

評価研究における定量／定性手法の統合へのアプローチ

橋本 圭多

概要

本論文では、評価研究における定量的手法と定性的手法の方法論について、方法論争から手法の統合へと至る経緯について検討している。評価研究の分野では、1970年代から1980年代にかけて定性的手法の側による定量的手法の側への批判が行われ、これらは定量／定性論争として整理されてきた。両手法の違いはそれぞれ実証主義と構成主義というパラダイムの対立として理解されてきた。社会調査をはじめとする評価研究の関連分野では、パラダイム間の相違を克服するために評価デザインや評価手法のレベルでの統合が模索されてきた。評価実務では、環境的、人的、金銭的、時間的制約を克服するために、定量的手法と定性的手法とを問わず複数の手法による評価が行われている。評価手法を選択する際には、これらの制約条件によって評価デザインが最適化されることになる。一方で、日本の政策評価制度は法律の中で定量的手法の重視を規定しており、客観性や成果志向の強調にみられるように定量的手法のパラダイムが優勢となっていると考えられる。定量的手法と定性的手法の統合に向けた評価研究のアプローチは、さまざまな制約条件が課せられる評価実務を進めていく上で、学術研究における方法論争とは異なる意義を有している。本論文では、定量／定性論争や手法の統合そして評価実務での事例を参照することで、評価手法をめぐる議論の全体像を明らかにすることを試みる。

1. はじめに

公共部門ではプログラム評価や業績測定と

いった評価が行われており、またそれらは行政の中のさまざまなアクター間で多元的に実施されている。ただし、アクターによって評価を行う目的や動機は異なっており、それが評価に対する不満や期待の温度差として現れる。日本の政策評価や行政評価における「評価疲れ」の問題はその典型である。

たとえば日本の男女共同参画政策では、女性センターの運営を担う指定管理者に対して地方自治体が評価を行っている。ここでは所管課が個別に行う評価のほか、総務課によるレビューや外部有識者委員会による外部評価などさまざまな評価が行われている。他方、実際にプログラムの実施に携わる指定管理者にとって、地方自治体が行う評価は指定管理者の次回選定を左右するものとして消極的に受け止められる。指定管理者が求めるのはむしろ、評価するプロセスを通じて現状の問題点を理解したり多様なステークホルダーの意見を把握したりするなど、プログラムの改善に資する評価結果を産出することである。評価研究では、前者はアカウンタビリティを目的とする総括的評価、後者は学習やプログラム改善を目的とする形成的評価として理解されている。

このように、評価の活用に対する理解が異なる原因として、定量／定性手法という評価の方法論に対する認識の相違を指摘することができる。総括的評価はおもに定量的手法によって、形成的評価はおもに定性的手法によって展開されてきた。応用社会科学として発展した評価研究においては、定量的手法と定性的手法とではその背景やルーツが異なっており、両者の方式の相違をいかにして克服するかがこれまでに議論されている。

地方自治体―指定管理者、あるいは中央政府

一地方自治体など、評価をする側とされる側に距離感や権力関係が存在する場合、評価は管理統制や意思決定の正当化など権力関係を保存する手段として用いられる恐れがある。その場合、評価をされる側のプログラム実施者が評価によるプログラム改善を期待しても裏切られることになる。両者のあいだに共通言語を構築するには、まず両者がそれぞれ認識する定量的手法と定性的手法の方式を理解する必要がある。

そこで本稿では、評価研究の分野における方法論争を参照し、それぞれで用いられる評価のデザインや手法を整理する。評価研究の分野では、評価が純粋な学術的調査とは異なることで生じる問題点についてこれまでに議論されている。政治との関係や科学的／実用的評価の対立は、評価が価値判断を求められていることから生じる問題である。1970年代はプログラム評価や定量的手法の隆盛を極めた時期であるが、1980年代以降はそれに対する批判や反動としてさまざまな定性的手法が提唱されるようになる。

2. 評価手法の変遷

2.1 第4世代評価

20世紀初頭から歴史を有するプログラム評価は、応用社会科学として教育学、心理学、公衆衛生など多様な学術領域における活動として行われてきた。プログラム評価の最盛期は、一般的にアメリカの会計検査院であるGAOがプログラム評価を採用した1970年代と言われているが、そこに至るまでやその後の評価のあり方は時代によって異なっている。そうした中で、定量的手法と定性的手法が依拠するパラダイム観も形成されていったのである。

グーバとリンカーンは、評価の時代を4つに区分し、当時における評価の趨勢を「第4世代評価」として定義した（Guba and Lincoln 1989）。グーバらは、第1世代を「測定」の時代、第2世代を「記述」の時代、第3世代を「判断」の時代とした上で、それぞれの時代における教育評価の特徴を述べている。

第1世代の「測定」の時代は、アメリカの学校教育におけるテストの普及を指している。テ

ストの目的は、生徒が授業内容を習熟しているかどうかを確認することにある。当時のテストに測定が導入されていったのは、ライス（Joseph Mayey Rice）によるスペリングの習得における学習時間と結果の関係に関する研究、ビネット（Alfred Binet）による知的発達の遅れた子どもを選別するための計量心理学的測定技術の開発、アメリカ心理学会の協力による第一次世界大戦徴兵時の精神テストの開発と学校での採用など、学校教育に関する多くの場面でテスト方法の開発と精緻化が行われたことが背景にある（Guba and Lincoln 1989: 23-4）。

グーバとリンカーンは、学校教育においてテストが採用されていった文脈的要因として、当時の科学的管理法の影響を指摘している（Guba and Lincoln 1989: 25-6）。テイラー・システムに代表されるように、科学的管理法は工場労働における人的作業の効率化を追求する手法である。科学的管理法の中核は動作時間研究（motion-time study）と作業の標準化（standardization）を基礎とした課業管理であり、作業計測の発想は産業工学の基礎として、標準化の原理は人事管理上の職務分類や財務管理上の標準コスト概念として広く適用された（西尾 2001: 35-6）。学校においても、校長の管理する学校という「工場」において生徒は処理される「原料」とであるとする科学的管理法の考え方が広く浸透していた（Guba and Lincoln 1989: 25-6）。

第2世代の「記述」の時代は、高等学校における新旧カリキュラムの長所と短所を記述する作業として現れた。第一次世界大戦の直後、アメリカの高等学校には初等学校程度の知識にとどまる学生が入学してきたためにカリキュラムの変更が検討されたが、それにより大学の標準的なカリキュラムをこなすための訓練が不十分な卒業生を受け入れざるを得なくなることを大学は恐れていた（Guba and Lincoln 1989: 27）。この大学側の懸念が妥当なものであるかを調べるために行われたのが「8年研究」と呼ばれる調査研究である。

1933年にはじまった8年研究では、高等学校の4年間と大学の4年間にわたる計8年間の修学状況を対象に調査が行われた。8年研究はアメリカの進歩主義教育協会が主導した実験的な研究であり、日本の指導要領や教育実践にも

多大な影響を与えている（浅沼 2010: 1-4）。この研究では、新旧カリキュラムにおいて定められた目標がどの程度達成されているのかについて情報を収集し、カリキュラムの強みと弱みを記述することが行われた（Guba and Lincoln 1989: 28）。

第3世代の「判断」の時代は、評価者が測定や記述のみならず判断を行うことが求められた。旧ソ連が1957年に世界で初めて人工衛星の打ち上げに成功したことを受けて、アメリカ連邦政府では全米科学財団や教育局のプロジェクトを対象に行われる評価が不十分であると考えられるようになったのである（Guba and Lincoln 1989: 29）。1960年代には、ケネディ政権で構想された「ニューフロンティア」、ジョンソン政権による「偉大な社会」「貧困との戦い」などを受け、教育、医療、保健などの分野で社会プログラムが実施されることになった。また、1967年の経済機会法修正によって、GAOが貧困プログラムの有効性に関してプログラム評価を行うようになった（渡瀬 2005）。さまざまな場面で評価への需要を促進され、評価者にはプログラムの判断を求められるようになったのである。

評価の妥当性は、それが科学的根拠に基づく客観的な判断であるかどうか依存すると考えられていた。キャンベルは1969年の論文で、実験的アプローチに基づいてプログラムの効果を検証することを提案している（Campbell 1969）。それに対して、クロンバックは1982年の論稿で、評価はアートであり科学的調査とは異なることを指摘している（Cronbach 1982）。ロッシらは、両者の主張を科学的評価と実用的評価の姿勢として対比的に整理している。「この両方の意見に賛成したいと思う人は多いだろう。すなわち、評価は科学研究の高い水準を満たすべきであり、同時にプログラムの意思決定者の情報ニーズに仕えるべきである。ところが、現実には、この2つの目標はしばしば両立させにくいという点が問題なのである。（中略）評価者は、知見の妥当性を確保するための手続きと、その知見をコンシューマーにとってタイムリーで、意味があり、有用なものとするための手続きとの間に、到達可能な均衡点をみつけだしていかなければならなくなる」（Rossi, Lipsey, and Freeman 2004: 23-5=2005: 23-5）。

グーバとリンカーンが提唱する第4世代評価もまた、評価の妥当性をめぐる議論から導き出されている。第4世代評価では、以下のようにこれまでと異なる想定が行われている（Lincoln 2005: 161-2）。第一に、第4世代評価では触知可能な現実のみならず、精神、価値、信念、意味の理解など社会心理学的な構成物を前提とする。第二に、評価者とステークホルダーとの関係は相互作用的で認識論的な実践として成立しており、両者は情報を十分に共有することができる。第三に、第4世代評価では用いられる手法の種類が拡大しており、統計モデルや数理モデルに加えて、定性的な手法を用いてデータの収集や分析を行う。第四に、第4世代評価では価値について考慮しており、人間の認知や意味の理解にとって不可避の問題であるとしている。

なお、キャンベルに対するクロンバックやグーバとリンカーンの立場をとったとしても、一般化すなわち当該プログラムに対する評価結果が他の状況においても同様の結果を導き出すことは可能である。佐々木は、キャンベル、クロンバック、グーバとリンカーンの議論についてそれぞれ一般化の方法が異なることを指摘している（佐々木 2010）。その理由は、一般化の過程で母集団をどこまで考慮するのがそれぞれ異なるからである。

佐々木によれば、キャンベルによる母集団の考え方は2つあり、ひとつがサンプルを引き出してきた母集団、もうひとつが前者の母集団を含む無限の母集団である（佐々木 2010: 78）。キャンベルの考えでは、サンプルに対する実験結果はサンプルを引き出してきた母集団に対して一般化可能であり、さらには一般法則に則することで無限の母集団にも一般化することが可能となる。クロンバックの場合、この一般化を政策対象となる母集団に対して行う（佐々木 2010: 79）。サンプルを引き出してきた母集団と政策対象となる母集団の特徴が類似していれば、両者はともに無限の母集団に属しているため、サンプルから政策対象となる母集団に対して推定法による一般化が可能となる。

それに対して、グーバとリンカーンによる一般化の考え方ではそもそも無限の母集団を考慮しない。人は各自が関わっている小集団に対してすでに一定の理論を有しており、他の小集団

を見る際にはそれにあわせて自らの理論を修正することが可能である（佐々木 2010: 81）。他の小集団を対象とした評価報告書を読むことは「追体験」（佐々木 2010: 80）をすることであり、その過程で自らの理論が修正され一般化されることになる。

このような一般化の特徴から、第4世代評価では認識論の多様性を指摘することができる。つまり、人種、民族、社会的性差、身体の健全さ、性的志向、言語の情勢といった特徴によって対象を認識し事実を構成するのである（Lincoln 2005: 162）。この指摘は、評価研究における定性的手法の構成主義的な特徴を示しており、このような認識論的傾向は質的社会調査として行われる調査研究にも現れている。

グーバとリンカーンの主張は、それ自体が特定の立場を形成しているものの、評価のあり方や考え方の多様性を示す上で重要な示唆を含んでいる。第4世代評価の考え方はその後、参加型評価の登場として現れることになる。

2.2 参加型評価

評価研究における定性的手法の隆盛とともに提唱されるようになったのが参加型評価（participatory evaluation）である。参加型評価にはさまざまな手法が存在するが、大きく分けて実用的参加型評価（practical participatory evaluation）と変革的参加型評価（transformative participatory evaluation）の2つに区別することができる（Cousins and Whitmore 1998）。両者は歴史的出自が異なるため、同じ参加型評価でも評価の目的には差異がある。実用的参加型評価では多くのステークホルダーにとって有用な評価結果を産出することが求められ、変革的参加型評価ではステークホルダーのエンパワメントや解放あるいは政治的社会的変化に焦点を置く。

参加型評価の手法は、アクションリサーチと呼ばれる調査手法と多くの共通点を有している。アクションリサーチは、調査活動を通じて対象に働きかけを行うことで実際の成果を生み出す手法である。アクションリサーチでは、調査者は問題解決への支援者であり、調査結果は調査に関わった人びとによる利用を想定しており、調査デザインは調査を進めながら組

み立てられていく（Merriam and Simpson 2000: 122=2010: 140）。アクションリサーチは、以下の点で従来の手法とは異なる特徴を有している（Merriam and Simpson 2000: 122-3=2010: 141）。

1. ある特定の状況（たとえば、教室、学校、社会的機関、地域社会など）に対して、直接に適用可能な知識の獲得を目的としている。
2. 調査上の問題は、効率的な教授方法やある地域に特有の汚染問題など、調査者を悩ませるような事柄から現れてくる。
3. 問題提起は一般的なかたちで示される。仮説はほとんど用いられない。
4. 一次的な資料よりも二次的な文献資料の方が、広く用いられる。調査者がとくに求めているのは、研究対象の現象に関するアイディアである。
5. 調査参加者は、体系的に抽出あるいは選択されたりはしない。調査参加者は、人間活動の自然な「流れ」の一部である。
6. 研究遂行の手続きは、研究の当初にはただ一般的に計画され、必要に応じて調査進行のなかで変更される。
7. 調査を行う際に、統制や実験条件にはほとんど注意が払われない。

さらに、参加型調査（participatory research）と呼ばれる調査手法は、アクションリサーチと比べて変革や社会正義といった点が強調されるようになる。参加型調査では、調査対象者が調査活動への参加を通じて知識を獲得し、自らのエンパワメントや社会変革を実現していくことが想定されている。参加型調査の手法は「社会科学の分野でよくみられる、高度に実証的で演繹的な調査法に対する反動」（Merriam and Simpson 2000: 126=2010: 144-5）である。参加型調査はソーシャルワークや公衆衛生学、開発学、社会学、人類学、看護学、教育学、コミュニティ心理学といった多様な分野において、エンパワメントやソーシャルアクション、アドボカシー、社会変革のための手法として関心が高まっている（武田 2015: 2）。

これらのアクションリサーチや参加型調査は

参加型評価の手法として用いられており、実用的参加型評価と変革的参加型評価をそれぞれ特徴づけているといえる。評価研究における定性的手法や参加型評価手法は、評価者による評価対象への接近、そして評価対象への働きかけを強調している点で、従来型の評価とは異なっている。こうした定性的手法や参加型評価の登場によって、定量的手法と定性的手法とのあいだで論争が行われることになる。

3. パラダイムの対立と手法の統合

3.1 定量／定性論争

グリーンとヘンリーは、評価をめぐる定量／定性論争について以下のとおり解説している (Greene and Henry 2005: 345-50)。定量／定性論争は科学哲学の世界で長年議論されており、評価をめぐっては同様の議論が1970年代から1980年代にかけて行われた。両者はそれぞれ異なるパラダイムに属していると考えられている。定量的パラダイムでは実証主義、すなわち正当な根拠が存在し、信頼でき、追試可能で、実験的にデザインされる客観的な手法が用いられる。このパラダイムの成功は物理学や医学において確認することができる。それに対して、定性的パラダイムでは構成主義、すなわち文脈的で価値依存のかつ偶然的な社会知識が調査の対象となる。応用社会学者として評価者の多くがこの論争に関わり双方への批判を展開した。

定量／定性論争には主要な要点が3つ存在する (Greene and Henry 2005: 346)。すなわち、哲学、実践、評価者の役割である。各要点において定量的手法と定性的手法は異なる姿勢をとっている。以下では、それぞれの姿勢について対比的に確認したい。

第一の哲学について、定量的手法の姿勢は演繹法と独立性の2つによって特徴づけられる (Greene and Henry 2005: 346-7)。演繹法では、はじめに理論から導き出された予測や仮説が求められる。評価にとってこのことは、政策の根底にある理論をもとにして、プログラム効果の予測が明らかにされることを意味している。この予測は評価の開始に先行することになる。演

繹法に代わるのが帰納法であり、この場合は観察から理論や予測を作り出すことになる。帰納法では一連の観察を説明するために理論が考案されることになる。したがって、定量的手法を採用する評価者は、理論が検証されない帰納法のアプローチに対して懐疑的である。もう一つの独立性は、プログラムや評価スポンサー、プログラム参加者からの距離感を意味している。独立性を確保するには、評価者が科学的根拠をふまえて厳格な判断を行うことが求められる。ここでは事実と価値を区別することが重要となる。

定量的手法の姿勢に対して、定性的手法の姿勢は、理論や仮説検証、演繹法といった定量的手法のモデルを採用していない。定性的手法では、事実と価値を二分する立場をとっておらず、独立性や客観性の確保は求められていない。すなわち「社会学者が生み出す知識は価値自由ではあり得ない。社会科学において『純粋な事実』は存在しない」 (Greene and Henry 2005: 347)。

第二の実践について、定量的手法では、社会プログラムの意図がアウトカムとして対象集団に対して反映されているかを確認することに重点がおかれる。「定量的手法を用いる評価者は、意図されたプログラムアウトカムの尺度を選択することや、プログラムの対象集団へのアウトカムの変化がプログラムにどの程度起因しているかを評価することに焦点をあてる傾向がある」 (Greene and Henry 2005: 348)。それに対して、定性的手法ではアウトカムを強調するのではなく、対象への理解に重点をおく。「プログラムの経験に関する特徴や本来の性質を理解することを強く主張することは、プログラムの関与が参加者の実生活にどのような違いを生じさせるのかを理解することと同じくらい重要である」 (Greene and Henry 2005: 348)。

第三の評価者の役割について、定量的手法と定性的手法では明確に異なる。定量的手法では評価者は「中立的なブローカー」であるのに対し、定性的手法では「政策アリーナ内」に位置づけられる (Greene and Henry 2005: 349)。定量的手法を用いる評価者にはプログラムに関する科学的根拠を提供することが期待されており、それらの情報はより有効なプログラムの採用を促すことになる。定性的手法を用いる評価

者にはプログラムの実施に伴う経験や困難について現場から情報を得ることが期待されており、それらの情報は政策決定よりむしろ現場レベルでの改善や市民への教育を促すことになる。このように、定量的手法と定性的手法とのあいだには特徴的な差異がある。

ライハルトとクックは、定性的パラダイムと定量的パラダイムの特性を対照的に整理している（表1）。定性的パラダイムは現象学と理解社会学に基づいており、アクター自身が有している準拠枠（frame of reference）から人間行動を理解することに関心を有している。そこでは自然的でかつ統制されない観察によって調査が行われる。調査は主観的に進められ、評価者はデータへ接近することで内部者として振る舞うことになる。グラウンデッドすなわちデータに根ざして調査や理論形成が進められることになり、発見志向で探検的、拡張主義的、記述的、帰納的なアプローチがとられる。調査はプロセス志向で行われる。データの収集では妥当性が重視され、現実を反映し内容が豊富で奥行きのあるデータが求められることになる。単一事例研究となるため、一般化ができないと考えられている。調査では対象を全体的に明らかにする傾向にある。定性的パラダイムでは現実を動態的なものとして想定している。

それに対して、定量的パラダイムは論理実証

主義に基づいており、主観的な状態や個人に対して関心を持つことはほとんどなく、社会現象の事実と原因を探究することに関心を有している。そこでは強制的に設けられかつ統制された環境における測定として調査が行われる。調査は客観的に進められ、評価者はデータから距離を置くことで外部者として振る舞うことになる。非グラウンデッドすなわち個々のデータに焦点をあてることはなく、検証志向で確認的、還元主義的、推論的、仮説演繹的なアプローチがとられる。調査はアウトカム志向で行われる。データの収集では信頼性が重視され、信頼ができ追試が可能なデータが求められることになる。複数事例研究となるため、一般化することができると考えられている。定量的パラダイムでは現実を安定的なものとして想定している。

ロッシもまた、定量／定性論争の背景を構成する2つの領域を提示している（Rossi 1994）。ひとつが玄人による評価（connoisseurial evaluation）であり、「評価活動は玄人の判断を提供することを軸として展開し、通常は小規模プログラムを対象に、適時かつ安価で行われる」（Rossi 1994: 33）。もうひとつが最終的なアウトカムの評価（net-outcome evaluation）であり、「評価活動はプログラムの有効性と効率性に関する見積もりを確立することを狙いとし、通常はかなり大規模のプログラムに適用される」（Rossi

表1 定性的／定量的パラダイムの特性

定性的パラダイム	定量的パラダイム
定性的手法の使用を支持	定量的手法の使用を支持
現象学と理解社会学 「アクター自身の準拠枠から人間行動を理解することに関心がある」	論理実証主義 「主観的な状態や個人にほとんど関心がなく社会現象の事実と原因を探究する」
自然的で統制されない観察	強制的で統制された測定
主観的	客観的
データへの接近、「内部者」の視点	データとの距離、「外部者」の視点
グラウンデッド、発見志向、探検的、拡張主義的、記述的、帰納的	非グラウンデッド、検証志向、確認的、還元主義的、推論的、仮説演繹的
プロセス志向	アウトカム志向
妥当性、「現実的で」「豊かで」「深い」データ	信頼性、「信頼のできる」追試可能なデータ
一般化できない、単一事例研究	一般化できる、複数事例研究
全体的	個別的
動態的な現実を想定	安定的な現実を想定

出所：Reichardt and Cook（1979: 10）

1994: 33)。それぞれの評価からは、前者が定性的手法、後者が定量的手法の特徴を含意していることがわかる。

ヘドリックは定量／定性論争について両者の統合の可能性を検討している (Hedrick 1994)。ヘドリックは両者をパラダイム、デザインアプローチ、手法の観点からそれぞれ区別している (表2)。グリーンとヘンリーの主張と同じく、ヘドリックの主張は両者のパラダイムを実証主義と構成主義に区別しているが、統合の観点から示唆を与えている。すなわち「もっとも広範なパラダイムのレベルでは、アプローチが相互排他的になる傾向にある。私たちがデザインアプローチや手法レベルの考察へと移るにつれて、相補的なやり方で両方が用いられるアプローチを想像しやすくなる」(Hedrick 1994: 49)。パラダイムのレベルでは難しくても、定量的手法と定性的手法が下位のレベルでは相容れる可能性がある。

定量的手法と定性的手法は、パラダイムレベルでは両者はそれぞれ対照的な特徴を有していた。定量的手法と定性的手法は、パラダイム間の相違を克服するために、評価のデザインや手法のレベルでの統合を模索してきたのである。

3.2 手法の統合

定量／定性論争では定量的手法と定性的手法の対立やパラダイムの相違が明らかにされたが、これらを強調しすぎることは両者の相互補完的な利用を妨げる可能性があるために適切ではない。定量的手法と定性的手法との統合につ

いては、社会調査法の分野で議論されている。

パンチは、定量的なアプローチと定性的なアプローチそれぞれの差異がステレオタイプの的に強調されており、両者の基底にある論理の類似性が覆い隠されていることを指摘している (Punch 1998=2005)。「アプローチ (量的か質的か) と目的 (たとえば理論検証であるとか理論生成など) とは相関するが、こうした相関は、完全な相関でも必要なものでもない。量的調査は多くの場合理論のテストのために用いられるであろうが、領域を探索し仮説と理論を生み出すために用いられても構わない。同様に、質的調査は、理論生成に一番向いてはいるが、仮説や理論を検証するために用いられてもまったく問題はない」(Punch 1998: 240=2005: 329-30)のである。両手法は、調査の目的に応じて適宜利用されることになる。

パンチは、ステレオタイプ化された2つのアプローチの区別について、ハマーズレイが示した二項対立図式のうち5つに言及している。すなわち、質的データ／量的データ、自然な状況の調査／人工的な状況の調査、意味への焦点づけ／行動への焦点づけ、帰納法／演繹法、文化型の確認／科学的法則の探究である (Hammersley 1992; Punch 1998: 240=2005: 330)。この図式は両極的なものではなく分布に幅があり、「それぞれの面での位置づけ方の選択は、哲学的考慮よりも、調査目的や調査環境に依存」(Punch 1998: 240=2005: 330-1)するのである。

手法の統合は、評価研究においても指摘されている。イアンニとオールは、評価者が定量的手法と定性的手法のそれぞれから多数の調査技

表2 パラダイム、デザインアプローチ、手法の例

用語	定量的	定性的
パラダイム	実証主義 (科学的手法) アプローチ	構成主義 (第4世代) アプローチ
デザイン	実験的 準実験的 代表サンプル 事例研究	解釈的で弁証的 パターンマッチング 事例研究
手法	本人へのインタビュー (構造化) 質問票 観察記録 行政記録	本人へのインタビュー (非構造化) フォーカスグループ 厚い記述 観察記録

出所: Hedrick (1994: 47)

術を引き出して利用することが可能であると指摘している (Ianni and Orr 1979: 93)。この調査技術には、対象への関与が最小限であるもの(非影響測定法や保管した記録のレビュー)から適度に関与するもの(尺度、テスト、サーベイの利用)そして参加者との積極的な関わり合いを必要とするもの(観察やインタビュー)までさまざまなものが用いられる (Ianni and Orr 1979: 93)。「ひとつの技術が特定の評価プロジェクトにぴったりと適合することはおそらくなく、なぜなら科学的探求に関するすべての基準が満たされることはあり得ないからである」(Ianni and Orr 1979: 93)。

それでは、定量的手法と定性的手法の選択や比重の置き方ではどのような点を考慮すべきであろうか。パンチの議論に戻ると、6つの点が考慮される必要があると指摘されている。第一に、つねに問題に立ち戻って本当に見つけ出したいことは何かを自問すること(問題が異なれば答えを出す方法も異なってくる)、第二に、定量的手法と定性的手法のアプローチの概略に立ち返ること(標準化されたシステマティックな比較を行うのか、それとも対象の全体像を詳細に明らかにするのか)、第三に、既存の調査文献の中から示唆を得ることができるか(既存文献中の定量的または定性的手法が自らの研究に与える影響)、第四に、資源の問題などに関する実用的な考慮(時間、資金、利用可能なサンプルやデータ、調査者の予備知識、状況へのアクセス、周囲の協力など)、第五に、知識のペイオフ(どちらのアプローチの方がより多くの有用な知識を得られるのかを調査費用と利益の観点から考慮)、第六に、スタイルの問題(どちらのアプローチを好むのかという哲学的問題や当人の潜在的気質の問題)である (Punch 1998: 244-5=2005: 335-8)。手法から考えるのではなく、そもそもの調査目的は何かという点に立ち返って、いずれの手法を用いるのかを検討することが求められる。「問題がどのようにして問われているかは、それに答えるために必要とされることが何であるかということに、影響する」(Punch 1998: 245=2005: 338)のである。

定量的手法と定性的手法を統合することで、調査目的に対して最適なアプローチをとることのできる可能性が広がる。ブライマンは、社会調査法において定量的手法と定性的手法を統合

するためのアプローチとして、以下の11種類を示している。すなわち、①トライアングレーションの論理、②定性的調査による定量的調査の促進、③定量的調査による定性的調査の促進、④全体像を描き出すための定量的調査と定性的調査の統合、⑤構造と過程、⑥調査者と対象者の観点、⑦一般性の問題、⑧定性的調査による変数間の関係性に関する解釈の促進、⑨マクロレベルとミクロレベルの関係、⑩調査プロセスにおける段階、⑪混合である (Bryman 1988: 131-52)。定性的パラダイムと定量的パラダイムとはそれぞれ異なる特性を有していることから、その二項対立的な区別を強調するよりも緩和することによって、調査のさまざまな場面において相互補完的に両者の強みを生かすことができると考えられる。

3.3 CIPP モデル

定量／定性論争におけるパラダイムの対立や手法の統合といった議論に先行して、評価研究では早くから手法の統合が提示されている。手法の統合として早くから用いられてきたのが、スタッフルビームによって1966年に紹介されたCIPPモデルである。

CIPPモデルは形成的評価と総括的評価を導く包括的なフレームワークであり、文脈 (Context)、インプット (Input)、プロセス (Process)、成果 (Product) の観点から評価のマネジメントを行う。それぞれの観点で注目される点は次のとおりである (Stufflebeam 2005: 61)。文脈の評価では、目標と優先順位の定義やアウトカムの判断のためにニーズや問題、利点、機会を事前に評価する。インプットの評価では、目標とされたニーズを満たしたり目標を達成したりするために代替的なアプローチや競合する活動提案、関連する予算を評価する。プロセスの評価では、活動の手引きやプログラムの業績判断あるいはアウトカムの説明に関する計画の実施について評価を行う。成果の評価では、事業を継続し成功を導くために、意図されたあるいは意図されていないアウトカムの評価を行う。

CIPPモデルはプログラムの改善に焦点を当てている。そのことは、スタッフルビームによる次の記述からも明らかである。すなわち「評

価においてもっとも重要な目的は、真実であることを示すことではなく、改善することである」(Stufflebeam 2005: 62)。この指摘は、評価研究が学術研究と異なることを端的に示している。学術研究では、科学的探求や検証を通じて命題の正しさを明らかにすることに関心がある。それに対して、評価研究の場合はむしろ、調査対象となるプログラムの改善に寄与するような評価結果を産出することが求められる。学術研究上の方法論争に比べると、評価研究ではよりプラグマティックな対応をとることで実務に寄与する傾向にある。このことは、手法レベルの統合として現れている。

CIPP モデルでは、文脈、インプット、プロセス、成果の4つの観点からさまざまな手法の利用が想定される(表3)。成果の観点は、さらにインパクト、有効性、持続可能性、移植可能性という4つのサブカテゴリに分割することができる。CIPP モデルの最終目的は、文脈、インプット、プロセス、成果(インパクト、有効性、持続可能性、移植可能性)の7つの観点

をバランスよく俯瞰した統合報告または最終報告を作成することにある。そのためには、個々の観点を確認するためのさまざまな手法を動員することで対応が可能となる。

この表では具体的な手法として、サーベイ調査、文献調査、資料調査、他のプログラムの見学、チームのアドボカシー、デルファイ法、プログラムの概略やデータベース、現場の観察者、事例研究、ステークホルダーへのインタビュー、フォーカスグループ、ヒアリング、費用分析、二次データ、ゴールフリー評価、写真記録、タスクの報告とフィードバック会合が挙げられている。これらの手法の中には定量的手法と定性的手法が混在している。それぞれの手法は、7つの観点のうちどの評価に強みを有しているかに違いがある。たとえば、費用分析では、インプットの評価として代替的なアプローチ間の費用を比較したり、プロセスの評価として実際に要した費用を計算したり、有効性の評価として要した費用に対して得られた効果を明らかにしたり、持続可能性の評価として将来にわたる費

表3 CIPP 評価において利用が想定される手法の実例

手法	文脈	インプット	プロセス	インパクト	有効性	持続可能性	移植可能性
サーベイ調査	✓		✓	✓	✓	✓	
文献調査	✓	✓					
資料調査	✓	✓	✓	✓	✓		
他のプログラムの見学		✓		✓	✓		✓
チームのアドボカシー		✓					
デルファイ法	✓	✓					
プログラムの概略やデータベース		✓	✓	✓	✓	✓	
現場の観察者			✓	✓	✓	✓	
事例研究			✓	✓	✓	✓	
ステークホルダーへのインタビュー	✓		✓	✓	✓	✓	✓
フォーカスグループ	✓	✓	✓	✓	✓	✓	✓
ヒアリング	✓	✓			✓		
費用分析		✓	✓		✓	✓	
二次データ	✓				✓		
ゴールフリー評価			✓	✓	✓	✓	✓
写真記録	✓		✓	✓	✓	✓	✓
タスクの報告とフィードバック会合	✓	✓	✓	✓	✓	✓	✓
統合報告または最終報告	✓	✓	✓	✓	✓	✓	✓

出所：Stufflebeam (2005: 64)

用や便益の見積もりを行ったりすることが考えられるだろう。

CIPP モデルでは、形成的評価と総括的評価それぞれにおいて文脈、インプット、プロセス、成果の4つの観点を考慮した評価が行われることになる。形成的評価では、「求められている介入の特定や目標の選択および順位づけの手引き」(文脈)、「プログラムや他の戦略を選択するための手引き」(インプット)、「業務計画の実施のための手引き」(プロセス)、「取り組みを継続し、修正し、採用し、終了させるための手引き」(成果)として評価結果が用いられることになり、総括的評価では、「事前に評価されたニーズや問題、利点、機会に対する、目標および優先順位の比較」(文脈)、「批判的な競合者のプログラム提案や対象となる受益者のニーズに対する、プログラムの戦略、デザイン、予算の比較」(インプット)、「実際のプロセスや費用の記録に関する十分な記述、デザインされたプロセスおよび費用と実際のプロセスおよび費用との比較」(プロセス)、「対象となるニーズやあるいは競合するプログラムに対するアウトカムや副次的効果の比較、取り組みについて事前に評価された文脈、インプット、プロセスに対する結果の解釈」(成果)として評価結果が用いられることになる(Stufflebeam 2005: 63)。総括的評価はおもに定量的手法によって、形成的評価はおもに定性的手法によって行われる傾向にあることが指摘されているが(Herman, Morris, and Fitz-Gibbon 1987: 26)、CIPP モデルでは形成的評価であるか総括的評価であるかを問わず、評価の目的に応じて定量的手法と定性的手法が用いられることになる。

3.4 混合手法

近年では、混合手法(mixed methods)と呼ばれる手法が注目されている。混合手法とは、2つ以上の手法を用いてデータの収集や分析を行う方法である。混合手法については学術研究における方法論としても議論されているが、評価研究からも注目を集めている。

評価研究において混合手法を用いる目的には、以下の5つがある(Greene 2005: 255)。第一が、トライアンギュレーションである。トライアンギュレーションとは、さまざまな手法を

用いることで評価結果の妥当性や信頼性を向上させることを意味している。第二が、開発である。開発とは、ある手法の結果を、別の手法で用いる標本や器具類の開発に用いることを意味している。第三が、相補性である。相補性とは、さまざまな手法を用いることでさらなる理解を促進し、評価結果の包括性を拡張させることを意味している。第四が、創始である。創始とは、相違や一致を求めるさまざまな手法による評価結果の中から新たな洞察を生み出すことを意味している。第五が、価値の多様性である。価値の多様性とは、評価における価値の次元についてさまざまな手法を用いることで広範な価値や意識を組み込むことを意味する。

混合手法を評価に適用した事例として、ダッタはアメリカの国際開発局(Agency for International Development: AID)が実施したインドネシアにおける乳幼児生存活動プログラムの評価を挙げている(Datta 1997: 36-7)。このプログラムは、5年のプロジェクトとして行われ、3年経った時点で取り組みの達成状況やインパクトについて中間的な評価を行うことになった。この評価ではさまざまな環境的制約が課せられており、それらを克服するために混合手法を用いた評価が行われるようになったのである。

ダッタによれば、プロジェクトによる変化をたどる上でいくつかの困難があった(Datta 1997: 36)。プロジェクトは、すでに進行中の活動について新たな支援を行っていた。これらの活動とはおもに、予防接種、下痢症状の抑制、栄養状況の改善、危険の高い出産の低減などである。また、それらの活動のいくつかは異なるドナーによって資金提供を受けていたり、それ以外の国内要因が影響を及ぼしたりしている可能性があった。評価は3週間という限られた期間で行われ、評価に宛てられる人員も4人と小規模のチームであった。当時のインドネシアの人口は1億7500万人であり、彼らは約6000ある島に広範にわたって分布していた。こうした背景事情や制約の中でも、混合手法を用いることで評価設問に対する回答を行うことができると考えられたのである。

表4では、プロジェクトを評価する際の設問に対して、それに答えるためのさまざまな手法が提示されている。この表からは、定量的手法と定性的手法とを問わずさまざまな手法が用い

られていることが理解できる。たとえば、「国際開発局の取り組みの文脈はどのようなものか」という設問に対しては、文書やインタビューあるいは過去の報告書による歴史分析の手法が用いられる。「全体的に、インドネシアは他のアジア諸国とどのように比較されるのか」という設問に対しては、世界銀行が提供するデータの二次分析が用いられる。「予防接種に対するプログラムのインパクトとは何か」という設問に対しては、各種ワクチンの適用範囲に関するデータを用いた時系列比較の手法が用いられる。「乳幼児の死亡率に対するインパクトとは何か」という設問に対しては、プログラムを実施した地域と実施しなかった地域との比較による準実験的手法が用いられる。

手法を選択する際の基準は、インドネシアと

いう環境によってもたらされる制約（たとえば地理的な要因）や評価者側の制約（人的、金銭的、時間的な要因）によって設定されることになる。今回の事例では、一からのデータ計測や長期間にわたる追跡調査などは物理的に実施できないだろう。「4人、3週間」という交渉の余地のない要件が、かえって評価デザインのトレードオフを最適化しているといえる（Datta 1997: 37）。

これらの手法からは、評価者自らが一から評価してデータを収集するというよりはむしろ、既存のデータや報告書を多分に活用していることがわかる。それらのデータや文書は、国際開発局や他の開発機関、類似プロジェクト、関係省庁、地域の病院、保健省、国連児童基金（United Nations Children's Fund: UNICEF）、第三者機関などによって提供されている（Datta 1997: 37）。

表4 インドネシアの乳幼児生存状況に関する評価の設問と手法

評価の設問	手法
AIDの取り組みの文脈はどのようなものか AID独自の貢献はどのようなものか	文書、インタビュー、以前の報告書を通じた歴史分析 米の自給自足、貧困の減少、教育の拡大といった変化の見込みのある貢献に関する質的事例研究による記述
全体的に、インドネシアは他のアジア諸国とどのように比較されるのか	世界銀行データの二次分析
活動はどのように実施されるのか	質的文書分析とインタビュー
母子サービスに対するプログラムのインパクトとは何か	家族計画や子どもの栄養状況に関する全国データを用いた時系列比較 終了した特定プロジェクトの報告書に関する質的分析
予防接種に対するプログラムのインパクトとは何か	DTP1（ジフテリア・破傷風・百日咳三種混合ワクチンの第1回接種）、ポリオ1（ポリオ予防ワクチン第1回接種）、BCG（結核予防ワクチン）、はしか予防、ワクチン新生児破傷風ワクチンの適用範囲に関する保健省のデータやAIDプロジェクトからのデータを用いた時系列比較 インフラの変化（たとえばワクチンの効能を維持するために中央から周辺部への低温流通システムを確立すること）に関する保健省のデータの質的分析 成功したワクチン輸送と地域診療所（インドネシア語で posyandu）の設立とを結びつける質的事例研究と一連の出来事分析
医療サービスの効率性に対するインパクトとは何か	時系列のワクチン生産に関する非公開データ（都市部と農村部において1985年から1986年までと1988年から1989年までのあいだの適用範囲の公平さの改善を示す）の比較分析
乳幼児の死亡率に対するインパクトとは何か	病院記録からの地域データを用いた、プログラムを実施する地域とそうでない地域、あるいは介入の対象となる病気とならない病気による準実験的な比較による事前事後分析（プログラムを実施する地域と実施しない地域のいずれにおいても、介入の対象とならない病気に変化が見られない場合、推論が著しく強化される）

既存のデータや文書を活用することで評価プロセスの充実を図るとともに、人的、金銭的、時間的な制約の中で最良の評価結果を生み出すことを企図している。

4. 日本の評価実務

日本では、中央府省の政策評価や地方自治体の行政評価などさまざまな評価実務が行われている。これらの現場では、本稿で論じてきた社会科学的な調査手法が積極的に用いられているとは必ずしもいえない。むしろ、中央府省や地方自治体のほとんどでは本稿で論じてきた評価ではなく業績測定（performance measurement）が主流化している。業績測定とは、事前に指標を設定し、それらを測定することで目標が達成されたかどうかを事後的に検証するための目標管理の手法である。日本の評価実務では、この業績測定が「評価」として用いられている場合が多く、本稿が参照してきた議論を援用するには困難が伴う。

このような事情から、日本の評価実務は定量的手法や定性的手法について広範に検討がなされている状況とはいえ、また講学上も定量的手法と定性的手法の普及状況に関する実証研究はほとんどみられない。本稿では最後に、政府開発援助（ODA）の評価と政策評価を例に、日本の評価実務における定量的手法と定性的手法の考え方について検討する。

4.1 ODA 評価

日本における ODA の分野では、政策評価制度が導入されるよりも古くから評価が行われて

おり、情報収集や分析を行うために定量的手法と定性的手法とを組み合わせる評価を行っている。外務省が発行している「ODA 評価ガイドライン」では、情報収集の手法として、インタビュー（構造化インタビュー、半構造化インタビュー、非構造化インタビュー）、フォーカス・グループ・ディスカッション、直接観察、文献調査、ベースライン調査、アンケート調査が例示されている（URL 1: 39）。また、分析の手法として、リスク分析、事例研究、費用便益分析、費用効果分析、産業連関分析、計量経済モデルを用いた分析、インパクト評価が挙げられている（URL 1: 40）。

独立行政法人国際協力機構（JICA）が 2004 年に策定した「プロジェクト評価の手引き—改訂版 JICA 事業評価ガイドライン」では、情報収集におけるデータの種類として定量データと定性データに関する記述がある。定量データは「農作物の収穫量、識字率、乳児死亡率、灌漑面積、建設された施設の数、参加した人数、テストの平均値など、データそのものが数値で表されているものである」（URL 2: 90）。定量データが適しているとされるのは、「実績や達成度などを測定する」「大人数を対象に調査する」「確立した測定手段がある」「統計分析を行う」といった場合である（URL 2: 91）。それに対して、定性データは「現象を記述的に把握したもので、特定の課題や人々の行動・認識をより深く詳細に知ることに適して」（URL 2: 91）いる。定性データが適しているとされるのは、「より深く、詳細な情報を調査する」「達成状況に影響を与えた周辺要因（阻害・貢献要因など）を探る」「あらかじめ分析方法を決めていない」「定量化する必要がない」といった場合である（URL 2: 91）。

表 5 データの種類と収集方法の関係

	質問紙調査		インタビュー調査	フォーカスグループ・ディスカッション	観察	
	選択肢式	自由回答式			チェックリストを用いた観察	視察、状況把握
定量データ	○		△※		○	
定性データ		○	○	○		○

※質問紙に準じる形で選択肢を用意したインタビュー（構造化インタビュー）を実施した場合は、ある程度定量化が可能

出所：（URL 2: 96）

JICA によるこの手引きには、理数科教師養成プロジェクトを事例に、定量データと定性データの例が示されている。定量データの例としては、「研修の参加者数」「養成された教師の数」「学生のテストのスコア」「教授方法の質を測定するインデックス」が挙げられており、定性データの例としては、「研修内容で不満に感じたこと・提案」「親の目から捉えた子供の変化」「教師養成プロジェクトに対する教員の認識の変化」「なぜ教授法が改善されなかったのかの理由」「教員養成コースの実施体制の適正度」が挙げられている（URL 2: 92）。

表5では、定量データと定性データそれぞれを収集するのに適した手法が示されている。収集しようとしているデータに応じて、適切な手法を選択する必要があることを理解できる。また、質問紙調査や観察のように、同じ手法でもその実施方法によって定量データを収集するものと定性データを収集するものが存在している。それぞれの収集方法が有する強みや弱みに留意し、またプロジェクトの状況や評価者の制約などを考慮した上で、複数の手法を組み合わせる用いることが求められている。手引きでは組み合わせの例として、質問紙調査結果の背景を探るためにフォーカスグループを行ったり、質問紙調査の中で選択肢式と自由回答式を併用したりすることが示されている（URL 2: 100）。

4.2 政策評価

日本の中央府省で政策評価制度が法制化されたのは2001年のことである。法制化に至るまでに各府省ではさまざまな研究会が設けられ、そこでは手法に関する議論が交わされている。日本の政策評価制度は、定量的手法を強調している点に特徴がある。

「行政機関が行う政策の評価に関する法律」（平成13年6月29日法律第86号）の第3条2項1号は「政策効果は、政策の特性に応じた合理的な手法を用い、できる限り定量的に把握すること」と定めており、政策評価の客観的かつ厳格な実施を謳っている。他方、定性的という言葉は法文の中には見あたらない。日本の政策評価制度はその根拠となる法令の中で定量的手法を重視する姿勢を示しているのである。

1998年3月から1999年6月まで当時の通商

産業省に設置された「政策評価研究会」は最終報告書の中で定量的手法と定性的手法に言及している。報告書では、定量的手法に対して多くの記述がなされており、費用便益分析、費用効果分析、コスト分析、統計解析法、対照実験法、行政指標を用いた評価について具体的な事例とともに紹介されている（政策評価研究会1999: 91-137）。定性的手法については、「精緻な」定量的手法に対する「簡便な」手法として簡単な紹介にとどまっており、数値的な分析による裏づけの重要性を強調している（政策評価研究会1999: 137-9）。

1999年8月から2000年12月まで当時の総務庁に設置された「政策評価の手法等に関する研究会」は最終報告の中で、「できる限り定量的な手法が望ましい」として定量的手法を重視する見解を示している（URL 3）。ただし、定量的手法の適用になじまない評価対象の場合には、あくまで客観性の確保に配慮する必要があるものの、定性的手法を用いることが適当であると示されている。とくに、総合評価方式を用いた政策評価においては、特定の評価手法によらず、評価対象に応じて定量的手法と定性的手法を用いて評価することが示されている。

5. おわりに

本稿では、評価研究における定量的手法と定性的手法について、定量／定性論争での議論を手がかりとしてその特徴を対比的に論じた。本稿の概略は以下のとおりである。

定量／定性論争は、定量的手法に対して定性的手法の側から批判が加えられたことを契機としていた。グーバとリンカーンが唱えた第4世代評価は、それまで定量的手法が優勢だった評価に対して定性的手法の重要性を説いたものとして注目され、その後の定性的手法の隆盛や参加型評価の提唱へとつながったのである。定量／定性論争では、両手法のあいだには実証主義と構成主義というパラダイムの相違が存在し、それぞれパラダイムの特性が異なることが認識されていた。定量的手法と定性的手法は、パラダイムの相違を克服するために、パラダイムの下位である評価デザインや評価手法のレベルで統合を模索していったのである。

手法の統合については社会調査法の分野でも議論がされており、そこでは両手法を組み合わせることで調査目的に対して最適なアプローチをとることができると考えられていた。また、評価研究においても、スタッフルームが提唱した CIPP モデルに見られるように、定量／定性論争が行われる以前から定量的手法と定性的手法を統合した評価プロセスが提唱されていた。この CIPP モデルが示唆するのは、定量的手法と定性的手法は、文脈、インプット、プロセス、成果についてそれぞれ異なる観点で評価を行っており、さまざまな評価手法を組み合わせることですべての観点を俯瞰した最終報告書を作成することができるということである。また、近年では混合手法の観点から評価を行うことについて議論がなされており、本稿ではアメリカ国際開発局が実施したインドネシアでの乳幼児生存活動プログラムの評価を例として挙げた。この事例からは、資金、人員、期間といった面でのさまざまな制約が、定量的手法と定性的手法の選択に関わる評価デザインのトレードオフを最適化していることを理解することができた。

本論文の最後では、日本の評価実務を例に、定量的手法と定性的手法がどのように用いられているのかを明らかにした。ODA の分野では、定量的手法と定性的手法を組み合わせることに合意がなされており、各種ガイドラインの中でデータの収集や分析にかかる定量と定性の特性の違いに関する記述を確認することができた。他方で、政策評価制度では可能な限り定量的手法を使用することが法律の条文中に定められており、また各省庁の研究会から出された報告書からも定量的手法の使用が強調されていることを理解できる。

本稿は、評価研究における定量的手法と定性的手法について、先行研究の整理を行う中で両手法の特徴の違いを明らかにした。また、評価研究では定量的手法と定性的手法の統合によるアプローチが早くから検討されており、評価実務においても手法の統合を観察することができた。課題として、本稿では手法の統合に関する具体例を分野ごとに検討することができなかったため、今後は各分野の事例研究を重ねることが重要であると考えられる。

参考文献

〈日本語文献〉

- 浅沼茂 (2010) 「八年研究の日本的受容の諸問題」『東京学芸大学紀要 総合教育科学系』61 (1)、1-14。
 佐々木亮 (2010) 『評価論—評価学の基礎』多賀出版。
 政策評価研究会 (1999) 『政策評価の現状と課題—新たな行政システムを目指して』木鐸社。
 武田丈 (2015) 『参加型アクションリサーチ (CBPR) の理論と実践—社会変革のための研究方法論』世界思想社。
 西尾勝 (2001) 『行政学 [新版]』有斐閣。
 渡瀬義男 (2005) 「米国会計検査院 (GAO) の 80 年」『レファレンス』55 (6)、33-61。

〈外国語文献〉

- Bryman, A. (1988) *Quantity and Quality in Social Research*, Unwin Hyman.
 Campbell, D. T. (1969) Reforms as Experiments. *American Psychologist*, 24 (4), 409-29.
 Cousins, J. B., and Whitmore, E. (1998) Framing Participatory Evaluation. *New Directions for Evaluation*, 80, 5-23.
 Cronbach, L. J. (1982) *Designing Evaluations of Educational and Social Programs*, Jossey-Bass.
 Datta, L. (1997) A Pragmatic Basis for Mixed-Method Designs. *New Directions for Evaluation*, 74, 33-46.
 Greene, J. C. (2005) Mixed Methods. In S. Mathison (ed.) *Encyclopedia of Evaluation*, 255-6, Sage.
 Greene, J. C., and Henry, G. T. (2005) Qualitative-Quantitative Debate in Evaluation. In S. Mathison (ed.) *Encyclopedia of Evaluation*, 345-50, Sage.
 Guba, E. G., and Lincoln, Y. S. (1989) *Fourth Generation Evaluation*, Sage.
 Hammersley, M. (1992) Deconstructing the Qualitative-Quantitative Divide. In J. Brannen (ed.) *Mixing Methods: Qualitative and Quantitative Research*, 39-55, Avebury.
 Hedrick, T. E. (1994) The Qualitative-Quantitative Debate: Possibilities for Integration. *New Directions for Program Evaluation*, 61, 45-52.
 Herman, J. L., Morris, L. L., and Fitz-Gibbon, C. T. (1987) *Evaluator's Handbook*, Sage.
 Ianni, F. A. J., and Orr, M. T. (1979) Toward a Rapprochement of Quantitative and Qualitative Methodologies. In T. D. Cook, and C. S. Reichardt (eds.) *Qualitative and Quantitative Methods in Evaluation Research*, 87-98, Sage.
 Lincoln, Y. S. (2005) Fourth-Generation Evaluation. In S. Mathison (ed.) *Encyclopedia of Evaluation*, 161-4, Sage.
 Merriam, S. B., and Simpson, E. L. (2000) *A Guide to Research for Educators and Trainers of Adults*, 2nd ed., Krieger. (= 2010、堀薫夫監訳『調査研究法ガイドブック—教育における調査のデザインと実施・報告』ミネルヴァ書房。)
 Punch, K. F. (1998) *Introduction to Social Research: Quantitative and Qualitative Approaches*, Sage. (= 2005、川合隆男監訳『社会調査入門—量的調査と質的調査の活用』慶應義塾大学出版会。)
 Reichardt, C. S., and Cook, T. D. (1979) Beyond Qualitative versus Quantitative Methods. In T. D. Cook, and C. S. Reichardt (eds.) *Qualitative and Quantitative Methods in Evaluation Research*, 7-32, Sage.
 Rossi, P. H. (1994) The War between the Quals and the Quants: Is a Lasting Peace Possible? *New Directions for Program Evaluation*, 61, 23-36.
 Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004) *Evaluation: A Systematic Approach*, 7th ed., Sage. (= 2005、大島巖・平岡公一・森俊夫・元永拓郎訳『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。)

Stufflebeam, D. L. (2005) CIPP Model (Context, Input, Process, Product). In S. Mathison (ed.) *Encyclopedia of Evaluation*, 60-5, Sage.

【URL】

1. 外務省 (2015) 「ODA 評価ガイドライン第9版」外務省ホームページ (2016年6月2日閲覧、<http://www.mofa.go.jp/mofaj/gaiko/oda/files/000083485.pdf>)。
2. 国際協力機構 (2004) 「プロジェクト評価の手引き—改訂版 JICA 事業評価ガイドライン」国際協力機構ホームページ (2016年6月2日閲覧、http://www.jica.go.jp/activities/evaluation/guideline/pdf/old_guideline.pdf)。
3. 政策評価の手法等に関する研究会 (2000) 「政策評価制度の在り方に関する最終報告」総務省ホームページ (2016年6月2日閲覧、http://www.soumu.go.jp/main_sosiki/hyouka/s_saihou.htm#100)。