

<研究ノート>

第4世代評価

—その背景と評価手法を中心に—

百海正一

序

日本経営情報学会「情報システム評価の研究部会」は当初（1998）、「情報システムの評価に関する研究」するという主旨で、情報システムの評価に関する文献を調査しようとして設立された。このなかで、我々メンバー¹⁾の目にとまったのは、英国ランカスター大学のホームページに掲載されている Evaluation of Cooperative Systems (Magnus Ramage, 1995)²⁾であった。この文献は、コンピュータ協調作業支援システム CSCW (Computer Supported Cooperative Work, 通称 CSCW) のなかで、評価のための新しい方法論を提案しているだけでなく、情報システムのみならず、社会一般の改革プログラムを対象としている。また、従来の評価方法の多くが、主に合理主義者の立場から、あらゆる状況に適応できる評価方法を求めていること、マネジメントや情報システム構築者の視点から情報システムの評価を捉えていることを問題点としてあげている。

この文献を通して、我々は評価に関する基本的な知識を得ることができた。この著者が評価に関して、著者の考えに近い文献として「第4世代評価 (Egon G.Cuba and Yvonna S.Lincoln, Fourth Generation Evaluation, Sage Publication Inc., 1989)」をあげている。「第4世代評価」に関する文献解説と討議を重ねた結果、情報システム評価に対しても、① 従来の評価法は、評価対象である「改革プログラム」に関する情報を評価者が意思決定者などに提供することを目的としているのに対し、第4世代評価では、「改革プログラム」の影響を受ける

全ての関係者が、評価結果に基づいてそれぞれ適切な行動をとるようになることを目指している、② 情報システムの開発は、組織におけるある種の改革をもたらすことになる、などの理由で有効であると考えに至った。

部会メンバーは、研究活動を通じて共有し得た成果を、2000年度経営情報学会秋期研究発表大会（於岩手県立大）のワークショップで報告し、2001年6月研究報告書³⁾としてまとめることができた。しかし、この報告書は紙面上の制約から、やや総花的に書かれている。そこで、本稿は筆者が第4世代評価の背景と評価ステップを中心にまとめなおしたものである。

1 評価の時代

1) 評価とは

この研究が、評価に関する研究であるならば、まず“評価とはなにか”，について定義する必要があるのではないと思われる。しかし、ここで「評価」の定義について説明することは、この研究の中心テーマではない。なぜならば、この研究は、評価活動をどのように進めて行くべきか、評価の目的は一体なにか、に関する論議が中心であり、かつ我々（注：グバとリンカーン達）は「評価を“正確に”定義する方法は存在しない」とは考える。さらに、我々は「評価は、人間の頭のなかに存在する構成物（コンストラクション）である」と定義する。

定義に関する論議はそのくらいにして、過去100年の間「評価」という言葉の意味がどのように変化してきたか、を簡単にスケッチしてみる。

「評価」の意味は、皆時代時代の歴史的な文脈、人々が実施しようとした目的、評価に関する研究者や実務家によって立つ価値前提、などを反映している。そして、評価活動の結果得られる構成物は、時代を経るにつれ、新たな情報に基づき、より洗練されたものになってきている。そして、今我々は第4世代評価と呼ぶ、新しい構成物を生み出す地点に立っている。もちろん、このような評価に関するコンセンサスが必ずしも得られたわけではない。それは、以前からある評価法に関する評価と同じである。しかし、我々が提案する第4世

代評価による構成物は、従来の評価法から得られた構成物が持っていた問題を克服する、あるいは少なくとも改善するものである、と思う。

2) 第 1 世代—測定の世界—

今日、我々が知っている「評価」はある日突然現れたのではなく、長い年月を経て、改良されてきた結果である。早くから評価に影響を与えたもののの中に、小学生の持っている様々な能力を測定する評価法がある。例えば、生徒が授業で学んだ内容をどの程度理解したかを知るためのテストは、既に 100 年以上の歴史を持っている。なかでも、テストを採用する要因の一つに、アメリカ産業界における科学的な管理法の出現があった。商品の生産やサービス活動において、人間が重要な経営資源であるとするならば、人間による作業を出来る限り効果的、効率的に管理するのは経営管理者の役割である、と考えられてきた。

このような考え方は、1939 年レスリスバーガーやディクソン (Roethlisberger & Dixon) によるホーソン工場の実験によって批判される頃には、科学的管理法の考えは、すでに学校全体に浸透していた。すなわち、学校に在籍する生徒は、校長が管理する学校という“工場”で処理すべき“材料”であると考えられていた。そして、テストは学校が設定した“仕様 (教育目標)”を生徒が満たしているかどうかを確かめるための手段である、と見なされた。当時の教育目標は、主として上級学校への進学のための準備のためであった。

これら諸要因により、学校におけるテストは著しく普及し、1920 年代から 1930 年代にかけて頂点に達した。特に、1908 年より始められたアチーブメント・テストは、その後様々な形で広く実施され、今日に至っている。

また、1922 年出版されたスタンフォード・アチーブメント・バッテリー (Stanford Achievement Battery) により初めて、一人の生徒が複数の科目のなかで相対的にどのような順位にいるか、を評価することが可能になった。1933 年になると、ミネソタ大学ゲルトロード・ヒルドレッヒ (Gertrude Hildreth) は、3,500 以上のメンタルテストと評価尺度に関するリストを公表した。その

後、1945年に出版された彼女の文献によれば、その数は5,200を越えていた。

この時代には、“測定”という言葉と“評価”という言葉は同じ意味で使われていた。評価者の役割は技術的なものであったから、調査対象とされるどのような変数でも測定することができるよう、利用可能な全ての測定方法に熟知していることが求められた。

注目すべきことは、評価を技術的なものであるとする第1世代の考え方が、今でも根強く残っている。例えば、生徒が高校を卒業するため、あるいは大学に入学するための手続きの一部として、テストに合格することが当然のこととして要求されることや、多くの国において学校や時として一人一人の教師をその能力をランク付けするために、この種のテストが使われていることから明らかである。

3) 第2世代—記述の時代—

第2世代の評価は、第1世代が持っていた重要な欠陥によって引き起こされた。すなわち、第1次大戦終了後まもなく、アメリカの高校に、中学校レベルの教育しか受けていない学生が大量に入学してきた。これらの学生達は、現行のカリキュラムでは、彼らの要求を満たすことができないとし、教育内容の変革を訴えた。なぜなら、学生達は高校を、社会で働く上で必要なスキルを身につける場所である、と考えていたが、当時のカリキュラムではこれら学生の要求を満たすようなプログラムは提供されていなかったからである。

これに対して、高校も履修単位や条件を変えるなどして、学生の要求に対応しようとした。しかし、この変更に対する障害は、現行のカーネギー・ユニット・システム—すなわち卒業に必要な科目と履修単位数を取得する規定—が大幅に変更された場合、大学の標準的なカリキュラムについていけない高等学校卒業生を受け入れざるをえなくなることを恐れた大学（単科大学および総合大学）であった。

大学側の懸念に対し、その懸念が妥当であるかどうかを調べるプロジェク

ト—8年調査プロジェクトと呼ばれた—が1933年にスタートした。調査の目的は、従来のカリキュラム基準を満たしていない高校生が大学に入学した場合、はたして大学の授業についていけるかどうか、を調査するためであった。ここで、8年という調査期間が選ばれた理由は、従来の高校教育を変更した場合、高校（公・私立）に入学した学生が、4年後大学に入学し、4年間の大学教育に成功裏に終了することが出来たかどうか、を調査するためであった。

調査を設計する担当者が最初に直面した問題は、新カリキュラムが果たして意図したように機能しているかどうか、を確かめるための評価法を、まず考案することであった。当時、調査プロジェクト本部のあったオハイオ州立大学にラルフ・テイラー（Ralph W. Tyler）がいた。彼は、その後この調査に携わることになったが、これまでの伝統的な評価法とは大きな違いが一つあった。それは調査目的、すなわち—新カリキュラムの開発が確かに機能しているかどうかという“プログラム評価”—を実際に確かめることであった。調査に参加した高校が、新カリキュラムを開発するに伴い、テイラーは新プログラムが定められた教育目標をどの程度達成しているか、を評価するための情報を収集した。これらの情報は、旧カリキュラムと新カリキュラムによる学習目標に対する達成度の違い、新旧カリキュラムの長所と短所を分析するとともに、旧カリキュラムの改善と見直しのために利用された。この評価プロセスは、今日形成的評価（Formative evaluation）と呼ばれているが、新カリキュラムが適切な成果を生み出しているかどうか、を確認されるまで、繰り返し実施された。

この調査結果から、今日第2世代と呼ばれる評価法が出現した。第2世代評価の特徴は、あらかじめ定められた目標に対する強みと弱みを記述することにあった。従って、評価者の役割は第1世代の測定者の役割から記述者の役割へと変わった。ただし、第1世代評価の特徴である技術者としての役割りも、第2世代でも同じように要求された。そして、第2世代では“測定する”という仕事は、もはや評価そのものではなくて、評価者の仕事のなかで使われるツールの一つであると再定義された。

4) 第3世代—判定の時代—

第2世代評価は、目標指向のアプローチであり、基本的には評価対象に対して正確に記述することにあった。

ステーク (Robert Stake) は 1967 年教育評価に関する論文 (The countenance of educational evaluation) のなかで、第2世代評価の問題点を以下のように指摘している。

従来の評価法は、評価における判定の役割を無視している。教育者による視点と評価者 (専門家) による視点とは同一の視点ではない。評価の専門家は、自分自身を記述者、すなわち生徒の素質、教育環境、成績などを正確に記述する者である、と考えている。これに対し、教師や学校の管理者は、生徒に対して評価者としての役割は持っているものの、生徒を外基準に照らして、生徒を適切に評価 (測定) されることを期待している。評価には、“記述”と“判定”は何れも不可欠なものである。ワースンとサンダース (Worthen & Sanders) も、1973 年出版の「教育評価：理論と実践 (Educational evaluation: Theory and practice)」のなかで、“記述”と“判定”の二つは、評価における基本的な行為である、と述べている。

“判定”という行為を評価に含めるべきだ、というステーク等の主張は、第3世代評価と呼ばれる評価の出現をもたらした。すなわち、評価者が評価における技術的・記述的機能の他に、“判定”という役割を併せ持つ世代である。評価における“判定”という機能は、当時の専門家から広く支持された。そして、彼らは従来の評価法が対処できなかった問題点を幾つか指摘した。まず第1に評価目標自身に問題がある。評価目標も評価結果と同じように評価の対象とすべきである、という主張である。第2に、評価活動の結果、評価を行う際に、なんらかの判定基準が必要になる。しかしながら、判定にはなんらかの価値観が伴う。従来評価のように科学的で、価値観から独立していると考えられてきた業務に、価値観を伴う基準を導入することは、当時の評価者にとって受け入れがたいことであった。さらに、評価に判定が必要であるということは、評価を行う判定者 (審判者) が必要になる。しかし、評価者は、評価者自身判

定する立場で行動するだけの能力を持ちあわせていない、と感じていた。さらに評価者は、判定者として行動した場合、その結果受けるかもしれない政治的リスクを恐れた。それにも関わらず客観的な理由で、評価者は判定するという役割を受け入れざるを得なくなった。

5) 第1世代から第3世代評価に共通する問題点

評価に関するこれまでの論議は比較的簡単であったが、評価の成果である構成物に含まれる内容は、その対象範囲の広さと洗練さ両面において、第1世代から第2世代へ、第2世代から第3世代へと一步一步前進している(図表1)。

1981年グバ・リンカーンらの「効果的な評価(Effective Evaluation)」によれば、第1世代を特徴づける適切な測定手法が開発されるまでは、各個人のデータを体系的に収集し、分類することは出来なかった。第2世代に入ると、評価プログラム、評価活動に必要とされるマテリアル(材料)、評価に関する政策と組織的な取り組みの必要性など、多くの人間以外を評価対象とする評価法を開発した。第3世代では、評価対象のメリット(内的あるいは本質的な価値)とその価値(外的あるいは文脈的価値)の両面にわたる判定を導くような評価が求められるようになった。

しかし、これら第3世代までの評価をすべて一つのグループと見なしたとき、さらなる改訂が必要ではないか、というある種の共通する問題点を抱えて

図表1 各時代における評価とその要点

	第1世代	第2世代	第3世代
評価目的例	与えられたカリキュラム下で学習目標に対する生徒の理解度	旧カリキュラムの問題点と改善点 新カリキュラム導入後の追跡調査	特定の政策の下で策定されたカリキュラムの有効性
評価者の役割	測定者 (測定)	記述者 (測定+記述)	判定者 (測定+記述+判定)
成果例	アチーブメントテストによる標準的な評価 尺度・評価項目	8年調査評価プロジェクト、報告書	幾つかの判定モデル 意思決定モデル 目標自由モデル

いた。

その第1は、管理主義への傾斜があげられる。ここで、管理者という言葉にはさまざまな人が含まれるが、この場合評価を依頼したり、資金を提供するクライアントあるいはスポンサー、または評価レポートに基づいて実行する責任をもつエージェントを指している。後者の例として、学校における理事会のメンバー、教育長、カリキュラム改革を検討中の学校長などがあげられる。

一般に、評価者が評価プロジェクトを引受け、指示を受け、そのアウトプットである報告書を提出する先は管理者である。管理者と評価者との関係は、時には以下のように、幾つかの望ましくない結果を生み出す。

① 管理者と評価者の関係から、管理者は評価結果に対して責任を負わなくてすむ。管理者が評価の外にいる限り、管理者としての資質や能力は問題にならないし、評価の結果、何をアウトプットしようが、その責任は問われない。もし、問題が生じた場合には、評価者に責任を転化すればよいからである。

② 典型的な管理者と評価者の関係は、権限を持っているものと、持っていないものの関係にある。管理者は、評価に際し、どのような質問をし、その解答をどのように収集し、解釈すべきか、そしてその結果を、誰に報告すべきかを決定する最終権限を有している。もちろん、これらの事項は評価者との同意に基づいて決められる。しかし、もし意見の一致を見なかった場合、最終決定権は管理者にある。これに対して、もし管理者が評価者の意見に合わない評価を下すことになった場合、評価者が意見を表明する唯一の手段は、管理者に対して拒否することだけである。

このような関係では、評価対象に対する質問項目に疑問を持ったり、別の解答を持っていたり、別の解釈をする可能性のあるステークホルダー (Stakeholder) は、事実上弱い立場に置かれることになる。第1から第3代の評価法では、管理者が評価に大きな権限を有している以上、評価に関心を持つステークホルダーの意見を反映することは不可能ではないが、一般的に難しい。

③ 管理者と評価者の関係との関係は、特権を有しているものと有していないものの関係にある。しばしば、管理者は評価の結果判明したことを公表すべ

きかどうか、そして公表するなら誰に対して公表すべきか、を決定する権利を有している。

評価者は適切な評価報告書を作成する権利を有しているが、その権利を情報開示する権限を有する管理者と取引することは別に珍しいことではない。しかしながら、本当の評価結果を知らされないステークホルダーは、評価の結果取ることが出来たかもしれない一連の行動—例えば、自分自身の利益を守るなど—をとれない。情報を知りうる立場にないステークホルダーは、結局馬鹿をみることになる。

第 2 に、多元的価値の問題に対応できていないことがあげられる。1960 年代後半以降、アメリカ人は政党間の対立だけではなく、人種、性別、そして簡単に解決できない価値観の対立があることを学んだ。

評価における判定を求める声が現れたのは、多元的価値に対する理解が出現したのとはほぼ同じ頃である。従来、価値の違いが問題として取り上げられない限り、“客観的な”評価手法の開発にも価値判断 (Value Judgement) が関わりあっていることや、明確な目標設定の際にも当然のことながら、価値の合意 (Value Agreement) を前提にしていること、などの事実が見過ごされてきた。

しかし、一度価値の問題が表面に現れてしまうと、それをもとに戻すのは不可能である。評価の際、一体誰の価値を優先するのか、あるいは価値の違いをどう扱うのか、が主要な問題となった。従来から、価値の問題が存在していたにも関わらず、評価結果が信用されてきたのは、評価手法が科学的であれば、価値の問題に無関係である (Value free)、という論議が行われてきた。

評価に客観性が求められる最大の根拠は、価値の影響を出来るだけ除去することであった。もし、科学的評価手法が価値と無関係でないとすると、評価結果に対して、様々な解釈が可能になるだけでなく、“事実”そのものも価値システムとの相互作用によって決まることになる。そうなると、評価におけるあらゆる行為、それが評価、調査あるいは政策分析であれ、政治的行為となる。

第3に、あまりにも科学的パラダイムに依存してきたことがあげられる。社会科学における調査に関わってきた人々は、自然科学における手法を積極的に取り入れてきた。しかし、科学的手法への極端な依存は、いくつかの望ましくない結果をもたらした。

① それは、“文脈剥離 (context-stripping)” と呼ばれる現象を導いた。文脈剥離とは、評価対象があたかも現実の文脈のなかではなく、人工的に注意深くコントロールされた環境下だけに存在し、評価対象をアセスすることである。

② 科学的パラダイムへの依存は、不可避免的に形成的定量的測定 (formative quantitative measurement), すなわち、評価プロセスで扱うデータの“ハードさ”に依存する。ここで、ハードなデータとは正確に定量化できるデータで、測定し、評価分析用モデルあるいは統計式にインプット可能なデータである。従って、第1世代から第3世代評価まで、測定しにくい質的データの問題はあまり取り扱われていない。

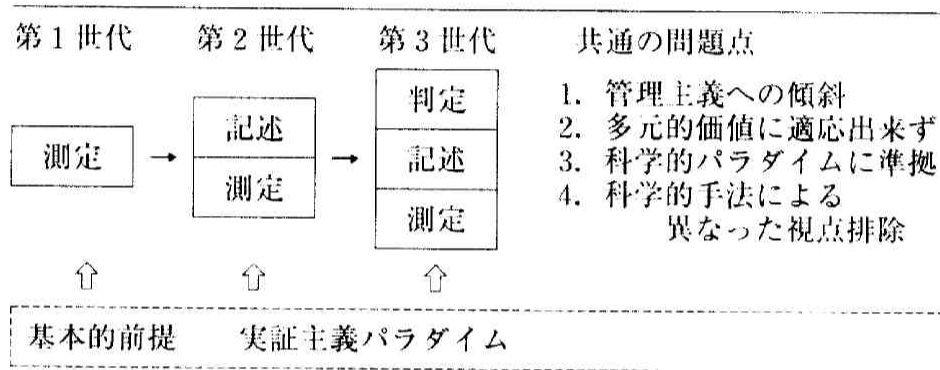
③ 科学的手法に基づいた評価結果は、ある種の権威を持っている。評価者は、被評価者が提出したデータをもとに科学的手法を使って処理した結果である、とクライアントに自信を持って言うことができる。従って、被評価者はその結果について議論することや、拒否することも難しい。

第4に、科学的手法を使うことによって、評価対象を異なった視点から考察する機会が失われる。科学の本質は、物事の真実を明らかにすることであるから、評価法に対する代替案を提案しにくい特徴を持っている。評価者、クライアント、そしてステークホルダーも一様に、その結果を“真実 (True believers)”として受け入れるよう強制される。すなわち、科学的に根拠のあるものはすべて額面通りに受け入れなくてはならない、ということになる。かりに、もし評価に問題があるとするならば、代替案も科学的な根拠がなければ、取り上げられることはない。

最後に、科学は価値観と無関係である、とされているから、評価者は科学的

パラダイムを信奉することにより、評価活動に対する道義上の責任を軽減される。また、評価者は評価結果がその後どのように使われるか、コントロールできないと言える。そのうえ、評価者の役割は報告書を提出した時点で終了するから、評価者は評価をフォローする責任を有していないと主張することもできる。何れにせよ、第3世代までの評価者は、評価からどのような結果が生じて、あるいは評価結果がどのように使われても、道義上の責任は問われることはない（図表2）。

図表2 共通の問題点



6) 代替的アプローチ

今日、第3世代までの評価法が抱えている問題を解決するための代替的なアプローチが求められている。その代替案の一つとして、「状況対応型構成主義的評価（Responsive Constructivist Evaluation）」と名づける評価法を提案したい。

第3世代までの評価モデルでは、評価のスペシフィケーション（仕様）であるパラメーターやその境界は、通常クライアントと評価者との協議によって事前に決められていた。ロバート・ステーク（Robert Stake）は、第3世代までの評価が事前に評価の仕様を決めていたことに注目し、「事前決定型評価（Preordinate evaluation）」という用語を考案した。これに対して、「状況対応型評価（Responsive evaluation）」という呼び方は、1975年ステークによって、初めて提案されたものであるが、「状況対応型評価」では、評価の影響を受けるステークホルダーも加わった多くの人達が、評価プロセスを通じて、パラメーターやその境界を多くの時間をかけて決めることに特徴がある。

このアプローチの背後にある基本的前提は、科学的パラダイムに代わる構成主義 (Constructivist) と呼ばれるパラダイムに基づいた評価を行う際に使われる手法である。ここで、「構成主義的」とは、端的にいうと「“真実”は人々の考えと無関係に存在するのではなく、人々の相互作用の中から社会的に構成される」という考えである。また、ある人は「構成主義的」のかわりに、「解釈学的・弁証法的 (Interpretive,あるいは Hermeneutic)」という用語を使う人もいる。名称はそれぞれ、パラダイムに対するある特定の洞察を表しているので、ここでは、状況対応的な焦点の当て方と、構成主義的なやり方のそれぞれの特徴を表す代表的な考え方を紹介する。

(1) 状況対応的な焦点の当て方：「状況対応型評価」の進め方は、評価プロセスに対するアルゴリズムも、まずどのような項目を質問し、どのような情報を収集すべきかを検討することから始まる。第1世代評価の場合は、幾つかの変数が特定され、収集すべき情報は、それら変数を測定すると考えられる測定モデル上で、生徒各個人の点数—例えば、アチーブメント・スコア—から構成される。第2世代の場合は、幾つかの評価目標が特定され、収集すべき情報は、例えば生徒個人の実績が、学校で記述された学習目標とどの程度強く結び付いているかの評価から構成される。第3世代評価の場合は、評価モデルによって必要な情報は異なる。例えば、意思決定モデルの場合、代替案を比較するための情報を収集することが求められる。また、ゴール・フリー (goal-free) モデルの場合は、経験された“効果”に関する情報が求められる。このような焦点となる要素 (変数, 目標, 意思決定など) はアドバンス・オーガナイザー (advance organizer) と呼ばれる。評価者は、どのような焦点となる要素を使っているかは、評価者が例えば、「あなたの目標は何ですか?」、あるいは「この評価は、どのような意思決定に対して、情報を提供する必要がありますか?」などという質問を提起した瞬間に明らかになる。

状況対応型評価もアドバンス・オーガナイザーを持っている。それは、ステークホルダー、すなわち評価によって何らかのリスクにさらされる可能性の

ある個人あるいはグループが、評価対象に関して認識している要求 (Claims)、懸念 (Concern) および争点 (Issues) を評価のアルゴリズムに組み込んである。ここで、「要求」とは、ステークホルダーによる主張であって、評価対象に対して肯定的な内容をいう。例えば、ある教育方法を導入すれば、学生の読む (reading) 能力は標準テストのスコアで一年以上改善されるであろう、という主張は要求である。「懸念」とは、ステークホルダーによる主張であって、評価対象に対して否定的内容である。例えば、コンピュータを利用する教育の導入した場合、学生の計算能力は後退するであろう、という主張は懸念である。また、「争点」とは、合理的に思考する能力を持った人が賛成しかねるような状況をいう。例えば、小学校に AIDS 教育を導入しよう、という主張は、争点である (図表3)。

図表3 評価の焦点

	事前決定型評価	状況対応評価
焦点の決め方	第3世代までは、何を調査するのは評価を実施する以前にクライアントと評価者との協議に基づいて決められる	ステークホルダーが評価対象に持っている評価の焦点、要求、懸念、争点は相互作用の進行と共に変化していくので事前に決めておくことはできない
焦点の内容例	第1世代：生徒の能力を表すと考えられる変数のそれを示すテストの点数 第2世代：教育プログラム目標と生徒の成績との関係を記述 第3世代：評価モデルにより異なる	ステークホルダーが評価対象、例えば教育プログラムについて認識している次の事項； 要求：肯定的意見 懸念：否定的意見 争点：常識的に賛成しかねるような意見

異なるステーク (利害) を持った関係者達は、それぞれ異なった要求、懸念、争点 (Claims, Concern & Issues, 以下 CC&I) を持っているであろう。これらステークホルダーの持つ CC&I を探し出し、解決のために対応することが、「状況対応型評価」における評価者の仕事である。

評価される人々には、多くの異なったステークホルダーが存在することを、グバとリンカーン (Guba と Lincoln) は、1981 年出版の「効果的な評価 (Effec-

tive Evaluation)」のなかで指摘している。そして、彼等はステークホルダーを大きく3つのクラスと、サブクラスに分類している。

- ① エージェント (agents)：評価結果をプロデュースする、使用する、あるいは実施することに関わる人々を指す。エージェントには、以下の人々が含まれる。
 - a) 評価プログラムを開発する人 (developers)
 - b) 地方と国を含む出資者
 - c) 現行プログラムに対して、何らかの改善 (例、地域のニーズに合った) あるいは変更の必要性を認識している評価者 (assessors)
 - d) 評価対象 (例、ローカルなプログラム) を開発、あるいは使用することを決定する立場にある意思決定者
 - e) 評価の結果、影響を受ける施設・補充品・材料などの供給者
 - f) 評価を依頼する人 (クライアント) あるいは契約者
 - g) 評価の結果、実行することに関係ある人 (例、授業を担当する先生、スタッフ、警察官、看護婦など)
- ② 受益者 (beneficiaries)：評価結果により何らかのかたちで利益を得る人を指す。受益者には、以下の人々が含まれる。
 - a) 直接的な受益者、評価を設計する際に“ターゲットとなるグループ”
 - b) 間接的な受益者、直接的な受益者との関係が改善されたり、強化されるなどプラスの影響を受ける人
 - c) 評価の結果、なんらかの利益を得る人、例えば、教材を提供する出版社、必要とされるサービスを供給する業者など
- ③ 犠牲者 (victims)：評価結果から何らかのマイナスの影響を受ける人々を指す。犠牲者のなかには、何らかの問題が生じた場合、不利益を被る受益者の一部も含まれる。犠牲者には、次の人々が含まれる。
 - a) 評価対象から排除される人、例えば、才能のある人向けの特別なプログラムから排除された普通の生徒など。
 - b) マイナスの影響を受ける人、例えば、元の学校施設が障害を持った生

徒達に使われることになったため、遠方の学校にバス通学することを余儀なくされた生徒達とその家族

- c) 評価結果により、政治的に不利益を受ける人、例えば、発言力や影響力の低下、権威を失う人
- d) 評価結果により機会損失を受ける人、例えば、教材を競合他社に奪われた出版社、あるいは必要な経営資源を別のベンチャーに投資していたほうが良かった人

「状況対応型」評価という名称は、評価対象に関して様々なステークホルダーの CC&I を探索するだけでなく、ステークホルダーから解決のための情報を収集し、対応することに由来している。つまり、評価は異なったステークを持った人達が、共通の構成物を構成するための相互作用のプロセスであるから、より洗練された構成物にしていくための手法が必要になる。そこで、構成主義的パラダイムに基づいた「解釈学的・弁証法的 (Hermeneutic Dialectic)」プロセスと呼ぶ手法を採用する。

「状況対応型」評価は、すべての CC&I に関し、ステークホルダー・グループ間で、コンセンサスに到達することを理想としているが、それは不可能ではないにしても、極めて稀である。通常、コンフリクトが残り、その解決のために、外部から新しい情報を入手することが必要になる。コンフリクトを解消するため、新たな利用可能な情報が、評価者よりステークホルダー・グループに与えられる。次に、評価者は交渉議題を準備し、交渉の場を設定し、そしてその進行役を務める。交渉議題に関係あるステークホルダー、もしくはその代表者は、争点となっている議題を解決するため、交渉に出席する。交渉の結果から生まれてくる最終的な結論は、ステークホルダー間の交渉の産物である。得られた結果は、もはや第 3 世代までの評価法の特徴である評価者やクライアントの専有物ではない。もちろん、交渉の席で解決できなかった項目は、依然争点として残る。しかし、各ステークホルダーは少なくともコンフリクトは何で

あり、他のステークホルディング・グループが、争点に対してどういう態度をとっているかを理解する。その後、再評価を行うための準備がなされる。このような繰り返しは、「状況対応型」評価の特徴である。評価は完了することなく、利用可能な時間など資源の物理的制約のために中断されるだけである。

状況対応型評価には4つのフェーズがある。それらは反復されることもあり、また重複することもある。

第1フェーズでは、まずステークホルダーが特定され、各ステークホルダーは主張したいと思っているCC&Iを文書で提出することを求められる。

第2フェーズでは、各ステークホルディング・グループが提出したCC&Iは全てのステークホルディング・グループに公開され、各グループはCC&Iに対する自分の考え、コメント、反論、同意などを提出するよう求められる。このフェーズで、はじめに提出されたCC&Iの多くは解決される。

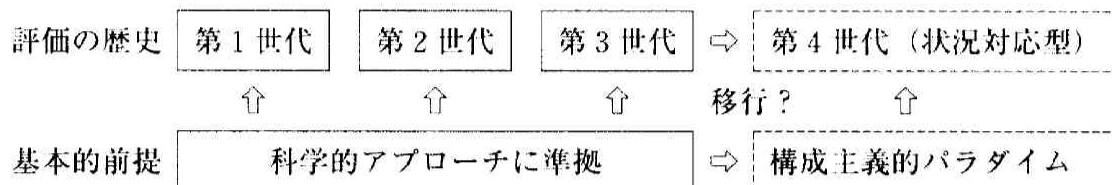
第3フェーズでは、前のフェーズで解決できなかったCC&Iを解決するために、評価者は新たに、情報収集活動を行う。情報収集用のフォーマットは基本的な論点が、クレームであるか（例えば、クレームをテストする情報が集められる）、懸念であるか（その懸念がどれほど正当なものであるかを示す情報が集められる）、それとも争点（賛否両側のそれぞれに対する支持と反論など）によって変わってくる。収集すべき情報は定性的かもしれないし、定量的であるかもしれない。状況対応型評価では、多くの人が誤って信じているように、定量的な分析をなんら排除するものではない。解決できなかったCC&Iに関する情報であれば、どのような情報でも収集する。

第4フェーズでは、評価者の指導の下で、各論争点に関するコンセンサスが得られることを目指し、ステークホルディング・グループ間の交渉が行われる。

しかし、必ずしも全ての論点が交渉の場で解決されるわけではない。残された論点は、次の評価のメイン・テーマになる。

(2) 構成主義的方法論 (Constructivist)：我々はこのアプローチを、今世紀に使われていたすべての評価を特徴づける科学的方法論に代わる方法論として提案する。それは従来の科学的アプローチとは全く反対の、構成主義的パラダイムと呼ばれる考えに基づいている (図表4)。

図表4 第4世代評価へ



存在論的には、このパラダイムは客観的な真実 (objective reality) が存在することを否定する。そのかわりとして、ここで言う“真実”とは、社会的な心の構成物であって、そのような構成物は、個人の数だけ (明らかに多くの構成物は共有されているが) 存在するという、主張である。我々は科学自身も、これら構成物の一つであると考ええる。従って、我々は科学だけが“正しいあるいは真の構成物”であるとして、受け入れるよう要求されるのであれば、反対はしない。

また、認識論的には、このパラダイムは主体-客体二元論 (subject-object duality) の可能性を否定する。その代わりに、探究の結果から得られる構成物が存在するのは、観察者と被観察者との間の相互作用があるからであって、構成物が評価プロセスのなかから現れてくるものは、文字通り創り出す、と考える。

方法論的には、このパラダイムは、科学を特徴づける操作的、実験的なアプローチの代わりに、解釈学的・弁証法を採用する。解釈学的・弁証法は、構成される“真実”を創り出す観察者 (評価者) と被観察者との間の相互作用の利点を十分に活用するので、観察する時点で得られるあらゆる情報に基づいた洗練されたものになる (図表5)。

構成主義的な方法論を使用した研究結果は、科学的な研究によって得られるであろう結果とは、全く異なる。例えば、

- ・ “真理 (truth)” とは、客観的な真実と一致することではなくて、十分な情

図表5 前提とするパラダイム

	従来の信念（科学的）	新しい信念（構成主義的）
存在論	<ul style="list-style-type: none"> ・ 実在論的存在論 ・ 観察者の意図と無関係に客観的な真実が存在する。それは、しばしば因果関係の形をとる自然の普遍的な法則に従っている 	<ul style="list-style-type: none"> ・ 相対論的存在論 ・ “真実”は、社会的な相互作用を通して構築される ・ この構成物は自然法則、因果等の普遍的な法則によって支配されるものではない
認識論	<ul style="list-style-type: none"> ・ 二元論的客観的認識論 ・ 観察者が、評価対象から独立して、かつ自分の価値観と無関係に、それを観察することは可能である 	<ul style="list-style-type: none"> ・ 一元論的主観的認識論 ・ 観察者と観察対象は独立ではありえない。研究成果は、観察者と被観察者との相互作用の創造物である
方法論	<ul style="list-style-type: none"> ・ インターベンショニスト（Interventionist）的方法論 ・ 正しい法則を発見するためには、観察対象をその文脈から剥離することが必要である ・ その結果、予測することや制御が可能になる 	<ul style="list-style-type: none"> ・ 解釈学的方法論 ・ 繰り返し、分析、クリティック論法を繰り返し、調査者と回答者の間から洗練された構成物をつくりだす

報を持ち、洗練された構成者の間でコンセンサスに達したものである。

- ・ “事実（facts）”とは、同じ価値観を持った人々のなかで意味を持つ。
- ・ “原因”と“結果”の関係は、相対的な関係であって、全ての相互作用のあるステークホルダーと関係がある。
- ・ 現象は、それが研究された文脈のなかでしか理解することができない。一つの文脈から得られた知見を別の文脈へ一般化することはできない。問題も解決案も一つの設定から別の設定へと一般化することはできない。
- ・ 評価の結果作られるデータには、事実とその言葉のもつ意味に価値の問題が含まれている。
- ・ 評価者はステークホルダーにとり、構成物を創り出す際の個人的パートナーとなる。さらに評価者はステークホルダーと共に洗練された構成物を構成する際のまとめ役になる。その上、評価者は交渉の場における、ステークホルダー間の仲介役にもなる。

このような主張は、一見非合理的に見えるので、多くの人から拒否されそうである。しかし、近年このパラダイムが提起する評価法を支持する人が増えていることも事実である。評価者を含む全ての人が、従来の科学的評価法こそ選択すべきパラダイムで、という意見に同意するとは思えないが、多くの分野で科学的実証主義に対する問題が提起されていることも、また事実である。

7) 状況対応型構成主義的評価を採用することの得失

ここでは、状況対応型構成主義的評価を、採用すべきかどうか、その得失を検討する。従来の評価法を支持する人が、明らかに状況対応型構成主義的評価への移行に対して、次の様な懸念を抱くであろう。

① 構成主義には、「物事には確実なものは何もない。探究が行われても客観的真理は得られない」、という暗黙の了解がある。物事が実際どのようなものであるか、またどのように機能するかを知ることは出来ない。従来の評価者は、このような不明確な評価にとうてい耐えられない。さらに、評価によって真理を見つけることができないならば、評価する意味は一体どこにあるのか、疑問に思う。

② 状況対応型構成主義的評価へ移行した場合、全てのステークホルダーが評価者やクライアントと同様に、評価のあらゆる段階で決定的に重要な役割を演ずることになる。従って、評価プロセスをコントロールすることが出来なくなる恐れがある。さらにコントロールできない場合、手法的にもまた政治的にも重要な影響をもたらす恐れがある。すなわち、あまり評価手法に詳しくない人が意思決定を行う立場に立った場合、その特権を行使することにより、調査の技術的適切性が脅かされる危険がある。さらに、ステークホルダー達に決定する権限が与えられたならば、政治的権力も手に入れることになる。調査の手法が、政治的な意見を異にするグループの間の主導権争いの対象になるかもしれない。

③ 状況対応型構成主義的評価にコミットすることは、アルコール依存症を改善する、成績の低い学生を再教育する、などの課題を達成するためのイン

ターベンション（すなわち処置、プログラム、マテリアル、戦略など）を見つけた可能性を放棄することを意味する。インターベンションの一般化を仮定するための基盤がない、あるいは明確な因果関係に基づいたインターベンションを考案するための基盤がないとするならば、社会が究極的には自分を取り巻いている様々な問題に対処できる、と信じる理由は殆どなくなる。逆に、状況対応型構成主義的評価の基本的前提を受け入れることは、いつか社会的な問題に対する解決を見いだすことができる、という望みを放棄することである。そしてもし、すべての問題を、特定の社会的慣行と文脈、文化、経済などに合わせて対処しなければならないとしたら、社会は不可能なほど大きくかつ困難な課題に呑み込まれてしまうだろうことを従来の評価者はおそれている。

これに対し、状況対応型構成主義評価を信奉する人は、社会の諸問題に対する有効な解決策を考案できないのは、一般的な解決策を見つけなくてはならないと考える先入観のためである、と答えるだろう。伝統的な評価法に付随して見えてくる確実性を、状況型構成主義的評価がもっている相対主義的な性質に置き換えることは、決して、“何でもあり主義（anything goes）”的な姿勢をとることではない。そうではなくて、その違いは、評価プロセスから得られる複数の構成物のうち、どの構成物が望ましいかをどのようにして決めるのか、である。伝統的には、このような比較は、どちらの構成物が、真実により近いのか、という基準で行われてきた。しかし、“客観的な真実”，が存在する可能性が否定されると、そのような基準は消滅し、もっと微妙で洗練された識別方法が必要になる。

状況対応型構成主義的評価者が心すべきことは、次のようである。すなわち、どれほど大勢のコンセンサスによって支持されていようと、現在普及している構成物に対し絶えず着目し、いつでも構成物を改善・変更する、更に、探索の結果、別の構成物の方がその場合にある人にとってもっと合理的で、適切であることが明らかになった場合は、現在信じられている構成物を否定してでも、新しい構成物を提案するだけの用意が出来ていなくてはならない。

状況対応型構成主義的評価にコミットすることは、伝統的な評価者が自分は

真実を発見した、と確信するときにあれほど容易に仮定した傲慢さを、我々は物事が、“真に”どのようなものであるかを知ることは決してできない、物事がそのようであるかについて、我々が持つ構成物は、探究自体によって創り出されるのであって、神秘的な“自然”によって決定されるのではない、という洞察に相応しい謙虚さで、置き換えることである。确实性を相対性に、コントロールを機能強化に、一般化された説明を局所への理解に、傲慢さを人間性に、それぞれ置き換えることは、第4世代評価者が得ることができる明らかな一連の利点である。

読者は従来の評価法か、それともこの評価法を採用するかの得失を、自ら決めなくてはならないであろう。

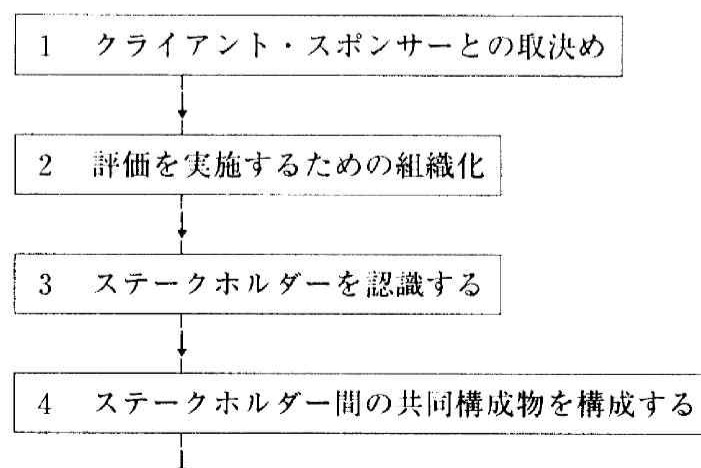
2 第4世代評価のステップ

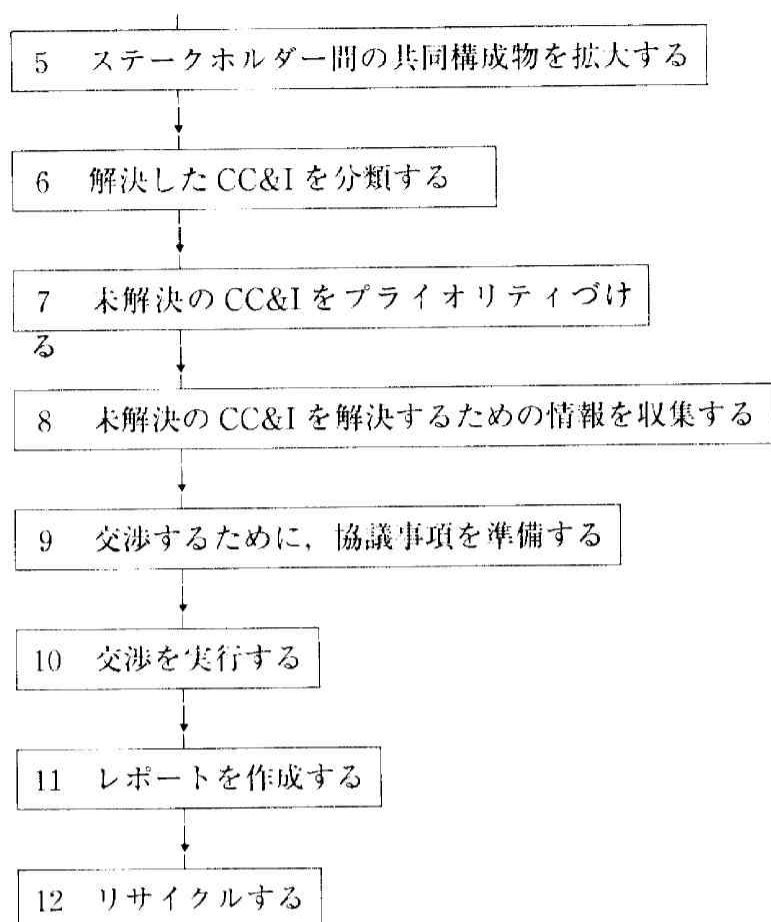
第4世代評価はステークホルダーの要求、懸念、争点(CC&I)という概念を使って、異なった意見を持ったステークホルダー間で合意を形成していく評価法である。ここでは解釈学的・弁証法的サークル(Hermeneutic dialectic circle)の原則と自然主義的探究法(Naturalistic inquiry)を使って、第4世代評価の実践ガイドラインを紹介する。

評価は、以下のフローチャート(図表6)で示されるように、12のステップから構成される。

[ステップ 1] クライアントとの取決め

図表6
第4世代評価のステップ





評価は、法的そして財政的立場に立って、評価を依頼する立場にあるクライアントとスポンサーと呼ばれる、あるパーティ（関係者）あるいは関係者達によって委任される。まず、評価を実施するにあたり、評価を実施する者は、どのような仕事を遂行するのか、に関する書類、すなわち契約書が当然のことながら必要とされる。契約書には、次のポイントをカバーする必要がある。

(1) クライアント（依頼人）あるいはスポンサーの確認：まず、評価を委任される人は、クライアントが誰かを認識する必要がある。クライアントの名前、タイトル、所属する組織は？、意思決定者か？それとも資金提供者か？あるいは評価の結果を利用する人か？などを認識する。

(2) 評価対象の確認：次に、評価される対象は何かを認識する必要がある。評価対象の性質や特徴は契約するドキュメント（書類）に記述してあるけれど

も、ドキュメントそれ自身からは何も情報は得られない。ドキュメントを読んだ人は異なった解釈をするだろうし、その結果異なった構成物を作ることになるだろう。クライアントあるいはスポンサーによって保持されている構成物は、公式上の構成物と呼ばれている。しかし、評価を実施しているプロセスで、異なった構成物が支配的になる場合がある。フィールド調査により、実施されている構成物と公式上の構成物間の違い、および実施されている構成物間の違いが明らかになるだろう。そして、もし評価対象がハッキリしていなかった場合、実行すべき構成物の内容 (implemented version) は、時間の経過と共に変わっていくだろう。例えば、もし評価を実施するエージェントが、教師の場合、評価対象をローカルな環境に合わせたり、資源の配分、スケジュール、マテリアル (教材) 入手の困難さ、などの理由で、変更を余儀無くされることもあろう。勿論、優秀な評価者は評価のプロセスで起こるであろう、そのような変化をモニターするためのメカニズム (措置) を事前に用意しておくだろう。事実、解釈学的・弁証法的評価法では、アウトプットである構成物も、変わってくることもありうるので、評価プロセス中にモニターするメカニズムを見つけることもありうるし、変化に寄与することもある。それにもかかわらず、まず①最初に、クライアントの最初の構成物を書類の形式で用意する、②次に、評価対象が評価のプロセス中に、変化する可能性があること、を契約書のなかに折り込む、③クライアントの最初の構成物は、評価の結果明らかになるであろう構成物に対するベンチマークになる。勿論構成物は、時間の経過とともに、依頼人の構成物に変化を与えるが、排除するものではない。

- (3) 評価目的の記述：グバとリンカーンは“効果的評価”の中で、評価に対する目的は次の4つの方法で概念化できることを示している。これらは評価目的によって異なるが、一般に構成主義者は要約的 (Summative) 評価よりも形成的 (Formative) 評価を採用する傾向にある (図表7)。

① 形式的・メリット評価 (a formative merit evaluation)；形成的メリット評価

図表7 要約的評価と形成的評価

	形成的 (Formative)	要約的 (Summative)
メリット (Merit)	① 形式的・メリット評価	③ 要約的・メリット評価
価値 (Worth)	② 形成的・価値評価	④ 要約的・価値評価

とは、ある評価対象の本質的価値をアセスすることに関するもので、かつ改善をはかる意図をもったものである。例として、提案された学校のカリキュラムが従来のカリキュラムと比較して、近代性、統合性、継続性、順序性、などの点で、従来のカリキュラムと比較して、より改善されたものであるとアセスすることができた、という例をあげておく。この評価法は、評価対象をデザインしたり、新たな評価を構築する開発チームにとって、大変興味がある評価法である。また、評価結果に対する関心があるオーディエンスは評価をデザインするチームである。

- ② 形成的・価値評価 (a formative worth evaluation)；形成的価値評価とは、ある評価対象の非本質的価値をアセスすることに関するもので、かつ改善をはかる意図をもったものである。例として、提案されたカリキュラムが望ましい結果、すなわち、どの程度応用能力が向上したか、などをアセスするものである。この評価法は、現評価対象がどの程度実用的か、そして実施するための手段に関心があるグループにとって、最も興味がある評価法である。
- ③ 要約的・メリット評価 (a summative merit evaluation)；要約的メリット評価とは、ある評価対象の本質的価値をアセスすることに関する評価法で、なおかつ近代性、統合性、などに対するミニマム・スタンダードを満たしているかどうか、を決める意図を持ったものである。この評価法は、評価対象を利用したり、マーケティングすることに関心あるグループにとって、最も関心があるものである。オーディエンスは、評価結果を利用する可能性のあるグループと実際に利用する人である。
- ④ 要約的・価値評価 (a summative worth evaluation)；要約的価値評価とは、ある評価対象の非本質的価値をアセスすることに関するもので、その結果

が実際に使われるかに関するものである。評価結果が、実際に採用された場合に、その結果によって影響を受けるグループに最も関心があるかもしれない。評価に関するオーディエンスは、評価結果に基づいてなんらかの意思決定をする人々である。このタイプのオーディエンスは様々である。

- (4) 依頼人あるいはスポンサーからの同意書（含む付帯条件）：評価を解釈学的・弁証法的方法で実施することに同意するクライアントあるいはスポンサーからの合意と、さらに付帯条件について記述する。同意書には、高潔な立場にいる人からの意見・考えが含まれていること、交渉すべきパーティのなかで、明確かつ効率的にコミュニケーションする能力のない人達（例えば、子供やメンタルにハンディキャップをもっている人、精神病者や自己管理できない人）を対象として除くこと、パワーを共有する能力のある人、説得的な交渉の場において、相手の意見を聞き・受け入れる能力のある人、価値観の違いを認めることができる人、コミットする能力のある人、などの記述が含まれる。
- (5) 評価者は、評価を実施中に、ステークホルディング・オーディエンスが出現する可能性があることを記述する：もし、ステークホルダーのCC&Iが、評価の主な対象であるならば、関連するステークホルダーには、どのような人々が含まれるのか、探し出したり、認識しておく必要がある。クライアントは、評価者がその意図について、知っていなければならない。評価を実施する以前に、ステークホルダー・グループを特定できた場合、契約の際に記述しておく場合もある。また、クライアントは、評価のための調査がどのように進行しているか、さらにその途中、新たにステークホルダーが出現する可能性もありうる、ということを理解しておくべきである。最後に、ステークホルダーも、評価を実施するメンバーに入るので、ステークホルダーも建設的に解釈学的・弁証法的条件に同意することを求められていることを、契約条文中に規定しておくべきである。

(6) 評価方法に関する記述：どのような評価方法を使うかを契約条文に簡単に記述する。

(7) 記録やドキュメントをアクセスする保証：記録やドキュメント類をアクセスすることに関する保証を契約条文に織り込む。ある資料は、法的に公開を禁じられてるものもあるし、またある資料は、公開されているものもある。しかし、多くの資料はその中間（グレイエリア）に属している。従って、評価者は契約の際に、これらの記録やドキュメント類が利用可能かどうか、について記述しておく必要がある。

(8) 情報に関する守秘義務：評価者は、法的に情報源の秘密と匿名を保証することを記述する。

(9) 利用されるレポートに関する記述：レポートは評価者の結論やリコメンデーションだけではなく、評価のフレームや実行を促す内容が含まれる。各オーディエンスは、それぞれ CC&I が異なっている。また、オーディエンスによって、その関心の度合いもまた異なっている。そのうえ、各オーディエンスの反応も異なっているので、異なったレポートが必要になる。評価者は各ステーク・ホルダーに対して、どの程度のレポートを必要としているか、あるいは欲しているか、を交渉するオプションを保持する必要がある。

(10) 技術的仕様をリストアップする：クライアントは、次のことについて知っている必要がある。

- ① エージェント：評価対象に関わる人々・出資者・アセスメントする人・評価結果を利用する人・施設等の提供者・クライアント・評価対象の実施に関わる人達を指している。誰が評価活動を実施するのか？、彼等の信任状は？、経験は？、など、クライアントはエージェントについて、その妥当性、信頼性などについて、知る権利を有している。

- ② テンタティブなスケジュール：評価プロジェクトには、なんらかのマイル・ストーンを示すことは不可能かもしれないが、しかしながら、イベントを推測することや、おおよそどのくらい評価活動に時間がかかるかを推測することは可能である。
- ③ 予算：評価者にとってもクライアントにとっても、現実には評価を実施するうえで、どの程度費用がかかるかを予想する。例えば、ある一定の作業を遂行するのに、Z時間の仕事をするのにY人の人数を必要とし、費用Xドル、とプロジェクトにかかる費用を推定する。
- ④ 得られるであろうプロダクトをリストアップ：評価プロジェクト終了後、どのようなアウトプットがプロダクト（構成物）として何がでてくるか、予想できる範囲でサンプルを提示する。

[ステップ 2] 評価プロジェクトを進めるための組織化

プロジェクト・リーダー（principal evaluator）は、評価プロジェクトを進めるうえでの4つの問題点（クライアントと契約する評価者にとって）、①評価者を選定し、訓練する、②評価プロジェクト組織を編成する、③評価プロジェクトを実施するために準備する、④実際に評価プロジェクトを進めるうえでの各種要因（ローカル・社会的・政治的・文化的）を考慮に入れる。

(1) 評価者の選定と訓練

今日行われている評価は、その活動を実施するためには多くの人数を必要とする。しかしながら、小人数で評価を実施するよりも、たとえプロジェクトが小さくとも、チーム編成による作業の方がメリットがある。全ての評価者は、インタビュー作業に、等しく従事するわけではないし、最初のゲート・キーパーに対してのインタビューに精通しているわけではない。また、全てのメンバーが過去のドキュメント類を読み・解釈する能力を備えているわけではないし、多様な事に精通しているわけでない。チームならば、異なったステークホルディング・オーディエンスや異なったCC&Iを

持った人達に対して対応することも可能である。また、チーム作業の場合、色々な評価方法や色々な分野に精通した人より編成することが可能である。

プロジェクト・リーダーの最初の仕事は、色々な背景を持った人達を評価者として選定し、チームを編成することである。編成されたチーム・メンバーは、必要とされる2つのコンティンジェンシー、すなわち①チーム・メンバーは評価活動に必要な知識と、実際に行う作業で必要とされるスキル（含む手法）、②リ・ソーシャライゼーション（個人作業からチーム作業へ）について学ぶ。

選ばれたチーム・メンバーの評価能力を一定のレベルに保つために、2つの方法がある。①相手の言っていることを真実かどうかを見分ける能力や評価する能力を向上するため、評価者自身をできるだけ、多くの新しい状況や評価対象である環境にできるだけ身を置いてみる。②経験者の指導のもと、クリニカル（臨床）あるいは訓練状況の下で、実際に評価を体験してみる。すなわち、訓練を受けた探究者による厳しい指導と批判的な目の下で、インタビューしたり、評価対象とされる分野を観察し、被調査者である相手の意見を聞き・まとめたり、ドキュメントを分析する、などの経験を積む。

ここで、チームによる評価は有利な点もあるが、また逆の面もあることを指摘しておく。メンバーが多数の場合、ミス・コミュニケーションの問題が発生するし、チームによる解釈学的・弁証法的方法は、共同構成物の構成に対しては有益であるが、異なった構成物が生じたり、極端な専門化、メンバー間の調整失敗（breakdown of coordination）、個人の確執のほか、人間の努力を無にするような困難な事態が生じる可能性がある。チーム・リーダーはこのような事態が発生することを常に予想し、困難な事態にならないように注意しなければならない。

(2) 評価プロジェクトのスタート

法的、倫理的な面で問題のない人（評価者）に、評価は委託される。評価

活動は、法的に認められているので、一見なにも問題がないように見える。当然のことながら、現場の応答者（Respondant）は、彼らの上司が評価活動に同意しているので、協力するであろう。しかしながら、そのような見方は極めて甘い考えである。各応答者はインタビュアーにとって、最初のゲート・キーパー（Gate keeper）となる。どの組織も、係や課などの組織単位に分かれている。そのうえ、各組織にはそれぞれフォーマルおよびインフォーマルなゲート・キーパーがいる。フォーマルなゲート・キーパーは職務上の権限を持っているし、インフォーマルなゲート・キーパーは、組織において何らかの影響力を持っている。

しかし、どちらのゲート・キーパーも評価活動を支持したり、妨げたりするパワーを持っている。さらに、この評価法の特徴であるが、ゲート・キーパーの背後に多くのステークホルディング・オーディエンス、例えば教育評価の場合、第2次受益者（secondary beneficiaries）として両親、子供、アルコール患者の連れ合い、罪人など、が考えられる。また、評価の結果によっては、犠牲となるであろう人々やフォーマルな組織では、発言する権限を有していない人々も考えられる。これらステークホルダー・グループやサブ・グループにも、それぞれゲート・キーパーがいる。これらグループと合意を得るためには、これらゲート・キーパーと交渉あるいは話し合いする必要がある。勿論、最初の交渉がうまくいけば、次のステップに進むことができる。

インタビュアーは各ゲート・キーパーに、必然的に同じ問い（調査の目的、特定のステークホルダーなど）を発するが、ゲート・キーパーは話した内容に対するリスクが発生したり、グループにとってのペイオフ（損得）が生じることから、これを問題視することもありうる。また、インタビュアーが各グループあるいは個人に、評価活動に協力するよう要請した場合、特別の報酬を要求される可能性もある。

評価活動を進めていくうえで、インタビュアーにとって重要な点は、まずゲート・キーパーから信頼を得ることである。ゲート・キーパーから信頼が

得られれば、自動的に信頼するに十分な情報を得ることが可能になる。応答者がインタビュアーを信頼し、尊敬できるようになれば、前向きにかつ本音で語るようになる。

[ステップ 3] ステークホルダーを認識する

ステークホルダーとは、評価によって、何らかのリスクにさらされる可能性のある集団（グループ）あるいは個人と定義するが、グバとリンカーンは“効果的な評価”のなかで、ステークホルダーを3つに分類している。

- ① エージェント：評価結果を使ったり、利用したり、あるいは実行する人々、すなわち、評価対象に関わる人々、例えば、出資者、アセスメントする人、評価対象を利用する人、施設などの提供者、クライアント、評価対象の実施に関わる人達を指している。
- ② 受益者：評価結果から何らかの利益や便益を得る人々、例えば、a) 直接的な受益者；すなわち評価対象者やターゲットとするグループ、b) 間接的な受益者；すなわち直接的な受益者との関係が改善されたり、強化されるなど、プラスの影響を受ける人々、c) 評価の結果使用された、という事実によって利益を得た人々、を指している。
- ③ 犠牲者：評価結果によって、なんらかの不利益を被る人々、例えば、a) 評価対象から排除される人々、b) 副次的に、マイナスの影響を受ける人々、c) 評価の結果政治的に不利になる人々、d) 評価の結果機会損失を受ける人々、などを指している。

しかしながら、評価途中、新たにステークホールディング・オーディエンスが出現する可能性もある。事実、評価活動当初見つけ出せなかったオーディエンスを見つげ出すことも評価者の役割の一つである。もし、新たにステークホルダーを見つげ出せたら、かれらの CC&I を早い段階で、評価の対象として考慮する必要がある。また、大きなステークホルダー・グループの場合、グループ内のサブ・グループの存在にも注目する必要がある。

例えば、新カリキュラムを導入した場合、ある子供の両親は不利益（犠牲）を被るし、別の人々にとっては、適切なカリキュラムである、という例はよくある話である。評価プロセスのなかで、新たに認識できた多様なステークホルダーに対応できるだけの能力が評価者にならない場合もあるかもしれない。これらオーディエンスを評価の対象とするか、それとも除外するか、あるいはどのような措置をとるのか、を決定するためにはなんらかの手段や方法が必要になる。オーディエンスの相対的ステークを分類するためには、論理的で、かつ倫理的な判断基準が必要になる。なかでも、第3あるいは第4レベルの受益者は、評価の対象外となる。一般的にオーディエンスのステークを相対化することは難しいと思う。なぜなら、各オーディエンスは、別のオーディエンスよりも自分達のステークのほうがより高いランクにある、と思っているからである。

いずれにせよ、新たなオーディエンスを評価の対象としないかは、評価者あるいはクライアントが、きまぐれに決めることはできないし、また人気投票で決めるわけにはいかない。多くの場合、オーディエンスが希望するかどうかは別として、第4世代評価では、考慮の対象とすることは交渉あるいは話し合いによって決めなければならない。この場合、評価者の仕事は交渉する場を設定することである。

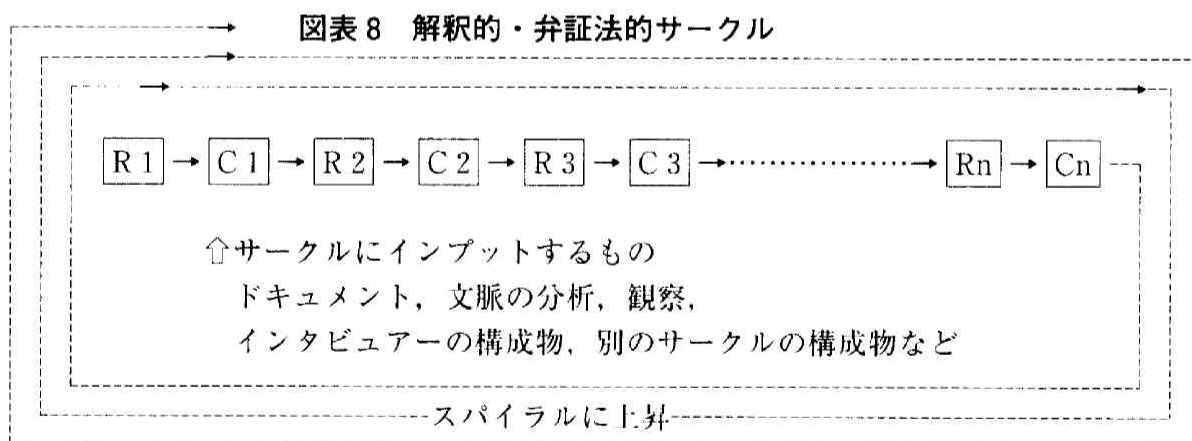
ここで、通常評価に適用している（ある）基準は新たなオーディエンスに対してふさわしいものでないということに言及しておく必要がある。それはまず第1に、多くの場合ステークホールディング・オーディエンスのパワーは相対的なものである。第2に、不十分な知識しか持っていないオーディエンスを評価の対象として、除外するのはリーズナブルである。但し、これらの人達を評価に参加できるレベルに達するに必要なステップを用意することは、評価者にとってモラル・オブリゲーションである。最後に、あるステークホルダーを評価対象とした際、技術的に評価するのが困難な場合には、ステークホルダーとして除外するのが適切であるかもしれない。

最後に、ステークホルダーと認識された場合、交渉に入る。そして、もし合

意がフォーマルに得られた場合、合意書はスポンサーとの契約時に同意を得て作成された部分と、評価の途中に発生してくる可能性のある問題に対して、リーズナブルな意義申立てをする余地のある契約部分と再交渉する余地のある契約部分から、同意文書は構成される。

[ステップ 4] ステークホルダー間の共同構成物を発展させる

ステークホルダー・グループが認識されたら、解釈学的弁証法的サークル・プロセスに入る（図表8）。最初のサークルを形成する目的は、最初の応答者から、できるだけ異なった構成物を見出すことである。第2の応答者は、できるだけ第1の応答者と異なった人を選ぶ。異なった応答者を探すには、最初の応答者に聞くことである。多くの応答者にインタビューした後、インタビューで得られた情報—応答者にとって特に重要なもの、勿論 CC&I も—を分析する。第3番目のサークルでは、特定のステーク・ホルダーグループが評価対象に対して、どのような目で見ているか、それと関連して、彼らにとって主要な CC&I がなにかを探しだすことである。最初は、懸念の範囲に、最後は埋路整然としたものに注意を払う。



以下、その手順を示す。

- ① まず最初に、誰がサークルの最初の応答者（R1）になるのだろうか？、という疑問がおきるかもしれない。その答えは、とくに誰でも良いといえる。なぜなら、最初のサークルを形成する目的は、できるだけ応答者達の異なっ

た構成物を見つけ出すことだからである。インタビュアーは、最初の応答者(R1)を選び、応答者が自由に応答する形式のインタビューを実施し、インタビュアーは、応答者の意見をもとに、最初の構成物(C1)を作成する。最初の応答者は、構成したい論点・争点について、本人の知っている言葉で述べ、記述し、コメントすることを求められる。評価者が記述する際、コメントには応答者本人のCC&Iに関する所見・発言や評価対象についての好き・嫌いなどの好みが含まれる。

- ② 次の応答者(R2)は、できるだけ最初の応答者と異なった人を選ぶことが大切である。異なった応答者を探すためには、最初の応答者に聞いてみることである。最初の応答者(R1)は、応答者(R1)の構成物から最も遠い人を、次の応答者(R2)として指名するよう求められる。第1の応答者(R1)とのインタビューで得られた情報—すなわち応答者にとって特に重要なテーマ、コンセプト、考え、価値、懸念や争点—は分析され、最初の構成物(C1)がつくられる。構成物(C1)はデータ比較分析手法(Constant Comparative Method)を使って、分析され、第2の応答者(R2)にインタビューする前には完成する。

- ③ 次に、応答者(R2)は、最初の応答者(R1)と同様に、自由に発言することが求められる。また、最初の応答者(R1)の構成物(C1)は、第2の応答者(R2)に紹介され、コメントを求められる。第2の応答者(R2)とのインタビューの結果、最初の応答者(R1)の構成物(C1)に対する批判と第2の応答者(R2)の意見を含んだ共同構成物(C2)が形成される。さらに、第2の応答者(R2)は、第3の応答者(R3)を指名するよう求められる。続いて、第1の応答者(R1)と第2の応答者(R2)の構成物(C1・2)をもとに、データが分析され、第3の応答者(R3)が批判する。このようなプロセスを通して、共同の構成物(C3)が必ずしも得られる保証はないが、しかし、コンセンサスを形成していくことが目標である。

- ④ 以上、 $(R_3 \rightarrow C_3, R_4 \rightarrow C_4, \dots, R_n \rightarrow C_n)$ のプロセスが繰り返し行われる。このプロセスを通して、最初構成物(C)の範囲を拡大することに努力が払われる。評価の(部分的)目的は出来るだけ情報を集め、構成物の範囲を拡大させること、そして多くのステークホルダーを出来るだけ認識し、彼らに構成物を形成する機会を与えることである。
- ⑤ その結果、幾つかの構成物(C)が形成されるが、その中でより高い重要な争点が応答者と別の応答者、インタビュアーと応答者との相互作用により認識され、選択され、より理路整然とした構成物が形成されていく。
- ⑥ 同様に、インタビュアーのストラクチャー(質問項目など)も徐々に変わって行く。インタビュアーの質問も当初、あいまいで非構造的なものであったが、徐々に構造的なものに変わっていく。
- ⑦ 最初の応答者のサークル($R_1 \rightarrow C_1, R_2 \rightarrow C_2, \dots, R_n \rightarrow C_n$)が終了したら、2回目のサークルに入る。初期応答者は、別の応答者の構成物に対して大した反応は示さないが、意見や感想を述べる機会を与えられるべきである。このサークルは、スパイラル状に進んでいくが、応答者達から得られる情報や、より洗練された構成物は特定の意見に偏ったり、集約されるわけではない。
- ⑧ また、このサークルに対するインプットとして、別のステークホルダー・サークルからの構成物をインプットすることも可能である。例えば、登校地区を中心に、新カリキュラム導入の調査を実施する場合、担任教師は両親のサークルあるいは教育行政に携わる人々、生徒、地区内の納税者、あるいはカリキュラムを担当した人々などから、インプットとして情報を受けるかもしれない。応答者から得られた構成物に関連した情報や、構成物に関連するドキュメントもあるかどうか調べられる。さらに、関連した資料や議事録などからピックアップされた記録も、よりよい構成物がつくるためのイン

プットとなる。さらに、構成物に関連する観測されたデータも、実際の中でテストされ、インプットとして利用される。このほか、観測されたデータは、インタビューの際、応答者にコメントを求め、その結果は次のフィードバックに利用される。最後に、インタビュアー自身の構成物も、より良い構成物を形成するために活用される。図に示されたプロセスは全ての自然主義的調査法、評価法においても、基本的アプローチは同じである。

[ステップ 5] ステークホルダー間の共同構成物をテストし、拡大する

解釈学的・弁証法的サークルがまわると、個人の構成物は、次第により良い構成物が形成されるようになる。また、この途中で構成物の形成に、プラスに影響を及ぼす資料やドキュメントを活用することにより、さらに洗練された構成物が形成される。評価を実施していく途中で、以下に示される幾つかの資料が必要になる。

- ① ドキュメントおよび記録：インタビューで得られた多くの CC&I は、現存するドキュメントや記録を参考にして明らかになる。
- ② 観察：インタビューで明らかになった因果関係から、より建設的な観察結果が導かれる。
- ③ 学術論文や文献：過去の文献は構成者にとって、構成物を形成する上で、有益なインプットとなる。
- ④ 他のステークホルダー・サークル：最も重要で、最もコンフリクトなマテリアルは、他のサークルからのパラレル・マテリアルである。他のステークホルダー・グループで明らかになった構成物は、評価対象の記述および CC&I と異なったものである。
- ⑤ 評価者自身：評価者自身も、当然のことながら、評価対象および文脈に対する前提知識や意見をもっている。

以上、5つの情報が構成物を形成するための情報源となる。

[ステップ 6] 解決済の CC&I を分類する

個々のステークホルダーのインタビューから表れてきた多くの CC&I は、次のインプットとなる。

[ステップ 7] 未解決の CC&I をプライオリティづけする

多くの CC&I は解釈学的プロセスのなかで、解決するが、一方未解決な項目も残る。CC&I のなかで、未解決項目は、あるグループにとって、高い価値を持っているものでも、別のグループでは全く受け入れられないようなもの、到底コンセンサスに到達しないようなもの、あまり合意の得られないようなもの、などがあげられる。これら未解決項目を、評価者はプライオリティづけする基準を決めなければならない。

- ① 可能なかぎり CC&I を分解する：多分 CC&I を分解すれば、ある種の解決が容易になると思われる項目や未解決な状態になっている項目がこれに該当する。
- ② 反対者もいるけれども、妥協点を見出す可能性のある CC&I：“結局、現時点では学校を閉鎖できない”という例に見られるように、多くの場合現実的で、急な状況のなかで、なんらかの決定が求められるもの。現実的に妥協して処理する必要があるもの。
- ③ 複数のステークホルダーの価値観に関連している CC&I：残った課題のなかで、最も困難で、かつ最も取り扱わなければならないものが、このカテゴリーに属する。しかし、ステークホルダー間で、妥協できないような CC&I は取り扱うことはできない。

[ステップ 8] 未解決の CC&I を解決するために、情報を収集する

CC&I のなかで、“未解決”項目とは、単に共同構成物にならないものを指している。交渉の場に乗らない価値観による違いにより、解決に到っていないのかもしれない。このような場合、解釈学的・弁証法的サークルが、まわる可能性が少ない。しかし、十分な情報が提供された場合、より適切なレベルの意見に集約できれば、あるいは共同構成物が得られる可能性もある。未解決項目

を少しでも減らすため、評価者はあらゆる手段を使い、あらゆる情報を収集する。例えば、あまり再構成できる可能性のないような情報でも、関連あるものであれば、収集する。現存する文献、他の情報源からのドキュメントや記録類も、参考になるかもしれないし、他のプロジェクトに参加した評価者の経験も、情報源として考慮すべきである。新しい手段や、既に出版されている検査による測定方法を使えば、ステークホルダー間の合意が導かれるかもしれない。

[ステップ 9] 交渉するために、協議事項を準備する

未解決の CC&I を認識し、関連する情報を収集する。次に、ステークホルダー間の交渉に対する協議事項を準備する。なお、交渉の際用いるデータは、価値からフリーではないことを十分理解しておく。

- ① 各ステークホルディング・グループの CC&I を注意深く定義する。その際、ある用語を翻訳すると、他のグループにとって必ずしも満足のいくものではないかもしれない。各言葉の定義には、サンプルを伴った十分な解説が必要である。
- ② 課題のなかで対抗する構成物、懸念あるいは要求のなかで、どれが特定の構成物かをハッキリさせる。未解決な CC&I とは、単に価値観の違い、などのような基本的な理由、あるいは情報不足によって生じたのかもしれない、ということを銘記すべきである。すべての集団は、未解決の CC&I が協議項目に載っているかどうか、を確認する。
- ③ 未解決の CC&I を少なくするため、情報を可能な限り収集する。争点の場合、ある争点を支持する、あるいは反駁する情報を、ハッキリと描写すべきである。懸念あるいは要求の場合、支持する、あるいは反駁する情報を準備すべきである。情報は、対決的に示すべきではない。解釈学的・弁証法的プロセスの目的は、ある集団を攻撃することでも、正当化することでもなく、異なった構成物を連結することである。対決ではなく、再構成することである。

- ④ より良い構成物を築き上げていくためには、交渉を上手く運ぶための準備、すなわち教育訓練、ロールプレイング、リハーサル、など準備が必要である。
- ⑤ 交渉の場で、ステークホルダー（あるいはステークホルダーの意見を代表する代理人）が、協議事項に対してアクセプトする可能性があるかどうか、をテストしてみる。

[ステップ 10] 交渉を実行する

交渉は解釈学的・弁証法的プロセスを通して実行されることが望ましい。このプロセスは、利点もあるが不利な点もある。プラスな点は、参加者達は協議項目と内容について、熟知していること。マイナスな点は、ある参加者は協議項目と内容を批判し、自分たちの意見を頑なに主張し、真の交渉を妨げることになるかもしれない。そこで、できれば新しいグループ、あるいは代表者（勿論、いろいろな要因を考慮したうえで）を使うことである。ステークホルダー・グループは、彼らが適当な人物だと判断した人を、交渉の代表者に選ぶ。その際代表者は、建設的かつ解釈学的・弁証法的サークルという条件の下で働くことを約束するという条件を、満たす人である。ステークホルダー・グループにより選ばれた代表者は、彼らのために働くことを委任されている。しかし代表者は、ステークホルダーにただ単に指示されて働くだけでなく、フリーな立場で交渉する場に、出席しなければならない。交渉自体は、解釈学的・弁証法的サークルの下で実施されるが、評価者は、このサークルの場では、ステークホルダーに対する仲介と促進の役割を担う。

交渉は、各未解決のCC&Iに関して、ある合意が得られた場合に終わる。その結果、過去に得られた構成物のもとに、新たに連結した構成物が得られる。しかしながら、多くの場合交渉しても合意に達せず、対立する構成物が残ることになる。交渉の結果、アウトプットとして、3つの可能性が考えられる。

- ① 解決：CC&Iが解決した場合、その後にアクションを伴う。
- ② 部分的に解決：部分的に未解決の状態には、情報が十分得られないな

ど、色々な理由が考えられる。交渉後も、ある CC&I は依然として、未解決の状態かあるいはペンディング（保留）の状態で残る。

- ③ 未解決：未解決な状態とは、相異なった、多分コンフリクトな構成物が存在する状態である。

[ステップ 11] プロダクトとしての共同構成物—ケース・スタディを通してのレポート—を作成する。

第4世代のレポートは、評価した結果と文章だけから構成されるだけでなく、構成者（評価者）がなぜ、このような結末になったかを読者に対してわかりやすく記述しなければならない。次の章で取り扱う評価に対する良いレポートに対する基準はないが、しかしながら、レポートは以下の4つの基準を満たしていなければならない。

- ① 公理的な、自明な (axiomatic criteria) 基準：調査結果はパラダイムを導くような基本的信念システム (basic belief system), に基づいているものでなければならない。
- ② 文体上の基準：レポートの文体は統一がとれていること、全体をカバーしていること、シンプルに書かれていること、およびクラフトマン・シップに基づいて書かれていること、一定の記述様式と構造で構成されていること。
- ③ アクション基準：レポートは、読者の一部にアクションを呼び起こしたり、起こしやすくなるように書かれていること。例えば、公平性、教育的配慮そして活動能力 (actionability) あるいは権限を付与する (empowerment) こと、などの基準である。
- ④ 適用あるいはトランスファー基準：読者に文脈上あるいは置かれている状況のなかで、応用するかもしれない、あるいは行動を起こしやすくなる基準で書かれていること。

[ステップ 12] リサイクリング

この評価法は、特徴として解答よりも、むしろ疑問を喚起する傾向にある。それは、未解決の CC&I が常に存在すること、そして、一定の時間内に、急いでつくられた合意形成物である構成物は、ほんの短い期間のみ保持される。しかしながら、評価として定着したと思われる CC&I は、新たな情報が得られた場合や新たな事例などが出現した場合には、また再評価すべき課題となる。このように見直しは常におこる。従って、第4世代評価による評価の終了は、決して終わりではなく、単に止まっているにすぎない。

3 第4世代評価の評価基準

ここでは、第4世代評価の良し悪しを判断するための基準について議論する。評価とは、統制のとれた探究の一つの形式であると認めるならば、探究の質を判断するための基準が重要になる。また、統制のとれた探究である研究や政策分析とは異なり、評価は独自のプロダクト、オーディエンス、そして成果を持つ。第4世代評価は、既に明らかにされたように、先行する3世代の評価と形式が大きく異なる。従って、評価の良さを判断するための基準を検討する必要がある。ここでは、第4世代評価の適切性を判断するための評価基準について整理し、その特性を明らかにする。

1) 第4世代評価の適切性を判断するための基準

第4世代評価の適切性を判断するには、3つのアプローチがある。第1は、並行基準すなわち信頼できるかどうかの基準である。この基準がこう呼ばれる理由は、伝統的なパラダイム内で長い間利用されてきた（基礎的）基準と相似になるように意図されているからである。第2は、第4世代評価の特徴である解釈学的プロセスそれ自身の特性によって、評価に対してもたらされる貢献を検討することである。第3は、構成主義それ自身の基本的信念体系に埋め込まれた基準を用いることで、第4世代に最もふさわしいものである。この基準は本物らしさ基準と呼ばれる。

1.1) 並行基準 (parallel criteria)

評価のための探究をきちんと実施したかどうかを評価する伝統的基準は、内的妥当性、外的妥当性、信頼性、そして客観性から成り立つ。これら伝統的評価基準は次ページの表の左半分に示す通りである。内部妥当性は、変数間の因果関係があるかどうか、調査によって得られた発見が本当に存在し、機能しているかの程度を評価するものである。外的妥当性は、因果関係の一般化が可能であるかどうかの程度を評価するものである。信頼性は、本質的に評価される現象と現象を評価するために使用される測定ツールの安定性の評価に関するものである。通常、現象は不変であると仮定されるので、その現象を評価する任意の道具は、本質的に同じ評価を生み出す必要がある。さもないと、信頼性がないと判断される。最後に、客観性は明らかに、先入観、価値観、偏見などから独立している程度を、評価するものである。

これに対し、伝統的な基準に出来るだけ近い距離を保ちながら、伝統的な4つの基準に並行する一組の基準を開発することを試みた基準が、並行基準である。それらは信用性、移転可能性、従属性、確証性から成る。その内容と特性は、対応する伝統的基準との対比の形式で、表の右半分に示してある。信用性とは、ステークホルダーによって構成された“真実”と、評価者によって表現された様々なステークホルダーに帰着する“真実”との間の程度を、評価するものである。移転可能性とは、例えば文脈の送信者であるインタビューされるステークホルダーと、受信者であるインタビュアー、あるいは評価者との間に存在する類似性の程度を評価するものである。また、移転可能性の立証には文脈の受信者であるインタビュアーが責任を持つ。従属性は、評価者による方法論の変更が構成物がどのように変化するか、追跡可能であるかの程度を評価するものである。確証性は、データの出所へ逆上ることおよびデータを解釈し、一貫した全体へ組み立てる論理の検証に関するものである。

しかしながら、並行基準によるアプローチは、結局のところ、伝統的基準に無理やり合わせた基準といえる。また、主として方法論的基準であり、方法に加えて成果、プロダクト、そして交渉基準も、等しく重要視する第4世代評価

図表 9

伝統的基準 (traditional criteria)	並行基準 (parallel criteria)
<p>(1) 内部妥当性 (internal validity)</p> <ul style="list-style-type: none"> - 変数間の因果関係があるかどうか - 探求による発見物が本当に存在し、機能しているかの程度 - 客観的世界 (“真実”) と研究で発見されたものとの間にある同型の程度 <p>(2) 外的妥当性 (external validity)</p> <ul style="list-style-type: none"> - 因果関係の一般化 - ある研究の発見物が、他の文脈の下であるいは他の被験者に対しても適用可能かどうかの程度 - 一般化の立証には評価者が責任を持つ - 無作為性と抽出条件が満たされれば、絶対的 <p>(3) 信頼性 (reliability)</p> <ul style="list-style-type: none"> - 研究の一貫性、安定性、正確性など - 妥当性が前提条件 - 同一現象が同一装置を適用すれば、同じ結果が得られるかどうかの程度 - 現象とそれを評価する計測ツールの安定性を仮定 <p>(4) 客観性 (objectivity)</p> <ul style="list-style-type: none"> - 中立性 - 先入感、価値観、偏見などから独立している程度 - 主観・客観の二元論が可能という認識論 	<p>(1) 信用性 (credibility)</p> <ul style="list-style-type: none"> - ステークホルダーによって構成された “真実” と評価者によって表現された様々なステークホルダーに帰着する “真実” との間の程度 <p>(2) 移転可能性 (transferability)</p> <ul style="list-style-type: none"> - 文脈の送信と受信との間に存在する類似性 - 移転可能性の立証には受信者が責任を持つ - 顕著な条件の重複一致の程度に依存し、相対的 - 作業仮説の提示と顕著な仮説が見いだされた時間、場所、文脈、文化に関する広範で注意深い記述の提供が必要 <p>(3) 従属性 (dependability)</p> <ul style="list-style-type: none"> - 評価者による方法論的決定と再構成物の成熟化で生じる変化を意に介さない (無視する) - 評価者による方法論の変更と構成物の変化は、従属性に対する脅威ではなく、むしろ成熟化し・成功した調査の証 - 変更と変化は追跡可能である必要がある (従属性監査) <p>(4) 確証性 (confirmability)</p> <ul style="list-style-type: none"> - データ (構成物、主張、事実など)、解釈、探求結果は文脈と人々に根ざしたもので、評価者の想像の産物ではない - データの出所へ辿ることおよびデータを解釈し、一貫した全体へ組み立てる論理の検証

にふわしくない (図表 9)。

1.2) 自己質管理としての解釈学的プロセス (hermeneutic process)

第2は、解釈学的プロセス自身の特性により、評価の質を判断することである。伝統的評価が、評価の良さを確認するために、その質の外部的客観的評価に事実上依存するのに対し、第4世代評価は解釈学的弁証法的过程によっ

て執り行われる。データの入力、受け取るや分析され、データを提供した応答者に、注釈、修正、改訂などのためにフィードバックされ、さらにまた、そのプロセスが継続するにつれて、創発する共同の再構成物へと組み込まれる。このプロセスでは、間違いが検出されない機会は非常に少ないし、情報の直接的継続的相互作用によって、成果の信頼性は高められる。また、解釈学的プロセスそれ自身が持つ公的な検査可能性によって、秘密性と情報の貧困さの類が予防される。さらに、評価が解釈学的弁証法的原理に従って執り行われるならば、評価者の先入観あるいは偏見が、結果を決めてしまう可能性は事実上はゼロになる。第2のアプローチは構成主義に根ざしているが、しかしこのプロセスには不明確さがあり、説得力のあるものとは言えない。

1.3) 本物らしさ基準 (authenticity criteria)

上記の2つのアプローチに対し、この基準は構成主義パラダイムの基本的仮定から直接導かれるもので、第4世代評価に最もふさわしいものである。本物らしさ基準には次のものから成り立つ。

(1) 公平性 (fairness)

公平性は、異なる構成物とその根底にある価値構造が、評価プロセスで公平な方法で提示され、明確にされ、点検され、そして考慮される程度のことをいう。評価者の役割は、これらの構成物をステークホルダーなどから捜し出し、伝達し、そしてそれらの構成物とその根底にある価値体系が対立している様子を明らかにすることである。

公平性を達成しているかどうかをチェックするために、2つの方法がある。第1は、評価者がステークホルダーを認識し、グループ内の構成物を明らかにしているかどうかである。インタビュアーが、評価活動中全ての潜在的ステークホルダーを識別し、ステークホルダー・グループの構成物を捜し出しているかどうか、監査対象の一部となる。ステークホルダーの構成物を提示することにより、CC&Iに関する対立や、その背景にある価値体系の違いがより明らかになる。公平性を達成するための第2のステップは、オープンな交渉と、そ

の結果として得られるリコメンデーションおよびアクション・プランである。未解決のCC&Iに関するプライオリティ付けを含む解決のためのステップおよび交渉のプロセスは、オープンに行うこと、また対等な立場に立った熟練した交渉人によって実行されること、全てのグループが同レベルの情報を所有している状況の下で実行されること、そしてステークホルダー自身が考案し、全てのステークホルダーが同意したルールに従って執り行われること、などの特性を備えている必要がある。

(2) 存在論的本物らしさ (ontological authenticity)

この基準は、個々の応答者自身の構成物が、改善され、成熟し、拡張され、そしてより精緻化され、その結果より多くの情報を持ち寄り、洗練された構成物になる程度のことで、文字通り個人あるいはグループの意識改善に関するものである。

存在論的本物らしさは、色々な代理経験を与えることを通じて高められる。個々の応答者（ステークホルダーとその他の人々）は、代理経験を通して、応答者自身の“世界”をより深く理解するようになる。応答者が、応答者自身の経験を再評価することを助けるような具体例（すなわち、自身の経験は他人の経験とどのように同じか、あるいは異なるかを見ること）、事例、そしてその他の材料を、評価者が利用可能に出来るのであれば、応答者自身の意識や自覚を高めることに役立つ。代理経験は、十分ではないかも知れないが、特に文脈あるいはコミュニティに対する応答者自身の意識や自覚を拡大させる強力なツールである。。

存在論的本物らしさの基準が達成されたかどうかを示す2つの方法がある。第1は、選ばれた応答者の証言である。個々のステークホルダーが、自分達は現在幅広い範囲の争点（イシュー）を理解しており、以前理解することに失敗した争点を理解できるようになった、という事実を証言できる時、それは存在論的本物らしさの証拠である。第2は、ケース・スタディーのための監査の際、これが評価プロセスの様々な時点で、個々人の構成物が記載された、という記載を持つことである。

(3) 教育的本物らしさ (educative authenticity)

この基準は、異なる構成物を持つ応答者が、あるステークホルディング・グループの構成物に対して、どの程度理解し、評価しているかの程度をあらわしている。ステークホルダー達は評価対象に関連する争点に対し、異なった価値体系を持った人達がいかに異なった解決案を持っているかを見る機会と、あるステークホルダーの構成物と大変異なったステークホルダーの構成物と向かい合う機会を持つべきである。

教育的本物らしさが達成されたかどうかを立証する2つの方法がある。第1は、評価プロセス中の参加者の証言である。証言は、交渉段階で煩雑に生まれるもので、自分と全く異なる人達の構成物を理解した、という事実が証明する。そして、証言は文書化が可能であるばかりでなく、公的に入手可能である。第2は、評価プロセスの最終段階で、解釈学的プロセスを通して、相手の立場を理解あるいは認める、という表現が監査報告に含まれていることである。

(4) 触媒的本物らしさ (catalytic authenticity)

この基準は、評価プロセスによって人々は刺激され、さらなる行動 (action) を促進させたか、の程度として定義される。新しい、そしてより洗練された構成物に到達したとか、異なった構成物を持った人の立場を多少理解するようになった、というだけでは十分とは言えない。評価の目的は、評価結果に基づき、人々が行動するようになったり、そしてあるいは意思決定するようになることにある。なぜなら、参加者が行動するようになって、はじめて第4世代評価は完成されたものになるからである。

この基準が満たされたことを立証する3つの方法がある。第1は、評価に興味をいだくようになった、という証言だけでなく、自発的に行動しよう、という参加者の証言が存在すること。第2は、交渉の段階でテーブルに乗った争点に関する解決策に依存する。つまり、解決策に対して、参加者間で合意に達し、参加者自身が自発的に行動するようになった、という調査結果が存在する場合である。第3は、評価活動終了後、ある一定期間内にシステマティック

なフォロー・アップ調査を実施し、人々の行動が変わった、というアセスメントが存在することである。

(5) 戦術的本物らしさ (tactical authenticity)

人々を行動に駆り立てようとするだけでは不十分である。行動しようと思っても、行動する権限を欠いている場合がある。戦術的本物らしさとは、ステークホルダーと参加者が行動する権限をどの程度与えられているか、に関する基準である。例えば、権限が付与されているかどうかの第1のステップは、全てのステークホルダーとリスクを持った人々が、インタビュー時に発言し、焦点となっている評価項目と、構成物の形成に関与する機会が与えられているかどうかである。さらに、行動する権限は、交渉プロセスの完了時点まで、参加者に十分に付与されていなければならない。

戦術的本物らしさが示しているかどうかをチェックする方法は、3つある。第1は、全てのグループから選ばれたステークホルダーと参加者の証言である。第2は、実際にどのグループが参加し、そしてどのような方法で参加しているかを判断するためには、ある程度フォロー・アップ調査が必要である。最後に、評価期間中どの程度権限が付与されているかどうか、参加者および評価者が判断する場合である。例えば、応答者は参加的であったか？全てのステークホルダーは、ステークホルダーあるいはその代表が交渉の場において、重要な役割を演じたと思うか？全てのステークホルダーは、交渉における能力とテクニックを理解し、活用することにおいて、以前よりも上達しているか？これらの質問に対する答えが一様にイエスならば、そのときは戦術的本物らしさは多分達成されている、といえる。

第4世代評価の品質を評価するには明らかに多くの方法があるが、最も基本的な質問は、“どの基準が適用されるべきか？”，である。ここでは、この質問に答えるため、3つの方法について述べ、各方法で提案された基準が何に由来し、どのように導きだされたかを示した。

各方法とも特定の評価目的に有効である。ここで、品質の良さを評価する基

準は、ある種の仮定に基づいている、ということを忘れてはならない。重要なことは、目的に合わせて、適切な方法を選択することである。第 4 世代評価は、構成主義に根ざした基準で評価されるべきであり、実証主義の基準で評価することは適切ではない。

4 第 4 世代評価のまとめ

ここでは、第 4 世代以前において実施されてきた評価と、第 4 世代として特徴づけられた評価との間の基本的相違について補足し、そしてそれ以前の評価者が演じた役割と比較して、第 4 世代の評価者がいかなる種類の役割を演じなければならないかを要約する。

1) 見解

まず最初に、第 4 世代評価についての見解を述べる。

第 1 に、これまで見てきたように第 4 世代による評価は、社会的・政治的なプロセスの中から生まれる。従来の評価では、社会的・文化的そして政治的要因を評価に取り入れることは、評価の妥当性を損なうものと考えられてきた。これに対し、第 4 世代評価には、文化や政治的要因に立脚している人間の活動全てが含まれている。なぜなら、構成主義的アプローチでは、社会的、文化的、政治的な特徴を資産のように扱い、さらにまた人間の持っている価値体系をも評価プロセスに組み入れるからである。

第 2 に、評価はステークホルダーと評価者との共同で、かつ協力的な評価プロセスの中から生まれる。第 4 世代評価では、一部の人達が評価を進めるのではなく、評価対象に関して、評価者といろいろなステークホルダーとの間で行う共同の相互作用のプロセスを通じて、合意された構成物が評価のアウトプットである。

第 3 に、評価はステークホルダーと評価者との間で、お互いに教え・学び合うプロセスのなかから生まれる。評価者を含め、誰もが評価という目的を追求

する限りにおいて、学習者であり教師である。ステークホルダーは評価者に対し、評価対象に関する自らの見解や構成物について意見を述べる。評価者はステークホルダーの意見に耳を傾け、その構成物を整理・分類する一方、同時に考慮すべき事項にも注意を払う。さらに、評価者は別のステークホルダーに、先の整理された構成物を説明する。そして、別のステークホルダーは、先の構成物に対する見解を述べる。このプロセスは、相互に教育的である。このスタンスこそ、従来クライアントあるいはスポンサーの意向に沿って評価を進めてきた従来の評価法とは明らかに異なっている。

第4に、評価は連続的かつ循環的であるが、評価プロセスからは異なった構成物が導かれる。評価から得られる全ての構成物は、常にある種の問題を抱えている。すなわち、評価された構成物は“真実”でも、あるいは最も“真実”に近いものではなく、ただ単に現存する情報や意見をうまく集約した構成物にすぎないからである。もし、新しい情報を入手したり、新しい事実が現れてきた場合には、これら構成物は更新されるか、廃棄されなければならない。

第5に、評価は評価プロセスからつくられる。一般的方法（例えば、評価手法など）を除けば、評価プログラムを立案するのは難しい。なぜなら、立案段階でどのような構成物が色々なステークホルダーから、評価プロセスに持ち込まれるか、誰も予想できないからである。さらに、あるステークホルダーの構成物が、他のステークホルダーの構成物よりも説得的か、そして合意が得られるか予想できない。また、評価者は事前に、どのCC&Iがステークホルダーにとって問題なのか、あるいは無理強いさせているのか、など事前に知ることも予想することもできない。

第6に、評価は予想できない結果を伴う。第4および第5項目で指摘しているように、評価活動が終了した時点で、どのような評価結果が得られるか、どのようなことを知りうるのか、断定できない。さらに、再構成のための合意が得られるかどうかの保証もない。また、複数の構成物が得られるかもしれない。例えば、評価すべき対象であるプログラムは“成功である”とか、“効果的である”という意見に集約されるかもしれない。一方、別の視点に立てば、

正反対の結果が導かれることもある。(ここで、注意しなければならないことは、“成功”とか、“効果的”という言葉の定義そのものも、合意に達する可能性があるかどうか分からない。思慮分別のある大人であれば、伝統的かつ制約条件の多いクラスとオープンなクラスとでは、子供の教育にとって、どちらが“適切か”，あるいは“効果的か”と軽々に言えないし、どちらの見解を正しいとも、間違っている、ともいいきれない。)ここで、我々は、“どの見解が正しい”，という表現は不適當である，ということを強調しておきたい。我々の見解では，“正しい”とも，“悪い”という表現よりも，どのようにして構成物や文脈が形成され，そして，構成物の裏に潜む価値体系によって，“異なっている”，という表現を使うのが適當である。

第7に，評価は，“真実”を創造するプロセスから生まれる。ここでは，構成物を“発見 (findings)”という用語を使わずに，最もらしいという意味で，“真実”という用語を使っている。評価プロセスを通してあきらかになった構成物は，参加者とステークホルダーが，さらに各ステークホルダーが交渉の場において，自らの構成物について意見を述べ，議論し，理解しあい，創りだされたものだからである。

2) 第4世代モデルによる評価結果

次に，第4世代モデルによる評価と従来の評価の違いについて要約する。

まず，第1に評価が限定的・絶対的評価の時代から，普遍的・相対的評価の時代が変わったことである。従来の評価では，ごく限られた能力のあるグループのみが，構成物を形成する役割を与えられ，そしてその評価結果が尊重されてきた。しかし，第4世代評価では，誰もが評価プロセスに参加できるようになり，評価結果は相対的なものになり，そして多くの見解が併存することが認められるようになった。このことは，従来の評価者の役割と異なり，科学者なかでも社会学者（インタビュアー，評価者などを含む）は，評価のアウトプットである構成物がその後利用されているかどうか，ステークホルダーや参加者と同様に，責任を負わなければならない，ことを意味している。

第2に、評価は、評価対象がどのような状況であるかを調査し、そしてその結果、受け入れがたい現状であると判断された場合、誰が何に対して責任を負わなければならないか、を決めなければならない。第4世代評価では、多くのステークホルダーは何らかの形で評価プロセスに参加し、そしてその意見がなんらかの形で、構成物に反映されるようになる。その結果、責任と同時に義務も共有化されることになる。例えば、従来我々は学校における教育、子供の教育に対する失敗に関して、多くの人々は教師の教育責任を追求してきたが、今日では、学校における子供の教育に関して、多くの意見が反映されるようになった結果、その責任を共有することになる。

第3に、従来の評価法では、その評価報告書のなかに、時には“興味あるデータが参加者から得られた。しかし、その取扱いは難しかった”，という但書きを伴うが、ステークホルダーを無視する傾向にあった。また、評価者がステークホルダーの持っているCC&Iをあまり深く認識しなかったり、ステークホルダーのCC&Iと評価者のそれとが一致せず、ステークホルダーのCC&Iの一部は切り捨てられる傾向にあった。評価者のなかには、ステークホルダーのCC&Iを何らかの形で評価に反映させようとする人もいたが、実際には、“技術的に難しい”などの理由で拒絶されたり、報告者のなかに反映されなかったこともある。さらに、クライアント、スポンサーや評価者などは、評価に関する質問事項や情報を、事前に知っていたり、あるいは共有していたりするが、ステークホルダーは、そのような情報を事前に与えられていないし、あるいは知りうる権利を奪われている。多くの場合、クライアントやスポンサーは情報を公開すべきかどうか、を決定する権利を有している。さらに、彼らは仮りに情報を公開するにしても、誰に、どのような手段で、そして何時の時点で公開するか、を決定する権利をも有している。これに対し、ステークホルダーは、評価に必要なデータだけを収集され、そして利用されるだけの存在である。これに対し、第4世代評価ではクライアント、スポンサーおよびステークホルダーはともに構成物をデザインし、そしてその結果を実行する立場にあるので、このような問題は未然に防ぐことができる。

第 4 に、従来の評価法では、多くの人々は評価結果をあまり知らない傾向にあるが、第 4 世代評価では、より洗練された構成物を築き上げるために、お互いに知識や考えを共有する。交渉の場においても、お互いに価値の共有化を進める努力をする。このような評価プロセスは、従来の評価法と比べて、誰もが評価結果をよりよく理解し、その真価を認めるようになる。

第 5 に、第 4 世代評価では、ステークホルダーも、評価プロセスに参加することにより、得られた評価結果に対して、責任をもつようになるし、評価結果を尊重した行動をとるようになる。過去、クライアントやスポンサーを除いて、ステークホルダーは評価に利用されたり、評価の責めを追わされてきた歴史をもっている。その結果、評価は信頼されたものにならなかったし、また結果に基づいて実行しようとしても、抵抗にあつてきた。しかしながら、第 4 世代評価では、多くのステークホルダーは評価プロセスに参加し、彼らの意見は構成物として反映されることから、実行にむすびつく。

3) 評価者の役割

ここでは、第 4 世代評価における評価者の役割の役割について要約する。第 4 世代の評価者は、各世代における評価者—技術者、記述者、裁判官—の役割を保持しているうえ、さらに再解釈を通して構成物を拡張したり、より洗練された再構成物へと変えていく役割を担っている。

第 1 に、第 1 世代における評価者の役割—技術者（測定者、テスト作成者、統計家）—から、インタビューされる人の話す文脈（言葉）を理解し、そしてその背景にある意味や、データを分析するアナリストの役割に変わる。この役割を担うためには、一般に使われている測定機器やデータ分析に関する知識だけでなく、これらを使いこなす技術的な熟練と技能を備えている能力が求められる。すなわち、第 4 世代の評価者は第 1, 2, 3 世代の評価者が評価に必要なスキル—テスト、測定、定量的分析（テスト、統計的处理など）、定性的分析および分析手法—と同等のテクニカル・スキルをを保有していなければならない。

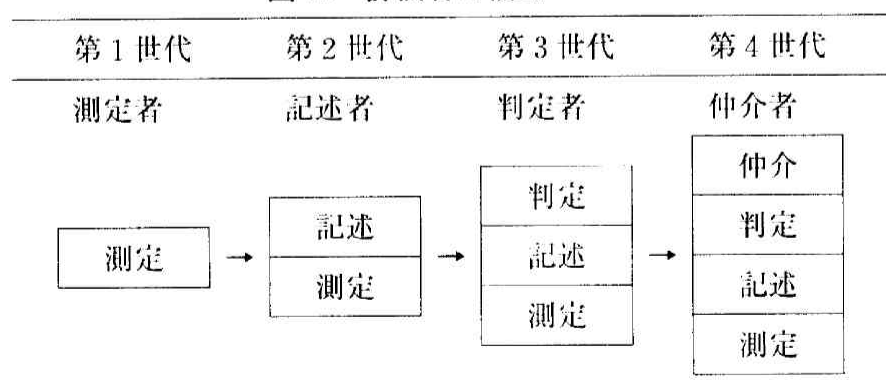
第 2 に、記述者という第 2 世代の評価者の役割から、啓発者と歴史家の役割

に変わる。第2世代の評価者の役割は、評価すべき対象の強みや弱みを正確に記述することであった。しかしながら、第4世代における評価者の役割は、より広範囲なものとなり、インタビューされる人の話す文脈を正確に記述するだけでなく、その背景に潜む要因をあぶりだす能力が求められる。

第3に、第3世代における評価者の役割は、評価対象、例えば、あるプログラムが成功するかどうか、そしてその効果あるいは効用について判定したり、そしていかにしてプログラムを改善するかに関するリコメンデーションについて、その結論を下すことであった。しかしながら、第4世代における評価者の仕事は、あるプログラムの是非を判断する立場から、評価プロセスにおける仲介的な立場に変わる。つまり、評価者の役割は、交渉の場における参加者やステークホルダーに、結論を導きやすくするための情報を提供したり、ステークホルダーが一つあるいはそれ以上の構成物を構成したり、合意を得やすくすることが、仕事の中心になる。

以上のように第4世代評価における評価者は、第3世代までの評価者の役割を果たすだけでなく、仕事の領域を再定義し、広げ、さらに統合したスキルの高い実践家になる。第4世代評価のアプローチでは、従来の評価者の役割のうえに、解釈学的・弁証法的プロセスにおける方法論など4要素が新たに加わる（図表10）。

図10 評価者の役割



さらに、第4世代評価における評価者は、第1に従来の評価におけるコントローラー（管理者）の役割から、参加型評価プロセスにおけるコラボレーター

(共同制作者)の役割へと変わる。

第2に、評価者は、評価対象に対する調査にたずさわる人の役割から、むしろ学びかつ教える役割に変わる。評価者は、ステークホルダーの質問に答えたり、ある仮説を検定したりするだけでなく、多くの構成物のうち、どの構成物がステークホルダー間で対立しているかを学び、そして全てのステークホルダーにその争点を教えなければいけない。

第3に、評価者は、ステークホルダーの持っている構成物を見つけ出す役割から、解釈学的・弁証法的サークルの構成物である新しい“真実”をつくりあげていく役割へと変わる。言い換えれば、評価者は、評価活動における技術者の役割だけでなく、ステークホルダーと共に構成物を形成していくリーダーの役割に変わる。

第4に、評価者は、従来の受身の観察者の役割から能動的な変革者の役割へと脱皮する。評価者の役割はあるステークホルダーの構成物や評価に関連する情報を別のステークホルダーに教え、紹介し、そしてより洗練された“真実”をつくりあげていく評価プロセスにおける仲介者である。その結果、ステークホルダーが彼らの評価結果である構成物に沿って行動するようになったとすれば、評価者はエージェント（仲介者）としての役割を果たした、といえる。

5 情報システム評価に関する若干の考察

最後に、これまで述べられてきた第1、第2、第3世代の評価と、情報システム評価との対応を述べ、現在研究されている評価の問題点を検討し、情報システム評価における第4世代評価の意味を私見を交えて考察する。

1) 第1、第2、第3世代の情報システム評価

グバ・リンカーンらによる第1、第2、第3世代という評価区分は、評価目的と目的を達成する手段の違いによる分類と考えられる。教育における第1世代評価は、学生がどの程度理解出来たか（目的）を、アチーブメント・テスト（手法）などにより測定する。第2世代評価は、新カリキュラムに移行した場合（目的）、そのメリットとデメリットを測定し、記述する。第3世代評価

は、何をどのレベルの学生に教えるべきか（目的）を判定する。すなわち、現状カリキュラムを分析し、その問題点や学習効果を記述し、新カリキュラムあるいはプログラムなりをリコメンデーション（判定）することである。

これを情報システム評価に当てはめてみる。第1世代評価は、事業活動のある機能を分析（測定）し、それに情報化した場合の費用を加味した費用対効果であると思われる。より具体的には、人事機能あるいは在庫管理活動などを分析し、その活動をシステム化する。例えば、人事あるいは在庫管理システムを新たに開発する、あるいは既存のシステムをメンテナンスする場合に必要とする費用、開発・メンテナンスに要する人員（マン）と時間（アワー）と、その費用対効果を計算（測定）し、情報システムの管理部門なりが評価する例が考えられる。第2世代の評価は、情報システムの改善などを目的とした評価であると考えられる。例えば、オンライン・バンキングシステム（On line Banking System）の預金システムと為替システムを統合する、あるいはコンピュータライズド・リザーベーション・システム（Computerized Reservation System）の予約システムと出発管理システム（Departure Control System）とを結合するような場合、そのメリット・デメリットを定量的に測定し、定量化できない効果を記述し、その結果を全社的立場から予算担当部門である経営計画室などが評価する例が考えられる。第3世代の評価は、情報システムの目的を判定することになると思われる。例えば、オンライン・バンキングシステムの勘定系システムと為替系システムとの統合、リアルタイム予約システム、出発管理システム、運航システム、整備システムとの統合など、複数部門に渡るシステムに統合する場合、クライアントより委託を受けた情報システム・コンサルティング会社がコンサルティング（測定+記述+判定）した結果を、クライアントである企業経営者、経営企画室などにリコメンデーションする例が考えられる（図表11参照）。

図表11 第1, 第2, 第3世代における情報システム評価の目的

第1世代	人事活動等、企業活動を分析し、情報化した場合の費用対効果を測定する
第2世代	情報システムの統合する、あるいは改善する場合、その効果を定量的・定

性的に記述する

第 3 世代 複数部門・企業間に渡る情報システムを統合化する，あるいは情報ネットワークシステムを総合化すべきかどうかをを判定する

2) 情報システム評価研究に関する若干の見解

この後，我々メンバーは，現在行われている，あるいは研究されている情報システム評価の関するいくつかの文献⁴⁾を読破し，小幡および筆者，柴，難波，田名部がその要旨を研究部会で報告した。その結果明らかになったことは，経営情報学会における論文の多くは大部分が第 1 世代と第 2 世代に関する研究であると言える。その論文を一二とりあげる。

(1) 鈴木・由良・リユービサ・松井・萩原「コンピュータ統合事業システムの多基準評価法に関する研究」1999

鈴木らは，コンピュータとネットワークを用いた情報を流通・処理し，業務を支援するコンピュータ統合システムを，経済性のみならず人的な側面を考慮して，効果的なシステムを計画・評価するかをテーマにしている。評価基準として，多基準評価手法の包絡分析法の問題点を改善するため，2つのパラメーターと階層的意決定の成果を取り入れた評価手法を提案している。さらに，提案方法の実行性を確認するため，コンピュータ統合事業システムの計画問題を数値例として取り上げ，システム評価への提案方法の有効性を論じている。

(2) 大前義次「情報ネットワーク投資の総合評価」1993

大前は，企業内情報ネットワークを対象に，企業がネットワーク情報システム投資を考える際，投資関連費用については定量化しやすいが，効果の見積もり，特に定性的な効果の評価は一般に困難であるとした上で，定量・定性の両評価項目を含めて，総合評価の方法を提案している。

これら学会誌に掲載された論文，特に第 1 世代に共通に見られる特徴は，実証主義的なアプローチである。研究の関心は，如何に客観的な評価を行うか，測定しにくい評価項目に対して，どのように数値化，客観化するかにある。

次に、実際の企業で働いている人からの情報システム評価に関する論文について触れてみる。

(3) 大浦勇三「情報化戦略と投資評価法」『情報化戦略と投資評価・システム運用管理の実際』研究業書シリーズ No.103, 1998, 企社団法人企業研究会

従来から情報システム評価の研究は、多くの企業で行われてきた。その評価法は、定量効果と定性効果に分類し、要因削減等の比較的明確に定量化できるものは定量効果として、その他のものはすべて定性効果として、スコアリングで評価するか、定量効果から類推する形が多い⁵⁾。稟議書や企画書のなかに、定性的評価の把握しにくい部分を、文章表現に知恵をしぼり、期待効果の不十分な部分をカバーしようとしてきた。しかも、「稟議書さえ通ってしまえば後は開発するだけ、そのプロセスや結果についてはあまり深く追求しない、かりに失敗しても誰も責任をとらない。」、ということになりがちであった。

これに対し、このような事態を避けるため、大浦は情報化戦略の視点に立って、情報システム評価に関して幾つかの提案をしている。情報化投資評価における最新の動向として、まず投資効果の評価がタンジブルなベネフィットからインタンジブルなものにシフトしつつあること、次に投資額の目安として、現在の売上／利益から現在および将来の売上／利益を加味した投資へ、さらに、従来は経営戦略とのリンクが不十分であったが、経営戦略とのリンクが不可欠であるとの観点から、以下のような情報化先進企業の情報化基本方針をあげている（図表12）。

図表 12 情報化先進企業の情報化基本方針

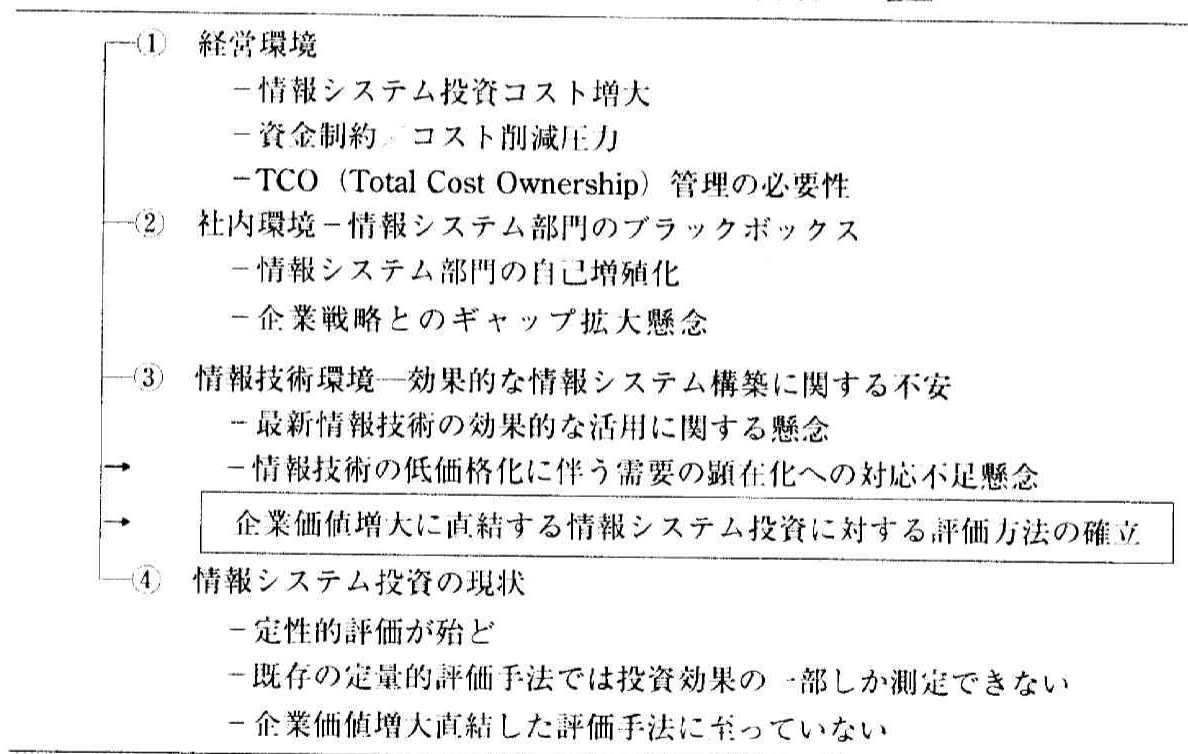
- ① 情報システムはビジネス創出／支援のために存在すること、
- ② 情報システムの企画立案はユーザーを含めた協働作業であること、
- ③ 限られた資源の優先順位付けを徹底すること、
- ④ 外部エクスペタイズ（Expertise）との戦略的パートナーシップを構築すること、
- ⑤ 情報システムはグローバル・スタンダードのもとで構築すること、
- ⑥ 情報システムを通じてのステークホルダー（利害関係者）への「付加価値」提供に

注力すること、

- ⑦ 信頼できる既存技術および最先端技術（検証済）を合わせた有効利用をはかること、
- ⑧ 活用・サポートするシステムやプロダクトを絞り込むこと、をあげる。

この方針の背景には、企業における情報化投資へのニーズは年々増加の傾向にある。例えばパソコンをふくむ通信ネットワークなどの投資をあわせて、ソフトウェア開発／購入、運用・保守、教育、社外委託などの費用も膨大な金額になる。こうした投資額の増加に対して、その経営上の効果が明確化しにくい状況が生まれており、その投資効果に対してトップの判断が難しい状況にある。また、膨大な開発コストをかけた情報システムが、経営面から見て、どの程度効果がもたらすか不明である（図表13）。

図表13 情報システム投資に対する評価方法の確立



従って、情報化投資に対する新たな認識として、情報化投資を、従来の経費や技術の問題として処理すべきではない。投資効果評価の視点は、従来のコスト削減（人員やスペース削減）など目に見える定量的効果から顧客サービス、高品質、クイック・リポンス、戦略的フレキシビリティ、リスク・マネジメントなど目に見えない定性的効果を重視する傾向に移っている。そのうえ、投資効

果は従来の「現在の売上げ／利益」から、「現在、および将来予想される売上げ／利益」をもとにしたものにシフトしつつある。さらに、経営戦略とのリンク強化、すなわち情報化戦略と経営戦略の統合化が重要である、と大浦は主張する。

次に、情報システム評価の基本的考え方として、3つの側面①全社の視点、すなわち「全体として」現状はどうなっているのか、どうすべきなのかを総合的な観点から評価する、②事前（企画案件）での投資評価、すなわち「期待効果対投資コスト」の評価や、優先順位づけが重要である、③事後（稼働）投資評価、すなわち「事前評価と差異分析」や評価手法・評価制度の見直しにつながる、ことをあげている。

そして、情報システム投資効果を享受するのはステークホルダー（顧客、社員、取引先等）全体におよぶことから、その効果は情報システム部門が評価するのではなく、業務（ユーザー）部門が評価し、その結果は、最終的には業績で見る視点が必要である、とする。このような意見の背景には、情報システム部門の意識と、経営トップの求めている情報システムとミスマッチを起こしていること、従来情報システム開発に関して、経営上の重要なテーマとして認識されることは比較的少なく、「経費」や「技術」の問題として扱われ、社内手続きにおいても、経営課題と離れたところで処理されてきたことがあげられる。

今後、経営トップから情報システムを評価していく仕組みをプロセスとして構築することが必要である。すなわち、①企画段階でのリスクと投資効果の検討、②審査段階でのスクリーニングと投資の適正かどうかの判断、③実行／モニタリング段階で計画逸脱かどうかの監視、④レビュー段階の差異分析とフィードバックによる軌道修正へ、と反映できる評価制度の仕組みが求められる（図表14）。

図表14 情報システム評価制度の仕組み



現段階で、このような情報システム評価を導入している企業は極めて少ない。彼の提案には、情報システムコンサルティング会社など、外部エクスパートイズによる意見の尊重（判定・勧告）、ステークホルダーや経営者層との合意形成が含まれており、さらに発展させれば、おそらく第4世代の情報システム評価を取り入れる可能性があると思われる。それがはたして可能か、有効かは今後の研究を待たなければならない。しかし、ここで取り上げた第4世代評価の持つ意味は、実証主義的パラダイムからの脱却であり、その観点から評価を見直さなければ、従来の評価研究の問題点は見えてこない。指摘できることは、客観化至上主義の評価に問題があることである。現状をよりよくするための評価でなければ、評価とは言えないのではないだろうか。

最後に、我々メンバーは情報システムにおける質的評価に関する研究が幾つかのプロシーディングス⁹⁾に掲載されていることを知り、現在「質的評価」の研究に取り組んでいる。

注

- 1) 文教大学（主査）小幡孝一郎 東京工業大学（現千葉工業大学・幹事）柴 直樹
 東京理科大学 難波和明 岩手県立大学 渡邊慶和
 神奈川大学 百海正一 就実短期大学 恩藤哲哉
 日本大学 齊藤敏雄 関東学院女子短期大学 松下倫子
 横浜国立大学 田名部元成 電気通信大学（現秋田県立大学） 住田友文
 尼崎産業技術短期大学 阿澄一寛 京都経済短期大学（現東邦学園大学）井上秀次郎
 なお、岩手県立大学での学会報告は小幡、柴、渡邊、百海、齊藤、難波、田名部の順
- 2) Magnus Ramage, "Evaluation of Cooperative Systems: First-year Report", Cooperative Engineering Group, Computing Department, Lancaster University, 10 October 1995 なお、「協調システムの評価」の詳細については、経営情報学会『情報システム評価研究部会研究報告書』, 2001年6月参照。担当は以下のとおり。第1章 序論：阿澄一寛, 第2章 評価法についての文献：渡邊慶和, 2章2節, 1 使いやすさの評価：難波和明, 2章2節, 2 情報システム評価：小幡孝一郎, 2章3節プログラム評価：百海正一, 第3章 CSCW に向けて：柴 直樹

- 3) 経営情報学会『情報システム評価研究部会研究報告書』 2001年6月
- 4) 情報システム評価に関する研究としては、以下の論文がある。
 - ① 鈴木 淳, 由良憲二, リュービサ・ブラチッチ, 松井正之, 萩原洋太郎,
『コンピュータ統合事業システムの多基準評価法に関する研究』経営情報学会
Vol. 8, No. 3 Dec, 1999
 - ② 力 利則, 藤野 喜一, 堀江 正之『システム監査における情報システムの
有効性評価モデルの構築』経営情報学会 Vol. 7, No. 2 Sep, 1999
 - ③ 社団法人企業研究会, 『情報化戦略と投資評価・システム運用管理の実際』
研究業書シリーズ No 103, 1998
金巻龍一「アウトソーシング, パッケージソフト活用の留意点と評価」
百瀬公明「情報化技術が急進展するなかでの情報化投資の考え方」
大浦勇三「情報化戦略の視点と投資評価手法」
増田有孝「投資効果を高める統合業務パッケージの適用方法」
小澤 博, 澤田博光「イントラネット活用による情報共有化環境の実現と知的
資産創造の推進—野村総合研究所のイントラネット—」
譲原雅一「情報化投資効果の評価方法」
 - ④ 山田善靖, 小橋康章, 砂川雅彦『仮想現実技術を応用した意思決定支援シス
テムの評価基準』経営情報学会 Vol. 3, No. 1, Oct, 1994
 - ⑤ 大前義次『情報ネットワーク投資の総合評価』経営情報学会 Vol. 2, No. 2,
Dec, 1993
 - ⑥ 富士通総研「情報化投資の経済的影響に関する定量的分析調査報告書」1998
- 5) 例えば, 拙稿「日本航空におけるコンピュータ・プロジェクト管理—過去・現
状・将来へ指向—」, 『電子計算組織利用の効率評価システムの調査研究—その2—
相対比較による費用効果分析 評価技法の体系化と運用研究効率性向上のための事
例研究—』財団法人 地方自治情報センター 昭和53年
- 6) また, 情報システム評価に関する研究のなかで, 以下の文献に見られるように,
近年定性的研究が盛んである。
 - ① Allen S. Lee & others, Information Systems and Qualitative Research
—Proceedings of the IFIP—, Chapman & Hall, 1997
その背景には以下の本に見られるような定性的研究の発展がある。
 - ② John W. Creswell, Qualitative Inquiry and Research Design Choosing among
Five Traditions, Sage Publication, 1998

引用文献

1. 大浦勇三「情報化戦略の視点と投資評価手法」社団法人企業研究会, 『情報化

- 戦略と投資評価・システム運用管理の実際』研究業書シリーズ No 103, 1998
2. 大前義次「情報ネットワーク投資の総合評価」経営情報学会, Vol 2, No. 2, Dec. 1993
 3. Egon G. Guba and Yvonna S. Lincoln, Fourth Generation Evaluation, Sage Publication Inc., CA, 1989
 4. 経営情報学会『情報システム評価研究部会研究報告書』 2001 年 6 月
 5. 鈴木 淳・由良憲二, リュービサ・ブラチッチ・松井正之・萩原洋太郎「コンピュータ統合事業システムの多基準評価法に関する研究」, 経営情報学会, Vol. 8 No. 3, Dec. 1999
 6. 拙稿「日本航空におけるコンピュータ・プロジェクト管理—過去・現状・将来へ指向—」, 『電子計算組織利用の効率評価システムの調査研究—その 2— 相对比较による費用効果分析 評価技法の体系化と運用研究 効率性向上のための事例研究—』財団法人 地方自治情報センター, 昭和 53 年
 7. Magnus Ramage, "Evaluation of Cooperative Systems : First - year Report", Co-operative Engineering Group, Computing Department, Lancaster Univerity, 10 October 1995
 8. Mason. R. O, MIS Experiments : A pragmatic Perspective, Benbasat, I., ed. Harvard Business School Research Colloquim Vol. 2, The Information Systems Research Challenge : Experimental Research Methods, Harvard Business School, 1989, pp. 3 - 20

「謝辞」このペーパーを作成するにあたり、小幡および齊藤両先生の報告資料（『第 1 章』, 『第 8 章』 1999.08.30 報告）を参考にさせていただいた。