日本評価研究

The Japanese Journal of Evaluation Studies

Vol. 5, No. 1, March 2005

研究論文

定量化された評定データのための分析手法の研究 一個人差による影響を受けない測定モデルの開発一

大久保智哉 中川正宣 牟田博光 前川眞一

研究ノート

行政評価の再構築

―理論着眼型評価思考の確立に向けて―

西出 順郎

実践・調査報告

国別プログラム評価の困難性を超えて 一実践的方法論に関する考察と試案―

三輪 徳子

成果志向の道路行政マネジメントの軌跡 ーマネジメント・サイクルのスタートから実践へ一

大庭 孝之 澤田 和宏 森若 峰存 塚田 幸広

日本評価学会 The Japan Evaluation Society

『日本評価研究』編集委員会

Editorial Board

編集委員長

三好 皓一(立命館アジア太平洋大学)

Editor-in-chief

Koichi MIYOSHI

副委員長

西野 桂子(ジーエルエム・インスティチュート)

Vice-Editors-in-chief

Keiko NISHINO

古川 俊一(筑波大学) Shun'ichi FURUKAWA

常任編集委員 Standing Editors 牟田 博光(東京工業大学)

Hiromitsu MUTA

山谷 清志(同志社大学) Kiyoshi YAMAYA

編集委員 Editors 大沢 真理(東京大学)

Mari OSAWA

賀来 公寛(東洋大学)

Kimihiro KAKU

河野 善彦(笹川平和財団)

Yoshihiko KONO

佐々木 亮(国際開発センター)

Ryo SASAKI

田中 弥生(東京大学)

Yayoi TANAKA

長尾 眞文(広島大学)

Masafumi NAGAO

松岡 俊二(広島大学)

Shunji MATSUOKA

Yasuko MURAMATSU

村松 安子(東京女子大学)

森 茂子(日本大学)

Sigeko MORI

事務局 Office 〒135-0047 東京都江東区富岡2-9-11 京福ビル

財団法人国際開発センター内

日本評価学会事務局

TEL: 03-3630-8031, FAX: 03-3630-8095

E-mail: yabuta.m@idcj.or.jp

日本評価研究

第5巻 第1号 2005年3月

目 次

研究論文 大久保智哉 中川正宣 牟田博光 前川眞一 定量化された評定データのための分析手法の研究 ―個人差による影響を受けない測定モデルの開発―	1
研究ノート 西出 順郎 行政評価の再構築	
一理論着眼型評価思考の確立に向けて— ······15	5
実践・調査報告	
三輪 徳子	
国別プログラム評価の困難性を超えて	
一実践的方法論に関する考察と試案― … 27	7
大庭 孝之 澤田 和宏 森若 峰存 塚田 幸広	
成果志向の道路行政マネジメントの軌跡	
—マネジメント・サイクルのスタートから実践へ— ·······45	5
委員会活動	
研修委員会57	
国際交流委員会	
学会賞審査委員会)
分科会活動	
社会・ジェンダー評価分科会 ····································	0
A SA SA CHEMINE STORY OF THE SAME OF THE S	
日本評価学会第5回全国大会報告6	
日本評価学会春季第2回全国大会のご案内	
特定非営利活動法人日本評価学会定款)
日本評価学会誌刊行規定 ····································	
日本評価研究投稿規定82	
日本評価研究執筆要領82	1

【研究論文】

定量化された評定データのための分析手法の研究 -個人差による影響を受けない測定モデルの開発-

大久保 智哉 中川 正宣 牟田 博光 前川 眞一 東京工業大学 大学院社会理工学研究科

okubo@ms.hum.titech.ac.jp

要 約

現在、内部評価報告書を、外部有識者などの第三者によってさらに評価するような2次評価がおこなわれるようになってきている。これら2次評価では、各報告書について評価項目ごとに評定者が評定付け (rating)をして、定量化した上で分析・検討されることが多い。しかし、このように定量化された指標は客観的でなければならないにもかかわらず、実際は評定者によって「甘い・厳しい」の個人差があるという指摘がなされている。そこで、本研究では評定データにおいて定量化された指標の客観性を向上させるために、評定の個人差を除去した指標を算出するため因子分析モデルを導入し、より客観性の高い測定モデルを提案した。また、特殊な欠損パタンにおいても因子分析モデルを適用できることを示し、実際場面での評定データの分析の幅が広がることを示した。また、それらの精度等についても検討を加え、有用な測定モデル・デザインについて提案した。

キーワード

評定データ、組織的欠損、 評定の個人差、 因子分析モデル、 欠損パタン

1. はじめに

従来より、評価手法においては定性的評価・定量的評価や自己評価・第三者評価などさまざまな位置づけの数多くの評価手法が提案され、またその有用性について論じられてきた(佐々木,2003)。

その中で、国際協力機構の事業評価年次報告書に見られるように内部評価報告書を外部有識者委員などの第三者によって、さらに評価するような2次評価がおこなわれるようになってきている。その他に、各独立行政法人に対する評価に関しても担当省庁がおこなった内部評価書について総務省がさらにその評価をおこなうという2次評価もお

こなわれている。これらの 2 次評価は「外部評価の拡充によって評価の客観性向上に資する (国際協力機構, 2003)」という認識のもとにおこなわれている。なお、このような 2 次評価の可能性については Rossi, et. al.(1999) においても論じられている。

これら2次評価は、各報告書について評価項目 ごとに評定者が評定付け (rating) をして、定量化 した上で分析・検討されることが多い。

しかし、このように定量化された指標は客観的でなければならないのだが(国際協力事業団, 2002)、 実際は評定者によって「甘い・厳しい」の個人差があるという指摘がなされている。 たとえば、渡部他 (1988) は、小論文の評価という枠組みにおいて、個人ごとの評定特性が異なることを示している。また、評定時の着目点についても個人差があると言われている (Engelhard, 1996; 山内, 1999a, b)。さらに、評定対象が文章を含む場合には、字の美しさ・汚さの影響 (渡部・曹, 1992) や、評定順序の影響についても報告されている (Bruin & Keren, 2003)。

したがって、このような定量化された評定データを扱う場合には、一つの評価対象に対して複数の評定者に評定させてその平均点をとるなどして客観性を高める必要がある(龍・佐々木, 2000)。

しかし、評定者によって評定する評価対象が異なってしまう場合には評定者それぞれの評定に対する厳しさなどの違いから、定量化された指標を単純に比較することができなってしまう。「ある評価対象には、甘い評定をする評定者が多く評定した」のでは、定量化された指標の客観性が失われてしまう。

評価対象の良し悪しを測定するためにおこなわれている評定が、評価対象の良し悪しによる影響よりも、評定者の甘い・厳しいという個人的要因によって左右されるようでは、信頼性のある指標にはなりえない。

2. 研究の目的・意義

そこで、本研究では評定データにおいて、定量 化された指標の客観性を向上させることを期待し、 評定の個人差を除去した指標を算出する測定モデ ルを示す。また、その上で方法の有用性について 検討をおこない、より客観性の高い測定モデルを 提案することを目的とする。

そこで、次の2点を通じて定量化された評定データの客観性向上を図ることにする。

1. 通常、われわれが手にすることのできる評定 データは、評価対象が持つ真の評価得点に加 えて評定者の個人差による誤差が混ざったも のである。そこで、真の評価得点と評定の個 人差を評定データから分離するために因子分 析モデルを導入する。

2. また、実際の評定場面においてはデータが全て得られない場合が多い。その場合には、因子分析モデルが適用できない。しかし、データが欠損している場合でも評価得点を推定する必要が出てくる。そこで、特殊な欠損パタンにおいては因子分析モデルが適用可能であることを示す。

1. に関しては評定データにおいて信頼性を下げている評定の個人差についてモデリングすることにより、各個人の評定付けに対する特徴を表現すると同時にその個人差による影響を除去した評価得点が推定されることになる。測定した評価得点をそのまま合計・平均するよりも客観性の高い指標が示されれば評価の質の向上にもつながるはずである。

2. に関しては先述のように時間的な都合上、評定対象によって評定者が変わってしまうため通常の因子分析モデルが適用できないといった場合にもいくつかの条件を満たせば因子分析モデルが適用可能であることを示す。また、その際にどのように評定を計画すれば(欠損のパタンを組めば)信頼性が高まるかを示す。これによって、欠損パタンにおいて自由度の高い測定が可能となってくる。

3. 個人差の影響を受けない測定モデル

(1) 完全データ

まず、測定モデルについて論じる前に完全データについて説明を加える。

通常、評定者が評価対象に対して、評定付けをすることによって評定データが得られる。本研究では、表 1 のように列に評定者を配しj で表し、行に評価対象を配しi で表す。つまり、評定者 2 は、評価対象 5 に対して「3」と評定付けしたことになる。

表1では、全ての評定者が全ての評価対象を評 定していることがわかる。このような欠損のない データを本研究では完全データと呼ぶ。

			0-11	1144		- J	•	
評定者	1	2	3	4	5	6	j	n
評価対象 1	1	3	5	3	2	4		5
評価対象 2	2	4	2	3	5	3		3
評価対象 3	4	3	3	5	1	2		3
評価対象 4	4	3	2	5	2	3		2
評価対象 5	2	3	2	3	3	4		1
評価対象 6	1	2	2	3	5	4		2
評価対象 7	2	3	2	4	3	3		1
評価対象 8	3	4	3	2	3	4		3
評価対象 i	:	:	:	:	:	:	:	•
評価対象 N	1	3	4	3	2	3	4	5

表 1 完全な評価データの例

(注 1) 「-」は欠損していることを表す

(出所) 筆者作成

(2) 因子分析モデルの導入による個人差のモデリング

ここで、完全データが得られた場合において、1 因子の因子分析モデルを導入することによって、評 定の個人差をモデリングできることを示す。

まず、因子分析のモデル式は

$$y_{ij} = \lambda_j f_i + \mu_j + e_{ij} \tag{1}$$

で表わされる。なお、データが標準化してあれば $\mu_j = 0$ であるので μ_j の項はなくなることになる。 しかし、本研究では欠損を含む場合も同時に検討するためデータを標準化しないで分析する。

ここで、(1) 式の中の y_{ij} 、 λ_j 、 f_i 、 μ_j 、 e_{ij} はそれぞれ次のような意味を持つ。

- *y_{ij}* は、評定者 *j* の評価対象 *i* に対する評定値を表わす。また、*y_{ij}* は標準化されている必要はない。
- λ_j は、全体の評定因子 (評定の対象としている構成概念) に対する感度を表わしている。 データが標準化されていれば $-1 \le \lambda_j \le 1$ の範囲を取る。 λ_j が負であることは、たとえば、全体が「良い」と評定する傾向にあるものに対して「悪い」と評定をする傾向にあるということである。また、その絶対値は程度を表している。
- f_i は、通常は因子得点とよばれ、評価対象
 i の真の評価得点を表わす。ここで f_i は平均 0、分散 1 である。先ほどの y_{ii} との違い

表 2 組織的欠損を含む評価データの例

評定者	1	2	3	4	5	6	j	n
評価対象 1	1	3	5	3	-	-		-
評価対象 2	2	4	2	3	-	-		-
評価対象 3	4	3	3	5	-	-		-
評価対象 4	4	3	2	5	-	-		-
評価対象 5	2	3	-	-	3	4		-
評価対象 6	1	2	-	-	5	4		-
評価対象 7	2	3	-	-	3	3		-
評価対象 8	3	4	-	-	3	4		-
評価対象 i	:	:	:	:	:	<u>:</u>	:	٠.
評価対象 N	1	3	-	-	-	-	4	5

(注1) 「-」は欠損していることを表す

(出所) 筆者作成

は、 y_{ij} が評定者j が評定対象i に対して評定した値であるのに対し、 f_i は評定者に依存しない、本来評価対象がもっている評価の量 (得点)を表わしている。評価の構成概念そのものと言ってよい。 $-3 \le f_i \le 3$ の範囲で値を取る場合がほとんどである。

- μ_j は、評定者 j が評価対象に対してつける 得点の基準点。つまり、評定者 j の評価の 厳しさ・甘さを表わしている (評価基準であ る)。この μ_j は、評価が 3 件法 (y=0,1,2)で行なわれれば、 $0 \le \mu_j \le 2$ となる。 μ_j が 高いほど評定者 j の評価が「甘い」ことを 意味している。
- e_{ij} は、モデルにおける誤差を表わす。モデルのパラメタで表現できない部分がここに含まれることになる。

つまり、(1) 式は、評定者 j の評定傾向 λ_j に評価対象 i の本来の評価得点 f_i をかけたものに、評定者 j の評定基準 μ_j と誤差 e_{ij} を足したものが、実際に評定者 j が評定対象 i に対して与える評価値 y_{ij} であることを意味している。また、評価値 y_{ij} は評価対象の本来の得点 f_i に評定者 j の様々な評定特性 (評価基準・評定特性) が加わったものであるという見方もできる。

通常、因子分析モデルにおいては項目の背後に 共通因子を仮定するが、本研究の場合には、評定 者の背後に共通の因子を仮定している。こうする ことによって、われわれが測定したい評定者間に 共通な構成概念を測定することができるのである。 また、この因子得点こそが評定データを分析する 際の客観的な指標となる。

(3) 個人特性を表すパラメタと測定モデルの関係

因子分析モデルを導入することによって、次の 3点を表現することが可能になった。

- 1. 評定者 j の評定基準 μ_j
- 2. 評定者 j の評定傾向 λ_i
- 3. 評定の個人差に依存しない評価対象iの評価 得点 f_i

ここで、上述のモデルは評定者の評定基準 μ_j と 評定傾向 λ_j が各個人で異なるという仮定を置いた 測定モデルとなっている。

しかし、場合によっては「評定の傾向について 個人差を認めない」とする評価モデルが妥当とな るかもしれない。また、「測定モデルの推定すべき パラメタ数を減らしたい」という理由から評定傾 向に関しては個人差を認めないことがあるかもし れない。

個人差を表すパラメタを調整することによって、 測定モデルは次のような表現も可能である。評定の 実施者の都合でモデルを選択することもできる。な お、因子分析と分散分析の関係は柳井・前川 (1999) が詳しい。

因子分析モデル 評定傾向の個人差と評定基準の個人差を認めた測定モデル。評定者ごとの μ_j と λ_j と f_i が推定される。なお、以下の(2)式は(1)式を再掲したものである。

$$y_{ij} = \lambda_j f_i + \mu_j + e_{ij} \tag{2}$$

分散分析モデル 評定傾向の個人差は認めずに (評定者間で一定; $\lambda_j = \lambda$)、評定基準の個人差だけを認めた測定モデル。評定者ごとの μ_j と評定者間で一定の λ と f_i が推定される。

$$y_{ij} = \lambda f_i + \mu_j + e_{ij} \tag{3}$$

標準化得点モデル 評定基準の個人差も評定傾向の個人差も認めない(評定者間で評定基準も評定傾

向も一定; $\lambda_j = \lambda$, $\mu_j = \mu$) 測定モデル。評定者間で一定の μ と λ と f_i が推定される。

$$y_{ij} = \lambda f_i + \mu + e_{ij} \tag{4}$$

(4) 特殊な欠損パタンにおける因子分析モデルの 適用

特殊な欠損パタン

先ほどまでは、表1のようにデータが完全であるという仮定のもとでの測定モデルについて論じてきた。データが完全であるということは、全評定者が全評価対象を評定しているということである。

しかし、実際の評定データを測定しようとした場合、評定に参加する評定者全てが評価対象を評価できるとは限らない。例えば、大量の評価対象について評定をおこなおうとすれば、時間的な制約から完全データを得ることはほぼ不可能であろう。

このような場合、評定者が評定している評価対象は実際にその評定値が観測されるが、その評定者が評定していない評価対象に関しては値が欠損していることとなる。そのような状態を表したのが表2である。ここで「-」は評定されなかったことを示す。つまり、評定者3と4は評価対象4までは評定しているが、それ以降は評定できなかったことを示している。

ここでは、Allison(1987) で示された方法を援用することにより、ごく特殊な評定値の欠損に限り、因子分析モデルが適用可能であることを示す。

また、本研究においては、時間的制約等の理由によって評定者が全ての評価対象を評定できなかったために作られた特殊な評定値の欠損を組織的欠損と呼ぶ。なお、この評定値の欠損は測定計画段階で実施者が意図的に作ることができるものである。

組織的欠損は次のような条件に基づくものと する。

- 1. 欠損がランダムに起きたものではなく、計画 者によって欠損部分が割り当てられたもので あること
- 2. 複数の評定対象の群を評定する共通評定者と呼ばれる評定者がいること (評定対象の群と群を結ぶ役割の評定者がいるということ) また、

共通の評価対象を配置する場合も考えられる が本研究では取り扱わない

まず、1点目についてであるが、これは欠損が 評定者の不注意によって起きたものではなく、測 定計画の一環として意図的に欠損させたものであ るということを示している。

次に、2点目についてであるがこれは、表2の 1、2列目のようにデータの固まりと固まりをつな ぐ共通評定者が必要であるということである。こ のことは、それぞれの評価対象を同一尺度上で表 現するための必要条件となっている。

また、図1は組織的欠損の例を示している。色が 濃い部分は評定者の中でも複数の評定対象群を評 定している共通評定者による評定箇所である。ま た、薄い部分は評定者による評定を表している。ま た、白の部分は評定者によって評定されていない 部分である。同様に図2も組織的欠損の例を示し ている。ただし、図3は共通評定者がいないので 組織的欠損になっていない。また、全ての評価対 象の群を特定の共通評定者が結ぶ必要はなく、群 によって共通評定者が異なるという図1のような 測定デザインでも組織的欠損の条件を満たす。

組織的欠損を含む場合の測定モデル

先に示した因子分析モデルは、組織的欠損を含 む場合の評定データには適用することができない。 そこで、組織的欠損をしている場合においても推 定値に偏りを出さずに測定モデルのパラメタ (λ_i)、

 f_i 、 μ_j) を出す必要が出てくる。

Allison(1987) においては、集団で単に個別に因 子分析をするのではなく「集団間で共通の評定者 の評定傾向 λ_i 、評定基準 μ_i は集団が変わっても 一定である」というような制約を置き、多集団間 で同時に推定値を計算している。つまり、いくつ かの評定集団の背後に単一(等質)な因子つまり評 価得点 (f_i) が仮定されるということになる。

ここで、図2のような評定データの場合を考え、 数式を用いて表現する。

たとえば、図2のデータを単に上半分の測定さ れた部分(1)と下半分の測定された部分(2)に分 けた上で、観測された部分のみを用いて

$$y_{ij}^{(1)} = \lambda_j^{(1)} f_i^{(1)} + \mu_j^{(1)} + e_{ij}^{(1)}$$
 (5)

$$y_{ij}^{(2)} = \lambda_j^{(2)} f_i^{(2)} + \mu_j^{(2)} + e_{ij}^{(2)}$$
 (6)

と、別々に因子分析をした場合、 $f_i^{(1)}$ と $f_i^{(2)}$ の 値を単純に比較することはできない。

なぜならば、評価得点 f_i がそれぞれのデータに おいて平均0、分散1を仮定された上で計算され たものだからである。つまり、別々に計算した場 合は(1)のデータで計算された評価得点と(2)の データで計算された評価得点は同じ値でも意味が 違うものになっているのである。さらに、因子得 点が比較できないのであれば、その他のパラメタ についても共通尺度上で比較することはできない ことになる。



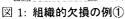




図 2: 組織的欠損の例②



図 3: 組織的欠損ではない例

- (注 1) 濃い色の部分は共通評定者が評定した y_i , である (図 1~図 3)
- (注 2) 薄い色の部分は評定者が評定した y_{ij} である (図 1 \sim \bigcirc \bigcirc 3)
- (注 3) 空白の部分は y_{ij} が欠損している部分である (図 1 \sim \boxtimes 3)

そこで、組織的欠損をしているデータの場合の分析では、(1) と (2) のデータで共通評定者となっている人 (濃い部分) の評価特性を $\lambda_1^{(1)}=\lambda_1^{(2)}$ 、 $\mu_1^{(1)}=\mu_1^{(2)}$ のように制約を課した上で同時に解を求める。

そうすることによって、得られる評価特性パラメタは、当然 2 つのデータ間でも共通尺度上で表すことができるため、因子得点 $f_i^{(1)}$ と $f_i^{(2)}$ も共通尺度上で比較することができるようになるのである。

こうすることによって、組織的欠損を含む場合においても、因子 (評価得点) の等質性を失うことなく推定できるのである。しかし、もし共通評定者がいない場合には、評定集団間で因子の等質性を確保することができないので、得られるパラメタが比較可能でなくなってしまう。

この Allison(1987) における仮定は、「群間で等しい因子がすべての群で同様に測定される」というものである。これを因子の測定不変性 (factor invariance) という (Mulaik, 1972)。扱う評定データが、この因子の不変性の仮定が成り立つものなのかは、実際に推定したあとで適合度指標を吟味すればよい。

期定モデル・欠損パタンの検討

これまでに、評定データに対して因子分析モデルを適用することによって、個人差のモデリングが可能であることを示した。また、その際に個人差のパラメタの数を調整することによって、いくつかの測定モデルも表現できることを示した。さらには、特殊な評定値の欠損状態を作れば(組織的欠損)、データが欠損している場合でも因子分析モデルが適用可能であることを示した。

ここでは、実際に実験をおこないデータを得た上で、組織的欠損を含むデータに対して因子分析モデルを適用したときに、どれくらいの精度で推定できているのかを確認する。また、その際にいくつかの欠損パタンを用意し、欠損パタンごとの推定精度についても検討をおこなう。

さらに、個人差を表すパラメタを調整すること でいくつかの測定モデルが表現できることを示し たが、ここでは、実際に評定データが得られたと きにどのようなモデルで測定した場合にもっとも 信頼性が高い結果を得ることができるかを選択す る方法について検討する。

前川真一

(1) 実験について

まず、測定モデルの妥当性検討のため、完全な 評価データを実験によって得た。

実験時期・対象 2003年8月に、大学生・大学院 生20名に対して実験を行なった。

実験内容 評定の対象に用いたのは、車の画像であった。評価対象から受ける「高級感」のイメージを5段階評価で評定してもらった。評価対象の数は100とした。車の画像はA5サイズの白い紙にカラーでプリントアウトしたものを用いた。また、画像提示の順序については、提示順序による影響を剰余変数としないために、評定者によって提示順序を変えた。

実験手順 実験ではまず、一人一人の評定者に回答用紙を渡した。次に、「今から、実験を始めます。これから渡す車の画像を見て、評定して下さい。一つの評定が終わったら、紙をめくって次の車の画像を評定して下さい。評定はあまり深く考えずに5段階評定で評定付けをして下さい」と教示をおこなった。そして、評価対象を渡し実験をおこなった。

検討用データ 得られた完全データは、20(評定者)×100(評価対象)のデータであった。本研究では、完全データから評定者をランダムに 10 通り並びかえ、さらに評価対象についてもランダムに 10 通り並び替えることにより、検討用データを作成した。したがって、一つの完全データから $10 \times 10 = 100$ の検討用データを作成した。

(2) 測定モデル・欠損パタンの検討内容

ここで、測定モデルを用いての次のような点を検討する。

欠損パタンごとの精度の検討

まず、実験によって完全データを得る。次に、そのデータから特定の意図的な欠損パタンに基づく 組織的欠損を含むデータを複数作る。その上で、完 全データに因子分析モデルを適用したときの推定 値と組織的欠損状態に因子分析モデルを用いたときの推定値のばらつき具合を欠損パタンごとに検討する。このことにより、組織的欠損を含ませる形で評定をおこなわなければいけないときに、より信頼性の高い推定値を得るには、どのような欠損パタンにした方がいいのかに示唆を与えるものと期待される。

共通評定者の特徴が精度に与える影響

組織的欠損を含むデータにおいて、共通評定者は、通常の評定者よりも多くの評定をおこなうことになる。そこで、共通評定者に評定が全体の傾向を反映していないような評定者を配置してしまった場合、推定の精度が下がるのではないかとの仮説について検討をおこなう。これにより、共通評定者を選ぶ際に何らかの示唆が与えられることが期待される。

分析する際の測定モデルの選択

先述のように個人差のパラメタの調整により、 複数の測定モデルを表現することができる。しか し、どのモデルが最も測定モデルに適しているか は、データに依存するので、状況に応じて使い分 けなければならない。ここでは、実際に評定デー タを得た場合、どのようにして複数あるモデルの 中から最も妥当なモデルを選べばいいのかを実際 のデータを用いて示す。これにより、測定モデル の選択に関して、客観的な基準が与えられること が期待される。測定モデルを評定者が恣意的に選 がよりも、客観的な基準に基づいてモデルが選択 されることは評定の信頼性を高めるためには重要 なことであろう。

(3) 測定モデル・欠損パタンの検討方法

欠損パタンごとの精度の検討

各欠損パタンの特徴 本研究において、意図的に 組織的な欠損を持つ評価データを8パタン作成した。用意した欠損パタンについて図4~11に示す。 また、表3においては欠損パタンごとの組織的 欠損状態において評定した数と完全データにおい て評定した数が示されている。その横には、1群 における共通評定者数と評定者数を示した。また、 組織的欠損のない完全データの評定数は 20(評定者)×100(評価対象)= 2000(評定数) となっている。 さらに、先ほどの 100 の検討用データは欠損パタ ンごとにある。

欠損パタンの A1~A3 までは、2 群を共通評定者によって結んだ欠損パタンである。A1~A3 までは順に評定者に対する共通評定者の割合が増えていっている。

次に、欠損パタンの B1~B3 までは、A1~A3 までの評価の群をさらに半数に減らした欠損パタンである。したがって、A1~A3 に比べて全体での評定数がほぼ半数になっていることがわかる。評定数とは、図 1~3 における色のついた部分の面積である。この欠損パタンは、A1~A3 に比べて評価対象も評定者の数も少ないが、1 群あたりの共通評定者の数は同じである。したがって、共通評定者の評定者数に対する割合は欠損パタン A1~A3 よりも大きくなっている。また、これらの欠損パタンについても、B1~B3 に対して順に評定者数に対する共通評定者の割合が増えていっている。

そして、欠損パタン C1、C2 は共通評定者が全ての評価対象を評定するような欠損パタンになっている。C1 と A1 を比較してみると、共通評定者の数と共通評定者が評定する評価対象の数がほぼ等しくなっている。異なる点は、残りの評定者が何群に別れていて、1 群あたりの評定者の数とその評定者が評定する評価対象の数が異なっている点である。また、C2 は C1 に対して共通評定者の数を増やした欠損パタンとなっている。

推定精度の指標 本実験では、組織的欠損状態から推定したパラメタがどれほどの精度によって推定されているかを検討するための指標として、完全データから推定したパラメタ (λ_j, μ_j, f_i) と組織的欠損状態から推定したパラメタ $(\hat{\lambda}_j, \hat{\mu}_j, \hat{f}_i)$ の差の絶対値を推定精度の指標とした。

また、 y_{ij} に関しては完全データの y_{ij} と組織的 欠損状態から推定したパラメタ $(\hat{\lambda}_j, \hat{\mu}_j, \hat{f}_i)$ を用いて計算した \hat{y}_{ij} の差の絶対値を推定精度の指標とした。ただし、各欠損パタンで 100 の検討用データがあるので、実際には、100 のデータでの差の絶対値の平均をとる。



図 4: 欠損パタン A1



図 5: 欠損パタン A2



図 6: 欠損パタン A3



図 7: 欠損パタン B1



図 8: 欠損パタン B2

図 9: 欠損パタン B3



図 10: 欠損パタン C1



図 11: 欠損パタン C2

- (注 1) 図の色は図 1,2 と同様に yii の欠損の有無を表す (図 4~図 11)
- (注 2) 横軸は評定者、縦軸は評価対象を表している (図 4~図 11)
- (注 3) 評定者は 20 名、評価対象は 100 個である (図 4~図 11)
- (出所) 筆者作成 (図 4~11)

表 3 各欠損パタンの全評定数と共通評定者数

	評定数	全評定数	共通評定者数	評定者数
A1	1100	2000	2	11
A2	1300	2000	6	13
A3	1500	2000	10	15
B1	600	2000	2	6
B2	800	2000	6	6
В3	1000	2000	10	10
C1	792	2000	2	8
C2	990	2000	5	10

表 4 各パラメタの範囲に対して誤差が占める割合

	$E[\epsilon_{\lambda}]$	$E[\epsilon_{\mu}]$	$E[\epsilon_f]$	$E[\epsilon_y]$
A1	9.6 %	7.0 %	3.8 %	16.7 %
A2	9.1 %	5.8 %	3.4 %	15.8 %
A3	8.8 %	5.7 %	2.9 %	15.5 %
B1	10.6 %	5.1 %	6.3 %	16.0 %
B2	9.1 %	3.0 %	4.9 %	14.5 %
B3	5.1 %	2.6 %	4.2 %	14.0 %
C1	7.1 %	3.5 %	4.9 %	14.9 %
C2	5.1 %	2.6 %	4.2 %	14.4 %

- (注 1) 評定数とは欠損パタンにおいて得られた y_{ij} の数を表す。 つまり色のついた部分の面積である (表 3)
- (注 2) 全評定数とは完全データにおいて得られた y_{ij} の数を表す。 つまり欠損パタンに限らず 2000 となる (表 3)
- (注3) 共通評定者数とは一つの集団において共通評定者の役割をしている評定者の数である(表3)
- (注 4) 評定者数とは一つの集団における評定者の数である。細かく集団を別けている欠損パタン B の方が少ない (表 3)

(10)

- (注 5) 割合を 100 倍して%表示にした (表 4)
- (出所) 筆者作成 (表 3, 表 4)

それぞれの推定値の誤差の平均を定式すると以 下のようになる。

$$E[|\epsilon_{\lambda}|] = \frac{1}{20} \frac{1}{100} \sum_{j=1}^{20} \sum_{k=1}^{100} |\hat{\lambda}_{jk} - \lambda_j|$$
 (7)

$$E[|\epsilon_{\mu}|] = \frac{1}{20} \frac{1}{100} \sum_{j=1}^{20} \sum_{k=1}^{100} |\hat{\mu}_{jk} - \mu_j|$$
 (8)

$$E[|\epsilon_f|] = \frac{1}{100} \frac{1}{100} \sum_{i=1}^{100} \sum_{k=1}^{100} |\hat{f}_{ik} - f_i|$$
 (9)

$$E[|\epsilon_y|] \quad = \quad \frac{1}{100} \frac{1}{20} \frac{1}{100} \sum_{i=1}^{100} \sum_{j=1}^{20} \sum_{k=1}^{100} |\hat{y}_{ijk} - y_{ij}|$$

共通評定者の特徴が精度に与える影響

ここでは「全体の評定傾向を反映しいている・い ない」ということを指標で表わすため、評価対象 ごとにi個ある評価得点 f_i と特定の評定者jの評 価対象ごとの評定値 y_{ij} の相関係数を用いた。こ の指標は評定者が全体の傾向を表わしてない、つ まり評価得点 ƒ と評定者の採点値の相関関係がな いと小さくなる。この相関係数を本研究では被説 明度と呼ぶ (因子分析の枠組みでは共通性と呼ぶ)。

ここでは、共通評定者となった評定者がどれほ どの被説明度をもっていたかをそれぞれ共通評定 者が同じデータ10個ごとに算出した。ここで、一 つのデータに着目するとそのデータには共通評定 者が複数いるので、共通評定者となった評定者の被 説明度の平均をとり、それをそのデータにおける「共通評定者の被説明度」とした。「被説明度」とは特定の評定者の評定値がどれほど評価得点fによって説明されているのかを表わしたものである

被説明度を算出した上で、共通評定者が同じデータ 10 個ごとに「共通評定者の被説明度」と「各データにおける完全データからの推定値と欠損状態からの推定値のずれ」の平均値を取って、10 種の欠損パタンそれぞれの「被説明度」と「推定値のずれ」を算出し、その 2 つの変数の共変性を見るために相関係数を算出した。たとえば、この相関係数が負であるということは「共通評定者が全体の傾向を反映していないようないわば「変わり者」が群と群を結んだ場合、推定の誤差は大きくなる」ということを意味している。

分析する際の測定モデルの選択

比較検討をおこなった測定モデルは先述の3つ のモデルである。

因子分析モデル 評定傾向と評定基準の個人差を認めた測定モデル ((2) 式)

分散分析モデル 評定傾向の個人差は認めずに (評 定者間で等値)、評定基準の個人差だけを認 めた測定モデル ((3) 式)

標準化得点モデル 評定基準と評定傾向の個人差も 認めない (評定者間で評定基準も評定傾向も 等値) 測定モデル ((4) 式)

モデルとデータの適合度 測定モデルにおけるパラメタ数の検討においては、AICを基準として検討をおこなう。AICとは、複数のモデルの中から最も良いモデルを選択するための基準として用いられ、単にモデル選択の基準をデータとの適合度だけにするのではなく、モデルの自由パラメタ数が増えることによる推定の不安定さを考慮に入れる基準となっている。そして、小さい値の方がモデルとして適切であると判断される。たとえば、標準得点化モデルのAICが同じデータの中で分析したモデルのうち、最も小さかった場合には、「因子分析モデル等で測定するより、データを標準化した上で評定値をそのまま使う方が妥当性が高い」ということになる。また、AICは(11)式で定義される。

$$AIC = -2\log L + 2r \tag{11}$$

ここで、L はモデルの尤度を表し、r はモデルのパラメタ数を表す (Akaike,1987)。

本研究では、それぞれの欠損パタンごと、100 個の検討用データでそれぞれ算出した AIC の平均値を計算し、その値を欠損パタン別のモデルのE[AIC] とした。

$$E[AIC] = \frac{1}{100} \sum_{k=1}^{100} AIC_k$$
 (12)

本研究では、それぞれの欠損パタンにおいて 3 つの評価モデルの *E*[AIC] の比較をおこなった。

(4) 測定モデル・欠損パタンの検討結果

本研究において、実験によって完全データが得られた。そして、その上で欠損パタンごとの推定値の誤差の平均と測定モデルごとの AIC が計算された。

欠損パタンごとの精度の検討

まず、各欠損パタンごとにそれぞれのパラメタ推定値の誤差が計算れた。この推定値の誤差から、先に示した式によって $E[|\epsilon_{\mu}|]$, $E[|\epsilon_{\lambda}|]$, $E[|\epsilon_{y}|]$, $E[|\epsilon_{f}|]$ をそれぞれ計算した。そして、それぞれのパラメタの取りうる範囲に対して、その誤差の平均がどれほどの割合を占めているかを表 4 に示した。各パラメタの取りうる範囲は、 $-1 \le \lambda_{j} \le 1$ 、 $0 \le \mu_{j} \le 4$ 、 $-3 \le f_{i} \le 3$ 、 $0 \le y \le 4$ であるとした。つまり、表 4 は $E[|\epsilon_{\mu}|]$, $E[|\epsilon_{\lambda}|]$, $E[|\epsilon_{y}|]$, $E[|\epsilon_{f}|]$ がその範囲においてどの程度の割合を占めているのかを示したものになっている。

この結果より、それぞれのパラメタの取りうる 範囲の中では5%から10%前後のずれでしかない ことが確認される。

今回の 8 つの欠損パタンのなかで、評定者特性のパラメタ推定値の誤差の平均が小さかったのは、欠損パタン B3 と C2 であった。他には、C1、B2 の値も小さい。 $A1\sim A3$ は値は大きい。また、評価対象の評価得点の推定値 \hat{f}_i の誤差の平均が小さかったのは、欠損パタン A3 であった。次に、 $A2\cdot A3$ と続く。そして、評定値 y_{ij} と復元した \hat{y}_{ij} の誤差の平均が小さいのは、欠損パタン B3 であった。C2、B2、C1 も値は他と比べて低目である。

10	汉此为及CIEC情况少闲床(旧闲际致)					
	λ	μ	f	y		
A1	0.067	0.510	0.501	0.658		
A2	0.200	-0.179	0.013	0.364		
A3	0.534	0.238	0.424	0.331		
B1	-0.408	-0.382	-0.511	0.053		
B2	-0.443	-0.368	-0.444	-0.552		
В3	-	-	-	-		
C1	-0.913*	-0.844*	-0.898*	-0.838*		
C2	-0.856*	-0.148	-0.450	-0.382		

表 5 被説明度と推定精度の関係(相関係数)

- (注 1)B3 は共通評定者の区別がないため計算せず
- (注 2)* は有意確率 1 %での有意差を表す
- (出所) 筆者作成

共通評定者の特徴が精度に与える影響

また、評価得点 f_i と各評定者の評定値 y_{ij} の相関係数つまり評定者の被説明度が計算され、被説明度と推定値のズレの相関係数について、ピアソンの積率相関係数の優位性検定が行なわれた。検定においては、

帰無仮説 H0 : 「母相関係数 = 0」 対立仮説 H1 : 「母相関係数 $\neq 0$ 」

とし、有意水準1%で両側検定を行った。

結果について表 5 に示す。太字で * がついているものが検定の結果 1 %水準で有意だったものである。

その結果、欠損パタン C1 において、 λ 、 μ 、f、y が有意水準 1 %で有意に相関が認められた。また、欠損パタン C2 においては、 λ のみ有意水準 1 %で有意に相関が認められた。

相関係数が認められた欠損パタン C1・C2 では、 共通評定者の被説明度と推定精度の関係には負の 相関が認められた。

分析する際の測定モデルの選択

次に、測定モデルごとの E[AIC] が欠損パタンごとに計算された (表 6)。表 6 において、括弧の中の数値は標準偏差を表している。また、各欠損パタンにおいて 3 つの測定モデルの中で最も E[AIC]の値が小さかったモデルに関して * で示した。

表 6 測定デザイン別の AIC

	因子分析モデル	分散分析モデル	標準化得点モデル
A1	2633 (35.5)*	2749 (49.7)	2904 (51.2)
A2	3151 (58.6)*	3242 (59.6)	3403 (59.3)
A3	3631 (62.5)*	3749 (63.3)	3904 (67.5)
B1	1516 (31.0)*	1542 (28.4)	1621 (32.2)
B2	1970 (32.4)*	2018 (32.7)	2130 (36.3)
В3	2409 (34.7)*	2486 (31.1)	2640 (37.6)
C1	1948 (44.6)*	1988 (49.7)	2115 (50.2)
C2	2396 (48.6)*	2477 (55.3)	2624 (58.0)

- (注1)()内の数値は標準偏差を表す
- (注 2)* はモデルの中で AIC が最も低いことを表す
- (出所) 筆者作成

その結果、本研究で扱ったデータの場合では因子 分析モデルの AIC が最も小さかったことがわかる。

5. 総括

(1) 結論

本研究において、因子分析モデルによって評定の個人差をモデリングできることを示した。また、その際に個人差による影響が除去された評価得点が表現できることも示した。さらに、特殊な欠損パタンを作ることによって因子分析モデルが適用できることも確認できた。

そして、実際に得られたデータを用いて、測定 モデルと欠損パタンについて検討を加えた。結果 から得られた知見については以下に示す。

欠損パタンごとの精度の検討

結果より、評定特性のパラメタの推定に関しては、共通評定者の数は増やした方が精度が良くなることが示された。また、評定者における共通評定者の割合を増やすと精度が良くなっているようである。そして、評定数(図1~3における色のついた面積)が多くなおかつ共通評定者の割合が少ないような欠損パタン(欠損パタン A1、A2)のような場合は、結果から推定精度がよくないことが示されているので、このような欠損パタンを組む場合には、評定集団の数を増やして評定者の数を減らし、共通評定者の占める割合を上げた方が評

価特性のパラメタに関しては、精度良く推定できることが示唆さらた。

しかし、1 群あたりの評価対象の数が少なくなると、評価得点の推定精度は下がっている。したがって、評価得点の推定精度を高めるには、ひとりあたりに評定させる評価対象を多くする必要があるようである。

まとめると、共通評定者の割合は多く、評定者 一人あたりに評定させる評価対象はできるだけ多 くするというパタンが精度良く推定させるために 必要な条件であろう。

また、本研究の欠損パタンでは、共通評定者の割付がすべてBalanced design になっている。しかし、Unbalanced な割付の場合 (例えば、共通評定者によって評定する群の数が異なる) について議論がされていない。例えば、共通評定者によって評定する群の数が異なることによる影響は、共通評定者の評定対象の数に応じて、推定に及ぼす影響が変わっていくように思うが、このことは今後精査していかなければならない問題であろう。

また、結果においては、推定値から復元した \hat{y}_{ij} の誤差が 15 %と比較的大きくなっているが、これは、 \hat{y}_{ij} が $\hat{\lambda}_j$ 、 $\hat{\mu}_j$ 、 \hat{f} を用いて計算した値であり、誤差が \hat{y}_{ij} に凝縮されてしまったとの原因が考えられる。

共通評定者の特徴が精度に与える影響

結果より、「共通評定者が他の評定者よりも多くの評定をおこなう場合には、共通者の被説明度が高いほど推定の誤差の平均は小さくなる」という共変関係が示唆された。

つまり、共通評定者には、評定に信頼の置ける 評定者を配置すると精度良く推定できる傾向にあ るということである。共通評定者が普通の評定者 に比べて多くの評定をおこなう場合の共通評定者 の重要性が示されたと言えよう。

分析する際の測定モデルの選択

測定モデルがいくつか表現できることを示したが、実際にデータを得たあとで、そのモデルの中でもっとも良いモデルを選択するために AIC という客観的な基準を用いて測定モデルを選択できる

ことの意味について示すことができた。たとえば、 本研究で用いたデータ・欠損パタンならば、パラメ タ数が最も多い因子分析モデルを選択し、測定モ デルとして採用するとよいことが示唆されている。

このことは、評定者による恣意的なモデル選択 を排し、客観的に評定データの分析をおこなう上 では重要な点となるであろう。

(2) 今後の課題

今後の課題として、欠損パタンの研究に関して は、シミュレーションをおこなうことにより一般 化できる結論を導き出すことがあげられる。その 際、結論で述べたような共通評定者によって評定 する群の数が異なる場合等についても精査してい きたい。

一方で、小論文試験などは大規模評定データの ため組織的に欠損を含む評定データとなっている が、このような実際にあるデータについても分析 をし、そこから知見を引き出すことも今後の課題 となってくるであろう。

さらに、今回の測定モデルでは最もパラメタ数の多かった因子分析モデルが AIC 基準では選択されたが、測定モデルにもっとパラメタを組み込んでいった場合はどのようになるのかも検討すべきことであろう。

また、評定者に対して理論的に潜在的なクラスに分類できる場合などが考えられる。その場合には潜在的なクラスごとの評定傾向や評定基準の値の分布なども考えることができる。つまり、それぞれの属性に関して分布を仮定したモデルについても提案することができるかもしれない。

付記

本研究の査読にあたって、担当編集委員・差読者 の方から研究に資する多くのご指摘とご提案を頂き ました。この場を借りてお礼を申し上げます。

参考文献

独立行政法人国際協力機構 (2003)『事業評価年次報告書 2003』、独立行政法人国際協力機構

国際協力事業団 (2002)『実践的評価手法』、国際協力出版会

佐々木亮 (2003) 「評価手法の現状と課題」、『日本評価研究』、3(2):1-5

柳井晴夫・前川眞一 (1999)『大学入試データの解析』、現代 数学社

山内香奈 (1999a) 「論文評定データの解析における多相 Rasch モデルと分散分析モデルの比較」、『教育心理学研 究』、47(3):121-130

山内香奈 (1999b)「多相 Rasch モデルによる論文評定データの解析」、『計量国語学』、22(3):104-115

龍慶昭・佐々木亮 (2000)『「政策評価」の理論と技法』、多 賀出版

渡部洋・平由美子・井上俊哉 (1988)「小論文評価データの解析」、『東京大学教育学部紀要』28:143-164

渡部洋・曹亦薇 (1992)「小論文における字の美しさの影響 について」、『東京大学教育学部紀要』 32:253-256

Allison, P. D. (1987). Estimation of linear models with incomplete data. Sociological Methodology 1987, San Francisco: Jossey-Bass, 71-103.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52, 317-332.

Bruin, W. B. & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. Organizational Behavior and Human Decision Processes, 92, 91-101.

Engelhard, G. (1996). Evaluating rater accuracy in performance assessments. Journal of Educational Measurament, 33, 56-70.

Mulaik, S. A. (1972). The Foundations of Factor Analysis, US: McGraw-Hill Inc.

Muthén L. K. and Muthén B. O. (2002). *Mplus User's Guide*. LA: Muthén & Muthén.

Rossi, P. H., Freeman, H. E., Lipsey, M. W. (1999). Evaluation - A Systematic Approach-, California: Sage Publications.

注記

MPlus(Muthén・Muthén, 2002) による測定モデルのプログラムを示す。ここでは、2 群に別かれていて、各群で評定者が 6名 (うち共通評定者 2名) いる場合の評定データの分析を仮定する (つまり、全部で評定者は 10名である)。

TITLE: /* タイトル */

DATA: FILE IS C:"データ指定";

```
FORMAT IS FREE;
TYPE IS IND;
```

VARIABLE: NAMES ARE GROUP U1-U6;/*評定者は6名*/
GROUPING = GROUP (1=g1 2=g2);/*2群*/

ANALYSIS: TYPE=MEANSTRUCTURE;

ESTIMATOR=ML; /*最尤法による推定*/ITERATIONS=1000:

MODEL: F1 BY U1-U6*;

[F1@0];

```
F101;
MODEL g1:F1 BY U1 (11) /* 因子パタン 1 */
              U2 (21) /* 因子パタン 2 */
              U3 (31) /* 因子パタン 3 */
             U4 (41) /* 因子パタン 4 */
              U5 (51) /* 因子パタン 5 */
              U6 (61);/* 因子パタン 6 */
            [U1] (12) /* 切片 1 */
            [U2] (22) /* 切片 2 */
            [U3] (32) /* 切片 3 */
            [U4] (42) /* 切片 4 */
            [U5] (52) /* 切片 5 */
            [U6] (62);/* 切片 6 */
              U1 (13) /* 誤差 1 */
              U2 (23) /* 誤差 2 */
              U3 (33) /* 誤差 3 */
              U4 (43) /* 誤差 4 */
              U5 (53) /* 誤差 5 */
              U6 (63);/* 誤差 6 */
             F101;
```

MODEL g2:F1 BY U1 (11) /* 因子パタン 1 */

[F1@0];

U2 (21) /* 因子パタン 2 */

U3 (71) /* 因子パタン 7 */

U4 (81) /* 因子パタン 8 */

U5 (91) /* 因子パタン 9 */

U6(101);/* 因子パタン 10 */

[U1] (11) /* 切片 1 */

[U2] (21) /* 切片 2 */

[U3] (71) /* 切片 7 */

[U4] (81) /* 切片 8 */

[U5] (91) /* 切片 9 */

[U6](101);/* 切片 10 */

U1 (11) /* 誤差 1 */

U2 (21) /* 誤差 2 */

U3 (71) /* 誤差 7 */

U4 (81) /* 誤差 8 */

U5 (91) /* 誤差 9 */

U6(101);/* 誤差 10 */

F1*;

[F1*];

OUTPUT: STAND RES;

```
標準化得点モデルの場合
  分散分析モデル・標準得点化モデルは MODEL ステー
トメントを書き換えれば良い。
                                                MODEL g1:F1 BY U1 (1) /* 因子パタンは同番号 */
                                                             U2 (1)
分散分析モデルの場合
                                                             U3 (1)
MODEL g1:F1 BY U1 (1) /* 因子パタンは同番号 */
                                                             U4 (1)
             U2 (1)
                                                             U5 (1)
             U3 (1)
                                                             U6 (1);
             U4 (1)
                                                            [U1] (2) /* 切片は同番号 */
             U5 (1)
                                                            [U2] (2)
             U6 (1);
                                                            [U3] (2)
            [U1] (12) /* 切片は違う番号 */
                                                            [U4] (2)
            [U2] (22)
                                                            [U5] (2)
            [U3] (32)
                                                            [U6] (2);
            [U4] (42)
                                                             U1 (3) /* 誤差は同番号 */
           [U5] (52)
                                                             U2 (3)
            [U6] (62);
                                                             U3 (3)
             U1 (3) /* 誤差は同番号 */
                                                             U4 (3)
             U2 (3)
                                                             U5 (3)
             U3 (3)
                                                             U6 (3);
             U4 (3)
                                                            F101;
             U5 (3)
                                                            [F100];
                                                MODEL g2:F1 BY U1 (1) /* 因子パタンは同番号 */
             U6 (3);
            F101;
                                                             U2 (1)
            [F1@0];
                                                              U3 (1)
MODEL g2:F1 BY U1 (1) /* g1 と同番号 */
                                                              U4 (1)
             U2 (1)
                                                              U5 (1)
             U3 (1)
                                                             U6 (1);
              U4 (1)
                                                            [U1] (2) /* 切片は同番号 */
              U5 (1)
                                                            [U2] (2)
             U6 (1);
                                                            [U3] (2)
            [U1] (12) /* g1 の U1 と同番号 */
                                                            [U4] (2)
            [U2] (22) /* g1のU2と同番号 */
                                                            [U5] (2)
            [U3] (72) /* 切片は違う番号 */
                                                            [U6] (2);
            [U4] (82)
                                                             U1 (3) /* 誤差は同番号 */
            [U5] (92)
                                                             U2 (3)
            [U6] (102);
                                                              U3 (3)
             U1 (3) /* g1 と同番号 */
                                                              U4 (3)
              U2 (3)
                                                              U5 (3)
              U3 (3)
                                                              U6 (3):
              U4 (3)
                                                              F1*;
              U5 (3)
                                                             [F1*]:
              U6 (3);
              F1*;
             [F1*];
```

Statistic Analysis for Rating Data —Suggesting Statistical Models to Eliminate the Personal Differences in Ratings—

Tomoya Okubo Masanori Nakagawa Hiromitsu Muta Shin-ichi Mayekawa okubo@ms.hum.titech.ac.jp

Tokyo Institute of Technology. Graduate School of Decision Science and Technology

Abstract

There have been reports suggesting that differences occur in ratings for each individual. Therefore the grading figures show the merits of the subjects of the evaluations, as well as the influences by the personal differences in ratings.

When it comes to rating data, the true rating scores can be expressed by showing the personal differences with the factor analysis. However, reality is that ratings often cannot be done alone, and in these cases, systematic missing patterns would be included to the analyses. Under the data in which systematic missing patterns occur, however, the usual factor analysis cannot be utilized because the rating scores cannot be shown on common scales.

Under these circumstances, the suggestion made by Allison(1987) could be applied to show the factor scores on common scales.

With this in mind, the purpose of this research lies in suggesting effective rating designs and rating models for the rating data which include numbers of systematic missing patterns.

As a result of this research, the following two points were proven to be effective: 1.) speaking of rating design, increasing the ratio of the common raters, and 2.) placing raters who well describe the tendency of the whole as the common raters.

Also, as a result of comparing the rating models using AIC, models that acknowledged the personal differences with both the rating difficulty and the rating discrimination were proven to be the most appropriate.

As shown in this research, efficient rating designs and highly reasonable rating models were examined and taken under consideration.

keywords

rating data, systimatic missing, personal differences in ratings, factor analytic model, missing patterns

【研究ノート】

行政評価の再構築

- 理論着眼型評価思考の確立に向けて -

西出 順郎

福井県

Junronishide@aol.com

要約

本稿の目的は、実用重視の行政評価、すなわち内部マネジメントの強化に資する行政評価を実現するため、現在の行政評価に内在する評価思考の深化を促し、行政評価の再構築に寄与することにある。評価思考の深化とは、従来の実績着眼型に加え、理論着眼型評価思考の確立を意味する。行政評価における実績着眼型の評価思考には、「評価手法の軽視」、「評価結果の軽視」、「ニーズ・アセスメントの軽視」、「事業と成果間における論理的連鎖の軽視」という4つの軽視、すなわち行政評価の機能不全をもたらす阻害要因を創り出す傾向がある。そのため、評価の本来の目的と評価の実施との間に不均衡が生じ、行政評価の目的に応じた有効活用は困難になる。一方、理論着眼型の評価思考は、事業の理論を論理的に検証するため、「説得力ある評価手法の導入」、「コミュニケーション力の向上」、「改革・改善志向の評価の強化」、「評価行為と意思決定行為の同一化」といった4つの効用をもたらす。そして、現在の行政評価に内在する阻害要因を克服するとともに、評価の目的と実施の最適化を促進する。行政評価は、理論着眼型評価思考の確立によって、実用重視に向けた更なる一歩を歩み出す。

キーワード

行政評価、実績着眼型の評価思考、理論着眼型の評価思考、 セオリー評価、ロジック・モデル

はじめに

総務省自治行政局の調査では、2004年7月時点で46の都道府県と13の政令指定都市の全てにおいて、行政評価は導入されており、また他の市町村3,109のうち、514団体が導入済、1,513団体が「試行中」もしくは「検討中」としている。1995年、三重県が全国に先駆けて行政評価を導入して以来、その勢いは留まるところを知らない。その一

方、行政評価の実効性について多くの問題点が指摘されており、その根本的な問題は、行政評価導入の本来の目的たる内部マネジメントの改革・改善に、評価が有効に機能していない点にある。

本稿では、その主たる原因を、現在の行政評価に内在する実績着眼型の評価思考にあるとする。 そして、もう一つの評価思考、すなわち理論着眼型の評価思考という概念に着目するとともに、行政評価における理論着眼型評価思考の必要性を訴

日本評価学会『日本評価研究』第5巻第1号、2005年、pp.15-25

16 西出順郎

える。

本稿の構成は、次のように整理される。第1節 では、行政評価の類型化において、評価思考とい う別の概念とその重要性を提示し、現在の行政評 価における評価思考を実績着眼型と定義する。第 2節では、行政評価の本来の目的を明確にすると ともに、実績着眼型の評価思考が誘引する問題点、 すなわち行政評価の機能を阻害する要因たる4つ の軽視について論じる。第3節では、もう一つの 評価思考として理論着眼型の評価思考を提示し、 その源であるセオリー評価について紹介する。第 4節では、理論着眼型の評価思考を導出する前段 階としてセオリー評価の特徴を整理し、第5節に おいて、行政評価における理論着眼型の評価思考 の効用について考察する。最終節では、理論着眼 型評価思考による行政評価の確立に向けた留意点 について論じる。

なお、本稿でいう行政評価は、その生い立ちに 鑑み、国が導入している政策評価もしくは公共 事業評価とは一線を画し、政策評価、施策評価、 事務事業評価等の呼称を問わず、行政活動全般の 改善を目指した地方自治体が導入する評価の総称 であることを付記する。

1. 行政評価と評価思考

(1) 行政評価の特性

評価は、その目的、主体、客体、時期、手法等、多様な角度から類型化される。行政評価も同様である。各地方自治体のHPを総括すると、現在の行政評価は、次のように類型化できる。評価の目的は、アカウンタビリティの強化、予算や計画への反映、職員の意識改革等、実に幅広い。要約すると、行政活動全般の改善を目指した改革・改善志向の評価となる。評価の主体は、行政職員がある内部評価である。評価の客体は、社会への介入手段である施策や事務事業(以下、「事業」とする。)が大多数を占めている。実施時期については、概ね単年度終了時の事後評価。となる。評価手法は業績測定を主流としている。

(2)評価思考による類型化

評価の類型化は、評価の特性を可視化することで、対峙する概念との比較分析に基づく更なる評価の発展に貢献する。そして現在、行政評価は、従来とは異なる視点による省察が求められている。なぜなら、ABM (Activity-Based Management)やBSC (Balanced Score Card)等、先進的な経営分析モデルが注目される一方で、行政評価の実効性は未だ疑問視されているからである。

解決できない行政評価の問題点とは何なのか。 その答えを導くものが、評価目的を確実に実現す るため、評価客体に対する評価の着眼点と方向性 を明示する概念である。

本稿では、このような概念を「評価思考」と定義する。行政評価における評価思考の意義については後節に委ねるが、それを深化させることによって、行政評価の再構築は可能となる。評価思考とは、行政評価の目的と評価の実施の最適化を導く羅針盤とも換言できよう。

(3) 行政評価の評価思考 実績着眼型の評価思考

評価思考の判断には、評価の目的を達成するため、評価客体に対し如何に接近するかが重要要素となる。現在の行政評価では、事業が実際にもたらす成果の実績値に着目して有効性や効率性を分析し、行政活動の改善に寄与することが想定されている。これは、評価の出発点を実際の業績とし、その達成度の比較分析を中心とした評価の実施を構築し、評価の目的達成を目指す評価思考、「実績着眼型の評価思考」によるものである。

行政評価における実績に着目した評価思考の源は、1990年代の、カラ出張に代表される度重なる不祥事、無駄と指摘される公共事業の乱発、そして成果を問わない予算査定偏重主義へのアンチテーゼにある。したがって、実績着眼型の評価思考は、各地方自治体が行政評価を構築する過程の中で、必然的に形成されたといえる。

2. 実績着眼型の評価思考の問題点

実績に着目した評価思考は、現在の行政管理から行政経営へという、行政の行動規範の転換を確かに促した。しかし、行政評価においては、決してプラスに機能したわけではない。それどころか、実績着眼型評価思考が行政評価の目的と実施との不均衡を導き、その不均衡が行政評価の機能不全を助長する場合がある。なぜなら、「木を見て森を見ず」の如く、実績に固執する余り、行政評価の目的に照合すると、必ずしも適当とは言い難い評価の実施を遂行し、評価目的の達成を阻害しているからである。

現在の行政評価は、中央地方関係の過度の融合性や成果の統制不可能性、現行事業の正当化(古川・北大路 2001) 評価結果の使い道の混迷、運用する職員の評価意識の希薄さ(西出 2003)等、数多くの問題点が指摘されているが、中でも評価それ自体に内在する問題を紐解けば、主たる原因は実績着眼型の評価思考なのである。

本節では、行政評価の目的を明確にした上で、 実績着眼型の評価思考の視点から、その目的達成 を阻害する具体的な要因について論じる。

(1) 行政評価の目的と阻害要因

行政評価の目的は、各自治体のHPが示すよう に、成果重視の行政運営、事業の改善、合理的な 予算配分、アカウンタビリティの強化、ひいては 組織風土や職員意識の改革等、多様化している。 しかし、「経営システムとしての評価という色彩 を強く持ってきた」(古川2004, p.40) こと、「成 果志向の政策評価システムを導入する目的の第一 は行政の生産性を高めることにある」(北大路 2001, p.242) ことからも窺い知れるように、その 本来の目的は、TQM的な改善志向の観点に立脚 している。勿論、地方自治体の多くもPDS(もし くはPDCA)サイクルの強化を掲げており、行政 評価による行政マネジメント・サイクルの実現 は、行政評価が真価を発揮するための条件といえ る(小野 2004)。計画策定、予算編成等のマネジ メントの強化、すなわち、PDSのP(プラン)領 域に対する、改革・改善志向の情報提供こそが、

行政評価の本来の目的である。

しかし実際には、改革・改善志向の評価結果が 供給され、P領域で有効に活用されているとは言 い難い。なぜなら、実績着眼型の評価思考が、P 領域での改革・改善を阻害する要因、すなわち 「評価手法の軽視」、「評価結果の軽視」、「ニーズ・アセスメントの軽視」、「事業 成果間におけ る論理的連鎖の軽視」という、4つの「軽視」を 誘引しているからである。

(2)評価手法の軽視

現在の行政評価は、評価手法の進歩に関心を払わず、専ら業績測定に依存している。行政評価が導入され始めた1990年代後半から、その風潮は変わらない。確かに、業績測定は、評価客体の成果を中心とした業績指標の実績を測定するとともに、目標値との乖離を可視化し、実績の良し悪し、ごのような手法への依存は、成果の増加率への依存は、成果の増加率への依存をもたらす。すなわち、業績測定の簡便かつ明快さのみを強調し、評価対象事業とその成果の因果関係を説明する評価手法をおざなりにするのである。このように実績着眼型の評価思考は、評価手法そのものを軽視し、実績をより深く検証する機会を自ら放棄する要素を含んでいる。

(3)評価結果の軽視

行政評価は、実績着眼型の評価思考によって、 改革・改善志向ではなく、判断・総括志向の性質 を帯びてくる。評価結果の多くが、極めて簡便な 数的処理による効率性・有効性を示す数値や担当 職員等の経験知に基づいた、表層的かつ叙述的情報と化すからである。事業の存否の判断に寄与すること自体は、有用なことだが。そのような傾向に拍車がかかると、行政評価の成果の証として、 廃止事業の数が喧伝されることになる。そして、 実績に着目した評価思考においては、評価イコール事業の価値判断という観念が蔓延し、事業の改善等、P領域の強化に寄与する具体的な情報は軽 視され、事業の存廃を提示する情報が重視される。 18 西出 順郎

制度上標榜しているにも拘らず、改革・改善志向の評価結果も形成されることはない。その結果、PDSサイクルにおける、S(シ-)領域とP領域の断絶をもたらし、時には、評価客体の改善ではなく成果目標を下方修正することによって、目標値と現状値の溝を埋めようとするモラル・ハザードを誘発する。

(4) ニーズ・アセスメントの軽視

P領域へ改革・改善志向の情報が提供されない以上、事業の立案もしくは再構築を行う場合、十分なニーズ・アセスメントが求められる。しかし、実績着眼型の評価思考は、評価客体たる事業を所与として捉え、事業の形成過程、すなわち政策的視点から事業そのものの必要性を吟味する観点は持ち得ておらず、ニーズ・アセスメントが強化を求めることはない。ニーズ・アセスメントが着実に履行されない限り、望むべき最終成果やターゲットが曖昧なまま事業が立案もしくは再構築され、ひいては評価の実施自体が困難になる。

(5)事業と成果間における論理的連鎖の軽視

実績着眼型の評価思考は、評価客体たる事業の 形成過程のみならず、評価客体の想定する論理的 展開、すなわち事業とその成果を繋ぐ因果関係に 対しても関心を払わない。その結果、次のような 問題をもたらす場合がある。第一に、問題点の究 明を困難にすることである。事業の実施に伴う市 民への影響過程は複層化している。事業の論理的 な曖昧さは、そのメカニズムを不透明にしたまま 評価を強いることになり、具体的な事業活動と市 民の行動予測を精査しない、信頼性の低い評価結 果をもたらす。第二に、実現可能性を無視した成 果目標の設定である。当初から実現可能性を無視 した耳触りのよい成果目標の設定は、暗黙のうち に、目標値を責任の問われない願望値へと転換す る。第三に、成果目標の改ざんである。事業の論 理的展開の曖昧さは、事業完了後、その実績を踏 まえ、事業の正当化に都合のよい論理的展開及び 成果指標の設定を可能にする。このため、偶発的 もしくは副次的成果を本来の成果として喧伝し、

本来あるべき成果への追求を曖昧にする。

3. もう一つの評価思考 理論着眼型の評価思考

内部マネジメントの改革・改善を志向し、PDS サイクルの強化を謳いながら、P領域の弱体化を もたらしている行政評価。その原因は行政評価の 目的と実施との不均衡にあり、その不均衡をもた らす4つの軽視の根底には、実績着眼型の評価思 考が脈々と流れている。その潮流に一石を投じ、 不均衡を解消するには、現在の行政評価における 評価思考を深化させ、もう一つの評価思考の確立 が必要である。

本節では、深化をもたらす評価思考として、事業の理論に着目した評価思考を提示する。次に、その有用性を認識する上で重要な概念であるセオリー評価を紹介するとともに、理論に着目した評価思考に基づく行政評価について考察する。

(1)理論着眼型の評価思考

もう一つの評価思考とは、事業の実績ではなく その可能性を説明する、事業の理論に着目するこ とであり、数値による表層的な判断・総括型から 論理力を駆使した改革・改善型の評価への拡大、 評価の目的と評価の実施との最適化状態への接近 を意味する。つまり、このような視点に立脚した 評価思考が「理論着眼型の評価思考」であり、評 価客体への論理的な接近によって、その着眼点を 実績から理論へと移行し、事業メカニズムの究明 に貢献する。

(2)セオリー評価とは

理論着眼型の評価思考は、理論重視型評価(Theory-Based Evaluation) いわゆるセオリー評価でに依拠しており、それを解きほぐした動的かつ実用的な概念である。セオリー評価とは、「事業費等や事業活動から成果までの連鎖における予測される流れを探求する評価」(Weiss 1998, p.338)で、事業の実施が、想定する成果をどのようにも

たらすのか、これを事業の理論でもって検証するものである。基本的には「多くの理論重視型評価は、特定のプログラム評価を導くために適切に活用されうる理論を展開する計画から始まる」(Stufflebeam 2001, p.37)との指摘があるように、ニーズ・アセスメントや業績測定と同様、プログラム評価の一部を構成するものといえよう。

(3)科学的アプローチと実用的アプローチ

セオリー評価における理論の典拠は様々であるが、Birckmayer and Weiss (2000)によれば、理にかなう科学的根拠の持った社会科学理論と、単に事業がうまく実施される否かの論理性を図表化したロジック・モデル(Logic Model)に整理される。

前者は、1980年代に理論主導型評価(Theory-Driven Evaluation)として、Chen and Rossiが提唱したものである(Chen 1990)。彼らは、従来の、理論のない事業、曖昧な成果、測定方法が誤った評価結果に失望し、事業と成果間における理論不在の状態(Missing Theory)に社会科学的理論の導入を唱えた。すなわち、事業と成果の因果関係を社会科学的に明確にし、事業の理論が存在しなかったが故に台頭した政治主導の政策形成の欠陥を、強力な社会科学の理論を用いて修正しようと考えたのである(Stame 2004)。

一方、後者は、科学的アプローチを追求する前者とは異なった実用的アプローチである。その先駆的役割を果たしたのは、1970年代におけるJ. S. Wholeyとそのグループである(Vedung 1997)。彼らは、評価可能性アセスメント(Evaluability Assessment)という評価のモデルを開発した。これは「徹底的な評価の対象となった事業が有益に評価されることを保証するために、厳密な評価の実施前に活用される技法」(Hatry 1999, p.xiv)のことである。

この背景には、米国におけるプログラム評価の限界からの脱却があった。すなわち、当時の評価は、判断・総括的な事後評価であり、関係者の政治的な抵抗にあったこと、また、正当性のある実験デザインを駆使できなかったことに起因して活用されなかったのである(Vedung 1997)。

また、評価可能性アセスメントは、事業開始時からある程度成功を保証する手段としても活用され、事業開始前に、事業の目的、方法、投入資源、成果目標等を論理的に説明可能とすることで、事業の成功確率の向上に資するものとされている(長尾 2003)。つまり、評価可能性アセスメントによって、質の高い事後評価の実施が確保されるだけでなく、評価対象の事業と政策の目標との連鎖関係、事業の数値目標、事業の論理性等が明確になり、説明のつかない事業計画は事前に排除されることになる。

(4) ロジック・モデリング

セオリー評価、特に、評価可能性アセスメントの代表的手法は、ロジック・モデルによる検証である。ロジック・モデルとは、それぞれの事業における想定される成果や一連の連鎖を図表化したものである(Love 1991)。評価研究の領域では、プログラム・セオリー(Program Theory)、コーザル・リンケージ(Casual Linkage)もしくはロジカル・フレームワーク(Logical Framework)と呼ばれている。立案された事業の論理的説明力、事業実施の物理的可能性、事業の具体的な成果目標は、ロジック・モデリングの中で検証される。

基本的なロジック・モデルの構成要素は、資源、事業活動、結果、成果、影響である。これらの要素は、「もし・・・ならばこうなる」(if then)という推論を基に、時系列的に結び付けられる((財)農林水産奨励会農林水産政策情報センター2003)。

(5)理論着眼型の評価思考による行政評価

セオリー評価を踏まえ、理論着眼型の評価思考による行政評価の基本形態は、実績着眼型と異なる様相を呈する。例えば、評価の主体は、同じ内部評価でも、実質的な主導権は、これまでの財政担当課から企画担当課へと移行する。予算と直接連動する既存事業の存廃の判断ではなく、新規事業や拡充等を伴う既存事業の論理的整合性の検証に重きを置くからである。評価の時期は、新規事

20 西出 順郎

	理論着眼型の 評価思考による行政評価	実績着眼型の 評価思考による行政評価
評価主体	内部(企画担当課)	内部(財政担当課)
評価客体	事業の理論	事業の実績
評価目的	理論上の成果検証(改革・改善型)	実績としての成果検証(判断・総括型)
評価時期	主に事前評価	中間もしくは事後評価
評価手法	ロジック・モデリング	業績測定

表1 評価思考による行政評価の類型

(出所)筆者作成

業の採否や既存事業の拡充等の決定に対しその効力を発揮することから、主として事前評価型となる。評価の手法には、ロジック・モデリングが導入される。

表1は、理論着眼型の評価思考による行政評価 を類型化するとともに、実績着眼型の評価思考に よる行政評価との比較を表したものである。

4. セオリー評価の特徴

理論着眼型の評価思考による行政評価では、P 領域の強化が求められ、その効用は、セオリー評 価の持つ特徴を行政評価に援用することで具現化 する。

本節では、セオリー評価の特徴を整理する。続く第5節で、行政評価における理論着眼型評価思考の効用について論じる。

セオリー評価の特徴については、数多くの研究者が指摘しており、Birckmayer and Weiss (2000)は、事業の立案と改善における利益、知識の創出における利益、そして評価研究の立案における利益の3点を示した。事業の立案と改善における利益とは、セオリー評価が事業の活動とその達成の間に介在するメカニズムについて情報を提供し、将来にわたって、効果的な戦略の展開に貢献することである。知識の創出における利益とは、人間の行動や行動の変化についての知識を増大することである。これは、事業の変化のメカニズムに関する知識が、特定の事業だけでなく、幅広く活用され事業の一般化にも貢献することをいう。評価研究の立案における利益は、セオリー評価が、事

業のメカニズムに対応した情報収集や重要な仮定 の選択を確実に実施することで、理論の連鎖を調 査するよう評価者を誘導することをいう。

特に、Weiss (1998) は評価の利害関係者に、また、Knaap (2004) は政策過程に着目して、セオリー評価の特徴を整理した(表2及び表3参照)いずれも、行政評価における理論着眼型評価思考の効用を考察する上で有意義な類型化である。

5. 行政評価における理論着眼型の評価 思考の効用

理論着眼型の評価思考によって、評価の目的と 実施が最適化状態に接近するにつれ、行政評価の 本来の目的たるP領域の強化は具現化される。

では、理論着眼型の評価思考は、どのような効用を行政評価にもたらすのか。考察すれば次の4つ、すなわち「説得力ある評価手法の導入」「コミュニケーション力の向上」、「改革・改善志向の評価の強化」、「評価行為と意思決定行為の同一化」に集約される。

(1) 説得力ある評価手法の導入

理論着眼型の評価思考は、業績測定とは異なった評価手法を要請する。ロジック・モデリングである。セオリー評価には、実験デザインなしで、事業の影響を評価する能力がある(T. C. M. Lam, qtd. in Knaap 2002) また、簡単に手段と目的のストーリーを提供でき、統計的な結果だけよりも、説得力があって記憶に残る(Weiss 1998)との特

表2 利害関係者の視点からみたセオリー評価の特徴

事業の立案者 (Program Designer)	・変化理論モデル(Theories-of-change model (いわゆるロジック・モデル)によって、自ら立案した仮定を明確にし、検討する機会を得る。
事業の実施者 (Practitioner)	・実施者達が、自らの行為とその目的を共有化することによって、実施するための原動力を増加させる。
事業の運営責任者 (Program Manager)	・評価によって、どの推論の連鎖が誤っているか、そのフィードバックを具体的な意 思決定に活用する。
類似事業に携わる運営 責任者及び資金提供者 (Managers and Funders of Similar Programs Elsewhere)	・事業の成功・失敗事例を検討し、自らの新規事業等に活用する。
政策決定者と市民 (Policy Makers and Public)	・政策決定者と市民に対し、簡単に手段と目的のストーリーを説明する。 ・ストーリーに説得力があり、記憶に残り易い。

(出所)Weiss(1998)に基づき筆者作成

表3 政策過程の視点からみたセオリー評価の特徴

政策の立案複雑性と焦点	・最も大切な問題に焦点を当て、複雑性を軽減する。 ・目的とその達成手段について、明確なビジョンを展開する。
政策の議論	・政策決定者の立場を明確にするとともに、政策論議の関係者すべてに判断の
開放性	準拠枠を与え、熟慮された議論を可能にする。
政策の実施と学習	・既存の政策やその実現方法を改善すべきか、政策自体から脱却し新しい視点
改善・革新	へと進化・革新すべきか、熟慮の上選択する。
政策の評価(活用) 関連性	・評価結果を政策論議と学習の一部に活用できる機会を持つ。 ・政策決定者の考えや願望に接する機会を持つ。

(出所)Knaap(2004)に基づき筆者作成

徴があるように、その流れを汲むロジック・モデルは、視覚的な表現によって、事業の長短を判り易く指摘する((財)農林水産奨励会農林水産政策情報センター 2003)。

数多くの事業の評価を実施する行政評価にとって、ロジック・モデリングは極めて現実的な評価 手法で、かつ、その判り易さは、理論に着目した 行政評価に対するインセンティブを一層向上させ る。

(2) コミュニケーション力の向上

理論に着目した行政評価は、政策論議、特に、評価対象となった新規事業の採否や既存事業の拡充等を決定する場において、事業部門と首長を含む意思決定に携わる部門(意思決定部門)とのコミュニケーションを強化する。それは、ロジック・モデリングが事業部門から意思決定部門へのレクチャーではなく、両者の協働によるディスカッションを創出するからである。コミュニケーションの強化によって、当該事業の目標から内在す

22 西出順郎

る課題に至るまで、意思決定に不可欠な情報が共有化される。そのため、評価困難な事業の採択は低減され、偶発的な成果による事業目的のすり替えも回避される(Weiss 1998)

(3) 改革・改善志向の評価の強化

理論着眼型の評価思考は、ロジック・モデルのフローに沿って議論が展開するので、論点の拡散を最小限にした議論が可能となり、改革・改善志向の評価を実現する。すなわち、事業の中で不要もしくは不適切な活動、もしくは論理の流れに沿って展開しない変化の発見によって、より現実味を増した事業の再構築に寄与することになる(Birckmayer and Weiss 2000)。また、そのような弱点を事前に除去するため、ニーズ・アセスメントをこれまで以上に強化する。

(4)評価行為と意思決定行為の同一化

事業の理論に対する評価行為は、事実上、事業の採否等を明確にする意思決定行為との同一化をもたらす。事業のメカニズムを検証することは、事業をより洗練すると同時に、その効果の是非を判断することであり、内部評価でかつ評価と意思決定の対象客体が同一である以上、両行為を区別できないからである。よって、前年度の事業の原本を事業の政策論議を実施可とは異なり、理論着眼型では、当該年度の事業の原案を事業部門と意思決定部門とが協働で議論する。議論の結果は、評価の結果であると同時に意思決定の結果でもあり、評価結果の軽視という概念は起こり得ない。

6. 理論着眼型評価思考による行政評価 の確立に向けた留意点

理論に着目した行政評価は、実在していないわけではない。近年、徐々に増加している。例えば、都道府県レベルでは福井県や和歌山県、市町村レベルでは加古川市、芦屋市、久喜市等が、評価もしくはマネジメント・システムの構成要素とし

て、ロジック・モデルを活用した評価を実施している。しかし、理論着眼型の評価思考による行政評価は緒についたばかりであり、その成否は、これからの取組み方にかかっている。

本節では、理論着眼型評価思考による行政評価の確立に向け、5つの留意すべき点について論じる。

第一に、事業の理論に着目した評価の使い道を明確にすることである。社会科学的に事業のメカニズムの追求を試みることなく、基本的なロジック・モデリングの定着を図るべきである。ロジック・モデルは、プロセス評価(中間評価)やアウトカム評価(事後評価)の準備行為にも活用できるが、特に、事前評価のコミュニケーション・ツールとして導入すべきである。

第二に、行政が介入する正当性を明確にした上 で、事前評価を実施することである。事業の理論 の検証は、その事業の採否等の決定と同時に進行 する。したがって、事業の論理的展開を議論する には、事業を実施すべき理由を確実に説明できる ことが前提である。しかし、事業部門が客観的な 情報を全く活用せず、過去の経験知のみに依存し て事業の立案を行なった場合、意思決定部門は実 施すべき論理的な説明を受けないため、事業その ものの必要性に議論を集中させてしまう。その結 果、議論の目的が拡散し、事業案とその成果の整 合性に納得することなく、議論は終了するのであ る。そのような状況を回避するため、ニーズ・ア セスメントを着実に実施し、その客観的な情報を 基に介入する正当性を明確にすることが不可欠で ある。

第三に、ロジック・モデリングの運用が軌道に 乗り始めたら、内部要因や外部要因等を十分配慮 した、より精緻なモデリングを実施することであ る。特に、外部要因が事業に与える影響は、決し て少なくない。そのような要因を考慮した成果の 設定は、より現実味を増すであろう。

第四に、知識の創出としてロジック・モデルを 蓄積することである。セオリー評価は、統計的な 結果だけよりも、説得力があって記憶に残り、将 来の政策に影響を与える可能性を秘めている (Weiss 1998)。実績の検証は、評価の対象事業の みがその結果を享受できるが、ロジック・モデル による検証は、成功原因等を論理的に分析し、その結果を一般化することで、類似事業に対する適用を可能にする。将来における関連事業の改善、プロセスやアウトカム評価に活用できるよう、検証で得た知識を十分に蓄積する必要がある。

最後に、理論着眼型と実績着眼型評価思考によるバランスの取れた評価の実施である。1970年代、既にPressman and Wildavsky (1984)が指摘するように、政策が意図する成果と実施による現実とは必ずしも一致するものではない。Birckmayer and Weiss (2000)も、プログラムの失敗なのか理論の失敗なのかを判断するためにも、同時にプロセス評価の必要性を主張している。理論着眼型評価思考の確立とは、実績ではなく理論着眼型評価思考の選択を意図するものではない。両者による評価が相乗効果を出すことによって、評価の質は向上する。理論に着目した評価思考を重視型や主導型ではなく、比較的中立的な語感のある「着眼型」としたのは、理論重視の風潮へ殊更誘うことを避けるためである。

おわりに

行政評価は、導入する地方自治体が増加する一方、現在、閉塞状況の中でうごめいている。行政評価に対する社会的関心は以前に比べ格段に低くなり、一種の流行が過ぎ去った気配さえ感じられる。そして、一部の研究者や実務者は、その発展のため更なる研究や試行錯誤を続けているものの、関心の低下は、儀式化への道を歩む行政評価を暗示しているかのようである。今こそ、行政評価を再構築し、その儀式化に歯止めをかけ、役に立つ行政評価へと変貌を遂げる第一歩を踏み出さなければならない。

理論着眼型の評価思考という概念はそのための 糸口であり、今後、多くの地方自治体が、実績に 着目した評価思考だけでなく、理論に着目した評 価思考による行政評価を導入するとともに、実用 重視の行政評価に向けて一層前進することを望ん でいる。

注記

- 1 行政評価の生い立ち及び定義については、西出 (2003)を参照。
- 2 政策、施策、事務事業と事業の位置付け、そして、 それらの定義や英訳語は必ずしも一様ではない。本 稿では、行政の介入手段すべての総称を事業とし、 Program Evaluationをプログラム評価と訳す以外、 Programの対訳をすべて事業とした。
- 3 一部の地方自治体では、複数年度にわたる既存事業の評価を中間評価と称する場合もある。しかし、単年度終了後にその成果を検証するわけで、まさに進行中の事業を形成的に評価しているわけではない。したがって、そのような中間評価は事後評価の範疇に入る。
- 4行政評価の実効性については、平成16年3月号『地方財務』に「行政評価の到達点と今後の展開」と題する特集、また、平成16年6月に開催された日本評価学会の春季第1回全国大会において「本物の行政評価、偽物の行政評価」と題した共通セッション等において活発に議論されている。
- 5 政策の成果指標の設定やその優先度を把握するため の県民意向調査を除き、回帰分析等の統計的手法を 評価に活用している地方自治体は皆無に等しい。
- 6 宮城県等の行政評価の先進自治体では、評価結果の 反映状況を詳細に公表している。しかし、公表して いる地方自治体の多くは、財源ベースの説明が中心 で、改革・改善志向評価結果といえない。
- 7龍・佐々木(2000)においてセオリー評価として紹介されて以来、実務者間でも、理論に着目した評価が認知され始めた。

参考文献

- Birckmayer, J.D. and Weiss, C.H. (2000). Theory-based evaluation in practice: what do we learn? *Evaluation Review*, 24 (4), 407-431.
- Chen, H.T. (1990). *Theory-Driven Evaluations*. Newbury Park, Calif: Sage Publications.
- Hatry, H.P. (1999). Performance Measurement: Getting Results. Washington, D.C.: The Urban Institute Press. (上野宏・上野真城子訳 (2004)『政策評価入門』、東洋経済新報社)

24 西出順郎

- Knaap, P.V.D. (2004). Theory-based evaluation and learning: possibilities and challenges. *Evaluation*, 10 (1), 16-34.
- Love, A.J. (1991). *Internal Evaluation: Building Organizations from Within*. Newbury Park, Calif: Sage Publications.
- Pressman, J.L. and Wildavsky, A. (1984). *Implementation* 3rd Edition. Berkeley and Los Angeles: University of California Press.
- Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation*, 10 (1), 58-76.
- Stufflebeam, D.L. (2001). *Evaluation Models*. San Francisco: Jossey-Bass.
- Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick, New Jersey: Transaction Publishers.
- Weiss, C.H. (1998) *Evaluation 2nd Edition*. Saddle River, New Jersey: Prentice Hall.
- 小野達也(2004)「都市自治体の行政評価の課題と展望」(財)日本都市センター『日本都市センタープックレットNo.10:行政評価の新展開』(財)日本都

市センター、2-15

- 総務省ホームページhttp://www.soumu.go.jp/(9/21/2004) 長尾眞文(2003)「実用重視評価の理論と課題」、『日本評価研究』、3(2):57-69
- 西出順郎(2003)「地方自治体における行政評価の実際と新しい評価者モデル」、『日本評価研究』、3(2):179-187
- (財) 農林水産奨励会農林水産政策情報センター(2003) 「ロジック・モデル策定ガイド」政策情報レポート 066(原著: W.K.Kellogg Foundation(2003) Logic Model Development Guide)
- 古川俊一・北大路信郷(2001)『公共部門評価の理論 と実際』 日本加除出版株式会社
- 古川俊一(2004)「評価の定着・成功への工程表:政 策形成への活用と経営システムとしての運用」、『地 方財務』、2004年3月号、ぎょうせい、597:34-52
- 龍慶昭・佐々木亮(2000)『「政策評価」の理論と技法』、 多賀出版

(2005.1.26受理)

A New Current of Evaluation for Management in Local Government Reform A Shift from Performance-focused to Theory-focused Evaluation Perspective

Junro Nishide

Fukui Prefectural Government Junronishide@aol.com

Abstract

The purpose of this paper is to reconstruct evaluation for management in local government reform through deepening evaluation perspectives inherent in the current practice, in order to establish evaluation for strengthening local government management. The performance-focused evaluation perspective tends to produce four impediments that bring evaluation dysfunctions in local government; disregards of evaluation approach, evaluation results, needs assessment, and logical linkages between programs and outcomes. These will produce an unbalance between the original evaluation purpose and actual practice, and thus, lose an opportunity for utilizing evaluation results reflecting on evaluation purpose. On the other hand, the theory-focused evaluation perspective produces four advantages through logical verification of program theories: introduction of persuasive evaluation approaches, improvement of communication, strength of improvement-oriented evaluation, and identification of evaluation action and decision-making. The theory-focused evaluation perspective overcomes impediments inherent in the current evaluation in local government and promotes best match of evaluation purpose and implementation. Evaluation in local government leads to a further step toward focusing on utilization through establishing the theory-focused evaluation perspective.

Keywords

evaluation in local government, performance-focused evaluation perspective, theory-focused evaluation perspective, theory-based evaluation, logic model

【実践・調査報告】

国別プログラム評価の困難性を超えて 実践的方法論に関する考察と試案

三輪 徳子

国際協力機構

miwa.satoko@jica.go.jp

要約

近年、主要援助機関では多くの国別プログラム評価が行われてきた。しかしながら、インパクトや帰属(attribution)の検証、さらに国別プログラムの評価可能性の問題もあり、方法論は確立されていない。JICAも過去に8件の評価を実施する中で、同様の問題に直面してきた。一方、結果重視や援助協調の進展は、援助効果やその評価の概念に変化をもたらしており、効果の検証においても貢献(contribution)の視点が新たに重視されつつある。本稿は、以上のような問題及び新たな議論をふまえ、国別プログラム評価の改善に向けた実践的な評価方法を考察するものである。具体的には、効果の検証と評価可能性の2つの主要課題への対応に向けて、対象途上国の開発戦略における協力の位置付けを整理したマトリックスを用い、プログラムの妥当性や戦略の適切性を検証するとともに、プログラムの実績を合算し、貢献の概念に基づいて協力効果を評価する方法を提示している。そして評価を活用して国別プログラムを改善し、評価可能性を高めていく漸進的アプローチを提案している。

キーワード

プログラム評価、アウトカム、貢献、評価可能性、結果重視マネジメント

1. はじめに

国別プログラム評価(Country Programme Evaluation 以下、CPE)は、国別プログラミングの導入に伴い1980年代末に開始され、1990年代後半から多くの評価が行われてきている。その背景には、国別プログラミングのもと、主要援助機関で「国」が援助の計画、予算、管理、報告の重要な単位となったこと、特に1990年代後半からの援助効果とセレクティビティの議論の中で、より選択的に効果の高い援助を行っていくためには国

レベルでの援助のマネジメントと評価が重要との 認識(World Bank 1998)が強まったことが挙げ られる。

こうした動きを受けて経済協力開発機構/開発援助委員会(OECD/DAC)評価作業部会(現・評価ネットワーク)では、1994年及び1999年にCPEに関するワークショップ(ウイーン・ワークショップ)を開催している。特に1999年のワークショップでは、多数の援助機関が参加して、方法論について広範な意見交換が行われた。そこで明らかになったのは、CPEの方法は機関によって様々であるが、多くの機関がインパクトや帰属の

日本評価学会『日本評価研究』第5巻第1号、2005年、pp.27-44

28 三輪 徳子

検証、さらに国別プログラム (CP) の評価可能性の問題を抱えていることであった(DAC 1999)

これらの問題は、日本のODAの政策機関(外務省)及び事業実施機関(国際協力機構(JICA)、国際協力銀行(JBIC))で実施されているCPE²にも共通する問題となっている。例えばJICAでは、国別事業実施計画に基づく国別プログラミングの強化にあわせ、1998年から本格的な国別の事業評価を導入し、2004年度までに8件を実施している。その目的は、特定対象国におけるJICAの協力実績を総合的に評価し、国別事業実施計画や協力事業の改善に向けた提言・教訓を得ることにあった。しかしながら上記のような問題から、目的を十分に満たす評価を行い得ておらず、CPEの有効性を高めるための方法論の改善が重要な課題となっている。

一方、1990年代後半以降、援助の世界では、国際開発目標の設定や結果重視マネジメント(RBM)の拡充など、結果志向のアプローチがさらに強まった。また、効果達成に向けたアプローチとして、途上国のオーナーシップとパートナーシップが重要視されるようになり、ドナーによる援助協調や調和化の推進、財政支援の導入等の一連の取り組みが急速に進展した。

こうした動きは、国レベルでの援助の効果の捉え方にも変化をもたらしており、これまでの特定ドナーによる援助の効果の視点に加え、開発目標達成に向けた途上国や他のドナーとの共同取り組みの効果の視点が重視されるようになってきた。後者のような効果をどのように評価するかは、CPEの新たな課題となっており、貢献の観点から方法論を見直す動きも見られつつある。

本報告は、CPEをめぐる以上のような問題や動きをふまえ、CPEの有効性を高めるための方法論を、日本のODAの主要実施機関であるJICAを事例に取り上げて考察するものである。当然ながら方法論には様々な選択があるが、本稿では、これまでの主要な課題への対応を主眼に置き、現実的な制約の中でより有用性の高いCPEを実施していくための方法論の提案を行う。ただし提案は、課題対応策として概念を整理した段階のものであり、実証及び改善は今後の課題である。

以下、第2節では、ウイーン・ワークショップ

の議論のレビュー及びJICAがこれまでに行った 国別事業評価の報告書の分析を通じ、CPEに関す るこれまでの問題を整理する。次に第3節では、 2004年に行った主要援助機関の調査に基づき、ウ イーン・ワークショップ以降の主要援助機関にお けるCPEをめぐる動向についてとりまとめる。そ して以上をふまえて第4節でJICAを事例にCPEの 改善に向けた試案を提示し、考察する。

2. 国別プログラム評価(CPE)の方法 論に関するこれまでの問題

(1) ウイーン・ワークショップでの議論

CPEに関して最も包括的な議論が国際的になされたのは、1999年のウイーン・ワークショップである³。そこでは、CPEは用途、アプローチ、手法ともに多様であるが、インパクトと帰属の検証、そしてCPの評価可能性が主要な課題ということが示された(DAC 1999)。

1) CPEの多様性

まず、用途については、CPの妥当性や効果の 検証、組織目標の達成度の検証、援助配分の見直 し、支持の確保等が主なものとなるが、CPが未 確立の機関ではCP導入に向けてCPEを活用する 場合もあることが指摘されている。

次にアプローチに関しては、分析深度・対象ともに多様であるとして、インパクトまでを盛り込み、長期間を対象に主な協力を網羅した「ヘビー型」と、妥当性・効率性等のより直接的な基準を用い、直近数年間の活動に焦点を当てた「ライト型」のタイプ分けが行われている。そして「ヘビー型」は、CPに統一性があり、適当な開発指標も入手できる場合、当該機関が対象国の開発に中心的な役割を占める場合、対象CPが当該機関の事業に大きな比重や重要な意味を有する場合等には適しているが、それ以外は「ライト型」の方が合っているとの議論がされている。

また、手法に関しては、マクロ・レベルでの変 化発現から要因をたどる形で当該機関の援助への 帰属を検証しようとする「トップダウン型」、CP を構成する個々のプロジェクトを積み上げてインパクトを検証しようとする「ボトムアップ型」の2種類があり、分析も叙述、計量経済、スコアリングなど様々な技法が用いられていることが示されている。

2) インパクトと帰属の検証、評価可能性

しかしながら、これら手法の違いにかかわらず、 インパクトと帰属の検証及び評価可能性は、共通 の問題となっている。

このうちインパクトと帰属に関しては、検証の 困難性の理由として、 当該機関の援助と他の機 関の援助、あるいは援助と援助以外の要因を切り 分けて変化(インパクト)を特定するのがプロジェクトやセクター評価に比べてはるかに難しいこ と、 特定の変化は援助以外を含む様々な要因か ら生じている他、援助も複数の複雑な変化を生む ことから、帰属の論理的検証に困難性が伴うこと、

反事実的状況(counterfactual)の構築も、長期成長モデル、計量モデル、類似国との比較等の手法は考えられるが、技術面でも運用面でも多くの難しさがあることが挙げられている。

また、評価可能性については、CPがゴール・戦略の明確性や、構成要素の首尾一貫性(coherence)を欠く場合には、評価が非常に難しいとの点が指摘されている。同問題は、国別プログラミングが確立しておらず、CPがばらばらなプロジェクトの集まりに過ぎないような場合は特に顕著となっている。しかしながら、CPEの最多実績を有する世銀でも、CPEの質の向上には国別援助戦略(CAS)のデザインの改善が必要との指摘をしており、米州開発銀行(IDB)では同問題に対応するために、CPEに際してCASの評価可能性をチェックしていることが報告されている。

(2) JICAにおける国別事業評価の経験

JICAでは、対象国の開発に対するJICAの貢献度を総合的に評価のうえ、 国別事業実施計画の改善及び事業(協力プログラム・案件)の形成・実施に資する教訓・提言を抽出することを目的に、国別事業評価を実施してきた。しかしなが

ら、過去8件の評価では、貢献度の総合的な評価は十分に達成できていない。その理由は、個々の案件の重点分野との合致及び成否は検証できても、課題体系におけるそれらの位置付けやプログラムとしての構成の適切性が検証できなかったことにある。そしてこれはCPの評価可能性の問題に起因するところが大きい。以下では、これまでの国別事業実施計画の手法に焦点を当てて、問題を分析する。

1)国別事業評価の概況

JICAでは1989年から国別評価の名称で、主要 セクターの代表的な案件を選び、積み上げる形で 国別の援助評価を行ってきた。しかしながら同方 法ではJICAの援助の重点セクターや課題の選定 の妥当性を検証できないとの問題認識のもと、 1998年度のバングラデシュ国別評価より、上記妥 当性の検証と援助戦略へのフィードバックを目的 とする評価が開始された(国際協力事業団、 2000a)。その後、1999年度に国別アプローチの強 化に向けて国別事業実施計画が導入されたのを受 け、国別評価も国別事業評価と名称を代え、メキ シコ(1999年度)、タンザニア(2000年度)、ボリ ビア(2000/2001年度) ホンデュラス・パナマ (2001/2002年度) スリランカ(同左) セネガル (2002/2003年度)の評価(2000年度から名称を国 別事業評価に変更)が行われている⁵。8件の評価 対象範囲(期間・調査項目) 評価手法の概要は 表1のとおりである。

過去の評価のアプローチをウイーン・ワークショップでの議論に照らして見ると、対象期間はバングラデシュの25年を最長に、概ね10年程度にわたり、対象分野もセネガル以外は主要援助分野を網羅するなど、対象範囲では「ヘビー型」に入る。ただし分析深度の方は、援助セクター・課題の選択の妥当性及び個別案件の積み上げに基づく実績の検証が中心であり、インパクトの評価はほとんど行われていない。

2) 評価手法とその問題

表1に示すとおり手法に関しては、援助セクター・課題の選択の妥当性を検証し、当該分野の個別案件の実績を検証することで、開発への貢献を

30 三輪 徳子

表1 これまで国別事業評価の対象(期間、項目)と評価手法の概要

国	期間	項目(報告書構成)	手法
バングラ デシュ	1971 - 1996	開発の推移・現状、マクロ評価 (重点セクター・課題)、プロジェクト評価、教訓	マクロ評価:開発の推移・現状分析に基づき、重点セクター及び主要4セクターにおける課題選定の妥当性を検証。 プロジェクト評価:既存の現況調査結果レビュー及び主要8案件の評価(評価5項目及び貧困・ジェンダー視点)により援助効果を検証
メキシコ	1988 - 1998	開発ニーズ、形態別評価、セクター別評価、総合評価、提言	形態別評価・セクター別評価:個別案件を評価 5 項目で 5 段階でレーティングのうえ、形態別・セクター別(7セクター)に平均点を算出、 総合評価:DAC 5 項目別に形態別・セクター別評価を平均
タンザ ニア	1991 - 2000	経済社会及び援助動向、事業評価(国レベル、プログラム)、教訓・提言	国レベル評価:経済社会及び援助動向に基づき妥当性、効率性の視点で評価、 プログラム:主要4セクターの個別案件を国別事業実施計画の課題別に12プログラムとして分類し、個別案件評価結果を積み上げる形でプログラムの妥当性、効率性、自立発展性を検証。
ボリビア	1985 - 1999	経済社会開発の推移と現状、 案件評価、セクター評価、横断 的評価(スキーム別、貧困・ジェ ンダー)総合評価、提言・教訓	案件別評価:主要4セクターの個別案件を評価5項目で評価、5段階でレーティング、 セクター評価:上記4セクターの推移と現状に基づき開発重点課題を体系図に整理、同図上にJICAの協力を位置付けて妥当性を検証、スキーム別評価:案件別評価結果を3スキーム別に分類、平均点を算出、貧困・ジェンダー:主要案件についてジェンダー・貧困配慮や効果を確認、総合評価:以上の結果を総括
ホンデュ ラス ・ パナマ	1991 - 2000	経済社会開発の推移と現状、 分野別評価、横断的評価(地域別評価、スキーム別評価、貧困・ジェンダー評価)提言・教訓注:地域別評価はホンデュラスのみ	分野別評価:主要5分野の協力の妥当性をポリビア・セクター評価と同じ手法で検証。あわせ各分野の個別案件をポリビアの案件別評価と同様の手法でレーディングするとともに、同じ上位目標の案件群を事後的にプログラムとして整理し、個別案件の平均点でプログラムの有効性を評価。 地域別評価:地域別の開発度と協力実施状況から地域選択の妥当性を検証するとともに、個別案件評価結果を地域別に分類して平均点を算出。 スキーム別評価及び貧困・ジェンダー評価:ポリビアと同じ手法を適用。
スリラ ンカ	1990 - 2000	社会経済開発の現状及び援助動向、個別案件評価、分野別評価、教訓・提言	個別案件評価:主要5分野の案件のうち、技術移転プロセス及び効果の観点から総合的に評価可能な案件群をプログラムとし、投入、活動、成果、目標を整理のうえ、評価5項目でプログラムを評価。 分野別評価:対象国の開発状況をレビューし政策目標・体系を整理のうえ、同体系にJICAの協力を位置付けて妥当性を検証するとともに、政策目標ごとの協力内容からパーフォーマンスを評価。
セネガル	1992 - 2001	援助重点分野の包括的評価、 水産分野評価、相乗効果評価、 提言	包括的評価:対象国経済社会状況、援助動向に基づきJICAの協力の妥当性を包括的に評価、 水産分野評価:プログラム・アプローチ・ロジック・モデル(PLM)を用い、目標を共有するプロジェクトを束ね、成果ごとに組み替えてプログラムを構築のうえ、各プログラムの実績、プロセス、インパクトを評価。また、住民への裨益の視点で評価。 相乗効果評価:統合的アプローチの効果を、水を中心に保健や教育を組み合わせた支援を実施の村落とそうでない村落との比較を通じて検証。

(出所)各報告書をもとに筆者作成

総合的に評価する方法が基本となっている。前者については開発の推移、セクター・課題分析、他ドナーの援助動向等から重点セクター・分野を明らかにし、右との整合性を検証する方法がとられており、主として叙述手法が用いられている。後者については個別案件評価結果の積み上げにより、当該セクター・課題分野の総合的な実績を検証する方法に拠っており、定性的な手法とレーティングを用いた手法の2種類がある。

ただし、重要セクター・課題への介入及び個別 案件の実績を検証しても、選択したセクター・課 題分野の中で課題解決にとって優先度の高い介入 が戦略的に行われているかが示されなければ、開 発への貢献度の総合的な評価は十分にはできな い。バングラデシュはセクター・課題選択の妥当 性、メキシコは個別案件実績の積み上げに主眼を 置いた評価が試行されているが、セクター・課題 における個別案件の位置付け(優先度、プログラムの構成の戦略性)の分析はなされておらず、セクター・課題レベルの評価と個別案件レベルの評価がつながっていない。

タンザニア以降の評価では、 開発課題体系へのJICA事業の位置付け及び プログラムの視点の導入を通じ、同問題への対応が試みられている。 しかしながら必ずしも試みは成功していない。

3) 開発課題体系への位置付け

まず開発課題体系への位置付けに関しては、ボリビア、ホンデュラス・パナマでは課題ツリーによる体系図を作成し、同図上にJICA事業を位置付けて選択の妥当性を検証している。同図には他ドナーの事業も記されており、課題のどこにJICAや他ドナーが介入しているかが一目で分かる利点がある。しかしながら、体系図は評価者が課題分析により「あるべき姿」として作成したものであり、対象国の実際の政策・施策・事業体系との関係が明らかでない。このためJICAも他ドナーも介入していない課題が体系図の大半を占める中、介入の優先度の判断が難しい。

一方、スリランカでは、対象国の開発計画等をベースに政策・施策目標を整理した体系図を用い、JICAや他ドナーの事業がどの施策目標に関連するものかを示している。従って相手国の現実

の政策・施策体系における位置付けを確認できる。しかしながら、整理が「新たな道路網の開発」といった大まかな目標レベルまでであるため、施策と事業との整合性を示すに過ぎない。また、JICA事業と他ドナーの事業も同じ施策目標分野であることが分かるのみとなっている。

4) プログラムの評価

次にプログラムの評価も、関連する個別案件群を束ねてプログラムとし、個別案件評価結果を積み上げて実績を示すに留まっており、プログラム・レベルの目標に対する効果やプログラムの構成等、プログラム・デザインの戦略性はほとんど分析されていない。その背景には、これまでの国別事業実施計画は、特定課題に関する案件を積み上げてグループ化したのみで、明確な目標の共有や構成要素間の一貫性が確保されている訳ではなく、プログラムとして評価が難しいことがある。

そうした評価可能性の問題があるにもかかわらずJICAがあえてプログラムの評価を実施したのは、評価を通じてプログラムの改善を図る意図があったからと推察される。しかしながら、結果としては、プログラム・デザインの戦略性の分析が不十分なために、プログラム改善のための適切な評価足りえていない。

プログラムの評価におけるもう一つの問題は、対象国や他ドナーの事業との関係が視野に入れられていないことである。プログラム目標レベルの課題に関しては、通常、JICAのみならず、途上国自身や他ドナーによる取り組みも行われている。そうした課題に対する介入の戦略性や貢献度の検証では、既存のJICA事業のみを対象としたのでは不十分であり、途上国や他ドナーの介入との補完性の視点が欠かせない。開発課題体系へのJICA事業の位置付けは、そのためのものでもあったはずであるが、プログラムの評価ではそれが活かされていない。

課題体系への位置付け及びプログラムの評価の2つの試みは、援助協調先行国のタンザニア及びボリヴィアを評価対象に選択して導入されている。同選択からは、援助協調枠組みでもある途上国開発戦略を課題体系として念頭に置き、JICAが同体系に他ドナーとの協調も含めて戦略的に介

32 三輪 徳子

入してきたかを検証する意図があったと考えられる。そして 重点セクターにおける優先課題の選択、 その中での、途上国や他ドナーの事業とも一貫性のある、戦略的な案件群の選択、 個々の案件の実績、の各段階の検証を通し、対象セクター・課題の開発の進展に対するJICAの貢献度の評価を行うことを狙ったのではないかと考えられる。しかしながら試みは十分には成功していない。 以上のように、これまでの国別事業評価では、援助公野と関発計画等の重点セクター・課題との

以上のように、これまでの国別事業評価では、 援助分野と開発計画等の重点セクター・課題との 整合性及び個別案件の実績は検証できている。し かしながら、 課題解決手段としての介入分野の 優先度、 案件群の組み合わせの戦略性、 途上 国や他ドナーの事業との補完性の検証が不十分な ために、JICA事業と開発の進展を結ぶロジック や貢献の度合いの評価が困難となっている。こう した問題への対応と、国別事業評価の評価可能性 の改善は、今後の国別事業評価の主要な課題と言 えよう。

3. CPEをめぐる新たな動き

「1. はじめに」に言及のとおり、援助をめぐる最近の動きは、CPやCPEをめぐる議論にも変化をもたらしつつある。JICAが2004年に行った主要援助機関の国別プログラミングとCPEの現状に関する調査結果」(概要、表2のとおり)からは、ウイーン・ワークショップ以降の主要な動きとして、 CPへのRBM適用、 開発アウトカムと特定機関の援助のアウトカムの区分、 「貢献」を用いた援助の評価、の3つが挙げられる。

(1) CPへのRBM適用

RBMの進展に伴い、主要援助機関では、RBM の単位はプロジェクトからCPレベル、組織運営レベルに発展し、戦略計画策定と業績測定を通じてこれらをマネジメントしていくシステムが導入されつつある。その先駆は米国国際開発庁(USAID)であるが、国連開発計画(UNDP)、世界銀行、英国国際開発省(DfID)において導入済みのほか、デンマーク国際開発庁(DANIDA)

やオランダ外務省でも導入中の状況にある。

このうち、世銀はCPレベルであるが、USAID、DfID、DANIDA、UNDPは組織レベルのRBMである。後者においてCPは、「組織全体 各局 在外事務所」の組織体制の階層と整合する形で整理された目標及び計画体系の一部となっている。すなわち組織目標の具現化に向けて、特定国において目指す目標と同目標達成のための事業をまとめた計画と位置付けられている。

RBMの適用により、CPやCPEの有様にはいく つかの変化が生じている。その一つは、CPの評 価可能性の改善である。これはCPが明確な目標 に基づき、戦略計画策定手法を用いて結果と手段 のロジックや指標を整理のうえ、策定されるよう になったこと、また、業績報告のために実績が定 期的にモニタリングされるようになったことに起 因する。その一方、RBMによるアウトカム・レ ベルの実績のモニタリング導入により、従来、 CPEで行われてきた実績の検証はモニタリングが むしろ担い、説明責任を果たす形になってきてい る。こうした中、モニタリングと評価の違い。を ふまえ、CPEの役割を改めて見直すケースも見ら れる。その一例がDfIDであり、CPEでは国別援助 計画(CAP)のデザインの適切性、プログラムの 妥当性及び効率性(プロセスの効率性や援助協調 を含む)等に焦点をあて、学習目的をより重視し ている。

(2) 開発アウトカムと特定機関のアウトカムの 区分

第二の動きは、国レベルでの開発効果に関し、他の援助機関や対象国との協力を通じて達成されるべきものと、当該機関が直接に達成し得るものとを区別する動きである。例えばDfIDでは、前者は「開発効果」(development effectiveness)後者は「組織の効果」(organizational effectiveness)として区分している。そして、「開発効果」はDfIDが最終的に目指し寄与するものであるが、帰属の検証が難しい、より長期のアウトカム及びインパクトであり、DfIDのみでは管理やコントロールが困難としている。また、「組織の効果」は、DfIDがある程度は管理・コントロールし得、

表2 主要援助機関における国別プログラミング及びCPEをめぐる動き

機関	国別プログラミング	国別プログラムのモニタリング・評価
世銀	2003年からリザルツ・フレームワー久 RF)を用いた結果重視CASを導入中。RFは対象国の開発目標、CASのアウトカム・指標、世銀の介入事業・実績をマトリックスに整理。	進捗・完了報告は実施部門がCAS Progress Report 及びCAS Completion Reportで実施。CASCRは悉皆的に実施。結果重視CAS導入にあわせCASCRを評価部門がレビューする体制を導入。Country Assistance Evaluationは評価部門が独立評価として実施。CAEは開発アウトカム、世銀の協力内容、開発への貢献をカバー。開発アウトカムと世銀によるアウトカムを区別のうえ、「貢献」で評価。世銀アウトカムについては計画・実施の適切さ等コントロール可能な事項につきレーティングを適用。
UNDP	1999年に組織レベルのRBMを導入、組織重点分野を設定。CPはUNシステムの国別援助枠組み(UNDAF)の下に位置付けられ、対象国の開発目標、UNDAFのアウトカム・指標、同アウトプット・指標、パートナーとの関係等をマトリックスに整理。	RBMのもと部署ごとに年次業績報告を作成。同一環としてCPのアウトカム・モニタリングを実施。CPの主要アウトカムを選んで実施部門が評価するアウトカム評価(OE)も導入(時期は目的に応じ任意。1サイクル2-4件)。OEは、アウトカムの発現状況、発現要因、UNDPの貢献をカバー。評価専門部署は独立評価として国別評価(Assessment of Development Results)を実施(対象期間:5-7年程度)。ADRはUNDP事業の戦略的位置づけ、開発アウトカムと、UNDPの貢献をカバー。「貢献」ではパートナーシップ、援助調整、政策助言も評価。
USAID	1997年に組織レベルのRBMを導入。組織目標、戦略目標(SO)中間目標(IR)の体系をRF(ツリー)で整理。国別戦略計画は地域・国の事情に基づきSOを選定のうえ策定、SO指標を設定。	組織レベルでの年次業績報告の一環としてSOのモニタリングを実施。SO終了時に評価を実施。国別戦略計画は年次業績報告をもってアセスメントを行っており、評価は行っていない。SO達成度は2004年からレーティングを廃止、指標自体を提示。USAIDの事業とSOの達成度の説明については「帰属」から「貢献」にシフト。
DfID	1998年にPublic Service Agreement 導入。国別援助計画(CAP)はPSA目標達成に向けた組織レベルのRBM体系(組織全体 各局 国 プロジェクト)の中に位置。CAPではPSA目標に向けて期待される変化、毎年の変化(指標)DfIDの介入をマトリクスで整理	年次業績報告の一環として組織、局、国、プロジェクトの各レベルでモニタリングを実施。報告書では上記結果をとりまとめ、PSA目標毎の指標達成度を記号及び進捗に関する叙述により提示。2001年からCAP終了時には実施部門によるレビューを実施。CPEは2002年から手法開発を含めて3カ国で試行(対象期間:5年間)。視点はCAPのデザインの適切性、プログラムの妥当性・効率性・有効性、貧困・ジェンダー等の配慮。評価に際しては開発アウトカムと組織アウトカムを区別、前者には「貢献」の概念を適用。また、業績モニタリングと評価の差異として後者では学習目的を重視。
DANIDA	2003年に組織レベルの業績測定枠組み(PMF)を導入。分権化のもと大使館がPMFの一環として年間業務計画に基づき国別戦略(CS)及びプログラムを管理。RFは特定化されていない。	年間業務計画に対する結果として大使館からAssessment of Country Programeと Assessment of Sector Programmeを本省に毎年報告。前者はCSの目標、横断的イシュー、DANIDA援助セクター、貧困削減、援助協調、等の進捗状況を、後者はプログラムの目標・指標、セクター政策、対象国努力、他ドナーの支援等の状況を 4 段階でレーティング。CPEは評価専門部署が独立評価としてこれまでヘビー型CPEを実施。今後はCPEは対象セクター・地域を絞る方向。評価では合同評価をより重視。
オランダ	2002年からRBM導入に向けて分権化 のもとでの国別の計画、モニタリング、評 価体制を構築中。	RBMのもと優先5分野の年次業績報告を2004年から導入。CPEは過去に対象20年間、二国間協力関係までカバーしたヘビー型を長期間かけて実施。CPEのコスト・タイミングに対する反省、CPEの技術的な問題、政策変更によるプログラム援助へのシフトから、当面はCPE実施は見合わせ、テーマ評価等を重視。

(出所)各報告書をもとに筆者作成

34 三輪 徳子

帰属の検証が可能、かつDfIDが説明責任を有するアウトプットあるいは中間アウトカム・レベルの実績としている。そして後者には、政策の質、プロセス及びリソースなど、オペレーション・レベルの方策が含まれるとしている(DfID 2001)。

同様にUNDPでもアウトカムを、 対象国の開 発計画目標等にあたる「国家開発アウトカム」 (national development outcome), 国連 (UN)シ ステムの国別プログラムにあたる国連開発援助フ レームワーク (UNDAF) のもとでUNシステムが 合同で実現を目指す「UNDAFアウトカム」、 UNDPがその国別プログラムサイクル(5年間) で実現しようとする「国別プログラム・アウトカ ム」に区分している。そして「UNDAFアウトカ ム」はUNシステムが集団で説明責任を有すると し、UNDPが説明責任を有する「国別プログラ ム・アウトカム」は、UNDPの結果に対する影響 が明確に特定化できるレベルを設定のクライテリ アにしている。

また、同様の概念のもと世銀では、CASのリザルツ・フレームワークにおいて 長期戦略ゴール (対象国の戦略開発目標) CASアウトカム (対象国のアクションや他ドナーの支援は必要であるが、世銀の介入が直接に影響を及ぼすことが期待されるアウトカム) 世銀の優先分野と実績の3段階に区分している。そして評価の対象とするのはCASアウトカムまでであり、対象国の開発アウトカムは直接的には対象にしないとしている (World Bank、2003a)。

アウトカムの階層区分は、より高次の結果達成が求められる中で、パートーナーシップを前提として事業を計画・実施することが必要になってきたこと、また、あわせ組織としての説明責任範囲を明確にすることが必要になってきたことに基づいて出てきた概念整理と考えられる。

(3) 貢献に基づく評価

第三が、援助効果と当該機関の国別援助の因果 関係を捉える視点の変化であり、調査対象機関の 大半では、これまでの帰属(attribution)から貢献(contribution)に視点を移しつつある。

帰属がアウトカムの変化と介入の因果関係を立

証可能な形で捉えるのに対し、貢献は、変化と介入の因果関係を可能性の高さ(plausibility)の観点から捉える視点である。アウトカムの変化と特定の介入の因果関係を厳密に検証するのは大変に難しいとの問題を背景に、帰属に代わる視点として出てきており、様々な角度からの分析を通じて、因果関係の可能性を判断するものである(Mayne 1996)。

国レベルの援助効果等に対する貢献の視点の適用は、従来の帰属の問題に加え、より高次のアウトカムの重視、そのためのRBMや援助協調の進展等と不可分と考えられる。ちなみにUNDP(2003)は、貢献に関しては「特定機関が何を貢献したか」に加えて「他のパートナーとどのように協力したか」が重要である旨、結果がより明確に定義され、開発の進展のモニタリングや協力のデザインが適切であるほど、因果関係は明確になる旨、述べている。

貢献の視点は、主要援助機関における最近のCPEの評価基準等にも反映されている。例えばIDBでは、2003年に改定したCPEの実施指針において、CPEの評価基準として妥当性、効率性、有効性に加え、一貫性(coherence)を盛り込んでいる。そして一貫性の視点としてプログラムのフォーカスとともに他の開発アクターとの連携をげている(IDB, 2003b)。また、DfIDでもCPEのTORの中で、DfIDの援助のインパクトを評価するのではなく、DfIDが開発プロセスにどのような付加価値を与えたかを評価すべきこととし、対象国におけるDfIDの行動及び活動の適切さ及び妥当性、対象国政府の国内政策との一貫性、並びに国際的な開発目標とのリンクを評価する概念的枠組みを作るように記している(DfID, 2004b)。

開発アウトカムと特定機関のアウトカムの区分及び貢献の視点に基づき、主要援助機関では、「当該機関のアウトカムの達成+他のアクターとの協力 開発アウトカムへの貢献」のロジックで開発アウトカムに対する当該機関の援助効果を評価している。例えば、世銀の国別援助評価の指針では、報告書の項目として「対象国の開発政策の変遷」に続き、「世銀の援助」(インプット及びアウトプット)、「世銀の国別援助の貢献」(世銀による他のドナーとのパートナーシップ、対象国と

の政策対話、援助調整等)「開発効果の評価(アウトカム及び結果)」(各分野)を記すことになっている。これは上記ロジックを報告書の構成に反映したものと言える。

しかしながら、ロジックは整理されても、当該機関の実際の援助と開発効果との関係を検証すったがある。例えば「開発アウトカム」及び「組織のアウトカム」の測定方法には、「開発アウトカム」レベルの指標の変化への寄まず確認し、そのうえで当該機関の変化への寄を検証していくトップダウン・アプローチの2つまき機関の活動、次いで国レベルへと結果を積がといくボトムアップ・アプローチの2つがある。このうち前者の場合は当該機関の援助ががわいに、このもの問題、また、後者の場合は活動レベルから積み上げていった結果が開発アウトカムまでなかなかつながらないという問題が指摘されている(DfID, 2001)。

こうした問題に対応するためには、開発アウトカム・レベルの目標及び指標と、当該機関の活動及びその成果指標の間をつなぐ、適切な中間レベルの目標及び指標の設定が必要である。このため多くの機関がCPでリザルツ・フレームワークを活用して概念整理を図っているが、国別プログラミングに際しての最も難しいところとなっている。ちなみに世銀でも、国別援助戦略の総合的なレビュー結果として、最終的な目標とそのための手段は詳細に記述されているが、両者の間のリンク、すなわち国別援助戦略の中間アウトカムとその指標が欠けた"missing middle"の問題がしばしば見られることを、指摘している(World Bank, 2003)。

4. 評価の困難性を超えて ー実践的方法論の考察

JICAのこれまでの国別事業評価では、重点セクターの優先課題に対し、途上国や他ドナーの事業とも一貫性のある戦略的なプログラムであったかどうかが検証できず、その結果、JICAの協力が全体として対象国の開発にどのように寄与した

のかが十分に評価できていなかった。また、評価 の困難さの背景として国別事業実施計画の評価可 能性の問題があった。

CPEをめぐる新たな動きは、こうした問題への対応に一つの解決策を示す。同観点から、国別事業評価の改善に向けた方法として考えられるのが、 途上国の開発戦略をベースに、途上国や他ドナーの事業を最初から念頭に置いて、JICA事業の位置付けや案件群の組み合わせの戦略性を検証し、 同検証結果を個別案件の実績を組み合わせることで、開発の進展に対するJICAの寄与を貢献の視点から評価することである。また、 の結果を活用して国別事業実施計画を改善していけば、評価可能性も高めることができる。

以下では、これまでの国別事業評価の問題に基づき、第3節に記した新たな動きをふまえ、国別事業評価の改善に向けた方法論を考察し、試案を提示する。なお、国別事業評価の方法論として検討すべき点は様々あり得るが、ここでは第2節で明らかになった主たる問題への対応に焦点を当てて考察する。

(1)問題に対する対応の方向性

CPEに関しては、これまでインパクト及び帰属の検証並びにCPの評価可能性が主要な問題とされていた。このうちインパクト及び帰属の検証は、援助効果の捉え方自体が変わる中、開発アウトカムと特定機関のアウトカムの区別及び貢献に基づく評価によって、一つの解決策が示されつつある。また、評価可能性に関しても、CPへのRBM適用により、目標、指標、及び結果に至るロジックがより明確化されたことが改善につながっている。これらは国別事業評価に関し、これまでの問題への対応と評価の有効性を高めるための方法を考えるうえでも参考となる。

国別事業評価の改善に向けては、いくつかの検討を要する点が考えられる。

1)評価可能性の問題

第一は、そもそもの国別事業実施計画の評価可 能性である。ウイーン・ワークショップでも指摘 36 三輪 徳子

されたように、CPが目標の明確性や構成要素の 首尾一貫性を欠く場合には、評価は非常に難しい。 過去の国別事業評価における問題も、国別事業実 施計画の評価可能性の問題が大きかったことが根 本的な要因となっている。

国別事業実施計画については、形としては「協 力実施方針 課題別プログラムー個別案件」の戦 略計画となっており、また、「開発課題の分析 協力プログラムの抽出 構成要素となる個別案件 の計画」のステップで策定し、戦略的アプローチ のツールとして活用することが企図されていた。 しかしながら、実際の事業の計画、予算、管理、 報告の単位が個別案件であったことからも、上記 アプローチは十分には進まず、個別案件をベース に同じような課題に関する案件群をまとめて課題 別プログラムとして整理するに留まってきた。こ のため課題別プログラムはプログラムとしての明 確な目標や構成要素の首尾一貫性が十分ではな く、また、「個別案件 プログラム 課題」の間 のロジックがつながらず、結果として課題に関す る開発の進展にJICAがどのように寄与したかが 容易には検証できない状況にあった。対象国の開 発の進展に対するJICAの寄与を検証するには、 さらに言えば、JICA事業が対象国の開発の進展 に全体としてより大きな効果をあげるようにする には、RBMのリザルツ・フレームワークのよう な枠組みの構築が必要である。

なお、この点に関してJICAでは、2004年度から個別案件ではなく、プログラムを計画、予算、管理の単位とする体制を導入し、プログラムをベースとする戦略的アプローチの強化に着手している。プログラムの構成要素の首尾一貫性及び戦略性等にはまだ課題の部分も少なくないため、プログラムの効果の検証はすぐには難しい面もある。しかしながら、評価結果を活用してプログラムを改善していくことは、戦略的アプローチを強化するうえで有益である。

国別事業評価の重要な目的は、対象国における 事業の総合的な効果の検証であるが、以上のよう に対象となる国別事業実施計画やプログラムが熟 度の点で発達途上にある場合には、それらの今後 の改善に向けた形成評価の側面も重要である。さ らに言えば、そうした評価可能性の低い段階では 形成的な側面により多くの比重を置き、国別事業 実施計画やプログラムが評価を通じて改善される にあわせ、効果の検証を中心とするといったアプ ローチの方が、評価可能性の低い中で困難、また は信頼性の低い評価を行うよりも実際的かつ有用 性が高いと考えられる。

2) 事業の位置付けの検証

第二点が、途上国や他のドナーの事業との関係の視点である。当然ながらこの中には、日本のODAにおけるJICA事業以外の事業、とりわけ外務省による無償資金協力やJBICによる有償資金協力が含まれる。

過去の国別事業評価では、途上国の重点分野や他ドナーの動向、あるいは同じ課題分野における他ドナーの事業の有無は見ているが、途上国や他ドナーの事業がJICAの事業とどのように課題解決に向けて関係しているかは十分には検証されてこなかった。そしてそれが、課題体系におけるJICA事業の位置付けの優先度、戦略性、構成の適切性の検証が十分にできなかった要因の一つになっている。

ウイーン・ワークショップの議論に見られるように、当該援助機関が対象国の主要ドナーかつ援助が対象国の開発に中心的な役割を占めるような場合を除き、単独ドナーの援助のみを見て開発の進展に対する開発アウトカム・レベルの効果を検証するのは難しい。そしてJICAの技術協力は規模の点では、程度の差はあれ、対象国の援助に占める割合は限られている。援助量から効果を見るのであれば、少なくとも日本のODA全体を視野に入れて、その中に位置付けて評価を行う必要がある。

言うまでもなく、援助が与える影響は、援助量のみで決まるわけではない。特定課題の開発戦略策定や、開発取り組みの中心となる仕組みづくりに対する技術協力のようなケースでは、投入自体は少なくても、当該課題分野の開発の進展に与える影響は非常に大きい。その意味でも、課題体系のどこに位置し、開発の進展にどのような影響を与える協力か、を明らかにすることが効果の検証には重要である。そして影響を見るためには、途上国や他のドナー等、他の開発アクターとの関係

を捉えることが必須である。

以上のような点を検証するためには、独自に作り上げた課題体系ではなく、実際に途上国や他のドナーが行っている包括的な取り組みの中にJICA事業を位置付け、その優先度や戦略性を吟味することが必要となる。そのベースとしては、援助協調の枠組みかつ、日本の国別援助政策及びJICAの国別事業実施計画の根拠でもある、途上国自身の開発戦略が最も適切である。

当然ながらJICAの介入は、途上国の開発戦略の中で日本の援助政策や、日本の経験・技術面の比較優位等に合致するものを選択して行われている。従って、優先度が高い分野であっても、介入が行われていないことはあり得る。それは他のドナーにおいても同じであり、援助協調の意義には比較優位を活かした補完もある。途上国の開発戦略における位置付けや途上国や他ドナーの事業との関係性では、そうした角度も含めて見る必要はある。

(2) 実践的国別事業評価方法(試案)

上記方向性に基づき、対象国の開発の進展に対するJICA事業の総合的な効果の検証及び 国別事業実施計画と事業の改善に向けた評価を行うための具体的な方法(試案)を以下に示す。

1)途上国開発戦略における事業の位置付け及びプログラム構成要素の戦略性の検証

事業の位置付けの検証のためには、 重点セクター・分野との合致、及び その中での介入分野の優先度の検証が必要となる。このうち は、開発計画等との整合性を見る形で、これまでも検証されてきている。問題があったのは の方であり、これを見るためには、特定課題における途上国の開発戦略の中でのJICA事業の位置付けや、途上国及び他の事業との関係性をより詳細に見ることが必要になる。

そのためのツールとして提案するのが、表3の「プログラム評価マトリックス」(以下、マトリックス)である。同マトリックスの特徴は、 特定課題に対する途上国の「政策 施策 事業」体系を課題達成のための目的体系、すなわちリザル

ツ・フレームワークとして整理のうえ、 JICA 事業ほか、途上国や他ドナーの事業をマッピング していること、さらに 施策や事業に対する途上 国の予算配分の概算を掲載していることである。また、後のステップでJICA事業の実績、開発の 進展及び貢献を検証する際に活用する、 個別案件の実績(レーティング)、 政策、施策レベルの開発目標とそのアウトカム指標、及び 貢献の判断結果(レーティング)も盛り込んでいる。

相手国側の予算を盛り込むのは、公的財政レビュー(public expenditure review)の考え方に基づくものである。これは予算配分は、当該事業が相手国側にとって本当に優先度の高いものかの判断基準の一つとなるからである。

表3は保健セクターを例としたものであるが、 上記のような整理及びマッピングにより、相手国 の政策、施策、事業との関係や、当該課題に関連 する事業の分散が一目瞭然であり、他ドナーの事 業との関連性も分かる。このため、 JICAが相 手国にとって優先度の高い分野に援助を行ってい るか、また、 他ドナーとの協調、あるいは JICA事業の分散ぶりから見て、プログラムの組 み方は戦略的と言えるかが、明快に検証できる。 ちなみに表3のJICAプロジェクトYはHIV/AIDの サーベイランス強化に関するものではあるが、相 手国側の政策体系のもとでは関連事業がなく、予 算も配分されていないことを示しており、一般的 には優先度が高いとは言えない例であるい。

相手国側の政策体系をベースとするのは、「対応の方向性」に既述のとおり、途上国側の優先度を明示するほか、途上国や他ドナーの事業との関係性を見る上で適当だからである。あわせ、国別事業実施計画においてリザルツ・フレームワークが構築されていない中、新たに体系を構築するよりも、簡単かつ首尾一貫性の点で適切であり、PRSP、セクター開発戦略等をベースに効率的にマトリックスが作成できることがある。

さらに、RBMが進む中、開発指標の設定やモニタリングの取り組みが行われてきており、開発の進展を検証するためのデータの入手が容易なこともある。これらの指標は、今後、JICAがプログラム・レベルで目標・指標を設定していく際にも活用し得る。

38 三輪 徳子

表3 プログラム評価マトリックス:A国健保セクター例

予算(A国予算:百万ドル)

政策		施	策		事	業	JICA	貢献						
保健セクター開発目標	課題別	刂目標	施策(予算)	関係ドナー	事業	関係ドナー	実績評点 (予算)	評点						
1.感染症疾病 ・死亡率の低下	1 - 1 . 北部辺境・ 重点対策対 象地域にお	指標 対象7州のマ ラリア死亡 率	北部地域マ ラリア対策 プログラム (15)	WHO (政策支援)	新薬導入・ 実証 (9)	WHO	-	-						
指標: マラリア死亡率 B% C%	_{家地域にの} けるマラリ ア死亡率の 低下	学 F% G%	(15)		母親と子供向 け蚊帳配布 (6)	UNICEF 日本 (無草 の根無償)	•	ı						
HIV/AIDS罹患率 D% E% (以下、保健セ	1 - 2 . HIV/AIDS サーベイラ	指標 対象人口 H万人 I万	国家エイズ 対策プログラム	UNAIDS (政策支援)	テスト・キッ ト配布 (35)	USAID ビルゲイツ財団	-	-						
クター各目標)	HIV/AIDS 内	内 サーベイラ ンス強化	UNAIDS ーベイラ DfID,EU ス強化 (財政支援)	定点観測実施 (15)	WHO JICA(プロジ ェクトX)	5 (5)	4							
		J% K%	(50) 感染率低下 (100)	感染率低下	感染率低下	感染率低下	感染率低下	感染率低下	感染率低下		該当事業なし	WHO JICA(プロジ ェクトX)	2 (3)	1
	1 - 3 . 母親から幼 児 へ の	指標 母親から幼 児への感染			コニュティ啓 蒙活動 (10)	UNICEF USAID, DANIDA	-	-						
	HIV/AIDS 感染率の低 下	率 L% M%								地域保健医療 人材育成 (20)	JICA(プロ ジェクトZ)	4 (6)	3	
					地域保健医療 機関整備 (70)	JBIC	-	-						
			計				計							

評点付け

JICA実績:終了時評価の5項目評価結果(妥当性、有効性、インパクト)に照らし、課題別目標に対する効果が「非常に高い(6)」、「高い(5)」、「満足できる(4)」、「やや不十分(3)」、「不十分(2)」、「非常に不十分(1)」

貢献:途上国開発戦略におけるJICA事業位置付け、構成要素の戦略性(一貫性、パートナーシップ)、事業実績等に照らし、対象国の課題別目標レベルにおける開発の進展に対する貢献が「非常に高い(6)」、「高い(5)」、「満足できる(4)」、「やや不十分(3)」、「不十分(2)」、「非常に不十分(1)」

(出所)筆者作成(ガーナ保健セクター5ヵ年計画を参考に、概念整理用に筆者が作成した架空の事例)

なお、途上国の政策体系の活用に関しては、かっては体系が構築されていない国も少なくなかったが、PRSPやセクター・ワイド・アプローチ(SWAPs)の進展により多くの国でベースとなるような体系づくりが進んできており、活用し易い状況になってきている。また、既に体系がマトリックス等に整理され、それに基づき援助協調が実施されている途上国も少なくない。既にそうしたツールがある場合には、新たにマトリックスに整理する必要はなく、使われているツールを活用して位置付けや戦略性を検証することが合理的であるう"。

なお、体系図としてツリーでなくマトリックス を適用したのは、ツリーの場合、 記された課題 が政策体系の階層(政策、施策、事業)のどこに 位置するかが分かりにくいこと、 マトリックス の方が各事業の予算や合算を明快に示せること、

同じく実績等をわかりやすい形で比べ、その見直しに活用できること等がある。ちなみに表3は単純化しているため、政策や施策を各々一階層のみで表しているが、施策の中にも階層があるところ、実態に合わせ階層を設定することが事業の位置づけを明らかにするうえでは必要である。ただし論理性にこだわり複雑化しすぎると実践性は低くなる。提案のマトリックスにおいて課題を軸とせず、因果関係を十分に満たせなくても、実際の政策、施策、事業を軸に置いたのは、課題にすると論理的な因果関係に囚われ、複雑化しすぎる可能性があることに因る。

この関連で、これまでのいくつかの国別事業評価では、プログラム・アプローチ・ロジック・モデル(PLM)を使い、同一の課題に関する複数のプロジェクトを成果ごとに区分・再整理し、プログラムとして捉えた場合の対象プロジェクト群の姿を明らかにしたうえで、妥当性や効果を検証する試みが行われてきている。また、JICA事業のみでなく他ドナーの事業も含めてPLMを使い、プログラムを構築のうえ評価した試みもある(三好ほか、2003)。PLMは、単発で実施されてきたプログラム・アプローチに基づくより戦略的な援助の検討に資する評価を行うためのツールとして有益である。しかしながら、成果レベルでの組み替

えは過度に緻密な作業となりがちであり、実践的でない面も見られた。このため本マトリックスでは単純にプロジェクトをユニットとしてマッピングのうえ、同一施策のプロジェクト群を協力方方のでして捉え、構成要素の戦略性を見るつりには、構成要素の戦略性を見るつつだっただしJICA事業の中には、一つではない、相手国側政策体系の一事業のは個別である。その場合には、何らかの形での要素の分離及び必要となる。ただし、その場合にも、みみ換えが必要となる。ただし、その場合にも、みみ換えが必要となる。ただし、変化の対象であるのりによび、より合理的と考えられる。

2) JICA事業の実績の検証

以上によりJICA事業の位置付け及び戦略性を 検証したうえで、施策毎の事業の実績を検証する。 同検証の目的は、プログラムの見直し(予算配分 見直し、構成要素のスクラップ・アンド・ビルド 等)及び開発への貢献度の検証の2つである。

まず、実績に関しては、JICA事業のみによるアウトカム実績の検証は、既述の評価可能性や、指標の設定・収集等の問題から難しさがある。そこで本方法では、アウトカムの実績は貢献により検証し、JICA事業の方は構成要素の案件が貢献し得る状況にあるのかないのかを検証することを提案する。具体的には、介入の優先度や戦略性が高いと判断できる場合に、案件が支援する相手国側施策または事業に貢献する効果を上げているかどうかを、個別案件評価結果から検証する。ただし将来的にプログラムの評価可能性が高まり、達成目標やその指標に基づき容易に評価ができるようになれば、特定の協力プログラムの実績を特定機関のアウトカムのレベルで検証していくことは考え得るであろう。

個別案件の評価は基本的に、既存の評価結果を 活用した2次評価手法にて行うことが適当と考え られる。その理由は、既に個別案件ごとにDAC 評価5項目に基づき評価を実施している中で、こ れまでのように国別事業評価において、再度、調 査を行うのは効率的でないからである。2次評価 では、妥当性は既に位置付け・戦略性で検証して 40 三輪 徳子

いることから、結果に関わる評価項目(有効性、インパクト)を中心に評価を行う。その際には、 案件間の比較や相手国側施策・事業レベルでの合 算・比較ができるように、そしてプログラム見直 しに活用できるように、レーティングを用いる。 ただし、レーティングはそれのみでは誤解を招き やすく、事業の改善に必要な情報も必ずしも十分 には得られないため、判定理由となる効果の現状 等について定性的な説明を付記することが望まし い。

3) 開発の進展と貢献の評価

開発の進展への貢献は、位置付け・戦略性の検 証及び実績の検証を組み合わせ、貢献の視点を用 いることで、厳密ではないまでも検証することが できる。すなわち、貢献の視点では、「相手国側 政策体系の優先部分への、他ドナーと協調しての 戦略的な介入 + 個々の介入における良好な実績 = 開発の進展への貢献」のロジックが成り立ち、開 発の進展は途上国及びドナーが共同で達成したも のと言えるからである。言い換えれば、事業が開 発の進展に効果があったことを検証するために は、貢献の視点を用いる場合にも、 介入が相手 国側政策体系において優先度が高く、 JICAO 他の事業、途上国の事業、他のドナーの事業との 関係において戦略的であり、 そのうえで個別の プロジェクト目標が達成され、上位目標が達成あ るいは達成に向かって効果をあげていること、の 検証がなされることが必要となる。

なお「貢献」の概念による開発の進展への貢献 度の評価では、他のドナーも主として叙述手法を 用いている。ただし、プログラムの見直しに向け てプログラム間の比較を行ったり、説明責任の観 点からよりわかりやすく結果を提示するために は、何らかの形での定量的な手法も有益である。 ここでは、その一つとしてレーティングを提案け ている。項目は、貢献の判断に必要な「位置付け」 (優先度)「戦略性」(一貫性、パートナーシップ)、 「実績」などを各々レーティングして合算するこ とが考えられる。また、開発課題に関係する相手 国側事業、他ドナーの事業、JICAの事業の予算 が算出できるような場合には、予算割合をもって 貢献の度合いを視点の一つに盛り込み、レーティ ングすることも考えられる。ただし、その場合は、「量」に加えて、他のアクターとの関係性から影響力が大きい介入かどうかといった「質」の面も勘案する必要はあろう。

4) プログラムの改善

以上の評価結果を活用し、プログラムを改善していく際には、いくつかの視点が考えられる。一つは介入分野の選択である。具体的には、マトリックスによる整理を通じて、相手国側の政策体系の全容及び優先分野を、体系の構成、予算配分、他ドナーの介入の現状等から検証し、JICAの得意分野も勘案のうえ、どの施策・事業を支援するかを吟味してプログラムに反映させていく。

第二がプログラムの戦略性の向上である。これについては、 JICA事業間、日本の他のODAスキーム、他ドナーとの協調の有無をはじめ、 課題別・階層(政策・施策・事業)別の分散状況、

各施策に関連する個別案件の実績や予算配分状況から、より効果を上げ得る構成にプログラムの構成を変えていく。その際には、実績が低く、優先度も低いような案件あるいは課題分野においては、サンセット方式を含めてスクラップし、より効果が見込まれる分野において案件形成・発掘を図っていく方法、あるいは予算配分をあらかじめ決め、それに従って案件を形成していく方法等が考えられる。

最後に、評価を通じた改善には、プログラムの 評価可能性の改善もある。評価可能性を高めるに は、目標の明確化、目標達成度を検証できる指標 の設定・データの収集、適切な因果関係のロジッ クの構築が、主たる取り組み事項となる。以上に 述べた第一及び第二の視点での改善により、貢献 に基づく因果関係のロジックは改善される。

また、JICAのプログラムでは既述のとおり、 目標及び指標の問題があるが、評価を通じて相手 国側の政策体系のどこに対象案件が属するかが明 らかになるため、同体系の同一施策に属する案件 群をまとめてプログラムと捉え、施策目標をプロ グラムの目標に設定、あるいは書き換えていけば、 相手国側の政策体系と合致した明確な目標設定が 可能となってくる。同様に、そのようなプログラ ムの場合には、相手国側の施策の目標達成度を測 る指標をJICAのプログラムの達成度を測る近似 指標 (proxy indicator) とすればよいことになる。

相手国側の指標との統合は、三好ほか(2004) も近年の援助潮流に基づき提唱しているが、指標 が設定されればモニタリングが可能となることか ら、評価の際の効果の検証は容易になる。また、 相手国側指標の統合は、合同でデータ収集ができ るため、入手可能性・実用性が高い。このように 評価を通じて不足を埋めていくことで、評価可能 性を含めて漸進的にプログラムを改善していくの は、プログラム・アプローチを着実に強化してい く実践的な方途と考えられる。

なお、評価結果を以上のようなプログラムの改善に活用していくためには、プログラムの見直しのタイミングと合う形で評価を行うことが望ましい。日本ODAに関しては、国別アプローチの強化に向けて2003年からODA現地タスクフォースが設置され、現地の大使館、JICA事務所、JBIC事務所がメンバーとなって、途上国と政策対話を行いつつ、援助計画案を作る体制が進みつつある。JICAの国別事業実施計画も同援助計画に基づき策定されるが、そうした現地タスクフォースによる援助計画案策定のタイミングにあわせ、国別事業評価を行うのは一つの有益な方途と言える。また、計画の一部を見直すような場合には、見直すセクターや課題のみを対象として、同じ方法を用いて評価を行うことも考えられる。

5. おわりに

以上、本稿では、これまでの問題や新た動きを整理し、問題への対応の観点から、JICAの国別事業評価の改善方法を考察してきた。これにより、主たる問題の解決に向けて、ある程度の概念整理はできた。今後の大きな課題は、試案の実証と改善である。

なお、今後の課題に関し、本提案は、これまでの主要な問題への対応に焦点を絞っている。従って、日本の援助政策との整合性の検証等、当然ながら検証すべき事項ではあっても、言及していないものがある。ちなみに日本の援助政策との整合性はこれまでも国別事業評価では検証してきてい

るが、実証では、これらも含めて実施のうえ、方 法論を試す必要がある。

また、位置付け、戦略性、実績については事実 の検証を中心としており、要因分析や反事実的状 況(例えばJICA事業がなかった場合はどうであ ったか等)は、今回の提案では掘り下げていない。 これらに関しては今後の検討課題である。特に JICA事業がなかったらどうであったか、あるい はそれだけの援助量が必要であったかのか等の点 は、説明責任の観点を含めて重要であり、検討し ていく必要がある。その際には、対象となる国別 事業実施計画の評価可能性をはじめ、評価のコス ト(費用、人員、期間)、得られた評価結果の活 用に照らした費用対効果、評価能力等から見た評 価のフィージビリティも見極めつつ、実践的な方 法論を考えていくことが必須と考えられる。評価 はツールであり、ツールとして適当か有用かが方 法論を議論する際にも最も重要な点ではないだろ うか。

謝辞

本稿は現在、JICAで実施中の国別事業評価の見直しに関する評価研究会(アドバイザー:上野宏神戸大学教授、三好皓一立命館アジア太平洋大学教授、神田道男JICA上級審議役)での議論及び主要援助機関に対する訪問聞き取り調査(筆者及び岩間望JICA企画調整部事業評価グループ職員)の結果に負うところが大きい。この場を借りて関係者に感謝申し上げる。ただし、本稿の内容は筆者個人の見解を示すものであり、JICAの正式見解を示すものではない。また、貴重な指摘をいただいた匿名の査読者に感謝申し上げる。

注記

- 1 国別プログラムは、援助機関によって国別援助戦略、 国別援助計画等、様々な名称が使われおり、評価も これに応じて国別戦略評価、国別援助評価等の名称 がある。本稿では、総称としてはDACでも使われて いる国別プログラム及び国別プログラム評価の表現 を使い、特定機関に関する場合のみ、当該機関で使 われている名称を使用する。
- 2 日本では、外務省の国別援助政策・計画に基づき、

42 三輪 徳子

JICAは国別事業実施計画、JBICは国別業務実施方針を策定して、国別アプローチによる事業を実施している。また、外務省が援助政策を対象に、JICA及びJBICが各々の事業を対象に、国別の評価を行っている。同評価は、外務省では国別評価、JICAでは国別事業評価、JBICでは国別レビューの名称が用いられている。

- 3 ウイーン・ワークショップ以降、DACではCPEに関する包括的な議論は行われておらず、IDBによる中南米4カ国でのCPEの経験に基づく非公式ワークショップ(2003年)などが単発的に開催されているに留まる。
- 4 この他、ウイーン・ワークショップでは、対象国とのパートナーシップの重要性、重複による無駄や対象国の加重負担を防ぐための援助機関間の協調の推進等も議論された。そしてCPEの目的や評価基準の明確化、適切な方法論の選択、合同評価や情報共有の推進、フォローアップ手続きの明確化等が、CPEのグッド・プラクティスに関する提言に盛り込まれた。
- 5 タンザニア、ボリビア、ホンデュラス・パナマ、ス リランカ、セネガルはコンサルタントに委託して 「外部機関による評価」として行われている。
- 6 これは評価の意図や開発課題体系、プログラム等の 鍵となる概念の共通理解に齟齬があったからと思わ れるが、その一因として、JICA側でも概念や方法論 がまだ整理・確立されていなかったために、業務指 示書の調査項目の記載が曖昧で意図が伝わりにくか ったことも考えられる。
- 7 文献レビュー及び訪問調査(2004年4月11日~17日:世銀、IDB、UNDP、UNICEF、USAID, CIDA。同5月25日~6月3日DfID、IOB、DANIDA, SIDA, BMZ, GTZ。いずれも団長・筆者、団員・岩間望JICA企画調整部事業評価グループ職員。)本調査結果詳細は、「総合分析・国別事業評価」報告書(2005年3月発行予定)に掲載。
- 8 実績モニタリングでは定量的な指標に基づいた結果 達成の有無が検証の焦点となるが、評価は定性的な 方法も含む、より掘り下げた分析を通じて要因やプロセスも検証するところに特徴があるとされてい る。(DAC、2000.pp.104-118)
- 9 貢献については、定義はまだ必ずしも確立されておらず、DACの「評価及びRBMに関する用語集」ほか、評価関連の専門用語辞典にもまだ載っていない。

- UNDP等では帰属と貢献の説明においてMayneの Addressing Attribution through Contribution Analysis (Mayne 1999)を引用しているが、帰属と貢献に関して言及した文献には、カナダの会計検査院や国際開発研究センター(IDRC)のものが多い。ただしMayneや会計検査院の議論の対象は政策・施策一般であり、政策・施策の業績測定のコンテクストでの議論となっている。このことからは、政策・施策に関する議論が、援助におけるRBMの進展も背景に、従来から帰属が難しいとの議論のあった援助評価に導入されたと考えられる。
- 10 ただし、援助が入っているために予算配分がされていないことも場合によってはあり得る。そうした場合は、オーナーシップと優先度の捉え方も含め、総合的な価値判断が必要になる。同様に、途上国側の優先度が低い分野にドナー側の関心で入っているような場合をどう捉えるか、といった判断もまた必要になる。
- 11 相手国側政策体系の活用に関しては、相手国側の開発戦略が取り組むべきと思われる課題を必ずしも含んでいないといった問題指摘があり得る。しかしながら、政策は最終的には選択であり、価値判断の結果である。仮にそうした課題があれば相手国側の政策に働きかけ、盛り込まれるようにするのが援助効果議論の中での流れであり、評価において「あるべき姿」に基づき、実際の事業と乖離した架空の価値体系を作り上げ、それをもとに妥当性や効果を議論しても意味があまりないように思われる。
- 12 他のドナーの場合、CPには5年あるいは3年といったサイクルがある。国別事業実施計画はローリング・プランをベースとしており定まった期間はない。目標達成度を検証する観点からは期間を設定することが望ましいが、相手国側の開発戦略のサイクルと様々なドナーのCPのサイクルが異なることによる問題も調和化の観点から指摘されている。右を考慮すれば、評価さらに国別事業実施計画のサイクルは、相手国側の開発戦略のサイクルと合致させることも一つの方法である。これは相手国側の開発戦略が、サイクルが変われば変わることからも適当と考えられ、サイクル終了前年に次のサイクルの優先施策などもふまえつつ評価を行えば、プログラムの改善には有益であろう。

参考文献

- アイシーネット (2001)「外部機関による評価 ボリヴィア国別評価報告書」 国際協力事業団
- アイシーネット (2002a) 「外部機関による評価 ホン デュラス国別評価報告書」、国際協力事業団
- アイシーネット (2002b) 「外部機関による評価 パナマ国別評価報告書」、国際協力事業団
- 上野宏(2004)「政策工学試論2:政策プロセス、政策評価、及び予算策定」、『日本評価研究』第4巻第1号: pp. 66-86
- コーエイ (2001)「外部機関による評価 スリランカ 国別評価報告書」 国際協力事業団
- 国際開発センター(2001)「タンザニア国別事業評価 報告書(第2次調査)」、国際協力事業団
- 国際協力事業団 (2000a)「バングラデシュ国別評価報告書」
- 国際協力事業団(2000b) メキシコ国別事業評価報告書」 国際協力事業団(2000c) 「タンザニア国別事業評価事 前調査団報告書」
- システム科学 (2003) 「外部機関による評価 セネガ ル国別評価報告書」、国際協力事業団
- 三好皓一(2002)「プログラム・セオリー・マトリックスの活用について」、『日本評価研究』第2巻第1号:pp.11-27
- 三好皓一・平田慈花・和田智代・中澤哉・喜多悦子 (2003)「プログラム評価の試み フィリピン感染症 対策分野の評価を事例として」、『日本評価研究』第 3巻第1号: pp. 43-56
- 三好皓一・アレハンドロ クルース (2004)「援助プログラム・プロジェクトの評価手法の構築:援助国の評価システムとの統合を目指して」、『日本評価学会春季第1回全国大会発表要旨集録』pp. 79-86
- 三輪徳子(2004)「国別プログラム評価の現状と課題に対する考察 国別プログラム評価をめぐるドナーの動きとJICAの経験から」『国際開発学会第5回春季大会報告論文集』pp. 187-192
- Binnendijk, Annette (2000). Results Based Management in the Development Cooperation Agencies: A Review of Experience, a background report presented to the February 2000 meeting of the DAC Working Party on Aid Evaluation.
- Cossee, Olivier (2001). A Review of Country Programme

- Evaluations in Selected UN Organizations and International Banks, commissioned by UNICEF.
- DAC (1999). *Evaluating Country Programme*, Vienna Workshop 1999. Paris: OECD.
- DAC (2003). Report on the DAC Partnership Forum Managing for Developments Results and Aid Effectiveness. (Paris, 11-12 December 2002)
- DfID (2001). How effective is DfID? Development Effectiveness Repor.t
- DfID (2004a). Departmental Report 2004.
- DfID (2004b). DfID Country Programme Evaluation, Draft Synthesis Report.
- IDB (2003a). Summary Report on the Informal Workshops on Country Program Evaluation: Lessons Learned from the Recent Evaluations of Bolivia, Haiti, Honduras and Nicaragua. a report presented at the DAC Working Party on Aid Evaluation on 27-28 March 2003.
- IDB (2003b). Protocol for the Conduct of Country Program Evaluations.
- Iverson, Alex (2003). Attribution and Aid- Evaluation in International Development: A Literature Review. International Development Research Centre, Toronto, Canada
- Mayne, John (1999). Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly,: Discussion Paper, Office of Auditor General of Canada.
- SIDA(2004). Looking Back, Moving Forward. SIDA
- Smutylo, Terry (2001). Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs. a paper presented at Block Island Workshop on 22-24 May 2001.
- UNDP(2002a). Handbook on Monitoring and Evaluating for Results.
- UNDP (2002b). Guidelines for Outcome Evaluators.
- UNDP(2003). Development Effectiveness Report 2003 Partnership for Results.
- World Bank (1998). Assessing Aid. Washington D.C.:
 Oxford Press
- World Bank (2003a). *Country Assistance Strategy*: Retrospective and Future Directions.
- World Bank(2003b). *Country Assistance Evaluation*: A Primer.

(2005.1.26受理)

44 三輪 徳子

Beyond Difficulties of Country Programme Evaluations A Proposal of Practical Methodology

Satoko Miwa

Japan International Cooperation Agency (JICA)
Miwa.satoko@jica.go.jp

Abstract

Country Programme Evaluations (CPEs) have been conducted by many donors over a decade. However, the methodology is not well established due to problems related to attribution of impact and evaluability of Country Programmes. While undertaking eight (8) CPEs, JICA has been facing similar problems and is now reviewing its methodology. Meanwhile, ODA is shift towards results-based management and improved aid coordination led to review and stronger focus on aid effectiveness including its assessment. This paper is based on the review of discussions on CPEs and on JICA is experiences. It proposes a new practical methodology for CPEs. Major features of this methodology are use of programme evaluation matrix to verify the relevance of interventions, aggregation of programme performance, and assessment of programme effectiveness in applying the concept of "contribution". This paper also proposes gradual improvement of Country Programme is evaluability.

Keywords

programme evaluation, outcome, contribution, evaluability, results-based management

【実践・調査報告】

成果志向の道路行政マネジメントの軌跡 マネジメント・サイクルのスタートから実践へ

大庭 孝之 澤田 和宏 森若 峰存 塚田 幸広

国土交通省 道路局 国土交通省 道路局 国土交通省 道路局 国土交通省 国土技術政策総合研究所 oba-t82ad@mlit.go.jp sawada-k259@mlit.go.jp moriwaka-m82ac@mlit.go.jp tsukada-y27x@nilim.go.jp

要約

行政機関が行う政策の評価に関する法律の施行以来、中央省庁においても、政策評価が取り入れられてきている。国土交通省道路局では、政策評価制度の導入にあわせて、ユーザーにとっての成果を重視する成果志向の考え方を、組織全体の基本と位置づけ、アウトカム指標を用いた業績評価の手法を核とする新たな道路行政運営のしくみ(一連の行政運営の取組みを「成果志向の道路行政マネジメント」と称する)を導入している。

本稿では、成果志向の道路行政マネジメントの立ち上げに関わった者として、以下のような様々な取組みや分かってきた課題を紹介する。

- ・アウトカム指標の設定と業績計画書・達成度報告書の公表を通じたマネジメントのスタート
- ・評価結果を施策や予算要求に活かすマネジメントサイクルの一巡
- ・日常業務への取組みに向けたガイダンスの策定など

キーワード

業績評価、道路行政マネジメント、成果志向、アウトカム指標、達成度報告書・業績計画書

1. はじめに

(1) 国土交通省における行政評価の導入

行政機関が行う政策の評価に関する法律の施行により、平成14年度から、中央省庁各府省において政策評価が導入された。国土交通省においても、国土交通省政策評価基本計画を策定し、平成9年度より取り入れられていた、費用便益分析等に基づく個別事業の評価システムとともに、政策アセ

スメント(事前評価) 政策チェックアップ(業績測定) 政策レビュー(プログラム評価)の3つの方式を基本とする政策評価が導入された。

その際、国土交通省においては、「政策評価の 導入を契機とする政策目標によるマネジメント と、業務運営面の改革は「車の両輪」である」と し、「省全体の運営に加えて各部局ごとに業務運 営方針(ビジネスプラン)を策定するという双方 向の展開が必要である」としている(国土交通省 (2004))。

このため、道路にかかる政策の立案・実施(道

日本評価学会『日本評価研究』第5巻第1号、2005年、pp.45-55

路行政)を担う部局である国土交通省道路局においても、政策目標によるマネジメントを通じて成果を目指していくことを、業務運営にかかる大きな課題と位置づけ、成果志向の業務運営への展開を図った。

(2)成果を重視した道路行政への転換

道路行政については、事業採択の不透明感や事業の効率性に対する不信感が存在しており、これらの不信感を払拭し、国民のニーズに応え、成果を重視した行政運営へ向けてさらなる努力を進めることは、最も重要な課題の一つである。

このような状況を鑑み、社会資本整備審議会により平成14年8月になされた中間答申「今、転換のとき」においても、「事業量確保のための道路整備ではなく、道路が提供するサービスにより実現される成果(アウトカム)を重視した道路行政に転換することが重要」と指摘した。その上で、具体的に、「道路行政を、アウトカム指標に基づき運営を行う方式に転換し、毎年度、指標に基づき業績の分析・評価を行い、その結果を予算編成等に適切に反映させるシステムを構築すべき」と提言している。

これらを受け、国土交通省道路局では、ユーザーの視点に立ち、より効果的、効率的かつ透明性の高い道路行政へと転換するため、平成15年度より、ユーザーにとっての成果を重視する成果志向の考え方を、組織全体の基本と位置づけた。アウトカム指標を用いた業績評価の手法を中心に、政策の評価システムを核とする新たな道路行政運営の仕組み(これら一連の行政運営の取組みを「成果志向の道路行政マネジメント(Outcomeoriented Performance Management of Road Administration)」と称している)を導入することとした。

つまり、道路を何km整備したかなどの、舗装率、改良率、歩道設置率などの整備の量に着目した運営から、交通事故死者数の減少、都市内の道路渋滞による時間損失の削減といった成果に着目した運営を目指すこととした。

もっとも、歩道を整備すれば歩行者事故が減少するという、整備量が事業効果に直結すれば、整

備の量に着目していた運営にも意義はあることに 留意が必要である。また、道路整備五箇年計画の 策定に際し、高規格幹線道路延長等、事業の量を 表す指標に加え、ユーザーにとっての成果を表す 指標(アウトカム指標)を用いた整備目標を併記 し、道路行政の目的を定量的に明らかにしてきた。 しかし、成果志向の道路行政マネジメントは政策 評価の導入にあわせた業務運営の改革につなげよ うとしていること、予算や政策立案と、その実行 までを視野に入れた道路行政に関わる組織全体へ の展開をにらんでいることが、特徴となっている。

平成15年度は、有識者からなる研究会の提言をうけ、業績計画書の策定するなど、本格的な道路行政マネジメントをスタートした年であった。また、16年度は、1年間の取組みを評価し次の施策に生かすマネジメントサイクルが一巡した初めての年であり、業務運営の仕組み、つまり、道路行政マネジメントそのもののあり方を含めた評価を行えた(表1)。

(3) 本報告にあたって

筆者は、国土交通省道路局などにおいて、成果 志向の道路行政マネジメントの立ち上げに関わっ た者として、これまでの取組みを紹介するととも

表1 道路行政マネジメント取組み

平瓦	成15年度は、道路行政マネジメントをスタート
平成15年3月	道路行政マネジメント研究会
	(委員長:古川筑波大学大学院教授)設立
平成15年4月	道路事業分析評価室の設置
平成15年6月	提言「『成果主義』の道路行政マネジメントへの転換 - 理論から実践へ - 」
平成15年7月	平成15年度 道路行政の業績計画書 策定、公表
平成15年10月	社会資本整備重点計画 閣議決定

平成16年度は、道路行政マネジメントの実践へ

平成16年4月 業績予算の導入

平成16年6月 平成15年度達成度報告書・平成16年度業績計画書 第定、公表

(都道府県ごと等地域レベルでも、達成度報告書・業績計画書を策定、公表)

平成16年11月 道路行政マネジメントガイダンス 策定

(出所)国土交通省道路局(2004a)をもとに筆者が作成

に、明らかになってきた課題、それらの解決のための新たな取組みの方向を紹介したい。他行政分野においても、政策評価を業務運営の改革に活かすための参考として頂くとともに、行政機関の政策評価の進化に役立てられれば幸いである。

2. 道路行政マネジメント研究会の提言

(1) 研究会の設置

国土交通省道路局では、平成15年3月、ユーザー 志向で成果を重視した新たな道路行政を実践する にあたり必要な事項について、主に行政マネジメ ントの観点から具体的な検討を行うため、道路行 政マネジメント研究会(委員長=古川俊一筑波大 学院教授)を設置し、検討を進めてきた。

その結果、平成15年6月に、「『成果主義』の道路行政マネジメントへの転換 理論から実践へ」と題する研究会の提言がとりまとめられた。(道路行政マネジメント研究会(2003))

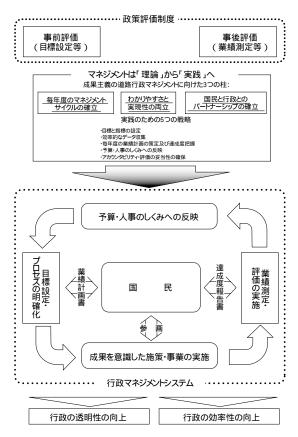
(2)提言の構成と背景

提言は、副題の「理論から実践へ」に謳われているように、形式的な理論を構築するのみではなく、実効性のあるマネジメントを継続的に実践することが重要であるという考え方が基本となっている。このため、提言の内容は、成果志向の実践に向けての基本的な方向や留意点等について提案されている。

提言では、まず、その背景として、我が国に限らず、各国ともに行政部門の効率化が共通の問題として認識されていることを指摘し、諸外国や我が国のこれまでの取組みを紹介している。一方、我が国の道路行政については、依然として事業採択の不透明感や事業の効率性に対する国民の不信感は根強く存在しており、これを払拭するため、成果を重視した透明性の高い効率的な行政運営へと転換することが急務であるとしている。

その上で、事業評価システム等、道路行政におけるこれまでの取組みも活用して、成果志向の行政マネジメントを実践に移すべきであり、平成15年度より、率先して、成果を重視した行政運営へ

図1 成果主義の道路行政マネジメントの実践に向けた3つの柱と5つの戦略



(出所)道路行政マネジメント研究会(2003)

の転換を目指すべきとしている。

さらに、成果志向の道路行政マネジメントを実践に移すために、成果主義の道路行政マネジメントに向けた3つの柱と実践のための5つの戦略を提示している(図1)。具体的には、毎年度のマネジメントサイクルの確立、わかりやすさと実現性の両立、国民と行政とのパートナーシップの確立を3つの柱に挙げ、また実践のための5つの戦略として、目標と指標の設定、効率的なデータ収集、毎年度の業績計画の策定及び達成度把握、予算・人事のしくみへの反映、アカウンタビリティ・評価の妥当性の確保が盛り込まれている。

3. マネジメント・サイクルを作ること からスタート:平成15年度の取組み

研究会提言の5つの戦略に沿って平成15年度の 取組みを概説する。

(1) 指標と目標の設定

施策がどのような成果を目指し、どれだけ達成したかを明らかにするためには、施策を目的別に体系化し、その成果を定量化するアウトカム指標を設定することが不可欠である。

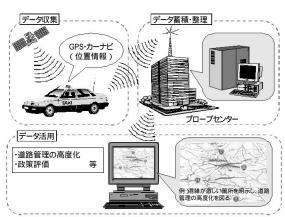
国土交通省道路局においては平成15年度業績計画書の中で、成果を表す指標として、17のアウトカム指標を設定し、平成19年度の目標値及び、その達成のための平成15年度、平成16年度の目標値を設定している。

これらの指標と目標については、社会資本整備 重点計画(平成15年10月閣議決定)及び国土交通 省政策評価基本計画(平成15年10月)の中でも位 置付けられている。

(2) 効率的なデータ収集

成果をあげるためにすべき施策、事業を一定の 技術的合理性のもとに企画・立案するためには、 現状の把握に必要なデータを、少なくとも毎年度

図2 プローブカー調査のイメージ図



(出所)国土交通省道路局(2004a)

収集することが必要である。そのため、道路局及び地方整備局等においては、情報通信技術等も活用し、より低コストに、高精度なデータを体系的に収集、分析する手法の開発・導入を進めている。

例えば、バス等にGPS等を取り付け、時刻と位置を調査することで渋滞状況を把握するプローブカーを用いると、時間帯別、方向別の渋滞状況を明らかにすることができる(図2)。これにより、従来の調査ではわからなかった詳細な渋滞状況を把握し、よりきめ細かい対策の立案が可能となる。

また、路線ごとの対策についても、渋滞損失時間や死傷事故率の状況を把握することにより、課題のある箇所を特定し、重点的に対策をすることが可能となる。

(3)業績計画の策定

道路局は、平成15年7月、平成15年度の道路事業による成果目標を示した道路行政の業績計画書を、我が国の公共事業で初めて策定、公表した(国土交通省道路局(2003))。業績計画書においては、指標の目標値を示すのみではなく、現状の課題を客観的データに基づき分析し、対策を示すとともに、判断の根拠となる客観的データについても併せて掲載している。

また、本省で作成する全国版に加え、地域の特性や、地域ごとのニーズに応じた成果目標を明らかにするため、都道府県、道路関係公団等、関係する道路管理者と協働し、地域住民や地域の有識者の意見も反映させた地域版の業績計画書も策定している。地域版業績計画書については、平成15年度は、32都道府県について公表した。

(4) 予算のしくみへの反映

成果志向の行政運営を真に実効あるものとするためには、予算の設定にあたり、成果目標に見合ったものとすることが必要である。しかしながら、筆者は予算の決定要因を全て定量的に表現し、目標や達成度に基づき、機械的に予算を立てることは不可能であると考えている。道路局では、成果目標に見合った予算設定を行うための第1段階として、渋滞の緩和や交通安全等の成果目標に対応



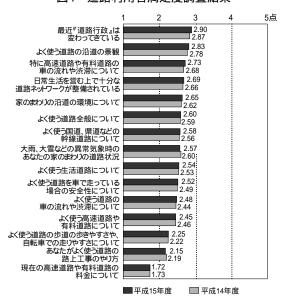


(出所)国土交通省道路局他(2004)

した業績予算を導入し、成果ごとの予算を明らかにする取組みを進めている(国土交通省道路局他(2004))。

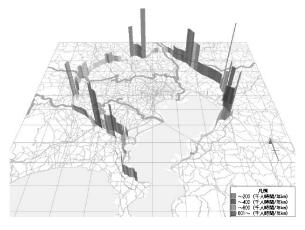
これまでは、国の管理する国道の改修にあてる「一般国道直轄改修費」、都道府県の管理する国道にあてる「一般国道改修費補助」、県道などにあてる「地方道改修費補助」といった道路種別予算であった。つまり、予算を執行する組織ごとの費目であった。平成16年度予算から渋滞対策にあてる「交通円滑化事業費」、都市間や地域内の効率的な移動を目指す「地域連携推進事業費」など、目指す目標ごとの費目分けをする業績予算への転換を図っている(図3)。このことにより、アウトカム指標による目標値と投入する予算を関係でカカム指標による目標値と投入する予算を関係のアカ

図4 道路利用者満足度調査結果



(出所)国土交通省道路局(2004a)

図5 渋滞3Dマップ(国道16号の例)



(出所)国土交通省道路局(2004a)

ウンタビリティ確保につなげることになった。これまでは、成果と結び付けなくても全体としての 予算を受けていた組織も、渋滞対策や交通安全な どの成果を意識した取組みを迫られ、道路行政マ ネジメントの推進の大きな仕組みになった。

(5)アカウンタビリティ・評価の妥当性の確保

透明性の高い道路行政マネジメントのためには、事前の目標設定、事後の評価結果に加え、背景のデータも積極的に公開することが重要である。そのため、前述の業績計画書については、全国版、地域版ともにインターネットを通じて公表し、広く利用者や納税者の視点からのチェックを可能としている。

また、業績計画書策定に際し、道路利用者の道路に対するニーズを収集するため、インターネットを通じて道路サービスに関する満足度を調査する道路利用者満足度調査を、平成14年度より実施し、不満の大きい分野に対する施策の企画、立案に反映する取組みを進めている(図4)、例えば、利用者から不満の多い事項は、高速道路や有料道路の料金、路上工事のやり方、歩道の歩きやすさや自転車での走りやすさが割り出された。これらは今後の取組みとして重点化していくべき事項として業績計画書に盛り込まれている。

また、国民に対し、数値目標や達成度のみでは

なく、現状を示すデータを公表することで、国民と行政が課題と目標を共有することが重要であると考えている。そのため、都道府県ごとの詳細な渋滞状況をわかりやすく明らかにした渋滞3Dマップなどを作成し、インターネットを通じ、広く公表している。例えば図5の国道16号の渋滞損失時間では、東名高速、中央道、246号、20号、17号、4号など放射方向の幹線道路との交差部周辺で激しい渋滞となっていることを表している。

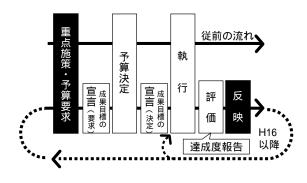
4. 道路行政マネジメントの実践へ:平成16年度の取組み

(1) 実践に向けた取組み

平成15年度業績計画書を公表することで、道路 行政マネジメントをスタートさせたものの、成果 志向への転換は理解できるものであるが、具体的 にどのように取組んでいくのか分かりにくく、地 域のニーズに合った国民が実感できる成果が見え るまでに至っていないとの意見が内外から出され た。

道路行政における成果志向に向けた取組みは緒についたばかりであり、職員への成果志向の意識の浸透や、納税者、利用者に対する分かりやすさを確保するための工夫など、未だ多くの課題を残している。実際の道路行政においても、アウトカム指標や目標の設定にとらわれすぎて、業績計画書と実務とが乖離しているため仕事は今まで通り

図6 新たな道路行政マネジメントの流れ



(出所)国土交通省道路局

という実態も否めなかった。

このような実態を踏まえ、筆者らは、道路行政マネジメントを日常業務に浸透させ、実効性のあるものにするためには、様々な課題に対して継続的な改善を図ることが重要であると考えた。このため、平成16年度からは日常業務への浸透を図ることを重視した道路行政マネジメントの実践のための取組みを行うこととした。具体的には、1)指標の現況値算出/達成度報告書・業績計画書の策定に加え、2)成果志向の浸透のための基礎的な仕組み作りを行い、3)日常業務への成果志向の組込みを図る。さらにこれらを通じて、4)国民と行政の信頼関係の再構築に努めることとした。

(2) 「達成度報告書・業績計画書」の策定

道路行政マネジメントの実践のための取組みの第1ステップとして、1年間の達成度を評価し、その結果を踏まえ、今年度の目標をまとめた達成度報告書・業績計画書を平成16年6月に公表した(図6、7)。このことで、事前に数値目標を宣言し、事後に達成度を評価し、評価結果を次の行政運営に反映する仕組みであるマネジメント・サイクルが初めて一巡したこととなった。

達成度報告書・業績計画書では、渋滞を減らす、

図7 平成15年度達成度報告書 ・平成16年度業績計画書



政策テーマ	指標	H14実績	H15目標	H15実績	H16目標	H19目標
	道路渋滞による損失時間		5.9億人時間/年 (H14比3%削減)			約1割削減
	道路工事時間	201時間/km·年		186時間/km·年 (H14比7%削減)		約2割削減
渋滞を減らす	全国	5%	15%	16%	50%	-
~ 道路交通の円滑化~	ETC利用率 首都高過	6%	20%	19%	60%	-
	阪神高道	3%	15%	11%	50%	-
	規格の高い道路を使う割合	13%	13% (H14より210万 _{台キロ/日転換)}	13% (H14より - 20万 台キロ/日転換	13% (H14より290万 _{台キロ/日転換)}	15%
地域を結ぶ ~地域間の連携~	拠点的な空港・港湾への道路アクセス率	59%	61% (青森港へアクセス)	61% (青森港へアクセス)	61% (中部国際空港へアクセス)	68%
~ 地球間の連携 ~	隣接する地域の中心の都市間が改良済みの国道で連絡されている割合	72%	73%	73%	74%	77%
	日常生活の中心となる都市まで、 30分以内で安全かつ快適に走行できる人の割合	63%	64% (約80万人増加)	64% (約60万人増加)	65%	68%
道路を守る	道路構造物保全率	86%	87%	87%	89%	93%
追給をする ~道路の維持管理~	這路桶這初休主 平 舗装	91%	H14の水準を維持	93%	H14の水	準を維持
	災害時に広域的な救援ルートが確保されている都市の場合	66%	68%	68%	70%	76%
	NO ₂ 環境目標達成率 ^{注1}	64%	67%	67%(53%)	-(77%)	-(約9割)
環境を改善する	SPM環境目標達成率 ^{注1}	-	約1割	9%(78%)	-(H15のオ	〈準を維持)
~ 環境の保全 ~	夜間騒音要請限度達成率	61%	63%	64%	65%	72%
	CO ₂ 排出削減量	261百万t-CO ₂			重輸部門における 万 t -CO₂まで削	
交通事故を減らす	道路交通における死傷事故率	118件/億台丰口	116件/億台丰口	119件/億台キロ	114件/億台丰口	108件/億台キロ (約1割削減)
~安全・安心な道づくり~	1日当たりの平均利用者数が5,000人以上の旅客施設の周辺等の 主な道路のパリアフリー化の割合	17%	21%	25%	30%	約5割
電柱・電線をなくす ~美しい景観の創造~	市街地の幹線道路無電柱化率	7%	8%	9%	10%	15%
道路行政の改革	道路利用者満足度	2.6点	2.7点	2.6点	2.8点	3.0点
担給行政の政事 ~アカウンタビリティの向上~	ホームページアクセス数	1,546万 アクセス/年	2,600万 アクセス/年	2,350万 アクセス/年	4,300万 アクセス/年	約1億 アクセス/年

表2 主な指標ごとの目標と実績

(出所)国土交通省道路局

地域を結ぶ、交通事故を減らすなどの道路行政の目指すテーマごとに分類し、成果を示す17の指標を用いて施策・事業の効果等を分析し、評価している。例えば、渋滞を減らすという政策目標は、道路渋滞による損失時間をどこまで減少しえたか、渋滞の原因となる路上工事時間をどこまで縮減しえたか、高速道路料金所での渋滞を解消する効果があるETC利用率がどこまで向上したかで、評価している。

達成度報告書・業績計画書のポイントは次の3点である。

b 目標の達成度を、都道府県別の達成状況などの データとともに、公表

1年前に宣言した成果の目標が、達成したか、達成しなかったかということを、成果を表す指標

注1:測定方法の定義の変更により、2通りの値を併記

の実績値とともに、公表する。(表2)

これまでは全国値でしか公表されていなかった ものも、施策を実施したところやそうでないとこ ろの比較や、都道府県や国道事務所別などの指標 のランキングなどのデータも公表。各主体の達成 度を公表することにより、施策や事業の責任主体 が明確になり、国民の視点からのチェックを可能 にする。

b 実施した施策・事業の効果を分析することに注力 成果志向の理念は、宣言した目標の達成度を分析し、次の施策・予算など行政運営に反映させる ことにある。数値目標を達成した、達成しなかったという点のみならず、なぜ達成したか・しなかったか、どのような手段が有効なのか、反省すべき点はどのような点であったのか、実施した施

策・事業の効果を分析することに注力する。

そして、改善策を検討し、平成16年度の目指す 成果の数値目標達成のための取組みに反映する。

b 反省を改善につなげる仕組みの導入

実施した施策・事業の効果を分析して得られた 結果から、反省を改善につなげる仕組みの導入を 図る。渋滞や交通安全などの課題の多い区間に重 点的投資を行うなど、今後の取組みを導き出す。

(3)評価結果を次の施策に反映する(「規格の高 い道路を使う割合」を例に)

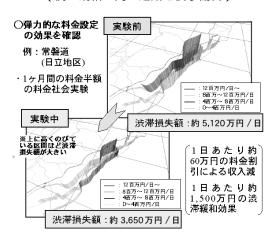
評価にあたっては、指標の値の結果のみで達成度を議論するのではなく、取組みの効果の分析まで行うことが重要であるとのスタンスでとりまとめた。17指標の中の規格の高い道路を使う割合を例に評価結果を次の施策に反映した取組みを紹介する。

b 指標の目的と結果

規格の高い道路を使う割合の指標は、高速道路のように自動車と歩行者が分離され、高速走行が可能な構造となっている規格の高い道路に自動車交通を分担させることで、幹線道路の渋滞、生活道路での事故、沿道の騒音等を改善すること等を目的として設定している。交通事故死者数を例にすると、規格の高い道路における死傷事故率を1としたとき、その他の道路におけるそれは2.3となる。我が国の規格の高い道路を使う割合は、約13%と欧米諸国の20~30%に比較して著しく低い状況であるが、指標値がドイツ並みの約30%になった場合、年間の交通事故死者数が約900人減少すると試算できる。

全道路の走行台キロに対する規格の高い道路の 走行台キロの比である指標値は、平成14年度の 13%から、平成15年度の13%へと、指標値として は横ばいの傾向であった。しかし、規格の高い道 路の走行台キロの増減量について、平成14年度と 平成15年度で比較してみると、無料区間の走行台 キロが約60万台キロ増加したのに対して、有料区 間の走行台キロは約80万台キロ減少(全国計で約 20万台キロ減少)とであることが分かった。また、

図8 評価結果を次の施策に反映させる (例:規格の高い道路を使う割合)



(出所)国土交通省道路局

平成14年度より道路局で行っている道路利用者満足度調査では、11の設問の中で、道路利用者の道路サービスに対する満足度が最低の結果となっている。割高感のある料金設定が有料道路の利用者離れにつながっていることが分かった。

b 施策・事業の効果を分析

国土交通省道路局は、平成15年度の重点施策と して、高速道路などの料金を引き下げ、利用交通 量の変動を確認する社会実験を22地区において実 施した。その中で、日立市内の常磐自動車道(日 立北インターチェンジ~日立南太田インターチェ ンジ間)では、1ヶ月間、料金を約半額に引き下 げたところ、実験区間内の各インターチェンジを 往来する交通量が平均で約70%増加。一方、周辺 道路の交通量は約4%減少した。それに伴う、1日 あたりの料金割引による収入減は約60万円であっ た。一方で、渋滞をしていた周辺道路の交通量が 減少し、渋滞が緩和したことによる時間節約で社 会便益は1日あたり約1,500万円と試算される。こ のことから、高速道路の料金を弾力的に設定する ことが道路網全体の効率化につながるという弾力 的な料金設定の効果を確認できた。(図8)

一方で、規格の高い道路を使う割合の伸びが全国で一番高い伸びの1.3%増を示した香川県では、 平成15年3月に高松中央インターチェンジ~高松 西インターチェンジ間が開通し、四国横断自動車道(高松自動車道)は全線開通した。この開通により、平成15年度は、全線にわたり交通量が大幅に増加した(平均30%増加(区間最大80%増加))。

b 反省を改善につなげる

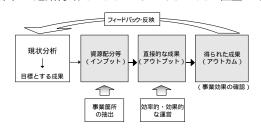
施策、事業の効果を分析した結果、以下の施策 を今後とも実施していくこととしている。

有料道路の弾力的な料金設定に向けた施策展開 高速道路等のネットワーク整備

また、別の課題として、無料で供用している諸 外国と比較して日本の高速道路のインターチェン ジ間隔が長く(高速道路の平均インターチェンジ 間隔:日本約10kmに対して、アメリカ約5km、 ドイツ約7km、イギリス約4km)、利便性を損な っている可能性があるため、

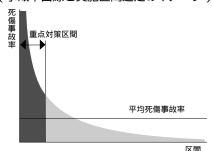
追加インターチェンジの設置に向けた施策展開の一つとして、平成16年度に一般道と容易に接続可能な既存のサービスエリア・パーキングエリアにETC専用の仮出入口を設置するスマートICの社会実験を実施する。(実験箇所として全国で29カ

図9 道路行政マネジメントガイダンスの位置づけ



(出所)国土交通省道路局(2004b)

図10 優先度明示方式 (事故率曲線と実施区間選定のイメージ)



(出所)国土交通省道路局(2004a)

所を採択。)

平成15年度の取組みの評価結果を踏まえたこれ らの施策・事業が平成16年度の業績計画となった。

(4)目標とする成果と予算配分の関連付け

データ整備を行い、ユーザーの視点に立った成果目標を指標の数値目標で設定しているが、それをどのように実現していくか、資源配分等の行政判断をどのようにしたらいいかが現段階では必ずしも明確になっていないのが実態である。そこで、道路局においては、目標とする成果を得るために、どのような施策・事業を実施すればいいかなどの行政判断(資源配分等)を支援する仕組みをまとめた道路行政マネジメントガイダンス(国土交通省(2004))を作成した。(図9)

ガイダンスでは、本省から現場までが問題意識を共有し、課題の高い箇所の抽出や施策の効果を 把握する標準的なプロセスとして使用する仕組み として、客観的なデータに基づく仕組みの導入と 情報共有・蓄積の仕組みづくりの二点について取 り上げている。

交通安全分野や渋滞分野では、目標とする成果をより効果的・効率的に達成するためには、課題の高い箇所を抽出し、かつ地域の実情を把握して対策実施箇所を選定することが必要となる。そこで、対策の必要性を示す客観的データ(死傷事故率、渋滞損失時間など)を、課題の高い順に並べて優先的に対策すべき箇所を明示する優先度明示方式を導入する。

例えば、交通安全分野では、死傷事故率の高い 区間を順に並べた事故率曲線を作成し、対策実施 区間選定の基本データとして活用する(図10)。 具体的な検討の流れは以下のとおりである。

区間の分割

区間の分割については、以下の考え方により分割した区間を基本とする。

・交差点と単路部に分類(事故の発生形態が異なる)

事故率曲線の作成

分割した区間毎の死傷事故率に着目し、区間を 死傷事故率に基づき順位付けを行い、事故率曲線 を作成する。

重点対策区間の抽出

死傷事故率に関する基準値(例えば、平均の2倍)を設定し、死傷事故率が基準値を超える区間をまず重点対策区間として検討対象とする。さらに、現場の状況に応じて対策実施箇所の判断や対策の工夫をすることとしている。

5. おわりに

本稿では、平成15年度当初より現在までの、成果志向の道路行政マネジメントに係る取組みを紹介した。

職員への成果志向の意識の浸透や、納税者、利用者に対するわかりやすさを確保するための工夫など、未だ多くの課題を残している。筆者は、これらの課題の解決に向けては、多くの方々の参加と努力による成功事例を積み上げ、少しずつでも日々改善していくことが重要であると考える。そして、それこそが住民の方々や国民から信頼される道路行政の実現に向けて最も重要であると考える。今後とも、新たな取組みの成果が得られた際は、適宜報告をしてゆく所存である。

これらの取組みについては、国土交通省道路局ホームページ「新たな道路行政マネジメント」(URL:http://www.mlit.go.jp/road/ir/ir-perform/ir-perform.html)において公開しているのでそちらも参照されたい。

参考文献

- 大庭孝之、澤田和宏、森若峰存(2004)「成果重視の 『道路行政マネジメント』の取組みについて」、『日 本評価学会春季第1回全国大会』
- 国土交通省(2004b) 国土交通省政策評価基本計画 平成16年7月改訂版
- 国土交通省道路局(2003)「平成15年度道路行政の業績計画書」
- 国土交通省道路局(2004a)「道路行政の平成15年度達成度報告書/平成16年度業績計画書」
- 国土交通省道路局(2004b)「道路行政マネジメントガイダンス」
- 国土交通省道路局、都市・地域整備局(2004)「平成 16年度道路関係予算概要」
- 道路行政マネジメント研究会(2003)「『成果主義』の 道路行政マネジメントへの転換 理論から実 践へ」
- 古川俊一(2004)「道路行政マネジメント概念登場の 意義と経営課題」、『建設オピニオン』2004年8月号、 建設公論社

(2005,1.26受理)

Outcome Oriented Performance Management of Road Administration From Theory to practice of Management Cycle

Takayuki Oba Kazuhiro Sawada Mineari Moriwaka Yukihiro Tsukada National Institute for Road Bureau Road Bureau Road Bureau Ministry of Land, Ministry of Land, Ministry of Land, Land and Infrastructure Management Infrastructure and Transport Infrastructure and Transport Infrastructure and Transport Ministry of Land, Infrastructure and Transport oba-t82ad@mlit.go.jp sawada-k259@mlit.go.jp moriwaka-m82ac@mlit.go.jp tsukada-y27x@nilim.go.jp

Abstract

National governmental agencies started adopting policy evaluation systems with the enforcement of the "Government Agencies Policy Evaluation Law" passed in 2002. Among them, the Road Bureau of the Ministry of Land, Infrastructure and Transport has shifted toward use of an outcome oriented system. With an emphasis on user-oriented performance as its organizational principal, it has introduced new public management systems for road administration, which use performance measurement based on outcome indicators.

The authors, who are also involved with the introduction of the Road Bureau s outcome oriented performance management system, describe the Bureau s innovative approaches, together with relevant background and emerging problems.

- 1) Start of performance management through setting of outcome indicators and publication of a Performance Plan and Achievement Report.
- Operation of the management cycle and use of evaluation results for subsequent measures and budget requests.
- 3) Development of the Guideline to Management of Road Administration.

Keywords

Performance Measurement, Performance Management of Road Administration, Outcome-oriented, Outcome Indicators, Achievement Report and Performance Plan

研修委員会

活動方針

研修委員会では、「評価の普及および評価に関する人材育成に寄与すること」を目的として、各種セミナーの開催、研修プログラムの開発、教材の作成、講師の派遣、その他人材育成に関わる活動の実施支援を行うことを基本方針としている。その方針のもと、2004年度は以下のような活動を展開した。

活動状況

1. 「評価インターン出前サービス」

「評価インターン出前サービス」は、学会の学生会員に対して「実際に評価を経験する」機会を提供すべく、「評価を外部に委託したい」と考えている地方自治体やNGOで実際に評価を体験させていただく事業である。今年度は、以下の3団体が受け入れ団体となり、計6人の学生会員をインターンとして派遣した。その途中報告は2004年12月の学会全国大会において報告され、2004年3月には評価作業を終え、最終報告書が取りまとめられた。2005年度も2004年度の反省を踏まえ、より効果的なインターン制度を実施していきたい。

1) 東京都練馬区役所

場所:東京都練馬区役所

評価内容:練馬区における行政評価制度の『合目的性』評価

インターン:横山麻季子(筑波大学大学院)

2) (特活) アフリカ地域開発市民の会

場所:ケニア共和国ナイロビ(団体事務所) および東部州ムインギ県ムイ郡 (事業地)

評価内容:地域保健事業に関する評価調査

インターン: 倉田野依 (立命館アジア太平洋大学大学院)、月岡悠 (関西学院大学大学院)

3) (特活) ラオスのこども

場所:東京都大田区(団体事務所)

評価内容: 『ラオスのこども』の組織運営に関する評価調査 インターン: 茂木勇、山田孝子、山村真紀(立教大学大学院)

2. 評価研修プログラム認証制度における審査委員会の開催

国内の評価慣行の推進を目的として、評価研修プログラムを対象とする認証制度を設立したが、学校 評価研修に係る第一回審査委員会を以下のとおり実施した。

1) 審査対象研修プログラム名及び実施機関

「広島県学校評価連携推進講座」/広島県立教育センター

2) 審査委員会の開催

日 時:平成17年2月1日(火) 18:00~21:00

場 所:国際開発センター 4階 会議室

出席者: 牟田博光 (委員長)、橋本昭彦、宮崎修二、竹内正興 (以上、委員)

源由理子 (事務局)

3) 審査結果

あらかじめ設定された審査基準に従い審査を行った結果、同研修を学会として認定した。

研修委員会共同委員長 長尾 真文源 由理子

国際交流委員会

- 1. 日本政府外務省とタイ政府外務省の共催による「ODA評価セミナー・ワークショップ」はバンコック Siam City Hotelで1月17日~21日間開催されました。前半の4日間にわたるプロジェクト・プログラム評価 を中心とした開発評価セミナーには、アジア諸国から50名の評価担当者が参加し、初日の廣野良吉日本評 価学会副会長と経済開発協力機構開発援助委員会評価ネットワークのルンドグレン事務局長による基調講 演に引き続いて、2日半に及び教育(ファシリテーターはタイ王国のチュランロンコン大学教授担当)、イ ンフラ (鴫谷JBICプロジェクト開発部開発事業評価室事後評価班課長)、農村開発と貧困削減 (湊直信 FASID国際開発研究センター所長代行担当)、環境(ジョージ・ピットマン世界銀行評価シニアオフィサ ー担当)、国別評価(ヌルル・アラム評価副部長担当)の5分科会が開催されました。4日間のセミナー最 終日には、分科会報告を中心に活発な討議が参加者の間で行われて、閉会となりました。続いて開催され ました評価ワークショップでは、アジアの17カ国の評価部局長と日本・国際機関の評価専門家が参加し、 開発評価セミナーでの成果とJICA企画・調査部事業評価グループの三輪徳子グループ長によるJICA開発評 価に関する報告をもとに、各国における今後の開発評価の課題と国際協力の方向について議論を重ね、廣 野議長による議長総括を採択して終了しました。アジア諸国では開発評価体制がかなり進んでいるインド、 マレーシアのような国々と未だ遅れている国々があり、ケース・スタディを中心としたセミナーと、評価 体制の構築・強化を主眼としたワークショップでの討論は参加者にとって非常に有益なものであったとの 報告がありました。本ODA評価セミナー・ワークショップの公式報告書は、近いうちに外務省ホームペー ジに掲載される予定ですので、ご関心のある会員の皆様には是非ご参照ください。
- 2. 4月12~14日国際開発評価学会(IDEAS)の3年に一度の世界大会がインドのニューデリーで開催されます。日本評価学会会員の報告・参加を側面から推進するために、旅費等に対する補助制度(10万円/名)を先の理事会で採択しました。なお、世界大会プログラムの最新版は、日本評価学会ホームページに掲載されていますので、ご覧ください。

国際交流委員長 廣野 良吉

学会賞審査委員会

日本評価学会では、学会の学術領域に関する研究および学会の発展に優れた業績があったと認められる 者を顕彰するために、学会賞の制度を設けています。学会賞には原著論文が評価学研究に大きく貢献した と認められる者に授与する論文賞、評価学研究の進歩に寄与する優れた研究発表を行い、なお将来の発展 を期待しうる者に授与する奨励賞、評価学の発展に関し顕著な功績のあった者に授与する功績賞がありま す。

2004年度は学会賞受賞者選考細則第6条により、2004年9月3付けで学会ホームページ上に学会賞候補者の公募を行い、10月31日まで自薦、他薦を受け付けました。さらに、11月12日に学会賞審査委員会を開催しました。選考に先立ち、選考基準の改訂について議論し、新しい申し合わせを確認し、ついで、論文賞、奨励賞、功績賞のそれぞれについて、申し合わせに基づき、功績、論文、発表原稿等公表された資料、アンケート結果、及び、日頃の学会活動などを総合的に検討しました。さらに、受賞候補者の意思確認のためメールにて議論を継続し、11月19日、学会賞受賞候補者選考について全員一致で下記の結論に達しました。

学会賞受賞者選考細則第7条では、論文賞は、選考当該年度の前年度に発行された本学会誌に掲載された原著論文等から推薦されるものとする、と定められています。昨年の総会で出た意見に基づき、今回から編集委員会から候補者の推薦をもらうことにしました。日本評価研究第4巻第1号及び第2号の原著論文の中から、編集委員会の推薦に基づき、論文賞受賞候補者として、上野宏会員を選考しました。第4巻第1号に掲載された論文が、評価学研究に貢献したと認められたものです。

奨励賞は、選考当該年度の前年度に開催された全国大会での発表等から推薦されるものとする、と定められています。第4回全国大会、及び春季第1回全国大会の場で発表された原稿等公表された資料、及び、それぞれの大会でのアンケート結果などを総合的に検討し、受賞候補者として田中啓会員、三輪徳子会員を選考しました。

功績賞は、選考当該年度の前年度までの学会活動への貢献に基づき選考されるものとする、と定められていますが、今回は該当者なしとしました。

これらの選考結果は、企画委員長を通じて理事会に諮られ、原案通り承認されました。また、この3人の受賞者には、12月4日の総会の場で、学会から賞状及び副賞が授与されました。

このような顕彰事業は、学会員の優れた業績を讃えると共に、学会員の主体的な活躍を支援するものです。表彰された3人の会員に心からお祝を申し上げると共に、今後ますますのご活躍を期待しています。また、今回惜しくも受賞を逃された会員の皆さんも、次回以降の受賞を目指し、学会大会での発表、学会誌への投稿等を積極的に行っていただきたいものです。

今後とも、学会賞規程、学会賞受賞者選考細則、選考基準などを適宜見直し、適切で透明性の高い学会賞運営を心がけたいと考えています。

学会賞審査委員会委員長 牟田 博光

社会・ジェンダー評価分科会

活動方針

社会・ジェンダー評価分科会は、評価に関する研究および応用の促進、情報交換、人材の育成等について、「社会・ジェンダーの視点」(階層・階級、民族・人種、年齢、健康等、さまざまな社会的属性や文脈との関連で、ジェンダーの平等を推進していく視点)から、これを推進することを目的とする。

2004年の活動状況

1. 研究会 ______

日 時:2004年7月7日(水)

会 場:東京大学

報告者1:目黒依子氏(上智大学教授)

題 目:「経済危機対応政策のジェンダー・インパクト評価に向けて-韓国事例をもとに-」

報告者2:石井菜穂子氏(財務省国際局担当参事官)

題目:「長期経済発展の制度分析ージェンダー・インプリケーションを探るー」

2. 第5回全国大会共通論題VII「開発協力とジェンダー視点」

日 時:2004年12月5日(日)

会 場:立教大学

報告者1:石井菜穂子氏(財務省国際局担当参事官)

題 目:「経済発展を支える制度とは何かージェンダーの視点とわが国援助への政策含意」

報告者2:古橋源六郎氏(内閣府男女共同参画会議議員)

題 目: 「男女共同参画の視点に立った政府開発援助(ODA)の推進について

今後の活動予定

2005年5月開催予定の春季第2回全国大会に社会・ジェンダー評価分科会として発表を行う予定である。

社会・ジェンダー評価分科会代表 大沢 真理 田中由美子

日本評価学会第5回全国大会報告

2004年12月4日(土)~5日(日)に立教大学において日本評価学会第5回全国大会が開催された。2日間を通して248名(内非会員は133名)の参加を得て実り多い議論がなされた。大会は共通論題セッション、自由課題セッションともに各8つずつで構成され、52の発表が行われた。

ここでは共通論題に焦点を当て紹介する。

日本評価学会では、評価慣行の普及と評価活動の質の改善において様々な取り組みを行っているが、その一つとして人材育成への取り組みが挙げられる。「認証制度のあり方に関する研究」では、人材育成のアプローチの一つとして資格・認証制度を評価分野に取り入れる試みについて、分科会が現在行っている調査の進捗状況を報告した後、評価分野の質の確保、評価分野の認知度向上、評価専門職のあり方等について議論を行った。

「地域に根づいた道路行政マネジメントと評価」では、国土交通省道路局が平成15年度より新たに開始した「道路行政マネジメント」についての実践報告に基づき、その狙いと運用上の課題について議論を行い今後の発展方向を明らかにするとともに、特定行政分野におけるマネジメント・モデルとしての先駆性について考察することを目的とした。

近年、評価の拡充が求められる中、外部評価が特に重要視されている。「外部評価の活用―課題と展望―」では、ODA評価を例にとり、被評価対象機関関係者および外部評価者双方の参加のもと、最近の外部評価に関する様々な取り組みに基づき、外部評価の有効性と制約、そしてそれらを踏まえた今後の外部評価の活用について議論を深めた。

「自治体における行政評価とマネジメント」では、常に先進的な努力を重ねてきた三重県の実践及び近年の自治体における成功事例を考察し、評価システムの進化、さらに経営システムへの転換を議論し、経営システムと評価の一般類型、それから導き出される一般理論を探求することを目的とした。

評価結果の事業改善への活用は、説明責任の遂行とともに評価の中心的な目的としてますます重要視されている。しかし評価結果のそのような活用は実際、必ずしもまだ十分に進んでいるとはいえない状況にある。「評価は役立っているか一評価結果の活用の推進に向けて一」では、ODA評価を例に、評価結果の有効活用に向けた課題と方策について、様々な立場からODA評価に携わる関係者による議論が展開された。

政策評価法が試行され、中央省庁において政策評価が本格的に実施されるようになってから3年が経過した。「国の政策評価の現状と今後の展望」では、政策評価に実際に携わってきた中央省庁の評価担当者や政策評価の研究者をパネリストにむかえ、国の政策評価のこれまでの運用の成果と今後解決すべき課題等について多面的な視点から議論を行うとともに、さらに国の政策評価が今後どのような方向を目指すべきかについて議論を深めた。

「開発協力とジェンダー視点」では、「開発を持続し推進する力は何か」について市場と制度の関連をジェンダー視点の有無を評価ポイントとして「経済発展を支える制度とは何か」、「男女共同参画の視点に立った政府開発援助(ODA)の推進について」のテーマで報告者が話題を提供し、議論を行った。

「なぜ、いまアカウンタビリティか NGOにとっての課題と取り組み」では、ハーバード大学ハウザー非営利センター実施「非営利組織評価理論」事業(笹川平和財団助成)において提示されたアカウンタビリティのフレームワークを実際に活用していくうえでの課題と取り組みについて議論を行った。「NGOがどのような形でアカウンタビリティを考察していくか」をテーマにハーバード大学教授ジェームス・P・ホーナン氏が報告を行い、それを基にパネルディスカッションを行った。

第一日目には、セッション終了後に総会を経て懇親会が行われた。クリスマスツリーやキャンドルに彩られた暖かな雰囲気の中、立教大学の学生による室内楽演奏のひとときを過ごしつつ評価研究者、実務者が一堂に回し親睦を深めるという一年に一度の貴重な時間を持つこととなった。

本大会が、会員各位の日頃の研究や実践活動の報告を基に、評価研究者、実務者の間の経験、情報、知識を共有化する場として、皆様にとって有益な機会となれば大変幸甚である。多くの方のご尽力により本大会を開催できたことに心より感謝申し上げたい。そして、今後もより効率的かつ高質な大会運営を目指す所存であるので、次回以降の大会に向けてもさらに多くの方々の発表と参加を願う。

第5回全国大会実行委員長 入山 映

インドで、特別ないの以の流のにあるがを出る。「必証制度のあり方に関する研究」。例此述及権権協制 リードリーでは、では格と認証制度を結構が野し取り入れる式みについて、分針会が現在行うではる ではなる込みを含まった液、活動分野の質の提供、添加が野の認知度向とよ評価専門職務場の誘導性つ

公場施及園園園園等基準網路線和資養經過學格及工作是表別機器等回到表演 はこれは予測は20mmでは、これを選挙の選問を開発します。

・重視が表現の場合を表現である。 ・重視が表現の場合を表現である。 ・重視が表現である。 ・重視が表現である。 ・重視が表現である。 ・重視が表現である。 ・重視が表現である。 ・重視が表現である。 ・重視が表現である。 ・重視がある。 ・重性がある。 ・重性がある。 ・更性がある。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・更性がな。 ・理性がな。 ・理性がな

12.20年の立文によりの信用は北美野養養の経済が返記をは解除の重心的を目的というませまが重要報告 20年に、20年の主義に結果の主のような活用は実際、の同じ、5まで十分に進入でいるとはいうかの最悪に

事態は没たった。ことも一定価格界の活用の推進に向けて一上では、ODA結婚を開送 階級結構を ほどがはた実際との第1十元でごめるを支援をよりDA条係に無数を確認をおりる路路を開送された。

文学部の記念が成立できた。中央名字において放棄が維持的に実施されるようになってから3年が経過 ここで同の攻壊所備の深次と今後の展察とでは、政策が新し、実際は維わってきた中央客庁の郵価担当を

決定計価の研究者をバラリスをにもかえ、国の政策評価のこれまでの運用の成果と今後解決すべき課題 このいで各面ので担当から資金を行うとともに、そとに国の政策評価が今後とのようたを向を目指すべ

(別名)はカビジェンダー視点」では、「開発を捧練し推進する力は何か」について市場と制度の関連をジ

戦争開発技術。ODA 万種連について1のテーマでは先者が結婚を設性し、議論を行った。

答列と、マー安部(非常報報集件額理論)等集(使用等限數因助成)において提示され<u>後でもウンタビ</u>

等一百畳に近、セッション等子後に転送を経て整理会が行われた。クリスマスフリーやキャンドルに終 年にはかる学用域の中、支援大学の学生による室内を造成をのひとときを過ごしつ心理価格を募。金器者

2011年上級陸を深めるという一年に一度の資源な時期を持つさせとなった。

日本評価学会第5回全国大会 プログラム(実績)

第1日:2004年12月4日(土)

9:00~10:00	受付 8号館2Fフロア			表代表的學品表。 1 E2a EE3	
午前の部 10:00~11:45	共通論題 I 認証制度のあり方に関する研究 (竹内正興)	認証制度のあり方に関する研究業績測定		自由論題 II 評価の新しい視角 (高千穂安長)	
午後の部 I 12:55~14:55	共通論題 II 地域に根付いた道路行政マネジメントと評価 (梅田次郎)		共通論題III 外部評価の活用 ―課題と展望― (杉下恒夫) (三輪徳子)		
午後の部Ⅱ 15:10~17:10	共通論題Ⅳ 自治体における行政評価とマネジメント (古川俊一)		共通論題V 評価は役立っているか 一評価結果の活用の推進に向けて- (牟田博光)		
17:25~18:25	· · · · · · · · · · · · · · · · · · ·		会	100 E	
18:30~20:00	************************************			(1) (経歴の) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	

第2日:2004年12月5日(日)

9:00~9:30	受付 8号館2Fフロア (***)				
午前の部 9:30~11:30	共通論題 VI 国の政策評価の現状と今後の展望 (田中啓)	共通論題VII 開発協力とジェンダー視点 (大沢真理)	共通論題VIII なぜ、いま アカウンタビリティか (田中弥生)		
午後の部 I 12:45~14:30	自由論題III 広島県の学校評価 (長尾眞文)	自由論題 IV 評価手法 (梅内拓生)	自由論題 V 政策評価 (山谷清志)		
午後の部 II 14:45~16:30	自由論題 VI ODA·事例評価 (河野善彦)	自由論題VII 評価インターンによる調査報告 (源由理子)	自由論題VⅢ 教育·人材育成評価 (三好皓一)		

() 内は司会者 自由論題報告の時間は原則として、発表20分、質疑5分、交替1分です。

日本評価学会第5回全国大会プログラム詳細 (実績)

12月4日 (土) 午前の部 10:00~11:45

共通論題Ⅰ	認証制度のあり方に関する研究
	司会者:竹内 正興 (国際開発センター)
	1. 評価分野における認証制度のあり方 (1) 各種認証制度の実態報告 (2) 評価分野への認証制度の適用可能性について に立った。 に立った。 (2) 評価分野への認証制度の適用可能性について に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 に立った。 について。 に立った。 にった
	2. (財)日本医療機能評価機構「第三者による評価の 滝沢 良明(日本医療機能評価機構) 必要性」
自由論題I	業績測定型評価
	司会者:渋谷 和久(国土交通省)
10:00~10:25	1. 地方自治体における評価の波及と生成過程の分析 〇古川 俊一(筑波大学) 森川はるみ(茨城県庁)
10:26~10:51	2. 業績測定型の行政評価・政策評価における計量に 小野 達也(鳥取大学) ついて 一目標達成度の把握を巡って一
10:53~11:18	3. 練馬区行政評価制度 一内部評価の限界点 調査から評価へ 現状と課題— 廣野 良吉 (成蹊大学)
自由論題II	評価の新しい視角
	司会者:高千穂 安長(玉川大学)
10:00~10:25	1. ローカルマニフェスト評価の設計と課題 田中 弥生(東京大学)
10:26~10:51	2. ホームページに見る地方行財政改革 道家麻衣子(立命館アジア太平洋大学 ~6県の比較研究~
10:53~11:18	3. ODAのマクロ評価:途上国のバランスの取れた発 渡辺 道雄(国際開発センター 展とODA
11:20~11:45	4. 地方自治体の行政サービスの市民満足度はこの程 大島 章嘉(市民満足学会) 度で良いのか?

12月4日 (土) 午後の部 | 12:55~14:55

共通論題Ⅱ	地域に根付いた道路行政マネジメントと評価
	司会者:梅田 次郎 (日本能率協会コンサルティング)
	1. 地域に根付いた道路行政マネジメントの実践報告 川崎 茂信(国土交通省)
	2. 「成果志向」の道路行政マネジメントの実践 大庭 孝之 (国土交通省)
	3. 道路行政マネジメントの概念登場の意義と経営課題 古川 俊一(筑波大学)
	討論者:家田 仁 (東京大学)

源 由理子 (国際開発コンサルタント)

共通論題Ⅲ 外部評価の活用一課題と展望一 司会者:杉下 恒夫(茨城大学) 責任者:三輪 徳子(国際協力機構) 1. JICAのプロジェクト終了時評価の2次評価について 杉下 恒夫 (茨城大学) 2. 外部評価の経験から 南条 俊二 (読売新聞) 3. 外部評価の活用 (国際協力銀行の事例) 鴫谷 哲(国際協力銀行) -JBICの事後評価における取組-4. 外部評価における学会と援助実施機関 松岡 俊二 (広島大学) 高橋 与志 (広島大学) 岡田 紗更(広島大学) 5. 外部評価と内部評価-評価の質と結果の活用の観 三輪 徳子 (国際協力機構)

12月4日 (土) 午後の部 II 15:10~17:10

点からの考察

6. 外部の評価者のスタンスに関する考察

~評価結果の想定利用者との関係性に焦点を当てて~

共通論題IV	自治体における行政評価とマネジメント			
	司会者:古川 俊一	(筑波大	学)	
1	1. 三重県の政策評価システムの見直しと現状	西城	昭二	(三重県)
	2. 評価と計画をリンクさせる施策マネジメント	佐伯	学	(日本能率協会コンサルティング)
	3. 日本的組織文脈におけるNPM:三重県改革を事例 にして 討論者		達郎次郎	(東京工業大学) (日本能率協会コンサルティング
共通論題V	評価は役立っているか一評価結果の活用の推進に向け	けて	中語	观察平(日)日alls
	司会者:牟田 博光 (東京工業大学)			NAT WEATE
	1. 評価結果活用の現状と課題	1000	k弘世 博光	
	2. ODA評価結果のフィードバックの現状と課題	北沢	寛治	(外務省)
対 中間 力 () カギ 中道 () () () () () () () () () (3. 評価結果の活用推進(国際協力銀行の事例) —JBICの事後評価における取組—	辻	一人	(国際協力銀行)
	4. 評価は役立っているか―評価結果の活用と推進に向けて	西野	桂子	(グローバル・リンク・マネジメント

12月5日 (日) 午前の部 9:30~11:30

共通論題VI	国の政策評価の現状と今後の展望
21346 -11	司会者:田中 啓(静岡文化芸術大学)
HE MARKET	政策評価制度の導入目的における「捩れ構造」 田中 啓 (静岡文化芸術大学)
大学 新聞 (新力銀行)	パネリスト:猪鼻 聡 (中央青山監査法人) 小野 達也 (鳥取大学) 渋谷 和久 (国土交通省) 玉木 伸介 (預金保険機構)
共通論題VII	開発協力とジェンダー視点
(学) (学)	司会者:大沢 真理(東京大学) 田中 由美子(国際協力機構)
	1. 経済発展を支える制度とは何か 石井菜穂子 (財務省)
	2. 男女共同参画の視点に立った政府開発援助 (ODA) 古橋源六郎 (内閣府) の推進について
共通論題VII	なぜ、いまアカウンタビリティか 日-英通訳有り NGOにとっての課題と取り組み
	司会者:田中 弥生(東京大学)
9:30~10:20	アカウンタビリティ・フレームワークを適用する際 James P.Honan(Harvard Graduate ONGOにとっての課題とその取り組み School of Education)
10:20~10:30	コメンテーター: 田中 弥生 (東京大学) パネルディスカッション
10:30~11:30	

12月5日(日)午後の部 1 12:45~14:30

自由論題VIII	広島県の学校評価
	司会者:長尾 眞文(広島大学) コメンテーター:橋本 昭彦(国立教育政策研究所) 東内 清孝(広島県立教育センター)
12:45~13:10	1. 中学校学年経営におけるロジックモデルの活用の 森田 哲(広島県立広島中学校) 試み
13:11~13:36	2. 中学校における成果評価の実際
13:38~14:03	3. 小学校における学校評価の実際:プロセス評価を 〇寺地 靖仁 (東広島市立八本松中学校) 中心として 高木美佐恵 (広島県立教育センター)
14:05~14:30	4. 学校評価者の人材育成に関する研究 寺地 靖仁 (広島県立教育センター)

12月5日(日)午後の部 1 12:45~14:30

自由論題IV	評価手法			
		(国際医	療福祉	:大学)
12:45~13:10	1. 評価の活用及び手法に関する考察	〇三輪 牟田		(国際協力機構) (東京工業大学)
13:11~13:36	2. 計量モデルを使った評価手法	〇末武		(あずさ監査法人)
13:38~14:03	3. ケイパビリティ・アプローチの適用による ロジック・モデルの再構成に関する一考察	池田 横井		(東洋大学) (立教大学)
14:05~14:30	4. ベキ法則に基づく健康と経済のマクロ指標を用いた 政府開発援助政策評価法の基礎研究(Ⅱ)	○正木 梅内	朋也 拓生	(統合社会学研究会) (国際医療福祉大学)
自由論題V	政策評価	かるとも	スキャ	15-55 - 16-50 3, 15
	司会者:山谷 清志 コメンテーター:窪			学院大学)
12:45~13:10	1. 公的機関のガバナンスとアカウンタビリティ評価	原田	隆	(産業技術総合研究所)
13:11~13:36	2. 規制に関する政策評価の手法に関して ー国内における制度化を踏まえてー	荒川	潤	(UFJ総合研究所)
13:38~14:03	3. 評価における倫理的な問題とスタンダード	○藤田 野本	伸子 綾子	(国際開発高等教育機構)
14:05~14:30	4. 参加型を採用した戦略的計画立案時の課題に関する 予備的考察	石田	健一	(東京大学)

12月5日(日)午後の部川 14:45~16:30

自由論題VI	ODA・事例評価
	司会者:河野 善彦(笹川平和財団)
14:45~15:10	1. 水道分野におけるODAプロジェクトの受益者に対 〇佐伯 健 (立命館大学) するインパクト評価手法の開発 山田 淳 (立命館大学)
15:11~15:36	2. The Impact of District Autonomy on Conservation and Environmental Management (Case Study on Bukit Soeharto Protection Forest (Case Study on Bukit Soeharto Protection Forest)
	East Kalimanan Indonesia) (欠席)
15:38~16:03	3. 政策評価マトリックス (PEM) を使ったセオリー評価: 佐藤由利子 (東京工業大学) インドネシアとタイに対する日本の留学生評価に 授業料収入の観点を加味して
16:05~16:30	4. 円借款事業の労働者に対するHIV/エイズ対策

12月5日(日)午後の部川 14:45~16:30

評価インターンによる調査報告					
司会者:源 由理子(国際開発コンサルタント) コメンテーター:長尾 真文(広島大学)					
1. 練馬区における行政評価制度の「合目的性」評価 ーインターンからみた第三者評価委員会の活動を	横山麻季子		(筑波大学)		
	:島添	悟亨	(練馬区)		
地域保健協力事業に関する評価調査	月岡	悠	(立命館アジア太平洋大学) (関西学院大学) (アースアントヒューマンコーポレーション		
3. 『ラオスのこども』の組織運営に関する評価調査	山村	孝子真紀	(立教大学) (立教大学) (立教大学) (東京大学)		
教育・人材育成評価	ton to the	開始古	93-1 0171-34-5		
司会者:三好善皓一	(立命館	アジア	太平洋大学)		
1. 海外技術研修の費用便益分析	〇牟田 菊田		(東京工業大学) (東京工業大学)		
2. 教育プロジェクトにおける信頼性評価	〇鈴木 佐治 本間	量哉	(豊橋技術科学大学) (東京大学) (豊橋技術科学大学)		
3. 評価者育成のための課題と対策 高千穂安長 - ODAプロジェクトの学生によるメタ評価の傾向 分析から-		恵安長	(玉川大学)		
4. 教員研修プロジェクトの効果に関する研究	○渡邊	美紀	(東京工業大学) (東京工業大学) (国際協力機構)		
	司会者:源 由理子コメンテーター:長月 1. 練馬区における行政評価制度の『合目的性』評価ーインターンからみた第三者評価委員会の活動を中心に一コメンテーター 2. ケニア共和国東部州ムインギ県ムイ郡における地域保健協力事業に関する評価調査コメンテーター 3. 『ラオスのこども』の組織運営に関する評価調査コメンテーター 教育・人材育成評価 司会者:三好 皓一 1. 海外技術研修の費用便益分析 2. 教育プロジェクトにおける信頼性評価 3. 評価者育成のための課題と対策一ODAプロジェクトの学生によるメタ評価の傾向分析から一	司会者:源 由理子(国際開:コメンテーター:長尾 眞文 1. 練馬区における行政評価制度の『合目的性』評価 横山麻ーインターンからみた第三者評価委員会の活動を中心にー コメンテーター: 島添 2. ケニア共和国東部州ムインギ県ムイ郡における 倉田 月岡 コメンテーター: 駒澤 コメンテーター: 駒澤 コメンテーター: 駒澤 コメンテーター: 駒澤 3. 『ラオスのこども』の組織運営に関する評価調査 茂木 山田 山村 コメンテーター: 田中 教育・人材育成評価 司会者:三好 皓一(立命館: 1. 海外技術研修の費用便益分析	司会者:源 由理子 (国際開発コンコメンテーター:長尾 真文 (広島 1. 練馬区における行政評価制度の『合目的性』評価 横山麻季子中心に一 コメンテーター:島添 悟亨 コメンテーター:島添 悟亨 2. ケニア共和国東部州ムインギ県ムイ郡における 倉田 野依 地域保健協力事業に関する評価調査		

日本評価学会春季第2回全国大会のご案内

春季第2回全国大会 実行委員会委員長 宮内 義彦 プログラム委員長 古川 俊一

来たる2005年5月21日(土)、日本評価学会春季第2回全国大会がJICA国際協力総合研修所にて開催されます。自由論題、共通論題の各セッションを予定しております。プログラムにつきましては、更新の都度、メーリングリストおよびホームページでご案内いたします。

参加申込の募集をホームページ上にて開始いたしましたので、会員各位におかれましては奮ってご参加 くださいますようよろしくお願い申し上げます。

記

1. 日時: 2005年5月21日(土) 9:45~20:00(予定)

2. 場所: 国際協力機構(JICA) 国際協力総合研修所

住所:東京都新宿区市谷本村町10-5

3. 主催: 日本評価学会·国際協力機構(共催)

4. 大会参加費: 会員 1,000円

非会員一般 3,000円、非会員学生 2,000円

5. プログラム(2005.3.11暫定版):

2005年5月21日(土) JICA国際協力総合研修所

時間(予定)	テーマ (予定)				
午前の部 9:45-11:45	自由論題 評価手法1	自由論題 II 保健衛生	自由論題 Ⅲ 教育	自由論題IV 政策・組織経営の評価	
11:45-13:00	お昼休み				
午後の部 I 13:00-14:00	公開シンポジウム (宮内義彦会長による講演)				
午後の部 II 14:15-16:15	共通論題 国の政策評価の現状と課題解決の視座 (古川俊一・山谷清志)		共通論題II MDGs達成に向けて 一開発成果マネジメントの課題― (牟田博光・三輪徳子)		
午後の部皿 16:30-18:00	自由論題 V 評価手法2	自由論題 VI 自治体評価	自由論題VII ODA評価	自由論題VIII ジェンダー	
18:15-20:00	懇親会				

()内は座長

自由論題報告の時間は原則として、発表20分、質疑8分、交替2分(計30分)です。 部屋割りについては確定次第発表致します。

The Japanese Journal of Evaluation Studies

Vol. 5, No. 1, March 2005

CONTENTS

Article

Statistic Analysis for Rating Data

- Suggesting Statistical Models to Eliminate the Personal Differences in Ratings -

Tomoya Okubo, Masanori Nakagawa Hiromitsu Muta, Shin-ichi Mayekawa

Research Note

A New Current of Evaluation for Management in Local Government Reform
- A Shift from Performance-focused to Theory-focused Evaluation Perspective -

Junro Nishide

Developments

Beyond Difficulties of Country Programme Evaluations

- A Proposal of Practical Methodology -

Satoko Miwa

Outcome Oriented Performance Management of Road Administration

- From Theory to Practice of Management Cycle -

Takayuki Oba, Kazuhiro Sawada Mineari Moriwaka, Yukihiro Tsukada

The Japan Evaluation Society