future cost reduction of secondary healthcare services |

Source: Baba (2016, pp.272-273), revised and translated by the author

Similarly, output indicators are used for the It's All About Me (IAAM) SIB project, a child adoption program in England. The ultimate objectives of this SIB project are to keep good relationships between adopted children and adoptive families, and the individual wellbeing of adopted children. However, the number of children registered for adoption and the number of placements with adoptive families are employed as output indicators to achieve the ultimate objectives.

The outcome metrics of Teens & Toddlers Innovation SIB project consist of improvements at school, qualifications, and employment. Improvements at school include improved attitude towards school, improved behavior at school, and a reduction in persistent truancy. Qualifications include attainment of an accredited the Qualifications and Credit Framework (QCF) Entry Level qualification, attainment of basic skills by young people over the age of 16, achievement of the National Qualifications Framework (NQF) Level 1, Level 2, Level 3, and Level 4. Employment consists of entry into first employment including a training element and entry into sustained employment. These outcomes are confirmed in a letter from the teacher or the school, a copy of the certificate, and a letter from the employer (Gustafsson-Wright et al., 2015).

Moreover, in the Ways to Wellness SIB project, Newcastle West Clinical Commissioning Group (CCG) adopted more obscure output indicators. This SIB project evaluates its performance by measuring self-reporting metrics, which are considered to improve the condition of patients with long-term diseases. These output indicators are related to the enhancement of health promotion according to some academic research. Secondary performance is measured as impact based on the savings on health care costs by comparing the health care costs in Newcastle West CCG with those of Newcastle North and East CCG.

4. Transparency of Impact Measurement

The term “impact” used in evidenced-based policy does not simply mean a quantitative scale of outcome. Although a large number of practitioners and researchers propose various definitions of “impact”, the following three concepts seem to be commonly adopted.

- (1) Impact is the difference achieved by a program, not merely numerical indicators.
- (2) Impact should be substantiated by reliable data, which shows a causal relationship between a program and outcome.

- (3) Impact includes not only final, but also initial and interim outcomes, and both economic and social outcomes.

The results of impact measurement are primarily used to judge whether a program has succeeded or not. Success of a project is normally judged based on a contract specification, which has been proposed prior to a tender. Thus, service providers should collect outcome evidence as prescribed in a contract and report it to the contracting authority.

In contrast, even if the expected outcome is achieved, it is not enough to prove effectiveness of policy because the level of effectiveness depends on program expenses. Thus, policies should be evaluated based on a wide range of aspects including economic and social perspectives, emphasizing value for money.

Evidence used in impact measurement should satisfy a certain level of reliability to prove outcomes. Evidence does not necessarily have to be quantitative; however, the contracting authority usually requires the quantitative data of service activities. Reliability of outcome evidence is influenced by data gathering and analysis methods as shown in Table 2.

The reliability of evaluation based on opinions by experts or simple numeric evaluation such as benchmarking, and before and after comparison is relatively low. In contrast, the randomized controlled trial (RCT) method is thought to be

Table 2: Reliability of Outcome Evidence

| Reliability | Method | Description |
|-------------|-------------------------------------|---|
| High | Randomized controlled trial (RCT) | Participants are randomly allocated to a target group, which receives intervention, and a control group, which does not receive intervention. Then, both groups are statistically compared. |
| Middle | Difference in differences | Pre- and post-treatment differences are compared between a target group and a control group. |
| | Cohort analysis | An observational study is conducted for a group, which is determined according to specified characteristics. |
| | Quasi-experimental trial | An empirical study is conducted for a group, which is not randomly determined, to examine the effect of a new program and technology on the group under a limited location or project term. |
| Low | Before and after comparison | Pre- and post-treatment states of a program are compared. |
| | Benchmarking | Outcome of an intervention are evaluated by comparing with standard performance levels. |
| | Opinions of experts and authorities | A commission consisting of experts and authorities discusses and assesses a program. |

Source: Baba (2018, p.37), revised and translated by the author

one of the most reliable methods for evaluation and is widely used in the U.S. (Rangan & Chase, 2015). In the U.K., more moderate methods seem to be used such as cohort analysis and difference in differences.

Impact measurement does not always require a high-level of reliability. The evaluation method used should be determined by considering efforts and expenses to measure impact. However, many service providers only have the data of daily activities, and do not possess the time-series individual data to explain to beneficiaries the changes induced. It is necessary to establish a data collecting system which does not disturb day-to-day operations of service providers.

For program evaluation, the evidence is ordinarily measured by outcome indicators which are specified by the contract to prevent cherry picking and cream skimming. In the U.K., a contracting authority and service providers sometimes negotiate details of a contract, including outcome indicators, through competitive dialogue schemes (Burnett, 2009). Service providers usually collect outcome evidence and report it to the contracting authority. This evidence should be constantly recorded, and the evidence for required outcomes should not be changed during a project, because it is sometimes used as payment criteria as shown in Table 1.

Reliability of outcome evidence is basically maintained by internal governance of the service provider, and the contracting authority monitors the outcome evidence collected as shown in Figure 2. To secure transparency of program evaluation, the contracting authority sometimes requires the service provider to hire a

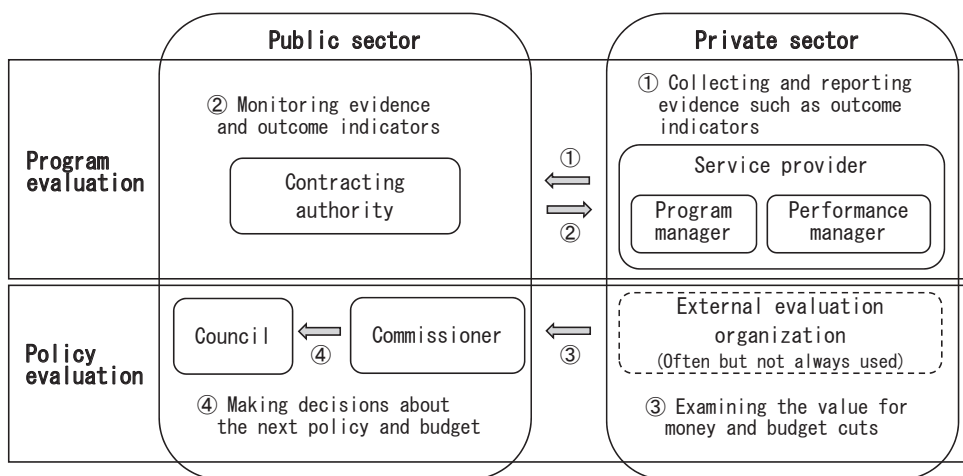


Figure 2: Program evaluation and policy evaluation

Source: Baba (2018, p.35), revised and translated by the author

performance manager who is not directly involved in service provision but is responsible for collecting and analyzing outcome evidence.

In contrast, for policy evaluation, a wider range of evidence is measured including economic and social performances. Based on collected evidence, the level of policy effectiveness is often evaluated by an administrative commissioner or an independent evaluator including a university as shown in Table 1.

In the Peterborough SIB case, the purpose of the independent evaluation was to estimate the level of reduction of re-offending, and the result was used for a payment decision (Anders & Dorsett, 2017). However, in the Essex SIB case, the external evaluation is not used for a payment decision but for the review of the policy from the economic and social perspectives (OPM, 2014). If an independent evaluation is used for a payment decision, it usually will take a long time to complete the outcome payment; thus, such independent evaluation seems to be suitable for multi-purposed policy evaluation.

5. Conclusion

Theoretically, the performance of SIBs should be measured as impact by applying a rigorous and reliable method such as the RCT. However, the RCT method is expensive and labor intensive. Thus, from a practical viewpoint, outcome indicators are useful when properly employing supplemental output indicators.

The primary objectives of a SIB impact measurement are to evaluate the performance of social services and to use such results to reduce government spending. Investors and evaluators usually require data on the impact, outcome and output of SIB interventions. Service providers and intermediaries we interviewed often mentioned that the evidence-based impact measurement is useful to enable innovation in the process of social service provision. Thus, performance-based management seems to be appreciated by service providers.

The Government of Japan is also interested in impact measurement (Japan Cabinet Office, 2016), and is now preparing to introduce it into some policies such as SIBs and the utilization of money in dormant bank accounts to promote public interest activities. Introducing an external evaluation organization could be useful to make sure reliability of impact measurement; however, heavy transaction costs and payment delays would be significant problems. Thus, to keep a balance between reliability and convenience of impact measurement, program evaluation and policy evaluation should be separately adopted. Consequently, it is better that

program evaluation is conducted by the service provider itself, and policy evaluation is conducted by an external evaluation organization.

Impact measurement is a useful tool in supporting evidence-based policy; however, outcomes of public services should not be oversimplified (McHugh et al., 2013, p.249). In addition, outcome-based contract budget cuts do not necessarily reduce actual spending, but it is just estimated spending reduction. Thus, if actual administrative restructuring is not enacted, there is the risk of increased government spending. Consequently, the appropriate practice of program evaluation and policy evaluation is important for impact measurement.

References

- Anders, J. & Dorsett, R. (2017) “A Review of the Matching Process for the Impact Analysis of the HMP Peterborough Social Impact Bond”, National Institute of Economic and Social Research.
- Baba, H. (2016) “Impact Measurement of Social Impact Bond [Sosharu Inpakuto Bondo niokeru Inpakuto Hyoka]”, Tsukamoto, I. & Kaneko, I. eds., *What is Social Impact Bond: Potential of Finance to Promote Social Innovation [Sosharu Inpakuto Bondo towa Nanika: Fainansu niyoru Shakai Inobeshon no Kanousei]*, Minerva Shobo, pp.260–278. (in Japanese)
- Baba, H. (2018) “Does Impact Measurement Improve the Quality of Public Services? [Inpakuto Hyoka wa Koukyousabisu no Sitsu wo Kaizensuruka?] ”, *The Business Review of Kansai University [Shogaku Ronshu]*, vol.63, no.2, pp.31–41. (in Japanese)
- Burnett, M. (2009) “Using Competitive Dialogue in EU Public Procurement: Early Trends and Future Developments”, *EIPASCOPE*, 2009/2, pp.17–23.
- Gustafsson-Wright, E.; Gardiner, S. & Putcha, V. (2015) “The Potential and Limitations of Impact Bonds: Lessons from the First Five Years of Experience Worldwide”, Brookings Institute.
- Japan Cabinet Office (2016) “Promoting Social Impact Measurement: Basic Concept of Social Impact Measurement for Solving Social Issues and Plans to be Taken (provisional translation)”, Naikaku-fu, Working Group on Social Impact Measurement.
- McHugh, N; Sinclair, S.; Roy, M.; Huckfield, L. & Donaldson, C. (2013) “Social Impact Bonds: A Wolf in Sheep’s Clothing?”, *Journal of Poverty and Social Justice*, vol.21, no.3, pp.247–257.
- MDRC (2013) “Financing Promising Evidence-based Programs: Early Lessons from the New York City Social Impact Bond”.
- New Philanthropy Capital (2014) “Building Your Measurement Framework: NPC’s Four Pillar Approach”.
- OPM 2014 “Evaluation of the Essex Multi-Systemic Therapy Social Impact Bond: Interim Evaluation Report, Version 2”.

Rangan, V. K. & Chase, L. A. (2015) “The Payoff of Pay-for-Success”, *Stanford Social Innovation Review*, Fall 2015, pp.27–39.

Social Finance (2010) “Towards a New Social Economy: Blended Value Creation through Social Impact Bonds”.

Social Impact Investment Taskforce (2014) “Measuring Impact”.

The Guild (2010) “Getting Started in Social Impact Measurement”.

Acknowledge: This research is partially based on the discussion of Baba (2016) and Baba (2018), and revised and translated for overseas readers. This research was supported by Kansai University’s Overseas Research Program for the year of 2017, JSPS KAKENHI (16K04021), and MEXT-Supported Program for the Strategic Research Foundation at Private Universities.