

インパクト評価は公共サービスの質を改善するか？

馬場 英朗

(要約)

エビデンスに基づく政策に対する関心が高まっており、公共サービスのイノベーションや財政削減に対する活用が期待されている。そのための成果測定手法として、インパクト評価の導入が諸外国で取り組まれており、その評価結果に応じて報酬の支払いを判断する成果連動型契約も広がりつつある。インパクト評価は、(1) 成果を単なる数値指標ではなく、プログラムが生み出した変化分として測定する、(2) プログラムが成果を生み出す因果関係を、データなどのエビデンスを用いて立証する、(3) 最終成果だけでなく初期的・中間的な成果を含み、経済的な成果だけでなく社会的な波及効果も含まれる、という点で従来型の成果測定手法とは異なっている。インパクト評価を機能させるためには、プログラムの達成状況を測定するための業績測定としての視点と、政策の効果をモニタリングして予算の見直しに生かす事業評価としての視点を区別し、各々の目的に応じて二段階の評価の仕組みを使い分けることが有用である。

(キーワード) エビデンスに基づく政策, 成果連動型契約, 業績測定, 事業評価, ソーシャル・インパクト・ボンド

1 はじめに

厳しい財政状況のもと、政府には歳出削減が求められているが、単に予算をカットするだけでは公共サービスの質と量を維持できない。そのため、小泉政権(2001～2006年)では「官から民へ」、鳩山政権(2009～2010年)では「新しい公共」、安倍政権(2012年～現在)では「共助社会づくり」という言葉を用いながら、民間にも公共サービス市場を開放し、地域住民が相互に支え合う仕組みを構築することを模索している。

このような公共サービス改革は、1980年代にイギリスなどで取り組まれたニュー・パブリック・マネジメント(NPM)に端を発している。NPMは公共サービスに市場原理を取り入れ、

業績と成果に基づいて公共サービス市場に民間事業者が参入することを広く認めるものであるが、「短期契約・協定，一般競争入札による高度な競争性と目先の調達価値確保が追求され，その弊害が広がっている」（北大路 2011, p.21）といった批判もある。

そのため，イギリスでも近年ではコスト削減のみを目的とするのではなく，一定水準以上の成果を出した場合に，行政からの支払いが行われる成果連動型契約（outcome-based contract）などが導入されるようになってきている。成果連動型契約は，イギリスではペイメント・バイ・リザルツ（PBR），アメリカではペイ・フォー・サクセス（PFS）と呼ばれているが，これらは単に「成果に応じた報酬を支払う」という支払基準を意味しているわけではなく，官民を問わず，より高い成果を効率的に生み出せる組織が公共サービスを提供する，という新しい公共調達スキームの一部を構成している。

そして，最近では公共サービスの成果を測定する手法として，インパクト評価に対する関心が世界的に高まっており，G8のタスクフォースにおいても議論が行われている（Social Impact Investment Taskforce 2014）。ただし，小関・馬場（2016）に指摘したように，現時点では標準化されたインパクト評価の手法は存在せず，その目的に応じて様々な評価方法が試行錯誤されており，実務現場に混乱を招いている部分もある。

そこで，本稿では民間事業者が公共サービスを担う場合に焦点をあてて，公共サービスの質的改善と財政健全化を両立させるインパクト評価のあり方について考察を試みる。

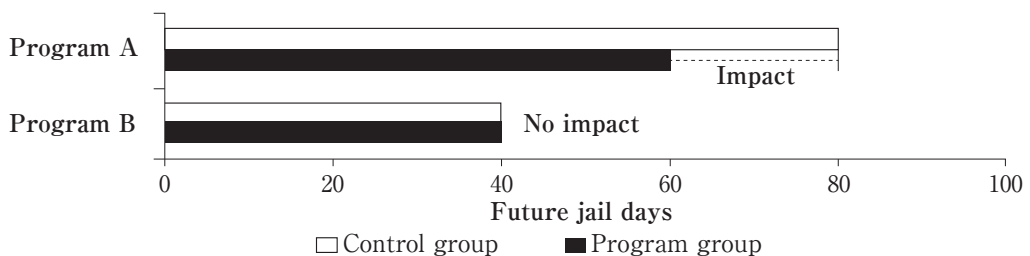
2 インパクト評価の意義

公共サービス分野にインパクト評価が導入されるようになった背景には，エビデンスに基づく政策（evidence-based policy）に対する世界的な関心の高まりがある。例えば，イギリスではブレア政権やキャメロン政権のもとで，医療，教育，貧困，産業政策，地域活性化など，様々な分野においてエビデンスに基づく政策が推進されてきた（家子ほか 2016）。また，アメリカでもオバマ政権のもとで「事実に基づいた証拠をベースとする政策決定と政策評価」が導入され，科学的に厳密なリサーチによる事実証拠を基盤としたプログラムが採択・継続されて予算を獲得し，有効でないプログラムは改革あるいは廃止することが取り組まれている（上野 2015）。

このようにエビデンスに基づく政策は，厳しい財政状況もとで政策の有効性を検証し，予算を取捨選択するために活用されている。そして，政策が想定する目標を達成できているかどうかを判断するための成果指標として，インパクト評価が用いられている。インパクト評価は，様々な主体によって数多くの手法が提案されているが，それらに共通する概念として，（1）成果を単なる数値指標ではなく，プログラムが生み出した変化分として測定する，（2）プログラムが成果を生み出す因果関係を，データなどのエビデンスを用いて立証する，（3）成果には最

終成果だけでなく初期的・中間的な成果を含み、経済的な成果だけでなく社会的な波及効果も含まれる、という特徴を有している¹⁾。

インパクト評価の考え方を図式化すると図1のようになる。これは短期受刑者に対して再犯防止プログラムを提供する事例であり、刑務所の収監日数が少ない（棒グラフが短い）方が高い成果を示している。そのため、絶対値ベース（各々の黒い棒グラフ）でアウトカムを比較する従来型の数値評価によれば、プログラムBの方が高い成果を生み出していると考えられる。それに対して、インパクト評価の場合は、プログラムを実施したとき（黒い棒グラフ）と実施しなかったとき（白い棒グラフ）との比較が問題になる。すなわち、プログラムBは実施しても、実施しなくても収監日数に変化はないが、プログラムAを実施すれば、実施しないときよりも収監日数が削減されるため、プログラムAの方が高い成果を生み出していることになる。



出所：MDRC（2013, p.18）

図1 インパクト評価の考え方

ただし、インパクト評価の測定結果は、目標とする成果が生み出されたか否かを判断することを目的とする、いわゆる「業績測定」(performance measurement) のためだけに用いられるわけではない。たとえ想定された成果が得られない場合であっても、インパクト評価を実施し、目標と実績との間に生じたギャップを分析することによって「正しい目標が設定されていたか」、「適切にプログラムが設計されていたか」、「プロセスが有効に実施されたか」ということを見直し、プログラムの改善に結びつけることが可能となる。

そして、目標とする成果指標をクリアして、プログラムが成功したと判断されても、その成果が投入した予算に見合った政策効果を本当に生み出しているかどうかは別問題である。そこで、「誤った成果指標が設定されていないか」、「当初想定されている社会的価値や財政削減効果が生み出されたか」、「プログラム終了後にどのような中長期的な変化が生じているか」を検証する「事業評価」(program evaluation) も併せて実施することにより、予算の見直しや取

1) 内閣府（2016, p.2）では、社会的インパクトは「短期、長期の変化を含め、当該事業や活動の結果として生じた社会的、環境的なアウトカム」と定義されている。

捨選択に関しても、インパクト評価を活用することができると考えられる²⁾。

3 公共サービスのインパクト評価

公共サービスにおいてインパクト評価が用いられる場合、(1)単年度ではなく、中長期にわたって効果が波及する、(2)社会的価値を生み出すと同時に、何らかの財政削減効果を求められることが多い、(3)インパクト評価による測定結果が、報酬支払いの可否判断や次年度以降の予算策定に反映されることがある、といった点が特徴的である。

通常、公共サービス分野であれば、評価の方法は入札段階で提示され、そこに定められた仕様に従って成果指標が設定される³⁾。そのため、業績測定スキームは外部評価機関などが考えるわけではなく、契約にあらかじめ定められた条件に従って、サービス事業者が成果指標に関するデータを収集し、発注者である行政機関に報告することにより評価が行われる。

例えば、マンチェスター市議会のメンタルヘルス分野及び子ども支援分野のコミッショナー(調達責任者)であるJock Rodger氏によれば、契約のなかに定められる成果指標は、入札に先立って実施される市場調査や実行可能性調査に加えて、サービス事業者や中間支援組織などの関係者との情報交換も行いながら検討される⁴⁾。さらに、成果指標はプログラムの成否の判断根拠になるとともに、成果連動型契約においては報酬支払いの可否を決定する基準となるため、プログラムの実施中は変更されることなく、継続的に用いる必要があるということであった。

一般的にプログラムの成果指標は、PDCAサイクルに基づいて柔軟に見直されるべきである。しかし、公共サービス分野においては、もしプログラムがうまく進んでいないとしても、その途中で成果指標が変更されると、ムービング・ゴールポストの状態に陥ってしまい、阻害要因

2) 田辺(2014, p.11)は、本稿でいう業績測定と事業評価に相当するものとして、業績測定(実績評価)とプログラム評価(総合評価)の用語を使用し、業績測定だけでは業績を正しく把握できないため、プログラム評価を併用することが有用であると指摘している。

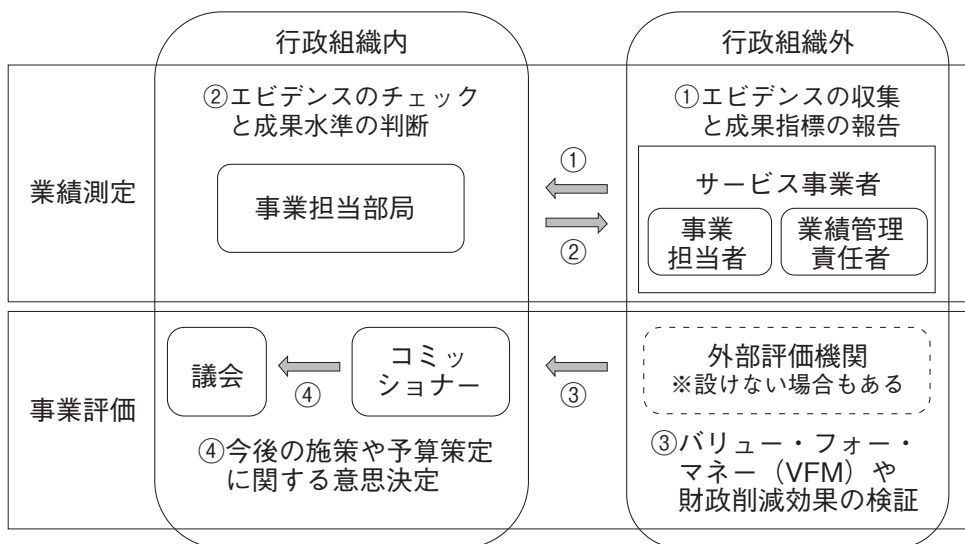
3) Burnett(2009, p.17)によれば、目標とする成果を達成するための方法が不明確なプログラムの入札条件を決める際には、利害関係者が集まって仕様を検討するコンペティティブ・ダイアログを活用することが有効である。コンペティティブ・ダイアログはEUの2004年公共調達指令に規定されており、2017年8月17日に筆者が見学した英国デジタル・文化・メディア・スポーツ省によるLife Chances Fund(薬物・子ども・若者など6分野への社会的投資に80万ポンドの予算を充てる)の研修会においても、発注元である地方自治体のコミッショナー(調達責任者)と、受注を希望する社会的企業がプログラムの仕様に関して一緒に検討するロールプレイが行われていた。

4) 2016年9月19日に、マンチェスター市議会の委嘱を受けたコミッショナーであるJock Rodger氏に実施したヒアリングに基づく。なお、マンチェスター市議会のコミッショナー事務局には50名ほどが所属しており、各事業の担当部局からは独立しながら組織横断的な監視・評価を行う「センター・オブ・エクセレンス」としての役割を担っている。そして、数人のコミッショナーが各分野を分担することにより、コミッショニング・ハブとしての機能を果たしているということであった。

は何かということが不明確になる。そのため、明らかに事業目標が間違っているような場合でなければ、安易に成果指標を変更するのではなく、事業のプロセスや実施体制を見直し、成果の達成を妨げている原因を探ることが、新たなイノベーションを生み出して、事業の改善にも結びつくと考えられている。

ただし、プログラムの実施期間中に成果指標を変更しないことが、インパクト評価の硬直化を意味するわけではない。なぜなら、ここでいう成果指標は、図2に示したインパクト評価の構造でいえば、プログラムが目標とした成果水準を満たしているかどうかを判断する業績測定に相当する。マンチェスター市において、このような業績測定は事業を直接担当する部局が実施しているが、さらに業績測定とは別に、当該サービス分野を統括するコミッショナーが、事業から独立した立場より成果指標をモニタリングし、期待される政策目標が達成されたかどうかを検討する事業評価を実施している⁵⁾。

したがって、公共サービス分野においては一概にインパクト評価といっても、プログラムの達成状況を測定するための「業績測定」と、政策の効果をモニタリングして予算の見直しに生かす「事業評価」という二段階の評価の仕組みが導入されている。そして、これらの評価が各々独立した権限に基づいて実施されることにより、プログラム自体の成否と政策的な有効性を検証する体制が整備されている。



出所：筆者作成

図2 イギリスにおけるインパクト評価の構造

5) 事業評価を大学やシンクタンクなどの外部評価機関に依頼し、その報告を受けて、コミッショナーが事業の有効性を判断している場合もある。

4 エビデンスの立証水準

公共サービスにインパクト評価を適用するためには、業績測定と事業評価のいずれにおいても、その成果を測定するためのエビデンスを収集する必要がある。ただし、前者においてはプログラムの達成度を測定するために、事前に合意された成果指標がエビデンスとして用いられるのに対して、後者においてはバリュー・フォー・マネー (VFM) の観点から政策の有効性・効率性・経済性 (3E) を判断するために、社会的・経済的な波及効果を含めた多面的な評価が行われる必要がある。

そして、インパクト評価に際して成果が生み出されたことを合理的に説明するためには、エビデンスが一定水準の立証能力を備えている必要がある。例えば、Social Impact Investment Taskforce (2014, pp.19-20) によれば、エビデンスの水準を高めるためには5つの質的要件(重要性・信頼性・比較可能性・追加性・普遍性)を備えなければならない⁶⁾。このとき、エビデンスは必ずしも定量的なものである必要はなく、上記の5つの要件を備えていれば、定性的なものでも用いることができるとされている。ただし、上述したようにインパクト評価のエビデンスは、プログラムが生み出した変化をプログラムとの因果関係に基づいて立証する必要があるため、現実にはサービス記録やアンケート調査などのデータを収集することにより、定量的な成果指標が用いられる場合が多い。

なお、データの収集方法については標準化されたものがあるわけではないが、表1に示すようにデータの集め方によって、エビデンスがもつ立証能力には大きな差異が生じる。例えば、日本でよく行われているようないわゆる数値評価や、有識者が検討委員会などで意見を述べる方法は、エビデンスの立証能力としては比較的低いと考えられているのに対して、RCT (ランダム化比較試験) により統計的に検証する手法は立証能力が高いものとされており、特にアメリカにおいて広く用いられている (Rangan & Chase 2015)。その一方で、後述するようにイギリスでは、RCTほどではないが中位的な立証能力をもつ、コホート分析や差分の差分法といった研究的な手法が多く用いられている傾向がある。

なお、インパクト評価に用いられるエビデンスは、必ずしも証拠能力が高ければよいというわけではなく、データ作成の手間やコストも勘案しながら、入手可能な範囲を検討する必要がある。エビデンスの入手に過度の負担がかかる場合には、「評価のための評価」に陥ってプロ

6) 重要性 (materiality) は資源提供者や組織の意思決定に影響を与えるデータを用いること、信頼性 (reliability) は高度な誠実性が確保された情報源からデータを得ること、比較可能性 (comparability) は整合性のある基準や手法を用いて他と比較できるデータを作成すること、追加性 (additionality) はその資源提供がなければ生じなかった追加的な効果を示すデータを用いること、普遍性 (universality) は市場や地域、活動分野にかかわらず共通に適用可能な手順によりデータを集めること、を意味する。

表1 エビデンスの立証能力

立証能力	データの入手方法	説明
高	RCT（ランダム化比較試験）	被験者を処置群と対照群へと無作為に振り分け、両グループのアウトカムを統計学的に比較分析する方法
中	差分の差分法	処置群における介入後と介入前の差と、対照群における介入後と介入前の差を比較する方法
	コホート分析	共通の性質をもつ集団を一定期間追跡して観察する方法
	社会実験	場所や期間を限定して、新しい制度や技術などを導入して有効性や問題点を検証する方法
低	前後比較	プログラムの実施前と実施後を比較する方法
	ベンチマーキング	基準点を定め、それに対して成果がどれくらい上昇あるいは低下したかを測定する方法
	有識者の意見	専門家や実務家によって構成される検討委員会を設置し、討議を行う方法

出所：筆者作成

グラム自体を阻害することにもなりかねないため、事業を実施する際に作成される活動記録をデータベース化し、特別な作業を加えなくてもインパクト評価に用いることができるように工夫をしておくことが望ましい。

特に、多くのサービス事業者では、活動記録のデータ自体は作成しているが、サービス対象者に生じた変化を捕捉できる個票データのような形になっていないことが多い。しかし、インパクト評価では、プログラムと成果との間にある因果関係を明確化する必要があるため、サービス対象者にどのような変化が生じているかを、時間軸に沿って測定できるように整理することが重要になる。

5 インパクト評価の事例

ここまで、公共サービス分野に対してインパクト評価を適用する際に考慮すべき点について整理してきたが、本節ではイギリスの事例に基づき、インパクト評価が実務現場でどのように用いられているかを考察してみたい。表2には、イギリスのソーシャル・インパクト・ボンド（以下、SIBという）において用いられている成果指標及びその分析方法などを示している。

このとき誤解されていることも多いが、インパクトとは必ずしも最終成果のみを意味するわけではない。イギリスのSIBの事例をみると、最終成果に至るためのロジックを立てることにより、その途中段階に位置する中間的成果や初期的成果についても、成果指標と認めて報酬が支払われるケースも多く行われている。また、イギリスではRCTのように厳密な成果の測定方法を追求するのではなく、差分の差分法やコホート分析、社会実験的な手法など、柔軟かつ多様な成果指標の分析が試みられている。

SIBは成果連動型契約の一種であり、投資家が拠出した資金を用いて民間のサービス事業者が公共サービスを提供し、一定以上の成果が認められれば、行政から投資家に対して元利金が

表2 ソーシャル・インパクト・ボンドにおける業績測定

実施場所	ヒアリング先	期間	事業内容	分析方法	成果指標	評価機関
ピーターバラ刑務所	The Young Foundation (2012.9.4) Social Finance (2012.9.5, 2015.9.16)	2010~2016年(政策変更により2015年に中止)	出所後の短期受刑者に対して支援・指導のプログラムを提供し、再犯率を低下させる	刑期が1年未満の短期受刑者3,000人を対照群と比較(差分の差分法)	プログラムの実施前後において、サービスを受けない対照群(全国平均データ)の再犯率の変化と、サービスを受ける処置群の再犯率の変化分を比較することにより、プログラムによって再犯率の低減率がどれくらい高められたかを測定する	<コミッション> Ministry of Justice <外部評価機関> 第1コホート: University of Leicester 第2コホート及び最終コホート: National Institute of Economic and Social Research
グレーターロンドン	St Mungo's Broadway (2014.10.30, 2015.9.14)	2012~2015年	短期宿泊施設への入居や就労支援、職業訓練などを実践したチームで行うことにより、路上生活者の自立を支援する	過去2年間に路上生活を6回以上発見され、かつ直近3カ月間に路上生活または一時収容施設で発見されたホームレスから416人を抽出(コホート分析)	(1) コホートにおける路上生活者の減少数 (2) 長期居住先の確保数 (3) 外国人路上生活者が母国で再定住した数 (4) 救急医療の利用削減数 (5) 就労数(フルタイム、パートタイム、ボランティア)として3カ月・6カ月以上継続 (6) 職業訓練参加者数	<コミッション> Department for Communities and Local Government/Greater London Authority <外部評価機関> なし(Greater London Authorityが成果指標を監査する)
グレーターマンチェスター	Teens and Toddlers (2016.9.19)	2012~2015年	将来、ニートになるリスクが高い若年者をターゲットとして教育支援を行い、失業者になることを防止する	教育省や国家統計局のデータに基づき10代の妊娠比率や、キヤンクやニートになる比率が高い地域より対象校を選定し、14~16歳の青少年1,150人以上を抽出(コホート分析)	問題行動の改善、学校の出席状況の改善、学業成績の向上、雇用機会の拡大などの指標を設定	<コミッション> Department for Work and Pensions (DWP) <外部評価機関> National Centre for Social Research/Insite Research and Consulting
エセックス市	Essex County Council (2013.11.1) Action for Children (2013.10.29)	2013~2018年	施設入所あるいは保護観察になる可能性が高い児童を対象に予防的支援を提供し、施設入所日数を削減する	380人の児童を20グループに分ける(コホート分析)	SIBが開始する前に集計された過去30カ月650件のケースと、MST(マルチ・システム・ミック・セラビリティ)サービスを受けた児童の4半期毎の平均施設入所日数を比較する	<コミッション> Essex County Council <外部評価機関> Office for Public Management (OPM)
イギリス全国	It's All About Me (2015.9.14)	2013~2023年	通常の方法では養子縁組先を見つけれない子どもたちに適切な家庭を見つけ、両親となる人にも支援と訓練を行う	年齢、人種、民族、障害、育児放棄などのトラウマがあり、養子縁組先を6カ月以上探している子どもたちを選定(社会実験)	(1) 養子縁組が必要な子どもの登録数 (2) 養子縁組の成数 (3) 1年後も養子縁組関係が壊れている数 (4) 2年後も養子縁組関係が壊れている数	<コミッション> 各地方自治体 <外部評価機関> 子どもたちの健康・心理面: The South London and Maudsley Hospital, Centre for Children 社会的インパクト及び組織と市場の変化: City, University of London, Cass Business School
ニューカッスル	Ways to Wellness (2015.9.11)	2015~2022年	地域社会への結びつきや孤立解消が健康と福祉の増進につぎ、長期的な健康関連のサービスを提供する	処置群であるニューカッスル西部地域と対照群であるニューカッスル東部地域との比較(差分の差分法)	(1) ツール(Wellbeing Star)を用いて患者の健康・福祉に関する8分野の改善を測定する (2) 処置群が利用した二次医療のデータと、比較群的に東部地域に居住する対照群が利用した二次医療のデータを比較する	<コミッション> NHS Newcastle West Clinical Commissioning Group (CCG) <外部評価機関> Newcastle University

出所: 筆者作成

支払われる。そのため、SIBの業績測定は入札時の仕様によって規定された成果指標をサービス事業者が測定し、行政の事業担当部局に報告することによって実施される。

ただし、成果指標をサービス事業者内に所属する直接の事業担当者が測定すると、成果をより高くみせようとして、クリームスキミング（成果が出やすい対象者だけを集める）やチェリーピッキング（自分たちにとって都合のよい成果だけを抽出する）が生じるリスクがある。そこで、通常は入札時に条件を定めることにより、図2にみられるようにサービス事業者に対して、データの管理や分析に関する専門スキルをもつ業績管理責任者（パフォーマンス・マネージャー）を雇用あるいは契約することが求められる。それにより、成果指標はサービス事業者自身が測定するものの、プログラムには直接携わらない内部的に独立した業績管理責任者によって、エビデンスが収集される体制を整備している。

日本でも現在、SIBが実際に導入され始めているが、エビデンスの透明性を確保するために、基本的には業績測定を外部評価機関に任せるべきという考え方がある（伊藤ほか 2016, p.46）。ただし、イギリスの事例では、業績測定を担うのは原則としてサービス事業者であり、自らが整備したガバナンスや内部統制に基づいて、一定水準の能力と独立性を有する業績管理責任者が成果指標の測定を行っている⁷⁾。そのため、エビデンスの透明性を確保するのはサービス事業者自身の責任であり、外部評価機関の役割はプロジェクトの有効性や効率性を測定する事業評価に重点が置かれている。

したがって、最初にSIBが導入されたピーターバラ刑務所のケースでは、外部評価機関による報告書（Anders & Dorsett 2017）に基づいて支払いの可否が判断されているが、その後に実施されたエセックス市のケースでは、外部評価機関がプロジェクトによって直接的に生み出した成果指標だけではなく、VFMの観点からみた政策スキームの有効性や費用対効果などを包括的に分析している（OPM 2014）。そして、このような外部評価は今後の施策や予算策定に対する判断材料として活用されるものであり、外部評価機関による評価結果が、サービス事業者に対する報酬の支払基準として用いられることは、原則として行われえないということであった⁸⁾。

なお、SIBの外部評価機関は必ずしも専門の評価機関に限られておらず、大学等が研究的な

7) 初めてSIBを実施するサービス事業者が自力で成果指標を収集することは容易ではないため、実際には中間支援組織等がデータ作成の支援などを行い、経験を積むに従ってサービス事業者が直接担う部分を増やしていくケースが多いと考えられる。なお、日本において厚生労働省が2018年度に実施する「保健福祉分野における民間活力を活用した社会的事業の開発・普及のための環境整備事業」に関する成果運動型契約では、事業者が測定した成果を検証する「第三者評価者」と、評価の枠組み及び支払いの妥当性を検討する「評価・運営機関」が設定されており、試行的にはあるが、概ね本稿が提示する業績測定と事業評価に沿った仕組みが導入されていると考えられる。

8) 2017年8月17日に、エセックス市など様々なSIBプロジェクトの外部評価を担当しているOPM（Office for Public Service）のChih Hoong Sin氏（事業開発マネージャー）に実施したヒアリングに基づく。

知見を用いながら取り組んでいるケースも多くある⁹⁾。また、イギリスにおいてもすべてのSIBに対して外部評価機関が設けられているわけではなく、グレーターロンドンの事例のように、行政内において当該サービス分野を統括するコミッショナーが、事業評価の役割を担っている場合もある。

6 結論と今後の課題

成果を単なる数値指標ではなく、プログラムが生み出した変化分として測定するインパクト評価は、公共サービスのあり方に大きな変革とイノベーションをもたらす可能性がある。特に、エビデンスに基づいてプログラムと成果との間に存在する因果関係を立証することができれば、有効かつ効率的な予算策定に資することが期待される。

ただし、その一方でMcHugh et al. (2013, p.249) が指摘するように、公共サービスには定型的な成果指標が存在せず、社会的な成果を測定することは非常に困難な作業であるにもかかわらず、インパクト評価は過度にモデルを単純化し、プログラムが簡単に成果をもたらすかのような印象を与えている、といった批判もある。また、諸外国で導入されている成果連動型契約についても、その結果として想定される財政削減効果は仮想的 (virtual) なものであり、人員や施設等に対する現実のリストラを伴わなければ、逆に財政支出が増大してしまうリスクもある。そのため、日本の制度下においてインパクト評価が有効に機能するかどうかは、今後のケースを積み重ねながら慎重に検討を加えていく必要がある。

さらに、業績測定と事業評価を混同したままインパクト評価が導入されると、適切な成果の測定が行われぬ可能性があることにも留意すべきである。なぜなら、業績測定はあくまでも入札時に設定された成果目標が達成されたかどうかを測定すべきものであるが、そこに事業評価の多面的な視点が加わってしまうと、評価の軸が歪められて何が成功要因や阻害要因となっているか不明確になるからである。そのため、成果連動型契約の支払基準となる業績測定は、事業評価とは切り離して事業担当部局が監視し、事業評価は事業の実施体制から独立した立場にある、コミッショナーなどの監督機関あるいは外部評価機関に実施させることがイギリスでは試みられている。

最後に、インパクト評価に要する人材とコストをどのように整備するかということも課題になる。インパクト評価が適切に活用されれば、公共サービスの質を維持・向上させながら財政削減に結びつけることも期待できるが、特にインパクト評価の導入当初には、適切なエビデン

9) 大学の役割は外部評価機関を務めることに限られず、Harvard大学 (Kennedy School, Government Performance Lab) は多数の地方自治体にSIBの支援要員として研究員を派遣しており、またオックスフォード大学 (Blavatnik School of Government, Government Outcomes Lab) は英国政府の委託を請けてSIBの導入支援を行っている。

スを選定し、データを収集するシステムを構築するための手間と費用がサービス事業者の負担となる。また、日本ではエビデンスを収集・分析できる能力を有した業績管理責任者も著しく不足していることから、インパクト評価に関わる人材の育成や、導入に向けた技術的な支援を提供することが、インパクト評価を機能させるためには重要となる。

謝辞

本稿は2017年度関西大学在外研究、JSPS科研費16K04021（研究代表者：馬場英朗）、文部科学省私立大学戦略的研究基盤形成支援事業（研究代表者：塚本一郎）の助成による研究成果の一部である。

参考文献

- Anders, J. & Dorsett, R. (2017) "A Review of the Matching Process for the Impact Analysis of the HMP Peterborough Social Impact Bond", National Institute of Economic and Social Research.
- Burnett, M. (2009) "Using Competitive Dialogue in EU Public Procurement: Early Trends and Future Developments", *EIPASCOPE*, 2009/2, pp.17-23.
- 家子直幸・小林庸平・松岡夏子・西尾真治 (2016) 「エビデンスで変わる政策形成—イギリスにおける「エビデンスに基づく政策」の動向, ランダム化比較試験による実証, 及び日本への示唆」三菱UFJリサーチ&コンサルティング政策研究レポート。
- 伊藤健・落合千華・幸地正樹 (2016) 「日本版ヘルスケアソーシャル・インパクト・ボンドの基本的な考え方」経済産業省。
- 北大路信郷 (2011) 「自治体における公共調達改革の課題—指定管理者制度活用のために」『ガバナンス研究』no.7, pp.21-54。
- 小関隆志・馬場英朗 (2016) 「インパクト評価の概念的整理とSROIの意義」『ノンプロフィット・レビュー』vol.16, no.1, pp.5-14。
- McHugh, N.; Sinclair, S.; Roy, M.; Huckfield, L. & Donaldson, C. (2013) "Social Impact Bonds: A Wolf in Sheep's Clothing?", *Journal of Poverty and Social Justice*, vol.21, no.3, pp.247-257.
- MDRC (2013) "Financing Promising Evidence-based Programs: Early Lessons from the New York City Social Impact Bond".
- 内閣府 (2016) 「社会的インパクト評価の推進に向けて—社会的課題解決に向けた社会的インパクト評価の基本的概念と今後の対応策について」社会的インパクト評価検討ワーキング・グループ。
- OPM (2014) "Evaluation of the Essex Multi-Systemic Therapy Social Impact Bond: Interim Evaluation Report, Version 2".
- Rangan, V. K. & Chase, L. A. (2015) "The Payoff of Pay-for-Success", *Stanford Social Innovation Review*, Fall 2015, pp.27-39.
- Social Impact Investment Taskforce (2014) "Measuring Impact", Impact Measurement Working Group.
- 田辺智子 (2014) 「業績測定を補完するプログラム評価の役割—米国GPRAMAの事例をもとに」『日本評価研究』vol.14, no.2, pp.1-16。
- 上野真城子 (2015) 「オバマ政権の社会政策における挑戦とCBO 40年の意味するもの」アジア都市コミュニティー研究センター。