

公民館評価研究序説

—「プログラム評価」の理論と手法—

小池 佐々木 源 保 孝 吾

はじめに

いまや、「評価の時代」といわれる。遠因には、キャッチアップを念頭に構築された戦後諸制度の疲弊を挙げることでできようが、直接的には、特に地方自治体において一九九〇年代中頃からすすめられた行財政改革の影響が大きい。新公共管理 (New Public Management) を掲げて、¹⁾ 行政の無駄を²⁾ 検証する切り札として評価の導入が急速に進んだ。二〇〇二年の評価法施行以後は、納税者へのアカウンタビリティを標榜しつつ、政策評価は各領域に浸透していく。

公民館については、二〇〇三年の「公民館設置及び運営に関する基準」で、事業の自己点検・評価の実施と評価結果の住民への公表が明記された。そうした方針は二〇〇八年の改正社教法でも堅持され、その第三二条では、「運営の状況について評価を行うとともに、その結果に基づき運営の改善を図る」よう公民館に求めている。すでに、その前年改正されていた地教法 (第二七条) は、教育委員会に、所管の事業に対する評価を義務づけていたこともあって、一般の部局と類似した事務事業評価シートの提出を公民館にも課す自治体の数は一気に増加をみるようになった。

しかし、もともと社会教育関係者のあいだには、評価そのものに対する不信や懐疑が根強くあったところに、行政はコストカットや事業仕分けの方便として評価を断行したものだから、評価をめぐる双方の理解はいっそう歪なものになってしまった。評価を無視できぬ時代の状況である以上、あらためて評価とは何かを問い、その意義や原理について関係者が正しく認識することが喫緊の課題とみなされる。とはいえ、わが国の場合、評価は、ようやく口の端にのぼりはじめたばかりである。したがって、この分野の研究の蓄積はいたって乏しい。そこで、本稿では、評価研究の端緒として、評価の理論と手法を、歴史的にはアメリカ合衆国で発展をみた「プログラム評価 (program evaluation)」に字ばうとしていく。

ちなみに、「プログラム評価」とは、「評価研究 (evaluation research)」と互換的に用いられるタームで、「社会プ

プログラム (social program) の働きと効果に関する情報の収集、分析、解釈、伝達を目指す「社会科学的活動」と定義される。³⁾ しかも、ここでいう「プログラム」は「社会問題を改善し、社会状況を向上させるための活動」⁴⁾、つまり「社会プログラム (social program)」を含意しているので、本邦社会教育で一般にいうところの個別事業はもとより、それよりも広範な施策レベルの事業を指すことが多い。

1. プログラム評価の理論

アメリカにおけるプログラム評価の歴史は、一九三〇年代に試みられた職業訓練用の読み書きプログラムや、感染症による死亡率の抑制を目指すプログラムなどまで遡る。⁵⁾ その後、社会科学の発達を背景に、ニューディール政策下における諸事業の効果を測ろうとする取り組みを経て、第二次世界大戦期には、兵士や市民のモラルをモニターする方法などが開発された。戦後になると、社会プログラムは、都市開発、住宅、教育、職業訓練、保健による予防活動などの分野へと拡大する。さらに、宇宙開発競争で、一九五七年にソビエトが、アメリカに先んじて人工衛星の打ち上げに成功した影響は大きかった。いわゆるスプートニク・ショックを機に、数学や科学に関する教育の強化がはかられた。その効果を測定しようとしたことも、プログラム評価の発達にすくなくならぬ影響を与えたと言われている。

一九六〇年代は、評価の問題についての国民的な関心が喚起された時代であった。「貧困との戦い」(War on Poverty) のスローガンを掲げたジョンソン政権下のもとで、失業、犯罪、都市荒廃、医療へのアクセス、精神医療など、社会問題解決のための諸プログラムが矢継ぎ早に実施され、そこに巨額の前算が充てられたためである。あまりに性急に実施されたために、それらのプログラムは期待されたほどの結果をもたらさなかった。その結果、一九七〇年代に入ると、費用対効果がプログラム評価の主要な関心事となる。一九八〇年代のレーガン政権・ブッシュ政権では、財政赤字削減と小さな政府を標榜したことで、政策としての社会プログラムの前算は縮小したものの、むしろ、プログラム評価は精度を求められるようになる。成果測定の厳密性、あるいは、プログラム改善のための実践

的な活用に耐えうる評価手法の開発などが、それである。

この間、評価がひとつの専門領域として確立をみたことの意味は大きい。それが、プログラム評価論の精緻化を支えた。専門雑誌の老舗『*Evaluation Review*』誌は、一九七六年に創刊されている。『*The American Journal of Evaluation*』や『*Evaluation and Program Planning*』などの専門誌も重要な役割を果たした。ロッシン (Rossi, P. H.)、ワイス (Weiss, C. H.)、パットン (Patton, M. Q.)、クロンバック (Cronbach, L. J.) とつた研究者らの著書における評価論の体系化が図られた。さらに、評価の専門職団体として、評価研究協会 (The Evaluation Research Society : ERS) と評価ネットワーク (Evaluation Network : ENet) の二つが、評価にかかわる専門職性と科学性の推進に多大の貢献をしてきた。なお、両団体は一九八六年に統合されてアメリカ評価協会 (American Evaluation Association) となり、現在に至っている。

いまや評価は、連邦政策だけでなく、州や地方政府、NPO団体等でも必要不可欠な活動として定着をみている。とは言っても、評価を行う目的をみても、利用者あるいはプログラム実施の意思決定者にむけて、当該プログラムの結果を合理的に説明するためであったり、あるいは、プログラムそのものを改善していくためであったり、評価を通じて新たな知識を獲得するためであったり、さまざまである。

しかも、評価の目的は、評価者の意図や関心と分かちがたく結びついている。したがって、成人教育における評価の実際を論じたステイール (Steele, S. M.) は、教育行政担当者、教師、プログラムに参加した学習者で、評価の目的はそれぞれ異なると、次のように指摘する。

「行政担当者は実施されたプログラム全体の質や成果、価値を総括できるような評価を求める。おなじ行政担当者でも、連邦と州の職員なら、各地で実施されたプログラムに関する包括的な情報に強い関心をもっている。教師の場合、基本的な関心は、みずからの実践がいかなる質を有し、どのような成果をもたらしたかを知ることにある。他方、参加者たちは、他者と比べたり、予期していたものに照らし合わせたりしながら、自分がどの程度向上したかを把握

することに関心をむけている。⁶⁾」

それゆえに、コーティネイ (Courtenay, B. C.) などは「評価に関する定義は、それを行う人間やグループの数だけある。」⁷⁾ と言って憚らない。だが、評価についていまだ合意がないとはいえ、将来に向けて決定を下すためと、プログラムの価値やメリットを考量するための情報収集に評価の基本目的を見出そうとする点だけはいずれの論者にも共通する。⁸⁾

評価の目的が、目標の達成度を測定し、現状の改善に資する有用な情報収集にあることが確認できたなら、次はそれをどのようにして実施するかである。

ドラッカー (Drucker, P. F.) らは、『非営利組織の成果重視マネジメント』のなかで、非営利組織のための「自己評価の手法」を提起している。興味深いのは、そこに示された5つの問い、すなわち「われわれの使命は何か」、「われわれの顧客は誰か」、「顧客は何を価値あるものと考えているか」、「われわれの成果は何か」、「われわれの計画は何か」である。⁹⁾ それらは、「自分たちがなにをし、なぜそうしているのか、またなにを行わねばならないのか」ということについて自問自答するための手立てにほかならない。効果的にマネジメントを行うための秘訣と彼は言うが、それにとどまらず、ここには、プログラムに対峙して何を問わねばならないかという評価の要諦が端的に示されているように思えるのである。

評価は、このように評価すべき課題を抽出し、それについて問いを設ける作業でもって始まる。これを、評価クエスチョンの策定とよぶ。評価クエスチョンは、当該プログラムの発案から、計画、実施、成果にいたる各段階ごとに設定される。このうち後二者の実施と成果については「プロセス評価」と「アウトカム評価」で詳述することにして、ここではプログラムサービスのニーズに関するクエスチョンと、評価デザインをガイドするためのクエスチョンの例をロッシらの著書から抜粋しておこう。¹⁰⁾

【プログラムサービスのニーズに関するクエスチョン（ニーズアセスメント）】

- ・扱われるべき問題の性質と重要性はなにか。
- ・ニーズのある集団の特徴はなにか。

・その集団のニーズはなにか。

・どのようなサービスが必要なのか。

・どのくらいの量のサービスが、どのくらいの期間必要なのか。

・そうしたサービスをその集団に提供するためには、どのようなサービスの提供方式が必要なのか。

【プログラムの概念化またはデザインについてのクエスチョン（プログラム理論のアセスメント）】

・どのような利用集団をサービスの対象とすべきなのか。

・どのようなサービスを提供すべきなのか。

・そのサービスにおける最良の提供システムはいかなるものか。

・どのようにして利用集団を同定し、集め、意図したそれであると確認することができるのか。

・そのプログラムはどのようにして組織すべきか。

・そのプログラムにはどのような資源が必要で、また適切か。

すでに明らかかなように、「ニーズに関するクエスチョン」は当該プログラムの意義や必要性を評価することを意図し、また「プログラムの概念化またはデザインについてのクエスチョン」では、プログラムが企図したサービスの対象と、サービスを提供するための全体構造を評価することをねらいとしている。したがって、それらクエスチョンが的是でなかったり、評価者が自身の見方だけを頼りに評価課題を設定してしまうようなことがあれば、意味ある評価は覚

束ないものになってしまう。それを防ぐには、主要な利害関係者の意見を積極的に聴取するなり、彼らの関与を促すことが望まれる。

評価クエスチョンの策定にあたってもうひとつ重要なことは、解答可能なクエスチョンを心がけることである。そのためには、曖昧な文言や抽象的な表現を避けたり、主観が混入する余地をできるだけ少なくする必要がある。意味ある回答を期待するなら、観察可能な指標をクエスチョンに含めたり、回答時に根拠となりうる信頼できるエビデンスを同定できるようなクエスチョンが求められてくるのである。

こうして評価デザインが作成されると、実際の評価に入る。一定の企図を有して実施されるのがプログラムであつてみれば、成果を評価することはプログラム評価の中核をなす。次節では、アウトカム（成果）評価について考察をおこなう。

2. アウトカム評価の原理と手法

評価学会などの動向をみるに、プログラムの「成果 (outcome)」については、どうやら「アウトカム」とカタカナ書きするのが慣例となりつつある。わが国では「アウトプット (output)」も「アウトカム」と近い意味で用いられることがしばしばあるものだから、不用意に日本語で「成果」と表記しようものなら、「アウトプット」と「アウトカム」のいずれを意味しているのか紛らわしくて、混乱を招きかねない。わざわざ「アウトカム」と表記するのは、「アウトプット」と区別しようとする意図がはたらいっているようだ。

アウトプットとは、プログラムが終了した際にプログラムによって発生した状況を、主として定量的に算出したものをいう。具体的には、参加者数や利用者数、施設の稼働率、講師に対するクライエントの割合、ユニットとして行った活動の数、組織されたグループ数や参加者に対する組織率などが、これにあたる。

他方、アウトカムは、プログラムが参加者個人に与えた影響であつたり、もつと射程を長めにとれば、プログラム

が社会に与えた影響を指す。プログラムによってもたらされた変化や変容と言い換えてもよい。これは、定量的に測ることもあれば、定性的に把握する場合もある。

アウトカム評価の目的は、当該プログラムが意図したものがどの程度達成されたかを測ることにある。したがって成果を漫然とひとくくりにして、たとえば「このプログラムは良かった／悪かった」論評してもあまり意味をなさない。こういった課題がどの程度まで達成されたかを厳正に測定することが不可欠となる。ロツシらにしたがえば、アウトカム評価を行うにあたっての評価課題は、以下のようである。¹⁾

【プログラムのアウトカムについてのクエスチョン】

- ・アウトカムのゴールや目標は達成されているか。
- ・そのサービスは受益者に対して利益となる効果をもたらしているか。
- ・そのサービスは受益者に対して不利益となる効果をもたらしているか。
- ・他のものよりもそのサービスから、より大きく影響されている参加者がいるか。
- ・そのサービスが目を向けている問題や状況はよくなっているか。

アウトカム評価で注意すべきは、その評価法である。アウトカム評価の考え方を図示したものが、図1である。図の縦軸には、プログラム実施前の状態（プログラム前アウトカムレベル）からプログラム実施後の状態（プログラム後アウトカムレベル）への移行を示している。事前の状態から事後の状態にどう変化したか、その較差を「アウトカム変化」という。

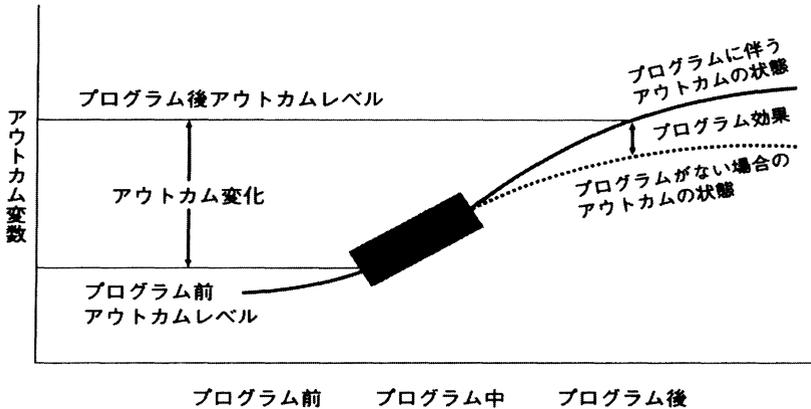


図1 アウトカムレベル、アウトカム変化、プログラム効果⁽¹²⁾

しかし、事後の状態が何によってもたらされたかを考えてみると、それが、プログラムによる効果だけではないことは自明である。知識の増加にたどって言えば、そこには、個人的な学習を含むプログラム以外の諸種の影響が混在しているはずだ。「プログラム後アウトカムレベル」をすべて当該プログラムの賜とみなすなら、それはあまりに早計に過ぎよう。言い換えるなら、「プログラム後アウトカムレベル」から、もしも「プログラムがなかった場合に想定されるアウトカムの状態」を差し引いたものが、実質、そのプログラムによってもたらされた効果（純効果）ということになる。厳密にいうと、この純効果を測定しようとするのがアウトカム評価なのである。だから評価者には、実際に得られたアウトカムにとどまらず、アウトカムのうちのどの程度がプログラムそのものによってもたらされたものなのか見極めることが求められてくるのである。

プログラムの純効果を測定しようとする、プログラムを与える介入群とそうでない対照群をあらかじめつくっておいて、両者のアウトカムを比較するのが理想的である。無作為化フィールド実験法と呼ばれるのが、それである。しかし、そうした実験が常に可能なわけではない。むしろ現実にはほとんど不可能に近い。実験による比較検討ができないとあらば、それに代わる別の方策を考え出さねばならない。

そのひとつが、完全とは言えないまでも何らかの比較グループを設定

するやり方である。同じ条件のグループが作れないなら、可能なかぎり性質の近似したグループを選定し、準備したり、場合によっては、対象集団そのものを統計処理によって介入群と比較群とに分けるやり方もある。それも難しくれば、全国平均など、既存のデータと比較対照させてみるのもよい。比較対象群がみつからない時の対応としては、同一のグループのアウトカムを経時的に測定して、その変化をもとに評価を行うという方法もある。さらに、もっと簡便なやり方としては、評価者が既有する観点（経験による基準）に照らして、成果を判断していくやりかたも考えられる。

いずれにしても、データをいかにして収集するかが、アウトカム評価の成否を決することになる。

ここまでくれば、後は評価の際に判断材料となるデータを収集するだけである。観察や記録、面接や質問紙による調査、標準化されたテストといったところが一般的である。信頼性 (reliability) と妥当性 (validity) の高い評価を行うには、それなりのデータをいかに収集するかと、そうして収集したデータを処理する際の評価尺度が重要である。それには、現象を的確に測定しうる尺度が設けられているか、同じ現象を同じ測定手順で扱ったなら、同一の結果が出るかどうかなどに絶えず意を用いることが肝要である。さらに、プログラム以外の諸要因がアウトカムに影響を与えていれば、それらを社会科学的に妥当と思われる方法で除去することも大切である。一定のトレーニングを積んだ専門家が評価に求められる所以である。

もっとも、社会調査の専門家に委せつきりでは、折角プログラム評価を行っても、そこから得られた知見が現場の改善に生かされる可能性は低い。それに、いみじくもアームストロング (Armstrong, P.) が述べたように、本来、評価は批判的思考によって成立する営為である。したがって、評価に継続的に携われれば、そのことを通して、批判的思考のスキルを習得するという効用も期待できる。だからこそ、利害関係者 (stakeholders) のそれぞれが、社会調査の専門家と協働するか、さもなければ彼らが提示する分析結果を理解し、それを活用する方途を主体的に見出そうとする努力が欠かせないのである。

3. プロセス評価の原理と手法

プログラム評価とは、「プログラムに投入した資源」、「実施された活動」および「活動から産出される結果」、そして「成果」（短期、中期、長期）の関係を構造的に把握する作業と言ってよい。したがって、プログラム評価では、図2のような「ロジック・モデル (logic model)」がしばしば活用される。

このロジック・モデルは、それぞれのカラムが右隣と「原因→結果」の関係で連鎖しているところに特徴がある。最終的なアウトカムとして意図する長期的成果を設定した際に、プログラムに関係した諸側面をロジック・モデルで構造化できなければ、それは、「理論上の失敗」を意味する。つまり、評価不能ということになる。さもなくば、次のような解釈も成り立つ。元来、プログラム評価論はプログラムの純効果を測定しようというものであるから、アウトカムにかかわる諸要因の因果関係が論理的に描ききれないとすれば、それはプログラム以外の要素が影響してもたらされた可能性が高いことになる。

ロジック・モデルに示されたように、アウトカムは、そこにいたる一連のプロセスの帰結である。だから、アウトカム評価を行っても、そこに至るプロセスが無視されていたのでは、当該プログラムの現状把握や今後の改善に役立つ情報の収集という面では片落ちの誹りをまぬがれない。つまりプログラム評価においては、アウトカムはもとより、それを規定するプロセスについても必要な情報を収集することが要請されてくる。

プロセス評価の意味や、それがプログラム評価の全体に占める役割を理解するには、プロセスにかかわる評価クエスチョンを一覧してみるとよい。ロッシらは、プロセス評価のためのクエスチョンとして次の16項目を提示している¹³⁾。多少分量は多いが、引用すると、次のようである。



図2 プログラム評価のためのロジック・モデル

【プログラムプロセスにかかわる評価クエスチョン】

- ・ 何人の人がサービスを受けているのか。
- ・ サービスを受けている人は意図された標的集団か。
- ・ その人たちは適切な量、種類、質のサービスを受けているか。
- ・ サービスを受けていない標的集団はいるか、あるいは、その標的集団の中にサービスを受けている人が少ないサブグループがあるか。
- ・ 標的集団のメンバーはプログラムのことを知っているか。
- ・ 必要なプログラム機能は十分に実行されているか。
- ・ 実行されるべき機能に対して、スタッフの数と能力は十分なものであるか。
- ・ プログラムはうまく組織されているか。スタッフは互いにうまく連携しあっているか。
- ・ プログラムは協力しあうべき他のプログラムや機関と効果的に協調しているか。
- ・ 資源、施設、そして資金は重要なプログラム機能を支援するのに十分であるか。
- ・ 資源は効果的、効率的に使われているか。
- ・ プログラムは管理委員会、助成機関、そしてより高次の行政機関より課された要求に準拠しているか。
- ・ プログラムは該当する専門的または法的標準に準拠しているか。
- ・ ある場所でのプログラム実績が他よりもかけ離れてよかつたり悪かつたりすることがあるか。
- ・ 参加者はプログラムの職員やその進行でのやり取りに満足しているか。
- ・ 参加者は自分たちの受けるサービスに満足しているか。
- ・ 参加者はサービス後の適切なフォローアップ行動に取り組んでいるか。

これらを通覧すると、クエスチョンは概して2つの問題関心に基づいて策定されていることに気づかされる。そのうちのひとつは、あらかじめ想定していた標的集団にサービスが届いているかどうか、つまり「カバレッジ(coverage)」を測定する質問群から構成される。アクセスメントには、プログラム記録、プログラム参加調査、コミュニティ調査が用いられる。ニーズがあるにもかかわらず、その標的集団のメンバーでプログラムに参加している人間の割合が低いときを過少カバレッジ、逆に、ニーズが大してあるわけでもないのに参加している人間がやたら多い場合を過大カバレッジと呼ぶ。ニーズのある人間に対するサービスの提供件数を最大化するのが理想であるから、過少カバレッジにしても、過大カバレッジにしても、何らかの対策が必要となる。

クエスチョンにみるもうひとつの質問群は「組織機能」に関係している。それらは、提供するサービスと、そのサービスをサポートするための支援機能がプログラムの設計仕様と齟齬がないかモニターすることを意図している。組織機能のモニタリングは、当該プログラムが、最重要課題を遂行するのにどれだけうまく資源を活用したり、取り組みを組織しているかを測ることを意図したものである。その際、サービスが標的集団にうまく届いているかどうか、届いていなければ、その原因を究明することに細心の注意が払われる。

他方、支援機能などというものは、とかく管理者の所掌事項とみなされがちである。したがって、支援機能をプロセス評価に含めること自体、奇異に感じる向きがあるかもしれない。しかし、たとえサービスを提供することが本来の機能であっても、そうした機能を維持し、またみずからを存続せしめようとすれば、プログラムは支援機能を併せもつ必要がある。実際、資金調達をはじめ、潜在的スポンサーや意志決定者、一般人を対象にしたプログラマイメージを高めるための広報活動、スタッフの研修、主要職員の募集と維持、サービスに必要な資材の調達などは、いずれのプログラムでも、運営上重要な側面を構成しているではないか。プログラムの存続にかかわる重要な機能である以上、プロセス評価においても支援機能を看過するわけにはいかないのである。

こうしてプロセスモニタリングを行ってみると、プログラムが当初のデザインどおりには展開をみていないことが

ある。その場合、原因を突き止め、プログラムを計画に近づけるよう改善策を講じるのが一般的である。しかし、原因は、もつと根本的なところにあることもある。そうした場合には、デザインそのものを再検討するという大胆な対応をすることも肝要である。

おわりに

戦後わが国の社会教育において中核的役割を果たしてきた公民館であるが、こと評価に関しては、事業が終わった時点で、参加者に感想を求めるのがせいぜいであった。しかし、相手は、ドロップアウトすることなく、最後まで学習を継続した人びとであるから、データそのものにもとと偏りがあるうえに、誰しも、みずからの行為を正当化したいという心情が働くから、よほどのことがないかぎり、「参加してよかった」、「満足している」などと肯定的な回答が相次ぐことになる。評価と呼ぶには、あまりにも杜撰で、情緒的にすぎる。

他方、昨今、各自治体が普及に余念がないものに、事業シートにもとづく評価法がある。シートには、事業の概要を記載する欄に続いて、当該事業の必要性や、見直しの可能性などをチェックする項目が設けられている。感想をもとめるやり方に比べると、一見、体裁はととのっているかのように見える。ところが、じつのところ、記載者は、かならずしも十分に信頼できるデータを持ち合わせているわけではない。だから、記載する人間の主観が紛れ込む余地はあまりに大きく、そのため、評価の厳密性という点でいまだ多くの課題を残しているのが実態である。

さらに、当該事業が終了した時の状況を、たとえば参加者数など、定量的なデータでもって示して、事足りりとする風潮も問題として看過できない。とりもなおさず、それは、関係者のすくなくからぬ者が、アウトプットとアウトカムとを混同していることに起因する。

地域住民の主体的な学習活動を支援するのが公民館の使命であってみれば、評価においては、当該事業に参加した人々の、学習による変容や、さらに長期的には地域にもたらされた変化にこそ注意を向けねばならない。

もつとも、そうしてアウトカムが把握できたとしても、問題は、それをどう評価（＝価値判断）するかである。その際、重要となるのが所期の目的であり、プログラムデザインである。アウトカムをそれらに照らしてみて、はじめて事業の成否を判定することができるのである。換言すれば、参加者の多寡でもって事業の成否を判定しようとするなどは、許し難き過誤をいくつも犯しているにひとしい。わが国における公民館評価の問題点は、およそ以上のよう

に要約できよう。

それにひきかえ、アメリカにおけるプログラム評価論は、プログラムの効果をあらんかぎり客観的、かつ正確に把握したいという欲求に支えられて、社会科学と手を携えつつ発展してきた。データの収集やその分析方法の面で信頼性や妥当性が重視されたのも、そのためである。

しかし、刮目すべきことはそれにとどまらない。プログラムをシステムティックにとらえようとする考え方は、当該プログラムによってもたらされたアウトカムのみならず、アウトカムを規定する諸側面を評価の射程に含めた。そのプログラムは、一体誰の、何に役立っているのかを明らかにするには、アウトカムを評価すればおよそ十分だろう。しかし、ニーズに対するアセスメントを考量することによって、プログラムの始原にまで立ち返って考える道を切りひらいた。さすれば、そもそも誰のための、何のためのプログラムであったか、つまり所期の意図をいま一度の確認作業を通して、実際のアウトカムと引き比べることができる。もし、そこに齟齬が見出されたなら、予算、施設・設備、職員、教師、提供物、参加者のうち、何が、アウトカムの導出を阻んだかを究明する。原因を明らかにすることによって、はじめて改善に資する方策を見出すことが可能となるのだ。しかも、評価クエスチョンを具体的に提示することによって、プログラムを改善するには、何を測るべきか、何を考えるべきか、微にいり細にわたって指し示さうとしているのは、さすが功利主義の国アメリカならではの発想と思考の賜と言えようか。

さらに、評価の当事者として実に多様な人びとが想定されていることにも驚かされる。評価に携わる人間の評価に対する意図は、彼らの立場や関心によって異なることは、本文中でも繰り返し述べた。そのため、評価のあり方に科

学性を重視する一方、多様な人々が評価に参画することを奨励する風土が醸成されてきたと言っても過言ではない。これら諸点に鑑み、理論と実践の両面で、いま、プログラム評価に学ぶ意味ははかりしれないように思われる。

注

- (1) 山谷清志『政策評価の実践とその課題―アカウンタビリティのジレンマ―』萌書房、二〇〇六年。
- (2) 社会教育における評価を阻んできた原因として、辻は、社会教育理論全体の脆弱性、頑固な地域主義、社会教育関係者の評価に対する悲観的、諦観的な態度、安寧希求的な考え方を指摘している。
- (3) 伊藤俊夫、河野重雄、辻功編『新社会教育事典』第一法規、一九八三年、五四六―五四七頁。
P.H. ロッシ、M.W. リプセイ、H.E. フリーマン著、大島巖、平岡公一、森俊夫、元永拓郎監訳『プログラム評価の理論と方法―システムティックな対人サービス・政策評価の実践ガイド』日本評論社、二〇〇八年、二頁。
- (4) 同上書、一七頁。
- (5) プログラム評価論の歴史については以下を参照した。
・ Shadish, W. R., Luellen, J. K., "History of Evaluation", Mathison, S. (ed), *Encyclopedia of Evaluation*, Sage Publications, 2005, pp.183-186.
・ 安田節之、渡辺直登『プログラム評価研究の方法』新曜社、二〇〇八年、一五―一八頁。
・ ロッシら、前掲書、八一―一五頁。
- (6) Steele, S.M., "The Evaluation of Adult and Continuing Education", in Merriam, Sharan B. and Cunningham, Phyllis M. eds, *Handbook of Adult and Continuing Education*, Jossey-Bass Publishers, 1989, pp.260-274.
- (7) Courtenay, B.C., "Evaluation of Adult Education", in Tujiiman, Albert C., ed, *International Encyclopedia of Adult Education and Training* (Second Edition), Pergamon, 1996, p.839.
- (8) Spaulding, S. "Adult Education Evaluation in Developing Countries," in Timmus, C. J., ed, *Lifelong Education for Adults: An International Handbook*, Pergamon Press, New York, 1989.
- (9) P.F. ドラッカー、G.J. スターン編著、田中弥生監訳『非営利組織の成果重視マネジメント』ダイヤモンド社

二〇〇〇年、xiii頁。

(10) ロッシら、前掲書、七四―七五頁。

なお、下記の原著も参考に一部訳文を改めた部分がある。

・ Rossi, P. H., Lipsey, M. W., and Freeman, H. E., *Evaluation A Systematic Approach*. (7th Edition), Sage Publications, 2004, p.77.

(11) 同上書、七五頁。(原著、Rossi, P. H., Lipsey, M. W., and Freeman, H. E. op cit., p.78)

(12) 同上書、一九五頁。

(13) 同上書、一六一―一六二頁。

(付言)

この研究は、日本社会教育学会によるプロジェクト研究の一員として「社会教育における評価研究」に関わったことを契機にしている。本稿は、同研究の一環をなすもので、アメリカで進展めざましいプログラム評価研究の成果に学びつつ、わが国の公民館が依拠すべき評価の基礎理論をさぐることを意図した。

なお、共著者の佐々木保孝は、本学大学院教育学研究科を経て、現在天理大学人間学部にて講師をつとめている。