

“社会のための科学”と研究開発評価

— プログラム評価の構造とSynthesiologyへの示唆 —

大谷 竜

“社会のための科学”が叫ばれて久しいが、そのような研究開発をどのように評価すればよいのであろうか。本稿では、研究開発評価のそもそもの考え方に立ち戻って概念整理することで、“社会のための科学”研究に有効な評価とは何かについて分かりやすく解説することを試みた。そのポイントは、評価はそれ単独では意味をなさず、研究開発を通じて実現させたいことへの道筋（戦略）と一体となって初めて機能すること、そして評価の役割は、戦略をより良く実行していくために実態をつまびらかにすること、などである。

キーワード：社会のための科学、研究開発評価、プロジェクト、プログラム、ROAMEF、ターゲットセッティング、戦略策定

“Science for society” and evaluation of research and technology development

– The framework of program evaluation and implication for *Synthesiology* –

Ryu Ohtani

This article intends to give a basic logical framework of “research & technology development (RTD) evaluation” for the realization of “science for society”. Although the RTD evaluation is to be designed for promoting the evolution of strategies that utilize the knowledge derived from science and technology for the well-being of society and for solving public issues, some confusion emerges due to misinterpretation of concepts and/or verbal abuse when putting evaluation into practice. Through conceptual mapping of previous studies with original daily-life examples, this article shows that the role of evaluation is to reflect the current state of affairs in order for strategies to be better reformulated cyclically and circularly.

Keywords: Science for society, evaluation of research and technology development, *project*, *program*, ROAMEF, target setting, strategy-making

1 はじめに

近年、ブダペスト宣言などに見られるように、科学研究において“社会のための科学 (Science for society)”の重要性が指摘されている^[1]。これは、人間活動に伴う地球環境の悪化への対処や、複雑化・脆弱化する社会の持続的成長のために、科学的知見を使って社会問題を解決していくことが社会から要望されていることを背景としている。我が国でも、独立行政法人（以下、独法）産業技術総合研究所（以下、産総研）ではその憲章の中心において「社会の中で、社会のために」を据え、本格研究という考え方を軸に、科学技術を用いて社会の持続的発展に貢献することを謳っている^[2]。また、2005年から開始された第3期科学技術基本計画において、科学技術を用いたイノベーション創出が大きな柱になっているが、これはまさに科学によってもたらされた知見を社会のために活用しようという試みに他ならない。

こうした“社会のための科学”を意識した研究開発が活発になる一方、研究開発から得られた成果を確認し、目的に照らして効果的な科学研究が遂行されているか、不断に見直しを行う必要性も言われている^{[3][4]}。そのため、我が国においては近年、公的研究機関や大学などにおける研究開発評価が盛んに行われるようになってきた。

しかし、こうした研究開発評価に関する各種制度や仕組みが確立される一方、それが“社会のための科学”の実現のために、有効に機能しているかどうかは必ずしも明確ではない。特に、評価を受ける現場の研究者側からさまざまな疑問が呈されていることが指摘されている^[5]。例えば、「自分の専門分野を評価できる程深い知識や経験をもった研究者が他にいるように思えない。そんな中、妥当な評価ができるのか？ いわんや、研究領域のことをよく分かっていない評価機関が何で口出しできるのか。」といった反論はよく聞かれる。また、“国の研究開発全般に共通する評価

産業技術総合研究所 活断層・地震研究センター 〒305-8567 つくば市東 1-1-1 中央第7
Active Fault and Earthquake Research Center, AIST Tsukuba Central 7, 1-1-1 Higashi, Tsukuba 305-8567, Japan E-mail: ohtani-ryu@aist.go.jp

Original manuscript received November 2, 2009, Revisions received January 4, 2010, Accepted January 15, 2010

の実施方法の在り方についての大綱的指針（以下、大綱的指針）”では事前評価の重要性についても触れられているが、多くの研究者は次のような疑問を口にするのではなからうか。「そもそも事前評価なんかできるのか？それは意味があるのか？研究も開始せず、何も出てきていない段階で、一体全体評価なんかできるはずがないではないか？」と。

これらの疑問に対して評価者と被評価者とが評価の意義と目的を共有し、共通の理解の地平に立って評価を実施していくことは、“社会のための科学”を実現させる上でもきわめて重要である。そのためには、評価そのものの考え方や思考の枠組みに立ち返って考える必要がある。そこで本稿は、こうした評価の論理構造について、親しみやすい例を使って分かりやすくまとめ、提示することを目的とする。以下、これまで研究開発評価論などで言われている基本となる概念を参考文献（入手しやすい日本語のもの）とともに示し、そうした概念や親しみやすい例を用いて筆者自らの論理展開を加えて説明していくことにする。

なお、本稿では研究開発の中でも、大学の理学系の分野で行われるような純粋基礎研究ではなく、科学技術の知見を使ってどのように社会問題の解決や政策課題の実現に結びつけていくかを目的とするような研究の評価を対象を限定する。こういった分野の評価の上でポイントとなるのが次の三つである。第一に評価には大きく分けて4種類の視点があること、二番目に、評価対象の階層性に関わる問題、三番目に、評価の局面（フェーズ）の位置づけ方に関する点である。

これらを考える際、研究開発評価においては注意深い用語の使い方が大変重要となるため、本稿で使用する用語の提示の仕方を以下のように定めることとする。まず、日常使われる用語であるが本稿においては独自の意味で使用され

ているものについては、斜体で示す（例えば、“プログラム”）。それに対して、特に独自の意味を有するわけではないが、それを強調したい場合や、ひとまとまりの用語や長い法令名等であって前後の区切りを明確にしたい場合には、クォーテーション(“”)で提示するものとする。一方、カギ括弧(「」)で囲われるものについては、文献からの直接引用、もしくは問いかけなどの“発言”に関するものに限るものとする。

なお、本稿で述べられている見解は筆者の個人的なものであり、筆者の所属する活断層・地震研究センターとは一切関係のないことをお断りしておく。

2 研究開発評価に必要な視点

2.1 四つの評価の視点

評価の視点について平澤氏^[6]は、達成度評価、価値評価、見直しと修正のための評価の三つを挙げているが、本稿ではこれに説明責任のための評価も加えた四つに整理した^註。それぞれは必ずしも独立ではないが、本稿では第一近似的に以下のようにその内容と特徴を整理する（表1）。

一般に評価と言われるとすぐに想起されるのが、達成度評価である。これは文字どおり、ある目標に対してどれだけ達成されたか、ということを見るもので、達成の可否について査定的になるのが特徴である。要するに達成具合を（多くの場合数値的に）管理し、査定することを、大きな目的とするものである。当然、達成していなければベケをつけられるので、これは“切る”ことと密接に関連する。

達成度評価は、達成すべき内容が分かっている、しかもその達成方法が既知であるような事象に対して、大変有効な方法である。典型的な例は、高校の定期試験である。高校の試験問題は、出題範囲も限定されていて、かつ出される問題の素材も教科書に書かれているなど、

表1 研究開発評価の視点とその特徴。評価の視点は主に^[6]による。

評価の視点	有効性: 方法 が既知	有効性: 方法 が未知	特徴
達成度	◎	×	・管理、査定(○×△付け)が目的 ・「切る」ことがキー ・「ルーチン」な対象・作業に有効
価値	◎	△	・価値を定めるのが目的 ・「比較」、「目的」がキー ・「目的や基準が共有」されれば有効
説明責任	◎	×	・出資者への説明責任が目的 ・納得されるかがキー ・費用対便益分析が可能ならば有効
見直しと 修正	必要なし	◎	・達成方法が未知の時のその方法の解明や、変化する環境への適応が目的 ・自分の行動から「学習・改善」がキー ・「分からない」対象に有効

既に答えもしくは答え方の分かっているものが出題される。試験においては、そうした内容を受験者がどれだけきちんと勉強して身に付けたかが問われるわけで、これはまさに達成度評価に他ならない。

ところが研究開発のように、未だ達成方法が検証されていない対象に関しては、達成度評価の視点から評価するのは適切ではない。なんとすれば、分からないから研究開発を実施するのであって、既に分かっている対象を取り上げて研究開発を行っても、それはもはや研究開発に値する行為とはいえないからである。それにもかかわらず、強引に達成度評価の視点を導入してしまうと、わざと目標を下げて、達成しやすいような安易な目標を設定して、想定目標に対して十分達成できた、と主張するような風潮が跋扈しかねない状況が生じる恐れがある。これは研究開発の促進のためには、はなはだ具合の悪いことである。

更に敷衍すれば、この達成度評価はともすると「どれだけ頑張ったか」という尺度が潜在的に入ってきて、たくさんの量を達成すればするほどいいという傾向を助長しかねない特徴を有する。即ち、達成方法は既知であるのだから、成果の量が多ければ、その分だけよく頑張った（だから素晴らしい）、という論理が潜在的に助長され得る。例えば、インパクトのある論文1本よりも、重要度は低くても100本も論文があれば凄い、という風潮の跋扈する土壌になりかねない。当該分野の専門外の人からみれば、1本しか論文がないよりも、100本も量があったらそれは凄そうに見えるからである。

二番目の評価の視点が、価値評価と言われるものである。これは文字どおり、対象や物事の価値を定めるのが目的であり、一般に、「～が非常に高く評価される」というような表現でよく使われる意味での評価である。価値評価は、評価者が評価対象を自らの価値観（価値基準）で判断する評価なので、極端な話、ノーベル賞を受賞した研究であっても、評価者にとって興味がなかったり価値がないと思われれば、“その人にとっては”価値評価は低いものとなる。よって、目的や判断の基準を共有しないと、評価者によってばらばらな評価になってしまうという特徴を持つ。「基礎研究と応用研究、どちらが評価できるか？」という議論が一つの例であり、これに対する答えは、全体の目的に即した議論が行われな限り、評価者同士の単なる価値観の押し付け合いに終始してしまう。また価値評価は、同じ対象でも、置かれている“状況”や“目的”によって、変化するというのも特徴である。例えば、かつて日本企業で基礎研究よりも応用研究の方が盛んに行われていたのは、後者の方が企業利益をより上げられるという価値評価のために他ならない。しかし、時代“状況”が変わってしまった現在、

同じ価値評価では立ちゆかなくなっていることは周知のとおりである。

三番目の評価の視点は、説明責任のための評価である。これは主に、研究開発費用の出資者（財政当局や納税者）に対しての説明責任を果たすことを目的とするものであり、出資してもらえるかを納得させられるかが重要となる。説明責任のための評価の一番わかりやすい形は、貨幣換算し費用対便益分析を行って、費用C（インプット）と便益B（アウトプット）の比(B/C)が1以上であることを示すことで、さまざまな経済学的算出手法が提案されている^{[7][8]}。しかし金銭に換算されにくい研究課題や対象については計算が困難であり、更に、この説明責任のための評価は「費用対便益比B/Cが1以上であるかどうか」という“指標”が達成されているかという意味で達成度評価の一つの形態に過ぎない。そして、B/C>1でない場合にこれを改善するためには次に何をすればよいかという必然的に出てくる疑問について、一切答えが得られないという根本的な問題を有している。

2.2 内省評価について

そこで重要になってくるのが、見直しと修正のための評価である。これは研究開発のように、答えの分かっている対象を取り扱う場合や、動的に変化する周りの環境に適応することを目的とするための評価であり、同じ“評価”という名前がついても他の三つの評価とは、全く様相の異なるものである。我々が今取り扱っているように、解決方法がよく分かっている場合、あれこれ考えては試し、考えては試し、というtrial and errorをやっていかねばいけない。といって無闇に行っても効率が悪いので、ある試みを行った結果、もしそれがうまくいかなかった場合、何がいけなかったのか自分の行為から学習して、次の行動への見直しのために活かしていくことが必要になってくる。しかし、うまく見直すためには、当然自分がどんな行為を行って、その結果発生した事象との間がどんな因果関係で結ばれているのか、ということをつまびらかにしなければならない。この“実態を「つまびらかにする^[3]」行為”こそが、ここでいう見直しと修正のための評価に相当する。

ここでいう見直しと修正のための評価に係わる内容とは、冒頭で述べた本稿のとる立場から、個別具体的な研究開発の内容の詳細に係わるものではない。こうした点を強調するのは、見直しと修正のための評価と言った場合、多くの研究者は次のような状況を想像しがちだからである。例えば、「ある夢の材料を作ろうと試験管を振ったができない。何故だろう？ そうだ、クエン酸を加えてみよう」、といったものである。研究者はこの状況において、達成方法が分からない事項（この場合は、ある夢の材料を作るこ

と) に対して、見直しと修正を行っている、と主張するのである。しかし、我々が取り扱っている研究開発の目的は何であったのか、今一度振り返ってみると、それは研究の“結果”、それを“使って”どのように社会問題の解決につなげるか、ということであった。つまり、見直したい内容とは、個別具体的なプロジェクトではなく(またそこに留まるのではなく)、それらの研究なりプロジェクトを実施した“結果”、その成果を“使って”、どのように政策課題を実現せしめるかを目的とした“システム”の設計が、我々が真に目指すべきことであった。そして、システムの改善へとつなげるため、実態をつまびらかにするものとしての評価があったはずである。科学技術研究の実施はそのための一手段であって、それ自体が目的ではない。もちろん研究開発はこのシステムのコアであり、よい研究結果は必須のものであるが、一方で、社会問題の解決のための時間も資源も有限であるのだから、「研究はやってみなければ分からない」ということでエンドレスに研究が続けられ、いつまでたっても必要とされる社会問題が解決されないのであれば、それは社会にとっては困った事態となってしまう。

このように、我々がここで対象とするところの評価は、社会問題の解決といった“意図した結果”⁹⁾を実現せしめる、“システム全体の見直しと修正”であることを強く認識する必要がある。もしシステムではなく、個別具体的な研究の専門分野(本例で言えば試験管の実験)を見直しと修正の対象とするならば、それは冒頭で述べた研究者の反論、「専門外なのにどういう資格で、当該専門分野を評価できるのか?」という疑問に突き当たることになる。それは当該分野の専門領域内のピアレビューに付すべきものであって、ここでの評価が口出しできるものでもないし、すべきことでもない。

ところで、一般に日本語の“評価”という言葉は、必ずしもこの意味を明示的に示すものではないという問題がある。即ち、評価というどうしても、どこかの偉い学者とか行政官などといった人達が高いところから見下ろして、

何かお墨付きを与えるような印象を拭いきれないのが現実ではなかろうか(例えば¹⁰⁾)。更に、それが査定という意思決定と潜在的に結びついて、評価される側からすると、通常とは異なることをすれば何か断罪されてしまうというような恐怖を無意識的に含んでしまう可能性がある。そうではなく、見直しと修正のための評価はより良くするために行うものであり、しかもそれは本来的に主体的な行為のほずである。そこで今後、見直しと修正のための評価を指すのに本稿では“内省評価”と呼ぶことにする。即ち、「自分の考えや行動などを深くかえりみる」(大辞泉より)という“内省”という言葉の有する意味を借用して、この言葉を使用する。

こうした内省評価という視点を導入すると、評価における“実績の把握”において、必ずから注目すべきものが達成度評価とは大きく違って来る。このことを文献⁹⁾の概念整理を援用して、以下説明する。一般に評価といった場合、我々はずいつい、成果(Product)に気を取られてしまい勝ちである。しかし実は成果(Product)というのは実績(Performance)の一部を構成しているに過ぎず、実績には他にも、その成果に至るまでの過程(Process)も含むことに注意しなければならない(図1を参照;⁹⁾の図を一部改変)。即ち、達成度という観点のみに限定した評価だと、目標値に対してきちんと達成したかをみればいいわけなので、どうしても成果、その中でも更に狭い、主題に直接かわる部分に限定した“主題的な成果”にしか着目しないことになる。先の試験勉強の例でいえば、要は合格するのに必要な点数をとれたかどうかしかみないことであって、そのためにどのように勉強したか、どれだけ勉強したかというのは達成度評価にとっては預かり知らぬもので、注目の対象とはならないのである。

しかし内省評価(見直しと修正のための評価)の視点に立つ場合、もし点数が目標よりとれなかった場合、どのように見直していけばいいのを知るためには、何が悪くてその点数を得るに至ったのか、そのプロセスをつまびらか

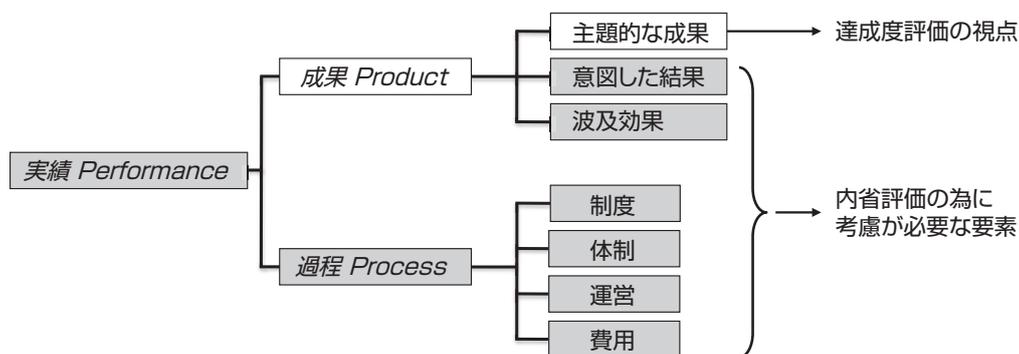


図1 研究開発評価の実績の分類。⁹⁾からの図を引用・一部改変。

にする必要が必然的に出てこよう。また成果の中でも、“主
題的な成果”だけでなく、それが“意図した結果”^[9]に最
最終的にどのようにつながっていくのか、そして得られた“主
題的な成果”はその中でどういった位置づけになるのかと
いう視点で捉えないと、試験の点数を得た“結果”、何を
実現したいのかという、より上位の目的の実現にまで結び
つかないことが往々にしてあり、注意が必要である。この
点については後述する。

3 評価対象の階層性—プログラム化の重要性—

さて、二番目のポイントは、評価対象を階層的な体系で
考えることである。これは一言で言うと、評価対象を一緒
くたにして扱うのではなく、階層性をもったシステムとして
捉え、それぞれの階層に応じてそれらの階層に固有の属性
を把握することが、効果的で有用な内省評価につながる、
ということである。

ある研究課題の評価を展開しようとした時に考えられる
階層構造とは、例えば図2に示すようなものである。まず
最上位には総合戦略があり、これは科学技術基本計画や
全体戦略に相当するものである。次にその元に“政策”が
展開され、更にそうした政策の元に、“プログラム”（後述）
があり、その元でようやく研究者がよく馴染んでいる個別の
“プロジェクト”が位置づけられる、という階層構造であ
る。ここでプロジェクトとは、“研究課題の集合体としての
研究事業”を指すことにする。それに対して、プログラムと
は、より上位の階層に属する政策と、個別事業としてのプ
ロジェクトを結びつける“仕組み”のことで、これが間に入
らないと、政策的な課題解決に資するような研究開発が効
果的・効率的に展開しにくい、というのがこの図の意味す
るところである。

このプログラムの重要性は、平澤氏^[11]の説明を借りて、
次のような思考実験を行うことで理解される。即ち、科学
技術に関する個別のプロジェクトの成果をかき集めただけ
で、果たして上位の階層にある、社会問題の解決や政策課

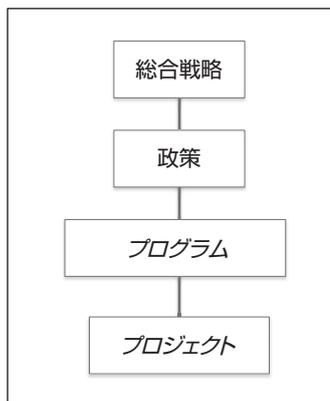


図2 研究開発評価対象の階層性。

題を実現できるか、と考える。身近な例を思い浮かべれば
分かるように、科学技術を使ってモノを開発したところで、
それだけでは単にモノがあるだけである。政策課題を実現
させるためには、政策と研究開発プロジェクトとの間をつな
ぐ仕掛けを両者の間に作り、「政策意図を具体化する仕組
み」^[11]を導入することが不可欠である。それは、個別プロ
ジェクトを構造化し、場合によっては技術以外の補助装置
を導入するなどして、研究成果がシームレスに政策課題へ
と受け渡しができるようにするといったことである。こうし
た仕掛けこそがプログラムと呼ばれるものの基本的な考え
方である。そうした仕掛けを作っておかないと、研究プロ
ジェクトを実施していわゆる“成果”はいっぱい出てきたけ
れど、それが社会問題の解決に全然貢献しない、というよ
うな事態が発生してしまう可能性がある。

さて、今与えられた研究開発課題に対して、プログラムと
いう仕組みが埋め込まれているかどうか、それを判定する
効果的なリトマス試験紙が存在する。それは次のような問
いを発することである。即ち、「研究開発プロジェクトを実
施したその結果、何？」という問いである（この問いかけは、
文献^[12]を参考にして定式化したものである）。研究開発が
下位のプロジェクトレベルにとどまっている限り、そこから
出てくるものは科学技術としてのモノや知見でしかない。
それが、より上位に属する政策課題へと、どのような道筋
で明確に結びつけられているかを明らかにしようというの
が、この質問の意図である。この問いに明瞭に答えられず、
あるいは抽象的で具体性に欠ける回答しか得られないとい
うことは、最終的に意図する結果（本稿では“社会問題の
解決”）への具体的な道筋が見えてこないということであり、
プログラム化が十分に行われていないことを示唆する。詳
しいプログラムの内部構造については後述する。

4 評価の局面の位置づけ

4.1 事前・中間・追跡評価の一体化

さて、評価の三つめのポイントは、評価の局面（フェーズ）
の位置づけ方である。今、図3にあるように、ある研究開
発のライフサイクルにあたって、戦略策定、実施、検証、
見直しというステージを設定すると、内省評価のためには、
事後評価や中間評価などをどのように位置づけたら良いか
を以下、考える。

まず、内省評価のためには、事後評価とは別に、“追跡
評価”と呼ばれる新たな局面を導入する必要がある。一般
に事後評価と呼ばれるものは研究開発の終了直後になされ
るので、厳密には直後評価と呼ぶべきで、この時にはまだ
成果もきちんと定まっていなく、特に成果が出てくるまで
に時間がかかるものは、直後評価で最終的な成果を見極め

るのは困難である場合が多い。また仮に成果がすぐ確定できる場合でも、どのような研究運営の結果そうした成果が出てくるに至ったのか、更なる改善のためにはどうしたら良いのかを知るには“分析”をする必要があるが、これもやはり時間がかかる作業である。これらのことを考えると、実は直後評価というものは、「当該プロジェクトの今後をどうするか」という“意思決定”が行われる場としては大きな意味を持つが、実態を「つまびらかにする」という内容においては、中間評価と大きな違いはないと言える。以上のように考えると、最終的な成果を“確定”し、得られた知見や教訓を“発掘”し、それを次にどう活かしていくかの“分析”をする作業は、事後評価とは別に改めて実施しなければならないことに気づかされる。こうした作業のことを、“追跡評価”と呼ぶ^{[3][5]}。

しかし追跡評価を加えても、内省評価としてまだ十分な設計になっていない。例えば、TOEICの勉強というプロジェクトを考える。ここでは、TOEICで800点以上とるという目的のために、英単語を1日100語覚える、という計画を策定したとする。この場合に理想的な内省評価とは、実際にこの計画を実施してみて、成果の確認を行い、改善点を探し出して、また次の計画に戻すことである。しかし、改善点を見つけ出して次の計画に反映させるといっても、何も見当がつかないままの状態やみくもに計画を改善しようとしても効果的な見直しは期待できない。例えば1日100語覚えられなかったとしたら、それでは少し覚える単語の数を減らしてみる(再計画する)ということがまず思い浮かぶが、ではいくつに減らせば適当であるかについては、実態を把握していなければ、直感的・感覚的にしか判断できず、このままでは有効な見直しが行えない。

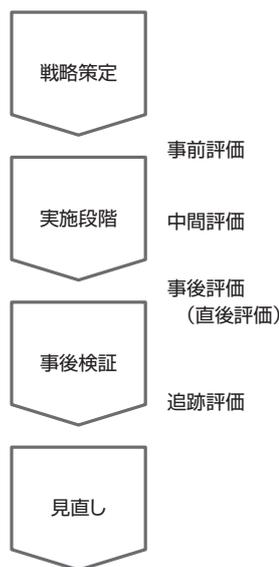


図3 研究開発のサイクルと評価の局面。

この見直しを真に効果的なものにするには、見直しを見据えて全体の評価のあり方を当初から有機的に設計する必要がある。つまり、図4にあるように、まず何故できなかったのか、どうすれば改善できるのかを“分析”するという行為が入らないといけな。本例で言えば、1日何語くらいしか覚えられなかったのか、それは何故なのかについて考察することに相当する。更に、より深い考察のためには、勉強しない日があるとしたら何で時間を確保できなかったのか、あるいは勉強しても全然暗記の効率が上がらなかった日はどういった日でどんな状況だったのか、という点について、分析を深めていかないと、何故できなかったのかつまびらかにならない。このデータに基づいた分析や考察に当たるのが、追跡評価である。しかしこうした分析を行うには、当然データに裏付けられていないと良い分析はできない。そうしたデータは、後から思い返したのでは信頼性のあるものは得られないので、課題を実施しながら、データも同時に集めなければいけない。これがいわゆる“モニタリング”といわれるもので、中間評価で実施する重要な内容の一つである。しかしこうしたモニタリングを行うためには、ではどういった項目についてモニタリングするか、あらかじめ実行課題（この場合は英単語を100語覚えること）を実施する“前”に設定されていなければならない。即ちこれが事前評価に相当するもので、事前評価は実行課題を策定・採択するだけでなく、追跡評価ができるようにモニタリング項目を定めることも含まれるわけである。このように、効果的な見直しに結びつけるためには、事前、中間、追跡の各評価がばらばらにあるのではなく、有機的な設計と運用が必要である。

4.2 ROAMEF

では以上の構造を満たせば内省評価として十分かという、更にもう一つ重要な要素が必要になってくる。先に説明したように、効果的な研究開発のためには、プログラムという仕組みが導入されていることも重要であった。そうした視点で今、図4を見てみると、これには「TOEICで800

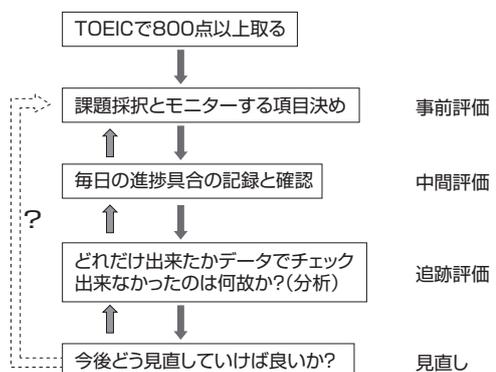


図4 プログラム化の例。

点とったその結果、それで何？」という問いに答えられていないことに気づく。そこで、上位階層としてここでは例えば、「英会話ができるようになりたい」という項目を設定するものとする、この目的のために、TOEICで高得点をとるのは確かに一つの方法であり、両者は意味ある関係として接合される（図5）。ただ、これは「英会話ができるようになりたい」という目的に対しての一つの方法であって、この目的の実現のためには、例えば“英会話資格の取得”など、ほかにも候補となりえそうな活動が考えられる。このようにして考えるとこれらは、“英会話の上達”という一つの大きな目的（意図する結果）の元、それを具現化させる指標となる、複数のより具体的な活動からなる構造になっており、言ってみればこれらは、“ターゲットセッティング”であると言える。そしてこれらのターゲットの実現のために、より下位の実行課題（英単語の暗記など）が位置づけられ、展開されている、と捉えることができる。

しかしこうしたターゲットを設定するためには、その根拠となる要素、即ち、“何故”そうしたターゲットを設定したのかという理由がないと、有効なターゲットセッティングができない。本例で、何故「英会話ができるようになりたい」のかの理由として、「アメリカ人の恋人を作りたい」という理由があるとしよう。それは、ここで設定した人物の人生設計から与えられるものとし、“アメリカ人の恋人を作る”ためには、会話ができないとコミュニケーションができないので、“英会話の上達”というターゲットが設定された、とするのは論理的な展開として十分想定されるものである。この何故（WHY）に相当するものは、英国の評価制度で言われるところの *Rationale* に相当する^[3]（図5）。

Rationale とはターゲットセッティングの WHY を説明するものであり、論理的根拠、正当性、根本的理由などといっ

たものが挙げられる。“論理的根拠”とは文字どおりその根拠としてロジックに基づくものである。本例で言えば、英会話は今後ビジネスでも必須の武器であり、国外に限らず国内の就職のためにも英会話がますます重要になる、という今後の世の中の状況を分析した上での論理的考察から、“英会話ができるようになる”というターゲットを設定する、といった場合のものである。一方、残りの二つについては必ずしもロジカルな根拠を必要とするものではない。“正当性”においては、規則や法律を根拠とするもので、決まりの上で道理にかなってさえいればよい。本例では、例えば産総研の職員は英会話ができないと解雇される、という内規がもし存在するならば、産総研職員に留まっていたいのであるならば、英会話を上達させる、というターゲットが導かれ、これは *Rationale* なものになる。最後の“根本的理由”というのは、ロジックにも規則にもその根拠を持たないが、行動者である個人あるいは集団の切実な理由（例えば情緒的なもの）があって設定されるものである。本例で挙げた“アメリカ人の恋人を作る”はまさにこのケースである。

このようにして捉えると、この *Rationale* と次の“ターゲット”の両者は、実は“戦略策定”にかかわることであり、これらが明確に設定されて初めて、その元でより有効で具体的な課題が選定できるという構造になっていることが分かる。英国の評価制度では、ここで言う“ターゲット”は *Objectives* と呼ばれ、プログラムの内容（WHAT）に相当するものに他ならない^[3]。それは今までみてきたように、実現させたい“意図した結果”であるところのもの（本例では“英会話ができるようになりたい”というような大きなターゲット）と、その指標となるようなもの（より具体的な小さなターゲット：本例で言えば“TOEICで800点以上

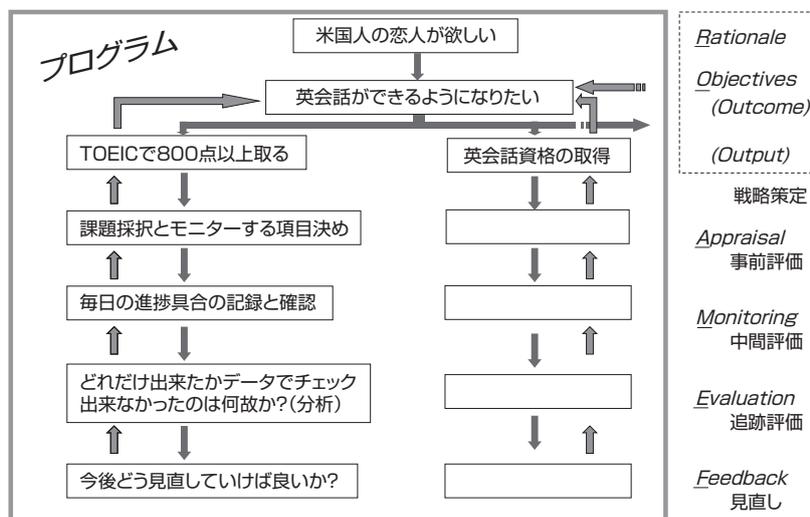


図5 プログラム化の例。図4の続き。

とる”ことや“英会話資格の取得”）とから成る（前者が *outcome*、後者が *output* と“定義”されるが^[9]、こうした用法は大きな混乱をもたらすことが多いのでここでは使用しない）。そして以下、これらをどのように（HOW）実現させるかについて、より具体的な活動事項（例えば英単語を1日100語覚える）たる実行課題が選択されると同時に、効果的なモニタリング項目を設定するという *Appraisal* が行われることになる（これは“事前評価”に相当する）。その上で、実際に実行しつつ、モニタリング項目のデータを収集するという *Monitoring*（“中間評価”にあたる）を経て、課題実施後にデータに基づいた運営上の教訓や知見の発掘などの分析を行うという *Evaluation*（ここでは日本語の訳語である“評価”よりは狭義の意味で、“追跡評価”にあたるもの）が行われ、改善点を見出して計画を見直していく、という *Feedback* につながっていく、という構造になっている。英国の評価制度では、この一連の流れは、それぞれの要素の頭文字をとって ROAMEF（“ロアモフ”と発音する）と呼ばれており^[3]、実はこうした構造を満たすものがプログラムの一つの必要条件となる。

4.3 プログラム評価の意義

ROAMEF に基づいたプログラム評価が優れている点は、ターゲットセッティングの有効性も含めた戦略策定の見直しを可能としている点である^[3]。これは ROAMEF を逆から辿ると分かりやすい。先に説明したように有効な見直し（*Feedback*）のためには、データに基づいた分析などの *Evaluation* が必要であり、*Evaluation* のためには、データの収集という *Monitoring* が必要である。そして *Monitoring* のためには、“実行課題の採択”と“モニタリング項目の設定”の二つからなる *Appraisal* が必要だが、実行課題の選択のためには、そもそも何を目的としてそのプログラムを実施するのが明確になっていないと設定のしようがないため、プログラムの *Objectives* が必要となってくる。そして有効な *Objectives* を設定するためには、そもそも何故そうした *Objectives*（ターゲット）を設定するのかという根拠が *Rationale* である必要があった。このように、ROAMEF には、*Rationale* と *Objectives* という大元の戦略策定まで問うことが構造的に埋め込まれており、ROAMEF に基づいたプログラム評価を行うのは、最終的には戦略策定の質的向上につなげるためのものであることが分かる。

このことが重要なのは、特に“社会のための科学”の研究開発においては初期最適化というのは非常に難しいためである。むしろ、最初から完璧な戦略策定ができていれば個別の課題の実施の上で全く問題ないのであるが、現実には戦略策定時においては調査が不十分であったり、戦略

策定に不可欠な情報が欠如していたり、また当初想定していたのは状況が変わってしまって、最初に設定した戦略策定では考慮に入っていなかった要素が出てくるなどからである。特に研究開発は、実際にやってみないと分からないという不確定な要素が強いため、将来予測は大変難しくなる。よって、実際に実行しつつ、周囲の状況や実行具合をみながら、軌道修正していくことが不可欠となる。

そしてその軌道修正には、プログラムレベルでのターゲット（ROAMEF においては *Objectives*）の設定の有効性まで含めることが非常に重要になってくる。*Rationale* な階層にまで遡って検討することにより、ターゲットそのものについての代替案を検討することが可能になるからである。本例で言えば、“英会話ができるようになる”ことの実現のために、TOEIC をはじめ英語に関する勉強をしたけれど、結果的になかなか上達しなかった場合を考えてみよう。そうした場合でも、“アメリカ人の恋人を作りたい”ということ自体を諦める必要は全然ないことに気づく。例えば、アメリカ人の中でも日本語を話せるアメリカ人を選べば本来の目的は達成される、という代替案が検討できるのである。こうした点がプログラム評価の非常に強力な点であり、何故 *Rationale* という要素まで含めて考えなければいけないかの理由はここに存在していたわけである。また、そこまで問わないと、いつまでもたっても非効率的な下位（プロジェクトレベル）の活動でもたもたしていて実効性が上がらず、結果的に無駄な投資ばかりして、かつ上位の政策実現に結びつかないというような事態が起きうる可能性がある。

プロジェクト評価だけでなく、プログラム評価も同時に実施する重要性は、特に産総研のような、“社会のための科学”を標榜する公的研究機関においては強調しすぎることはないであろう。我が国の公的研究開発では、「どうしてこういうプロジェクトが施策として出てくるのか」というような研究課題が選ばれることがあるとも言われる^[10]。よって、もし行政から与えられたターゲットが妥当でなかったり、戦略策定が適切でない場合、プログラム評価を実施することによって、そこで得られた明確な証拠とともにそうした点を政策担当者などに指摘し、上位の政策側が変わる契機を作っていくことが非常に大切になってくる。そうでないといつまでもたっても、社会問題の解決という社会からの期待に対して、個々の研究活動が非効率的にしか結びつかないことになりかねない。そうした意味で、社会から公的役割を付与された研究開発型独法は、独法化によって獲得した自由度を最大限に活用し、そうした点についても示していくことが、研究自体の遂行とともにもう一つの重要な役割であると考えられる。

5 プログラム評価としてのSynthesiology

実はこうした ROAMEF の原理に基づいた研究開発のプログラム化の実例集として、本誌 *Synthesiology* が挙げられる。*Synthesiology* に求められる論文の要件を、その「研究論文の記載内容について」からまとめてみると、以下のようになる。

- (1) 研究目標の設定→(2) 研究目標と社会との関係(社会的価値)の説明→(3) 研究目標を達成するためのシナリオの提示→(4) 要素の選択と要素間の記述(要素技術群の選択とその選択の理由)→(5) 要素間の関係と統合(構成)の考え方およびプロセスの記述→(6) 結果とその自己評価→(7) 将来の展開

これらの項目は、これまでみてきたプログラムの構造と以下のように対応する。まず冒頭の二つは、プログラムを始める上での *Rationale* な理由(2)と、それに基づく *Objectives* の設定(1)に相応する。そして、(3)のシナリオとは、*Objectives* 達成までの道筋の設計においてしばしば使われる“ロジックモデル(ロジックチャート)”に他ならない。次の(4)、(5)は、プログラムの *Objectives* を実現するために、具体的な課題(プロジェクト)を選定し、組み合わせ、プログラムとして有効な仕組みを作りあげる作業(*Appraisal*)に他ならない。(6)の“自己評価”は文字どおり *Evaluation* に相当し、最後の(7)は、展望として一種の *Feedback* を記述していると言える。

特に注目したいのは、*Synthesiology* が、社会の中の研究の“目標”(Objectives)の提示と、そもそもそうした研究課題を設定した意義を社会的な文脈の上で説明する(WHYを説明する *Rationale*)という点を強調し、冒頭にそれらの記述を求めているという記載順位である。というのは、これはまさにプログラム設計や評価の順番と対応する重要な点だからである。即ち、社会の中での *Rationale* な理由が設定できて、有効な研究の目標(Objectives)が決められる、というプログラムの構造である。その元で初めて、社会のために有効なものとなりうるプロジェクト選定の基盤が整うのである。別の表現をすれば、最初に研究開発プロジェクトがあって、そこから社会的な価値が何かを探すのではなく、*Rationale* に基づいた目標(Objectives)が設定できて初めて、有効な研究プロジェクトが選定できるのである。*Synthesiology* の求める論文の執筆要項は、まさにこの論理展開と軌を一にしている。

こうした視点からみると、*Synthesiology* の持つ意義は次の3点に集約される。一つめはプログラム化を通じて、個々の研究開発プロジェクトを社会に接続する試みが実践され、その結果が“記述”されていることである。これはまさに、産総研の憲章にある「社会の中で、社会のために」

という理念の具現化そのものである。二つめは、研究開発のプログラム化の具体例の提示である。筆者の知る限り、研究開発の実際の例について ROAMEF を軸として体系的に記述された文献は余り見られない。よって *Synthesiology* は、研究開発評価の方法論を深める評価研究者のためだけではなく、プログラムの実践的な深化に取り組むプラクティショナーにとっても大変貴重なサンプルを提供するものと考えられる。そして三つめは、研究開発の深化のためのフィードバックである。*Synthesiology* がプログラム化の実践と分析を提示する場と捉えれば、*Synthesiology* へ研究成果をまとめることで、当該研究の内省評価が半自動的に行えることを意味している。そのことで、当初の研究戦略の見直しと修正が図られ、プログラムの更なる深化が促されることになろう。このように *Synthesiology* は、プログラムという視点から研究開発を記述することで、“社会のための科学”に向けて自らの研究実践から学び、見直しと修正を可能とする場を提供していると考えられるのである。

6 おわりに

以上見てきたように、研究開発評価というのはそれ単独では意味をなさないものであり、戦略策定と一体となって互いに作用しながら進化していくものであることが分かる。

むしろ、社会問題の解決や政策課題の実現は、科学技術のみでできるものではなく、行政や社会などによる幅広い取り組みも必要であることは確かである。しかし、今日的な意味での科学研究において重要なことは、単に研究の成果を“ごろり”と外に出すのではなく、目的に応じて使いやすいように加工した上で、研究開発の成果を受け取り手が積極的に使うようになり、それを通じて社会構造の変革が促されるような“仕掛け”をあらかじめ作っておくことが研究開発の範疇に含まれてきていることである(例えば、^[13])。でないと、研究の論理のみが先行してしまっ、得られた成果のその後の使い方については「後はどなたかにおまかせします」ということにもなりかねない。そうではなく、研究成果の受容者、あるいは顧客(カスタマー)のニーズを見越し、先取りした“仕掛け”を作っておく必要があるが、その“仕掛け”こそプログラムに他ならない。

以上のように、今後、“社会のための科学”をより強力に推進する上において、研究そのものの質的・量的な向上とともに、戦略策定と一体となった評価が重要な役割を果たすと考えられる。我が国においては、先進主要国と比較してこうした評価人材の集積や養成に立ち後れがあることが指摘されており^[14]、今後、戦略策定や評価において一層の充実が必要だと考えられる。

なお本稿の更に詳しい内容について、<http://staff.aist.>

go.jp/ohtani-ryu/ に掲載する予定なので興味のある読者は参照されたい。

謝辞

本稿は、経済産業省産業技術環境局技術評価室の私的懇談会である「アウトカム懇談会」で筆者が行ったレビュー発表を元にしたものです。発表に対する大井健太氏、中村修氏、大久保泰邦氏、小笠原一紀氏のコメントは本稿の作成の上で大いに参考になりました。また産総研の小玉喜三郎特別顧問、小野晃副理事長、加藤碩一フェローおよび、岡村行信活断層・地震研究センター長と小泉尚嗣同チーム長には初稿を読んでいただき、貴重なフィードバックをいただきました。平澤冷未来工学研究所副理事長には、過去の報告書からの図の転用を快く許可いただきました。最後に、担当編集委員の小林直人早稲田大学教授（産総研特別顧問）には本稿を丁寧に読んでいただき、大変本質的な指摘をいただけたことで、本稿の飛躍的な改善ができました。ここに記して感謝します。

注 その他にも例えば、“ダイオキシンの健康への影響評価”や“アウトカム評価”などといった、“算定”や“見積もり”といった意味で“評価”という言葉が使われる場合があるが、本稿では考察の対象とはしない。

参考文献

- [1] 日本学術会議科学者コミュニティと知の統合委員会: 提言: 知の統合—社会のための科学に向けて— (2007).
- [2] 独立行政法人産業技術総合研究所: 産総研憲章 制定の意味, 産総研TODAY, 5, 8-11 (2005).
- [3] 文部科学省(委託調査): 研究開発評価の質の向上のための調査・分析 (2008).
- [4] 日本学術会議研究評価の在り方検討委員会: 我が国における研究評価の現状とその在り方について (2008).
- [5] 経済産業省(委託調査): 研究開発プロジェクト等の評価手法に関する調査 (2002).
- [6] 平澤冷: 研究開発における戦略策定と評価, シンポジウム「戦略的な研究評価について」報告書, 13-28, 独立行政法人産業技術総合研究所 (2006).
- [7] 文部科学省(科学技術振興調整費): 経済性効果分析手法とコスト算定手法の開発 (2004).
- [8] 新エネルギー・産業技術総合開発機構: 米国における定量的研究開発評価手法に係る調査報告 (2005).
- [9] 文部科学省(科学技術振興調整費): 研究開発のアウトカム・インパクト評価体系 (2006).
- [10] 総合科学技術会議: 総合科学技術会議第31回評価専門調査会議事概要(案) (2004).
- [11] 平澤冷: 政策のプログラム化とアウトカムの把握・評価, 経済産業省第5回研究開発評価フォーラム配布資料 (2008).
- [12] 産業技術総合研究所研究評価検討委員会: 産総研の研究開発評価のあり方(中間まとめ) (2004).
- [13] 吉川弘之: オープンラボによせて, 産総研TODAY, 9, 2-9 (2009).
- [14] 文部科学省(科学技術振興調整費): プログラムオフィサー等の資質向上に資する国内セミナー等の開催 (2005).

執筆者略歴

大谷 竜 (おおたに りゅう)

1999年3月東京大学大学院理学系研究科地球惑星物理学専攻博士課程修了。博士(理学)。同年4月通商産業省工業技術院地質調査所入所。2001年4月独立行政法人産業技術総合研究所地球科学情報研究部門研究員。2009年4月より、同活断層・地震研究センター研究員。この間、2003年2月から2005年2月までスタンフォード大学客員研究員。これまで、主にGPSを使った精密計測手法による地震や地殻変動、大気圏変動に関する研究に携わってきた。



編集委員からのコメントとその回答

コメント1 整理と主張の区別

コメント (小林 直人: 早稲田大学研究戦略センター)

本論説の一番のポイントになると思うのですが、①すでに研究評価論などでいわれていることを整理して提示した部分と、②著者の独自の視点で新たに述べた他の部分が、混在して分かりにくいと思われます。特に後者の部分、たとえば「反省評価」などは著者の独自の主張としてとりわけ強調して述べる点だと思いますが、今のままの記述だと単に他の人の主張を整理して述べた「解説」と受け取られかねないので、表現を工夫してはいかがでしょうか。

なお、「反省評価」という言葉は、どうしても「反省」という言葉に含まれる倫理的な印象が入ってくると思います。ここでは研究開発をより進化発展させるという意味で「進化のための評価」(Evaluation for evolution) などというのは如何でしょうか？

回答 (大谷 竜)

一研究者として日頃感じるのは、研究開発評価では、基本となる概念や言葉の使い方に注意しないと容易に混乱する要素が数多く含まれ、それが研究現場において適切な評価を遂行する上で時に大きな障害になっているのではないかと、ということです。そこで本稿では、研究開発評価の中でも特にプログラム評価に関する基本的な“考え方”を著者なりに整理し、分かりやすい言葉で表現し直すを試みました。紹介した個々の概念については既知のものですが、筆者の知る限り、それらを首尾一貫した形でコンパクトに親しみやすい形で概観したものがなく、そういった形で論説化することに多少なりとも意義があるのではないかと思ひ、本論説を著した次第です。この試みがうまくいっているかどうかは読者の判断を待たなければなりません。本稿を通じて、多忙な現場の研究者がプログラム評価の考えを理解する一助になり、被評価者と評価者と同じ理解のグラウンドに立つための足がかりになれば、と思っております。

一貫しているテーマは、“社会のための科学”研究の実現に資するような評価とはどういうことかであり、そのためには、実態をつまびらかにすること、最初に目指すべき戦略をもつこと、実態に応じてそれを柔軟に修正していくこと、の重要性を強調しております。以上の趣旨を冒頭で簡単に説明し、ご指摘の①と②の違いが分かるよう書き直しました。

なお「反省評価」というネーミングですが、確かに「何か悪いことをした」というような倫理的なニュアンスが含まれる場合もあります。一方、ご提案していただいた“進化のための評価”は、まさに究極的に目指すものであり、本稿で述べている評価の理念そのものですが、もう少し具体的で、実際の行動を表すような言葉の方が、特に評価の現場でイメージをもつためには良いように思えます。“改善評価”も考えましたが、“改善”という言葉には、そのインクリメントの意味が転じて、頑張っつつこつこつと努力するニュアンスが含まれ、内容よりもその姿勢の意味でとられる可能性があることから、あまり適切ではないように思えます。

ここで原点に立ち返って、“反省評価”の内容である見直しと修正

のための評価は何のために行うかという、「より良くする・なる」ためであったわけですが（その意味でまさに“進化のための評価”という理念になるわけですが）、「より良くする・なる」の主体は誰かといえば、その行為を行う当事者に他なりません。“社会のための科学”研究でいえば、研究実施者であり、それは他者に強要されるものではなく、自らが主体となって行うものだと考えられます。以上のことから、多少なじみの薄い言葉ではありますが、“内省評価”という言葉を使いたいと思います。大辞泉によれば、“内省”とは「自分の考えや行動などを深くかえりみる」とあります。よってこの言葉には、“反省”という言葉がもつ倫理的な色彩が少なく、暗黙的に含む“謙虚に”、という意味で「つまびらかにする」ということにも通じ、何よりも実施者自らが行う行為としての意味も含まれていると考えます。多少哲学的で奇異な響きもありますが、それ故に常にその意味に立ち返って考えさせられるメリットもあります。

コメント2 価値評価

質問（小林 直人）

第2章に「価値評価」の説明が出てきます。この「価値評価」の内容について記述をもう少し充実していただけるとよいと思います。平澤先生の資料にあった言葉だと思いますが、実はなかなか重要な言葉です。是非執筆者なりの考えを述べていただくことを期待します。

回答（大谷 竜）

研究開発評価には、通常の研究開発とは異なる固有の論理体系がありますが、こうした考えに慣れていない評価の現場で時々混乱が見られるように思えます。その中でも、自らの個人的な価値観を（しばし無意識に）援用した評価が行われている場面があるようにも思われます。本文でも説明したように、価値評価は「～が高く評価される」といった文脈で使われる評価ですが、「高く評価するかどうかは評価者たる自分」であるからには、自分の価値観、極端な場合、その人の人生観を反映したものが色濃く出る場合があります。問題は、こうした価値評価が研究開発の促進と深化の観点から有効かどうかで、それを問おうというのが価値評価という概念を導入した狙いです。その旨を本文に追加しました。

コメント3 追跡評価

質問（小林 直人）

第4章に研究開発直後に行われる事後評価とは異なる「追跡評価」の重要性が述べられています。しかしその内容は、「時間をおいた上でデータに基づいた事後評価」という内容のように見受けられます。内容は結構なのですが、一方、通常 NEDO などで行われている追跡評価とは、プロジェクト終了後1年から5年の間にプロジェクトの成果をどう活かして事業化にもっていったのか、などの波及効果を評価するいわば「アウトカム評価」に近いものと考えられます。そこには、そのプロジェクト以外の促進要素が多く入ってくるのが通常です。本稿でいう追跡評価は評価論で通常いわれているものでしょうか。そうでなければ、別の言葉で述べた方が誤解を招かないと思われませんか。

回答（大谷 竜）

2001年に内閣総理大臣決定された“国の研究開発評価に関する大綱的指針”によれば、「研究開発においては、終了後、一定の時間を経過してから、副次的効果を含め顕著な成果が確認されることもまれではない。こうした点を踏まえ、学会などにおける評価や実用化の状

況を適時に把握し、必要に応じて、研究開発施策、研究開発課題などについて追跡評価を行い、成果の波及効果や活用状況などを把握するとともに、過去の評価の妥当性を検証し、関連する研究開発制度などの見直しに反映する。」（下線は著者によるもの）とあります。

また、NEDOが2004年に定めた“追跡調査・評価の進め方について”にも追跡評価の観点として、「1）成果の説明責任の観点：研究開発の波及効果、売上の発生、市場の形成など、2）運営管理の見直しの観点：各種評価、実施体制、基本計画などの妥当性など、3）技術開発戦略への反映の観点：国としての取り組みの必要性など」（下線は著者によるもの）、を挙げております。

以上のことから追跡評価は、“研究開発成果の波及効果の把握”という側面とともに、少なくとも国レベルでは、研究開発の運営（制度）や戦略策定の改善に資することも目的としていると考えられ、本稿でもこの立場を採りたいと思います。

本文中でいわゆる事後評価（直後評価）と、追跡評価との区別が分かりにくいという点についてはご指摘を踏まえ、新たに直後評価についての内容の説明を加え、上記のような意味での追跡評価の独自性を浮かび上がらせるように書き直しました。

コメント4 ROAMEF

質問（小林 直人）

第4章に、「評価の局面の位置づけ」として4.1事前・中間・追跡評価の一体化、4.2 ROAMEF、などの、重要な論点が記されています。ここでプログラムの中の戦略が極めて重要であることを指摘したいと思います。特にRationaleで問うWhy?のみでなく、Objectivesを設定するWhatとHowも重要ですね。これをどのように設定するか記述があるとよいと思います。なお、プログラムの戦略は、いわばプログラム戦略と呼ばれるもので、それはより上位の戦略—いわば政策戦略の中に含まれるという理解でよいでしょうか。

回答（大谷 竜）

ご指摘のとおり、Objectivesはプログラムのターゲットそのもの（そのプログラムで“何を”実現させたいのか）なので、まさにWhatに相当します。またAppraisal以下の要素は、ターゲットを“如何に”実現させるかという意味でHowそのものになります。その旨分かるよう、書き加えました。

後者の質問に答えるには、評価の階層構造に立ち返る必要があります。評価対象が階層構造をもつということは、実は上位階層に対して下位階層が目的—手段の関係として位置づけられる、ということです。プログラムの場合、その上位階層である政策に対して、プログラムがその目的達成の手段になっている、ということに他なりません。したがって、ROAMEFの内部構造と同じように、プログラムは、上位にある政策の目的を境界条件として定められることとなります。プログラムの“戦略策定”（RationaleとObjectives）自体はプログラム内部に含まれますが、その根拠は、上位の政策から由来するという事です。

このように考えると、政策—プログラム—プロジェクトとは全てが関連していて、本来切れ間がないものですが、そこに政策やプログラムなどといった“区分”を設けるのは、“目的—手段”の間で区切ることで評価の重複を避けるとともに、担当者の所掌を明確にすることで、効率的で分かりやすい評価を実現するためです。別の角度から見れば、プログラムを設定する際、できるだけ政策を展開させる際の“一つの単位”となるように設計することが、効果的な評価体系を構築する上でポイントになると考えられます。