

大学教育評価の課題と展望^(注1)

大 山 泰 宏

(京都大学高等教育教授システム開発センター)

The Themes and Perspectives of University Educational Evaluation

Yasuhiro Oyama

(Research Center for Higher Education, Kyoto University)

Summary

Evaluation is now introducing into Japanese University at severe speed, especially into it's education. But regrettably, it should be pointed out that this introduction is often being carried out without inquiring into the purpose or context of the evaluation. Hence, evaluation is sometimes misunderstood as a mere political intervention or pressure, which agency (the evaluated sides) let pass or disregard.

In this paper, the status quo of the diverse types of evaluation are discussed, and is proposed what should be done for the actualization and substantialization of the truely effective evaluation. Through the overview about the Participatory Evaluation and Meta Evaluation, as well as the action research of evaluation at the "Research Center for Higher Education of Kyoto University", the social constructivist evaluation theory is showed as one of the effective evaluation strategies.

はじめに

ここ数年のうちに、わが国の大学において、評価に急速に関心が向けられるようになった。いわゆる評価ブームが生じ、授業評価、教育業績評価、大学教育評価など、教育活動のさまざまなレベルにおいて評価が計画・実施されている。言うまでもなく、平成13年度より発足した大学評価・学位授与機構が、この動きを加速させた。しかも、教員の教育業績評価については、やや性急とも思える導入へ向けての動きが、多くの大学で生じつつある。このように大学における評価には、ポリティカルな色あいがつきまとっている。こうした政治的な評価は、評価理論の立場から評価本来の意義や目的を考えた場合、きわめて特殊で変形された評価である。しかしながら、時には誤解され、評価とはそもそもこのような政治的意味合いの強いものだと考えられている。

一方的に査定や管理の手段とされたり、資源配分に結びつけられりする評価に対しては、多くの大学教員が「何かおかしい」と感じている。しかし、それに対して代替となる自律的な説明や反論をしっかりと打ち出すのではなく、「何かおかしい」評価が台風のように通り過ぎていくのを待ち、また、そうした評価に対しては、すべてを開示してみせるのではなく、とりあえずは不利にならないよう、小手先の技術を使い辻褄を合わせるといった実状があることは否定できまい。こうした状況のもとでは、評価は、その本来の目的である組織を活性化するということからほど遠く、組織を疲弊させる「雑用」がひとつ増えたにすぎないことになってしまうのである。

このような状況のもと、喫緊の課題としてなすべきことは、評価とは本来どのようなものであるのかという、評価の本質というべきものを明らかにし、外発的に課せられる評価に対して抗する理論、もしくは、外発的な評価を、組織の活性化に寄与する実質的で実りものにしていくための方策等を、考究していくことであろう。

評価に関しては、現在さまざまな方法、発想、適用領域がある。料理の素材や目的に応じて、さまざまな調理器具

が発達したように、評価も、その目的や領域に応じて、さまざまなものが工夫考案されている。こうした多様な評価に関して、それぞれの特徴および限界を明らかにし、評価という道具の使用に関する留意点を、まずは整理することが必要であろう。というのも、適切に道具を使用するには、選択可能な道具はどんなものがあるのか通覧し、目的や特徴に応じて適切なものを選び、場合によっては用い方自体に工夫を加えることが必要であるからである。いかなる道具も、あらゆる操作や対象に一律に適用可能なものはない。それぞれが、目的や対象に応じて、特化し分化している。評価もこれに同じである。評価の適用可能性をさぐることは、逆に言えば、その評価が何のためにあるのか、何に適しているのかという、特徴が見えてくるはずである。

さらに、本論では、そのような評価の特徴と限界をふまえた上で、できるだけ評価を生産的に活かそうとすれば、どのようなデザインが考えられるのか、考察していきたい。

1. 評価をめぐる考え方

1) 評価の暫定的三区分

日本語で「評価」と一括して表現される事柄は、英語ではいくつかに区分される。例えば、measurement, evaluation, assessment が挙げられ、これらはいずれも、その発想においても方法においても異なっている。この3つの区分は必ずしも厳密なものではないが、本論ではまず、これらの区分をあえて強調し、その差異について考察することで評価を分化してとらえ、後の考察の準備としたい。

measurement は、まさに「メジャー（ものさし）」という言葉が当てはめられるように、特定の測定尺度を基準として、対象を測ることである。この場合、測定尺度自体は、評価が行われる以前に規定されており、測定という行為がなされる時点では尺度自体が反省化・対象化されたり問題化されたりすることは、ほとんどない。

これに対して、evaluation は、「evaluate（値踏み）」というニュアンスを含むように、尺度の価値づけが問題となり、尺度構成自体に関するリフレクションを含んでくる。すなわち、ひとつの尺度ではなく、さまざまな次元の情報を考慮しつつ、評価対象を記述し値踏みするのである。たとえば、一個のりんごに関して、重さや糖度などを測定するのは measurement であるが、evaluation では、さまざまな指標を用いた測定をおこなった後に、それらの測定値に重み付けを行い、りんごの「質」を総合的に判定し、等級や値段という情報に縮約していく手続きを含む。この場合、evaluate へ至るために、どんな指標をとるのが適切であるか、総合的にはどのように判定するのが適切であるかなどが、問題に上ってくる。

assessment に至ると、さらに事態は複雑となる。例えば、このアセスメントという言葉が使用される「環境アセスメント」の例を考えれば分かるように、まずは、評価対象自体が持つ複雑さが増大してくる。しかも、評価対象の現況ばかりでなく、時間を経るとどうなるのか、さらには評価対象と連関している他の事象への影響はどうなのかといった、評価対象を超えたさらに広い文脈において評価しなければならないのである。ここでは、評価の情報は、必ずしも数値化されたりひとつの指標に縮約されたりする必要はなく、質的な記述に限られることさえある。何よりも、評価対象に働きかける主体、その評価情報を利用する主体にとって必要な情報であることが、重要である。医学や心理学において患者や来談者の状態を、何らかの測定メソッドで把握することもアセスメントの一つであるが、ここには対象への働きかけや治療・教育行為への連動を内包したものであることを考えてみれば、容易に分かるであろう。

上記3つの評価において、それぞれに留意される事柄は異なってくる。たとえば、measurement では、対象を「正確」に測定することが、その評価の正否を決定する。すなわち、尺度自体は問題化されないのが、既定の指標による測定が、どれだけ誤差やバイアスなく行われるかといったことが、主要な留意点となるのである。これは、測定における信頼性 (reliability)^(#2) が主要な問題とされていることである^(#3)。evaluation では、測定指標自体が問題化されるので、対象の記述に適切で妥当なものであるか、最小限の指標で最大限の情報が得られるように測定ポイントが設定されているかといったことが、留意点へと付加される。すなわち、評価指標の妥当性 (validity) も問題となってくるのである。assessment に至っては、ここにさらに、測定対象の改善にとって有意義な情報が得られるのかどうかといった、実践へと連動するプラクティカルな意味も、重要な関心事項となるのである。

ここまで述べてきたように、measurement, evaluation, assessment となるに従い、評価のあり方は複雑となり、

評価をめぐる価値や文脈も無視できなくなる。しかしながら、測定評価（measurement）が価値や文脈とは無縁であるわけではない。そこでは、価値や文脈に関する言及が回避されているにすぎないのである。それが何のために行われる測定であるのか、何を目指しているのか、どんな限界があるのかといったことが、本来ならば、measurementにおいてこそ、問われるべきであろう。この問いかけと反省化がなければ、評価の指標・尺度の実在性を素朴に信じ、評価を単なる一方的な裁断や差別化の手段として終わらせてしまう危険性が高い。

実際、評価の歴史を見てみれば、測定評価が、さまざまな差別へと結びついてきたことが見て取れる。たとえば、さまざまな身体的特徴の測定が、人種的な偏見差別を正当化するために用いられたり、あるいは優生学や育種学と連動し断種や不当な閉じこめがなされてきた不幸な歴史は、枚挙にいとまがない（Gould, 1981）。これらが、非理性的な差別意識からなされてきたかといえそうではなく、「人類の向上のため」といった高邁な理想のもとに行われてきたのである。本稿では詳述する余裕はないが、19世紀末のイギリスで（さらに驚くべき事は、米国では20世紀半ばまで）、ラマルクの選択淘汰説、遺伝学、評価・統計が結びつき、人類向上のために、評価によって人為的に選択淘汰し、劣性で病的な種を絶やし優秀な種を残すといった、評価を通じた改善計画が存在していたのである。大局的に見れば、現在の評価のあり方も、その延長上にあると言わざるを得ない。このような過ちは、評価指標そのものを問題化していなかったこと、あるいは、評価が使用される改善計画、すなわち、評価の目的自体を反省化していなかったところに、その悲劇がある。昨今のわが国の大学評価をめぐるポリティカルな動きにも、このような危険性を感じずにはいられない。特定の測定指標にもとづく評価、そしてそれはしばしば、きわめて単純なアウトカム（達成指標）やアウトプットにもとづく評価がおこなわれ、資源配分によって優秀な種（大学）には良好な環境を与え、劣性な種は統廃合や断種をおこない、わが国の大学を世界水準の大学にまで向上させようとする「高邁な」優生学的計画がある。これに抗するためにも、評価が使用される文脈に関するリフレクション、評価の指標自体に対するリフレクションが、より意識的におこなわれなければならないのであり、そうしたリフレクションを含む評価の方法が、考えられなければならないのである。

2) 評価の4世代

Guba & Lincoln (1989) は、評価の4つの世代区分を提唱している。これは、自らが第四世代として位置づける「参画協調型評価（participatory evaluation）」を提案するうえでなされた区分である。この区分は、さまざまな形の評価の特徴と限界を整理し吟味するのに適切であるので、少しここで言及しつつ、評価の区分をおこなってみよう。

1. 第一世代から第三世代の評価

評価の第一世代とされるのは、測定主義の世代である。ここでは、ビネの知能テストなどのように、対象はあらかじめ設定された基準や指標によって測定される。これは、先述した評価の3つの区分では、measurementに当たる。この世代では、評価における測定基準は、評価対象や観察者とは独立に存在するものであると考えられており、実証主義（positivism）の発想のもとにあるという。評価指標や評価の文脈へのリフレクションが乏しく、測定されたものは、評価対象や文脈とは独立であると考えられているので、点数化された指標によって相互に比較される。そして、評価から得られた情報は、評価者が利用するものである。

第二世代に位置付けられるのは、形成的（formative）評価である。形成的評価とは、評価をあくまでも被評価者へフィードバックし、その活動の改善へつなげるという発想のもとにある。かつて、わが国の初等・中等教育で盛んにおこなわれた小テストの流行は、この発想のもとにある^(註4)。この評価では評価は必ずしも数値にする必要はなく記述正式であってもよく、質的評価でもよい。評価の結果は、評価対象へフィードバックされ形成的に使用されることが、前提とされるからである。評価から得られた情報は、被評価者自身が利用したり、評価者の教育的介入等を計画する際に使用されたりする。いずれにしても、評価によって被評価者にとって意味ある情報を得ることが、重要とされるのである。しかしながらここでも、その評価指標は、評価者や観察者からは独立したものであると見なされている。従って、この評価においても、単一の評価指標が、異なる対象や文脈に当てはめられているのである。

第三世代は、基準認定（accreditation）の方式である。これは、法廷での裁判の方法を取り入れたものであるという。評価者は、被評価者（評価対象）を、多様な観点から総合的に評価する。被評価者には、その評価に対し反論

をおこない、自己の在り方に関して自己の視点から説明をする機会が与えられている。この反論をも含めたうえで、評価者は被評価者に対する最終的な評価（決断）を下すのである。この方法の場合、評価者からの一方的な裁断とならぬよう、被評価者からの反論という契機を設けた点に意義がある。しかし、この方式が、大学基準認定において主に用いられていることからわかるように、たとえば「基準認定をするかしないか」というごとく、依然として外的に客観的な基準が想定されている。評価－反論－再評価といった手続きは、評価のバイアスや偏りをなくし、評価という判断を「客観的」かつ「公正」になすために行われるのである。

2. 第四世代の評価

これらの3つの評価方法を超越するものとして、グーバとリンカーンは、参画協調型評価（participatory evaluation）を提唱している。この方法は、現在では、ユネスコやNGO等のプロジェクトやプログラムを評価する際に、しばしば使用される方法である。これは、評価者の観察世界や意味構成の文脈を離れた「客観的」な評価基準の存在を否定する。いかなる評価も、その観察者すなわち評価者が意味づけている価値観や世界観の制約を逃れられない。それは、被評価者のものとは異なっていたり、評価への様々な参加者（stakeholders）間で、しばしば異なっている。すなわち、同じ事象も評価の視点や評価者の意味世界から異なった評価がなされうるということを、前提としているのである。

例えば、国際協力・援助の一環として、あるアフリカの村にポンプ付きの井戸を掘ったとしよう。このことにより、村民は炎天下数キロ離れた井戸にまで毎日水を汲みに行くという、過酷な労働からは解放されることとなる。これは、協力者の側から見れば、村民の生活水準を向上させるために善意でなされることであり、ポジティブに評価されるべき行為である。しかしながら、井戸を設けられた村の住民にとっては、たしかに数キロ離れた井戸まで水を汲みに行くという労は持たずに済むことができるようになるが、そのことによって、毎日の生活のきわめて中心に位置している労働、生活の根幹であった労働が失われてしまい、新しい生活をどのように組み立てていくべきか、途方に暮れてしまうこともあるのである。また、水を汲みに行く合間に交わされていたおしゃべりの時間が失われることで、共同体としての基盤が揺るがされるという事態が生じることとなるかもしれない^(註5)。

この場合、村民にとっては、むしろ井戸がない不便な生活のほうが、生活の幸福感を保証し、「豊かな生活」をもたらすのである。だが、国際協力者が間違っていて村人の側が正しいとか、逆に、国際協力者が正しく村人が間違っているのだとかは、言えない。それぞれの主観の文脈や世界観によって意味づけが異なっているのであり、どちらが正しいとか間違っているとかは、本質的に問えないのである。必要なのは、それぞれの主観や世界観を交流させ対話をおこない、協調して何をなすべきか、何がもっとも適切であるのか、共通理解を醸成していくことなのである。結果として、井戸は取り払われることになるかもしれないし、あるいは逆に、井戸を村に持ったうえで生活をどう再構成していくかというプロジェクトがスタートするかもしれない。このような、異なる世界観の対話と共通理解の醸成のプロセスこそが、評価の過程であるというのである。

グーバとリンカーンによれば、彼らが第四世代と呼ぶところの参加協調型評価は、ポストモダン思想や社会構成主義（Social Constructivism）の考え方を基盤にしている。すなわち、いかなる世界の記述も観察者の主観的構成を離れてはありえないと考え、それを単なる相対主義にするのではなく、対話し協同し、共通理解を社会的に構成していくことをめざすのである。この評価過程の中には、評価基準や評価の文脈をリフレクションするという契機が含まれており、それ以前の世代の評価が持つ欠点を克服しようとするものである。このような評価観を採用するとならば、バイアスをなくし客観性を志向しようとする測定主義が基礎となったこれまでの評価を超えて、別の形での評価を模索し洗練していく必要があるだろう。すなわち、評価はどのように意味づけ直されるのか、また、そのように意味づけ直されたとき、評価がその本来的な目的を最大限に実現するためには、どのような工夫がなされるべきかを考究する作業が、次に必要となってくるのである。

この作業に取りかかるまえに、現在実施されている評価のうちで、とりわけ授業評価に関して、具体的に批判的検討をおこないたい。これにより、測定主義的な考えにはどのような不都合が生じるのか、明らかとなろう。そして、評価はどのようなものとして捉えられどのようにデザインされるべきかが、より明確となることであろう。

2. 授業評価の意義と課題^(注6)

授業評価は、大学にける教育評価のなかでもっとも基本的なものであり、わが国でも急速に導入されるようになってきた。とりわけ、この数年での普及はめざましい。その基本的意図は、個々の教員の恣意性に任されていた教授活動の実態を明らかにし、教育改善ひいてはF D (Faculty Development)につなげていくということである。授業評価は、その導入に際してさまざまな困難がありうるが、大学の因襲を変えていくのに一定の役割を果たし得ていることは疑いない。しかしながら、現行の授業評価には問題点も多いのも事実であり、この点をまずは意識化しておく必要がある。

1) 現在の授業評価の問題 ― 測定尺度としての問題

授業評価の多くは、尺度項目に対する回答、すなわち、授業評価フォームを使用したものである。これは、心理学で言う心理測定法的一种である。しかし、測定法としての厳密な議論がなされぬまま、単なるアンケートのように使用されている事も多い。授業評価が、教育活動の測定として重視されるのであれば、これを測定法の観点から検討することは、多少細かい議論とはなるが、なされなければなるまい。さらには、授業評価がどのような問題点をはらんでいるかを明らかにしておくことは、その適切な使用のためにも重要なことである。その際、とくに信頼性 (reliability) と妥当性 (validity) が問題となる。

1. 授業評価の信頼性の問題

信頼性とは、その尺度が「物差し」として安定したものであるかどうかということである。授業評価であれば、その質問項目の意味が明確で誰に対しても同じ意味を持つかどうか、また、測定対象の科目や測定時間、測定場所にかかわらず、一貫したものであるかどうかということが、信頼性として問われる。授業評価の信頼性に関しては、現在いくつかの問題が指摘されている

i. 測定時期の問題：信頼性に関する問題の一つは、授業評価を実施時期によって、それぞれに固有のバイアスがかかってしまうということである (梶田、1996)。授業評価は、授業の最終回で施行される場合が多いが、この場合、次のようなバイアスがある。1) 最終回には、これまでほとんど出席しなかった学生もやってくることが多く、適切な評価がされないおそれがある。2) 最後まで残ったのは、真面目で授業に適應した学生であるので、評価が甘くなる。3) 一年の労をねぎらって、「ご祝儀点」とでもいうべく甘い評価が下される傾向がある、ということである。これらのバイアスをできるだけ防ぐため、一般には、最終回から1、2回前に評価をおこなうのが、良いとされている。

このような信頼性の揺らぎを逆手にとって利用すれば、授業評価の数値を上昇させる、小手先の技術も工夫が可能となる。これは、アメリカでの教授経験の長い、ある教員から聞いたことであるが、授業評価実施の数回前から授業の進行スピードを落とし丁寧に教え、時々おもしろい冗談を挟み、授業最後の回では学生の労をねぎらうような言葉をかけると、授業評価の総合点数は、いとも簡単に上昇するという。この皮肉なエピソードは、授業評価の信頼性について考慮することがいかに大切であるか、端的に示している。

ii. 評価者の固有適應の問題：授業評価の信頼性の問題は、測定の揺らぎばかりでない。授業評価を行う集団がどのような集団であるのかによって測定値は異なってくるのであり、授業をその受講生から切り離して、客観的に良いか悪いかを論じるのは、本来無意味なことである。いかなる授業であれ、それは学生と教員の相互行為の組立ての上に成り立っている。学生の受講態度が、授業の進行に大きな影響を与えるばかりでなく、授業のあり方によって、学生もまさにその授業に固有の適應形態を示す。たとえば、早口でスピードの速い授業は、当初は理解しにくくても次第に慣れてくる場合も多い。あるいは逆に、丁寧な教え方の授業でも、情報を整理して聞こうという学生の主体的な努力が失われ、だんだん難しく感じてしまうこともある。このように、授業のあり方とそれに対する学生の認知や適應との関係は、単純なものではない。複数の授業を単一の尺度で単純に比較することは、例えば、熱帯の気候とサバンナの気候のどちらが生物の生存に適しているかと論ずるようなもので、ナンセンスな側面もある。このことは、次のような例を考えても分かる。すなわち、同じ授業であっても、大学院生と新入生とでは評価が異なる。大学院生にとっ

て知的興味をかき立てるおもしろい授業が、新入生にとっては意味不明の退屈な授業であることは、いくらでも生じうる。授業評価を、その評価者の特性から切り離して、「客観的に」良い授業であるかどうかは、本来問うことはできないのである。従って、米国で授業評価の結果が公表される時には、何年生が何割出席していたか、学生に対するその授業の成績評価はどうであったかなど、授業評価の置かれている文脈や背景を示しており、評価結果を立体的に解釈できるように工夫しているところが多い。また、科目の配置というカリキュラムの面にも検討が加えられるべきである。

iii. 評価視点の誘目に関する問題：授業評価の信頼性の問題に関して、もう一つ指摘しておきたい。同じフォームの授業評価を数年にわたって用いることは、年度ごとの授業を比較して授業改善の進行を調べるうえで、重要なことである。しかしながら、同じ項目であっても、他の項目との関連で、その意味が異なってくる。例を挙げるならば、ある理系の実験で、老朽化した設備のもと基礎実験をおこなっていた。授業評価をおこなったところ、やはり設備の不十分さを指摘する声が多く、この声に応えるべく、予算を獲得して設備の改善をおこなった。これで次年度の授業評価の結果は、さらに良くなるだろうと予測していたものの、たしかに、設備の不十分さを指摘する声は激減したものの、授業の内容に関する評価が前年度よりもやや下がってしまい、落胆したという。

この現象は、次のように説明することができる。設備が悪いとき、評価者は設備のほうに注意が向き、内容面はそれと比較すると「ました」と評価されていたのであるが、設備が改善されると内容面へ評価視点が誘目され、評価が厳しくなったのである。ちょうど、家族4人で狭い賃貸住宅に住んでいるときは騒音など気にならなかったのに、広い一戸建てに引っ越してからは家族のたてる生活音が気になってくるのと同じようなものである。さらに、一般的に、評価対象が限定され分析的な評価を求められるほど私たちは「あら」をさがすように評価の視点が変化するという傾向がある。たとえば、料理の味の評価を求められるとき、「(全体として) 味はどうか」ときかれるときには、「おいしい」という評価であったとしても、例えば「肉の焼け具合はどうか」と視点を限定し誘目すると、「少し硬すぎる」などと、否定的な評価が出現してくるのである。

このように、同じ評価フォームであっても、評価の文脈や条件が異なれば、その意味は変化する。したがって、前年度の評価の平均が4.2点で次年度が4.0であったとしても、この数値から単純に「授業が下手になった」などと結論することは、必ずしもできないのである。せいぜい言えることは、「評価の数値が下がった」ということにすぎない。なぜそうした数値の下降をもたらしたのかを、綿密に考察していかなければ、評価の意味がないのである。

2. 授業評価の妥当性の問題

上述のように、授業評価に例をとってみると、測定法としての信頼性には問題が多いことがわかる。測定尺度の適切さに関するもうひとつの概念は、「妥当性」である。妥当性とは、用いる尺度が、本当に測るべきものを測定しているかどうか、適切な尺度設定がなされているかどうかを示すものである。例えば、果物のおいしさを測るのに、糖度ではなく脂肪度を尺度としても意味がないように、どんな尺度を用いるかは、評価の成否を決定するきわめて重要な要因である。授業評価の妥当性に関して問題となるのは、評価項目の指標としての妥当性や学生によって評価された内容の妥当性などである。

i. 授業評価フォームの問題：現在、さまざまな授業評価フォームがあるが、項目の構成はおおよそ似通っている。すなわち、「①教育技術に関する質問：板書、しゃべり方、レジュメ」、「②教育内容に関する質問：わかりやすさ、知的好奇心、体系性」、「③情意に関する質問：熱心さ、人柄、学生への配慮」、「④成績評価に関する質問：成績評価の方法の適切さ、公平性」、「⑤責任性に関する質問：シラバスとの整合性、質問への応答」、「⑥学生自身の自己評価：どれくらい熱心に参加したか、予習・復習はしたか」などから構成されている。米国の大学では、「⑦授業改善への開かれ具合に関する質問：学生の質問にきちんと答えたか、学生の意見を採り入れて授業を改善しようとしていたか」という、授業改善の努力に注目したプロセス評価をおこなうところもある。しかし、これはあくまでも、学生と授業者が相互形式で授業を進めていく授業形態にふさわしい項目であり、講義形式が主流のわが国の大学事情にはあまりなじまず、採用しているところは、ほとんど見あたらない。

以上のような定型化された質問に加えて、各教授者が独自に項目を付け加えることができるように工夫をこらして

いる場合も多い。すなわち、評価シートには質問欄を空白にした項目が数個用意されており、評価実施のときに、教授者が自分で知りたいと思う事柄を質問し、学生に答えてもらうというものである。さらには、自由記述を最後に設ける場合も多いが、これ自体は数量化されることは少ないにしても、その評価に関して詳しく分析する際の補助情報となるばかりでなく、授業評価を教育改善に連動させていく場合に、極めて重要な情報となりうる。

いずれにしても、授業評価フォームの項目の選定に際しては、「授業評価によって何を知りたいのか」、「授業評価をどのような活動と結びつけていくのか」を考慮しておくことが必要である。従って、評価項目は、それぞれの評価主体において、しっかりと吟味して作成されるべきである。しかしながら、最近導入されている授業評価の中には、他校の事例や既存の項目を無反省に転移してきたという話も多い。既存の項目が特別の意図もなく、つなぎ合わされるだけでは、そもそも、何のために授業評価がなされるのかということさえも、忘れられてしまう。授業評価の妥当性は、低いものにならざるをえないであろう。

ii. 情報縮約の問題：授業評価の項目は、上述したように多次元的なものであるが、最終的にデータを要約するときは、全項目の平均を算出し、これを最終的な評価指標として用いることが多い。しかしながら、このような情報縮約の方法は、統計的処理の観点からは、極めて乱暴であり、妥当性・信頼性ともに低いと言わざるをえない。というのも、この方法は、多次の情報を無理に単次元化し平均しており、喩えてみれば、バスト、ウエスト、ヒップの平均値を算出しスタイルが良いかどうかを評価するようなものであり、きわめてナンセンスである。しかも、たとえば「授業内容」の次元と「授業技術」の次元とでは、「よい授業」を規定する要因としての重要性が異なるはずであるが、そうした重み付けに関しても、まったく考慮されていない。確かに、この方法によらなければ、複数の授業を直感的に比較することが困難であるのも事実である。とはいえ、単純平均はあくまでも蓋然的な目安にすぎず、わずかな上下に教員が一喜一憂したり、教育業績評価のための指標として用いられたりすることは、無意味なことである。

授業評価が与えてくれる多次元的な情報を無理には要約せず、評価の結果を項目の領域（次元）ごとに要約する方法も、良く用いられている。この場合は、次元ごとに異なる代表値を算出したり、食品成分グラフのように、レーダーチャートで描かれたりする。この方法では、情報量が多いため、授業間の比較が難しくなるが、多次元性が保存されており、ひとりの教授者が継続的に評価をおこないその時系列的变化を追ったり、教員個人へのフィードバックを重視し教育改善に連動させたりするときには、有意義な処理方法である。

iii. 学生の評価自体の妥当性の問題：教育業績評価は研究業績評価より難しいと言われる。その理由のひとつは、評価者が評価の対象に関する専門家でないことだと言われている。研究業績であれば、それを評価する同僚は当該分野の専門家であり、評価するだけの能力を持っていると想定されうる。それに対して、教育評価、とくに授業評価では、評価者は学生であり、教員の教育を正当に評価する能力を持ち合わせているかどうか、確かに疑念を向けることは可能である。さらには、評価される教育の良否を、どの時点でどのような基準に従って判断するのが、極めて難しい。コース終了後直後にその効果が出るのか、それとも何年か後に効果が顕在化するのか、実のところ測定や判断は難しいのである。

実際、授業評価を導入する際には、おおよそどこでも、次のような疑念が出てくる。たとえば、「授業評価をすると、学生に迎合するような授業が増えて、結果として授業の質を落とすのではないか」とか、「出席率の低い学生もいるわけで、彼らに果たして妥当な評価ができるか」という疑念などである。時には「学生がそもそもまじめに評価などするのか」といった、頭ごなしの懐疑もある。たしかに、授業評価がすべての科目について一斉におこなわれる場合、学生のなおざりな回答が生じやすい。しかし、これも工夫によってはかなり回避できる。例えば、学生の自己評価を問う項目を初めに配置した方が、コミットの深いまじめな回答が得られるということが、経験的に知られている。すなわち、まず自分自身を振り返り、次に相手のことを振り返るという順番である。また、授業評価を実施する際のインストラクションも重要である。授業評価の結果が教員に対して重要な情報であり、学生にはその責任があるということを、しっかりと認識されるようなインストラクションが必要である。

さらには、たとえ出席率の低い学生であっても、授業技術に関する質問項目、教授者の情意に関する質問項目などでは、出席率の良い学生と比べてもほとんど回答に差がないということも、知られている。逆に言えば、これの評価次元は直観的な判断が可能であるということであり、今後、授業評価の研究において、その構成概念を詳細に検討

していくための、ひとつの示唆を与えてくれる。

2) 授業評価の効果

授業評価を実施すると、各大学の教育現状に関して、これまで隠れていた様々な個別事情が、たしかに明らかになってくる場合が多い。それと同時に、授業評価によって一般的に明らかになった事柄もある。

例えば、研究能力と教育能力は逆比例するというような憶説は、まったく否定される。良い研究業績を上げている教官の授業評価は、概して高い場合が多い。ただし、この逆は必ずしも真ではないようである。また、授業評価の結果を、職階間で比較した場合、非常勤講師が一番高く、次いで講師、助教授、教授の順に低くなることが多い。これは必ずしも、教授の教育が一番まずいということを意味しているわけではない。年齢に逆比例しているのである。確かに若い教員には、概して教育に熱心である者が多いかもしれないが、同時に若い教員のほうが学生に感覚的になじみやすい。しかも、最新の教育機器に習熟しており（これが必ずしも優れているというわけではないが）、学生の情報受容のスタイルに適しているという理由が推測されよう。

さらに、「授業の良さ」を構成している要因は複雑に絡み合い、授業評価の各領域を単に高くするだけでは、必ずしも全体的に授業自体の質が向上するわけではない、ということも明らかになっている。例えば、ある授業者は全般的に高い授業評価結果を得ていたのだが、板書やOHPに関してのみ、評価が低かった。そこで、授業を改善すべく、やプリントを用いて情報を伝達するようにした。ところが、授業評価結果は上昇するどころか、逆に、他の領域まで低下してしまったのである（香取他、1995）。その原因を、授業者の自己分析も交えて調べてみると、プリントにしたことで、板書を使っていたときのコミュニケーションスタイルが失われ、結果として学生に授業者の熱意や意図が伝わりにくくなっていたのである。しかも、プリントで整理された情報を伝えられることで、学生が自ら考え情報を同化していく能動的な過程が失われ、学生の動機付けが低下していたことが判明した。この例が示すように、単に授業評価の結果を出すだけではなく、授業評価の結果が何を意味しているのかを解釈するための知識や技術、さらには、その結果を教育改善のためにつなげていくシステムを含んででなければ、授業評価の意義は薄くなってしまう。

3) 授業改善との連動

教育の成功とは、教える側の成功ではなく、学ぶ側の成功を第一義に考えるべきである。授業評価とは、学ぶ側の学生の視点を教員に知らせてくれる。このような学習者中心の観点に立つと、授業評価の意味づけも、自ずと明確となる。評価において重視されるべきことは、授業評価の単なる信頼性ではなく、教員が授業評価によって学生の視点を知り教育改善につなげていくという、形成的な側面である。この場合は、評価の測定としての正確さよりも、評価結果をどのように次のステップへ活かしていくかということが重視されるのである。

1. 授業評価の役割

授業評価が、人事に関わる教育業績の査定に直接用いられることは、現在のわが国ではほとんどない。授業評価には、先述したように信頼性と妥当性に対して解決すべき問題が多く、単独で教育業績評価に使用することは、望ましいとは言えない。教育業績評価に用いる場合には、ポートフォリオ評価など、別の評価の方法と組み合わせたり対比させたりする。このような補償的な方法で、確かに信頼性や妥当性は高まるが、それはあくまでも暫定的なものでしかない。本論の冒頭で述べたごとく、評価は測定のためではなく、被評価者の活動の改善のために、被評価者のために利用されるべきなのである。したがって、授業評価も、あくまでも教育改善に連動させ、形成的に用いていくものでなければならない。すなわち、教授団の教育技能を中心とした資質向上であるFD（Faculty Development）の営みの中のひとつとして位置づけられるべきものである。

もちろん、授業評価を導入するだけでも、教育改善の誘因となりうる。例えば、授業評価は、あまりにもひどい教育を行っている教員を検出するアラームとしての役割を果たす。また、自分の授業に他者の視点が入るということで、授業者には自然にリフレクションが展開される。さらに、先述したように授業評価の平均点自体に測定的な意味はないとしても、それを向上させようと目標を立てることが、授業の質を改善することもある。

しかしながら、このようなカンフル剤としての効果は、数年のうちに失われてしまう。授業評価を導入した当初は

物珍しく高揚していたが、数年後には形骸化し、「毎年やっても変化なし」ということで中止する大学も多い。また、ひどい教育を検出するアラームとしての授業評価の役割は、大半の教員にとっては無縁である。このように、授業評価が実効性をなくしていく現象は、授業評価を教育改善に結びつけていくためのシステムや営みが欠けている事に起因している。

2. 教授者へのフィードバック

授業評価を、教育改善の営みに連結させていくための基本は、授業評価の結果を教員にしっかりとフィードバックすることである。先述したように、単なる平均値を知らせるだけでは意味がなく、少なくとも領域別の情報を残した上で、学生の自由記述も合わせてフィードバックするべきである。さらには、個人への直接的なフィードバックだけでなく、同一科目群での相対的な位置を示し教授者にフィードバックする例もある（慶應義塾大学湘南藤沢キャンパスなど）。

しかしながら、いずれにしても、評価の結果によって自らの授業を振り返り、授業者の授業に対する自己省察を促し、具体的な授業改善の方法を提言するような、支援組織が必要である。例えば、米国ハーバード大学では、授業評価の結果が一定以上低かった場合は、教授学習センターで支援を受けるようにと、学部長を通して勧告がある。このような支援組織があってこそ、授業評価の結果を、授業の改善に活かすことができるのである。あるいは、相互授業参観などによる同僚の評価やアドバイスと連動させていくような努力が、必要となってくるのである。

3. カリキュラムや教育環境改善への連動

個々の授業は、授業をとりまく様々なシステムや環境の制約を受けている。したがって、どのような文脈に置かれているかを考慮せずに、その授業だけを単独に抜き出して評価することは、危険性が多い。授業評価は、個人の授業改善だけでなく、授業をとりまくシステムの改善にも連結していくべきである。そうでなければ、授業改善を単なる個人的な努力にだけ期することになり、結果的に授業評価は、個人を抑圧する制度となってしまうであろう。

まず、授業評価はカリキュラムの改善に連動すべきである。次のような例を考えてみよう。実質的に良い授業であっても、適切な時期に学生に与えられなければ、評価は低くなってしまう。内容の深い高度に専門的な授業でも、初年次の学生にとっては、難解で学生のことを無視した独善的な授業と映るであろう。幅広い知識が統合され学問の導入として非常に適切な授業でも、大学院生にとっては浅薄な授業と映るかもしれない。その授業がどの時点でどのように学生に提供されるべきかという、カリキュラム全体の構成を視野に入れつつ、個々の授業の組立ては工夫されていくべきなのである。

次に授業評価は、教育環境の改善にも連動されるべきである。どんなに劣悪な教育環境であっても、見事な授業をする教授者も確かにいるが、それを教員全てに期待することは、筋違いである。個々の授業者が、余分な努力をしないためにも、本質的なところに授業改善の努力を注ぐことができるためにも、授業評価の結果を、授業環境の改善にもつなげていくような全学的な工夫と活動が望まれる。

このように、もっともミクロレベルの教育評価である授業評価は、さらにそれより上位の階層との関連から、意味づけられ解釈されるべきなのである。

4) 循環的な授業評価

授業評価を教育改善のために利用するという、形成的な側面を重視すると、授業一回一回に評価を取り入れる方法に行き着く。すなわち、毎回毎回の授業で評価をしてもらい、それをもとに教員が自分の授業を振り返り、次回の授業の改善へ結びつけ、さらにそれを評価してもらうという、循環をおこなっていく方法である。

この場合、授業評価フォーマットによらず、自由記述形式によるほうが、ずっと有効な情報を得ることができる。例えば、授業に関する感想やコメントを学生に書かせている教員は多いが、それこそが循環的な授業評価の一例である。

循環的な授業評価の代表的なものは、ハーバード大学をはじめ米国の多くの大学で行われている、一分間テスト（minute paper）である。これは、各回の授業の最後に、「今日の授業のポイントは何か？」「今日の授業で解らなかつ

たことはどこか？」という、たった2つの簡単な質問に回答してもらうものである。これによって教授者は、その日の学生の理解度や自分の授業の成功（学生の学びの成功）の度合いを知る事ができ、次回の授業の組立てに活かすことができる。これは、学生の成績評価と学生による授業評価が一体となった方法とも言える。さらに、この一分間テストを行うこと自体が、学生の記憶定着や理解促進に役立ったり、積極的な受講態度を喚起したりするという、副次的効果もある。

同様の試みは、京都大学でもおこなわれている。同大学の高等教育教授システム開発センターの公開実験授業では、毎回の授業の最後に、感想や意見などを学生に書き込ませる「何でも帳」と呼ばれる一種の評価シートを使用している。一人一人にあてがわれたこの綴りに、学生は授業に対するコメントを書き込む。それに対して授業者は、毎回、返答コメントをつけて返却する。さらには、学生のコメントをいくつか取り上げて、次回の授業時に学生に配布し教材として利用している。

「何でも帳」は、一分間テストと同じく、大がかりな仕掛けもない単純なツールであるが、授業改善に大きな効果をもたらす。学生の視点を知ること、授業者はその日の授業を立体的に振り返ることができ、次の授業の改善のアイデアも得られる。こうして授業を日常的にリフレクションしていくことが、教育改善の基本となる。

何でも帳から抜粋され、授業時に配布される他の学生のコメントは、学生にとって大きな刺激となる。学生は、他の学生の視点と自分の視点を照らし合わせ自らを振り返るとともに、コメントを行なうことへの動機付けも高くなる。さらに、学生の評価に授業者が返答コメントをする（すなわちこれは、学生の評価に対する評価である）ことで、学生が授業を評価する目も次第に鍛えられていく。学生の授業を見る評価する目の成長自体が、学生の学びなのである。このように、授業者と学生が循環的に評価を続けていくことが、評価の形成的な性質を最大限に活かすものであり、評価の本質的な役割と言えよう（田中他、1999）。

5) メタ評価の必要性

ここまで詳細に検討してきたように、授業評価において信頼性・妥当性を十分に確保するのは容易なことではない。すなわち、それを測定としての評価として用いるには、決して超えられない限界があるのである。もちろん、信頼性と妥当性を求めて、測定としての厳密性を志向することは重要であり、その評価に基づく後の活動の成否は、その厳密性にかかっているといってもよい。しかしながら、厳密性への志向が自己目的化してしまえば、評価の本来の目的が忘れ去られてしまうであろう。

いかなる評価にも、限界がある。大切なのは、評価をする際に、その評価の特徴や限界に関して、十分に意識化しておくことなのである。すなわち、いかなる評価過程にもその評価自体の評価を内包させていくことが重要なのである。これは、その評価に対する「メタ評価 meta-evaluation (Stufflebeam, 1994)」の契機を組み込むことである^(註7)。

メタ評価とは、評価実践に対する反省意識を、制度的に具体化したものであるといえる。これは評価過程に対する一種のモニタリングである。実は、こうした評価過程に対する反省は、どんな評価であれ実際に働いているものであり、それが評価の過程を修正しつつ進展させ、また、評価が完了した段階では、評価過程を振り返り、次の評価へ向けての改善点などを考えるものである。メタ評価者とは、こうした評価に対する反省と評価を、制度化したものなのである。

そもそも評価とはひとつの反省作業である。すなわち、自己評価であれ外部評価であれ、それは主体（自己評価者）の実践への反省である。ところが、この評価という反省行為は、メタレベルではひとつの「実践」である。これまでの評価システムでは、こうした評価行為という「実践」に対して、「反省」を行なうことを意識的にはしてこなかった。メタ評価とは、この評価という反省実践への、さらにメタ的な反省であるといえるのである。

メタ評価を組み込むことで、評価の過程やシステムは、生成的に発展させられていく。また、第四世代の評価理論が示したことを考慮すれば、評価過程とメタ評価過程が描き出す世界を交錯させそこから醸成される共通理解こそ、新しい活動への開きの機会がもたらされるのである。

次には、こうしたメタ評価をはじめ、できうるかぎり評価を生成的に活かす仕組みを工夫した、京都大学高等教育教授システム開発センターにおける自己点検・評価／外部評価の試みをの検討を通して、この論をさらに具体的に発

展させていきたい。

3. 生成的な組織評価をめざして

京都大学高等教育教授システム開発センター（以下、本センター）では、設立5年目である1999年度に、その組織的活動を振り返り、今後の活動の計画立案に活かすべく、自らの組織評価をおこなうこととした。本センターには、大学教育評価研究部門が設置されており、このような評価研究を中心とした部門は全国でも類をみないものであった。1999年度の自己点検・自己評価／外部評価では、この研究部門を中心として、自らの評価の実施を、評価研究のフィールドとして、評価実施・評価研究が一体となった活動をおこなった。これは、評価研究という点からは、組織評価というさまざまな要因が関係してくる活動に関して、実践の場面を通して、その問題や課題を明らかにし、その解決を模索することで、評価に関する実践的な知を構築し蓄積する、アクションリサーチでもある。また、実施される評価活動に内在的な観点からすれば、できるだけ意義あり実りあるものとなるよう工夫された評価システムの実効性を試す実験の場でもある。この成果は、大山他（2000）にまとめられている。

以下、この評価研究／活動において何が意図されてきたか、またそれは実際にどのようなものであったかを記すことで、評価を生成的に活かしていく可能性に関して、考察したい。

1) これまでの組織評価の問題点

大学設置基準の大綱化以来、自己点検評価は努力目標、ついで義務となり、自己点検評価が多くの組織でおこなわれ、報告書があいついで出版されていた。また、外部評価も盛んとなり、組織評価はひとつのブームというべきものを迎えた。しかしながら、公刊されている評価報告書を通覧するに、自己点検評価、外部評価とも、その多くにおいては評価の本来の目的が必ずしも実現されていないと、指摘せざるをえない。それは、以下のような理由による。

自己点検評価とは、本来ならば、組織が自己の活動に組織的な反省の目を向けるものである。しかしながら、もっとも酷い場合、その組織の構成員の業績や日誌の羅列にすぎないものも散見される。これは、「評価のための資料」にすぎず、「評価」であるとは、とてもいえない。また、大学全体の自己点検・評価など、大きな組織の評価の場合、各部局などの活動単位ごとに書かれた自己評価文書が掲載されている場合も多いが、これも結局は、各活動単位がほぼ自己アピールのような形で自らの活動を記しているにすぎず、宣伝パンフレットと寸分変わらないものも多い。これでは、組織の活動に対して評価によるメスは何ら入れられず、形骸化した文書主義に墮するだけである。このような形態の自己点検評価では、しばしばヒエラルキー型の組織において、上部から下部に対して事務文書として作文が課せられており、建前主義的な回答を招きやすい。

外部評価の問題としては、ときにそれが談合的文書に墮していることが指摘できるであろう。すなわち、外部といっても、その評価される組織にとってきわめて身近な人物であったり、あるいは評価に際して事前に、どのようなことを外部評価として指摘してもらいたいかを、依頼していたりすることさえある。これでは、外部の視点を導入することで組織活動を見直すという外部評価の目的からは、きわめて遠くなってしまう。これとは逆に、できるだけ「外部」の人に評価を依頼した場合は、別の形の問題が生じてくる。たとえば大学の評価を、経済界の人物に依頼したような場合、その評価の適切性に関して、疑いが持たれることも多い。大学の人間でないものに大学の事情がわかるのか、といった類の疑念である。時には、苦笑まじりにその評価が扱われたり、的はずれな指摘として実質上無視されたりもする。この不幸な状況は、評価をおこなった外部者に責めがあるのではない。そうした異なる観点からの評価を活かしていく仕組みが欠けているという、評価の設計上の問題であり、評価実施の主体のほうに責めがある。先述した参画協調型評価の例にみるとおり、異なる世界観や価値観を持つ者が適切に評価の中に活かされると、非常に実り多い成果をもたらすことができる。外部者からの評価を、組織の自律的な営みに連結していく方法こそ模索されるべきなのである。そのためには、どのようなデザインの評価にするべきであろうか。

2) 授業のカンファレンスから組織評価へ

異なる視点を錯綜させ、ひとつの事象を多様な観点から描き出し、その多様性にさらされることを通してこそ、活

動主体が生成的に発展していくと考える方略は、京都大学高等教育教授システム開発センターにおいて長い歴史がある。これこそが、授業検討会において採用してきた方法論であると言える。

本センターが平成8年度より実践している公開実験授業において、授業終了後に授業者と参観者は、その日の授業に関して、それぞれの「見え」の世界を語りあう。同じクラスに参加していても、授業者が見て感じ取っていたクラスの様子と、参観者が見て取った様子は異なる。また、参観者も、それぞれの興味関心や立場によって、それぞれが異なった「読み」をしている。どの見方が正しくどの見方が間違っているということではない。かといって、どの見方も相対的なものにすぎないというでもない。それぞれの「見え」を語り合うことで、参加者それぞれが自分の授業観、認知枠などを意識化し、状況の多様な解釈の力量を養い、これを授業における発見的な問題解決力につなげていくことが意図されるのである（石村他、2000）。本センターでは、このような相互研修の方法を、FDの中心に据えておこなってきた。これは、稲垣（1996）の「授業のカンファレンス」に類するものであり、「羅生門アプローチ」と命名されている。黒沢の映画『羅生門』において、ひとつの事象に対する描き方や意味づけが、それぞれの主観によって様々であるように、授業においても多様な解釈が存在しうるのである。授業の場にいる学生も、誰一人として同じ世界を体験している者はいない。それぞれの主題的関心において、それぞれの世界の受け取り方、すなわち個別的な「授業の体験」をしているのである。したがって、授業者が、多くの授業の見方を知り、その多様性を自由に生きていくとき、学生の学びの在り方が見えてくるのである。

このような方法論の蓄積に基づき、評価の方法が模索された。以下に、工夫された評価の方法を具体的に記述しつつ、評価の理論について述べてみよう。

3) 組織改善に連結する評価のアイデア

1. 評価のプロセスの重視

本稿では、先に、「何でも帳」を使用した授業評価において、授業評価のプロセスこそが学生の目を鍛えていく「教育」のプロセスでもあることを述べた。同様のことは、組織評価にも言える。組織の成員が評価活動に参加することで、彼ら自身が理念を再認識したり、自らの活動に関するリフレクションをおこなったりすることとなり、組織成員としての力量を高めていくばかりでなく、組織の活動そのものの改善につながっていくのである。

しかしながら、多くの組織評価では、評価報告書という最終的なアウトプットを作成することに力点が置かれ、そこへ至るまでの過程は、できるだけ早く切り抜けていくべき「作業」として捉えられていることが多い。これは、授業評価において最終的な数値ばかりを見るのと同様、倒錯した事態である。本センターの評価においては、評価のプロセスそのものを出来るだけ実り多いものにするため、次のような工夫が凝らされた。

2. 組織の理念・目的の見直し

現在の組織評価において、主流となっているものは、アウトカムズ評価である。これは、組織自身が定めるその目的や理念、あるいは達成目標と照らし合わせて、その組織の成果がいかなるものであったかを評価するものである。これは、それぞれの組織によって異なった目標設定ができ、評価の基準（評価指標ではない）が多様に設定できる。大学評価・学位授与機構が主体となる評価においても採用しているものである。

しかしながら、アウトカムズ評価には、いくつかの大きな欠点がある。第一に、成果にもとづき評価をおこなうので、そのような成果を規定している要因がどのようなものであるかがつかみにくく、組織の改善には活かしにくいということである。したがってアウトカムズ評価の場合、「評定」になりやすく、生成的に活かす評価には遠くなりがちである。第二に、既存の理念や目的と照らし合わせて評価するので、組織はきわめて機能主義的に捉えられる。すなわち、組織は特定の目的に供するためのものであり、その価値は、できるだけ機能的にその目的を達成することで測られる。ここでは、当初の目的からはみ出た意外なものや思いもしなかった結果は、低く見積もられやすい。本来はこうした逸脱こそシステムが生成的に発展していくきっかけがあることは、セカンドサイバネティクスを提唱する Maruyama (1963) をはじめ、多くのシステム論者が説くところである。しかし、アウトカムズ評価の場合は、いわゆる逸脱を回収していくネガティブ・フィードバックとして働きやすく、システムの生成的な変化には結びつきにくい。アウトカムズ評価の第三の欠点は、アウトカムが照らし合わされるべき理念や目的自体に対する問いかけが、

ほとんど含まれていないということである。これは、アウトカムズ評価が機能主義的な問いかけであり、価値の問題を含んでいないということとからの必然の帰結であると言えよう。喩えてみれば、銃の目的は人を殺傷することであり銃はその殺傷能力によって評価されるのであるが、「人を殺傷する」という目的の妥当性自体は、アウトカムズ評価では問いにくい。ここまで極端でなくとも、組織はその設立当初の理念が、いつまでも有効であるものではない。とりわけ現代のように変化の急速な社会においては、理念や目的に関しても、絶えざる見直しが必要である。そのような、評価の文脈自体の問いなおしをしなければ、早晚システムは硬化してしまうことであろう。

本センターの評価においては、以上の点から、理念や目的自体に関しても、問い直しをおこなうこととした。この問い直しという作業は、構成員によって理念を作り出していくことでもある。すなわち、本センターはどのような機関であり、どのようなことをめざすべきであるのか再確認し、主体的に再構成していくことでもある。

3. 自己評価者の評価能力の評価

評価は、組織の日常的な活動へとつながり、それを改善していくときにこそ、その本来の意義がある。これを最大限に発揮するためには、組織の自己評価能力自体を重視する必要があるだろう。そもそも主体は、その活動において、意識的にも無意識的にも、自らの位置づけをつねにモニタリングし確認しながら、活動をすすめている。これは、個々人の授業者の授業実践においても、組織の実践においても同様である。評価とは、このような日常的な反省の営みを、制度化したものにすぎない。逆に言えば、活動主体は、自己評価活動を日常的におこなっているのである。従って、制度化されたイベント的な評価においては、このような日常的な自己評価活動へと連結していくような仕組みを考えるべきであり、それこそが、一回限りの評価が継続的な力を持ちうるための、不可欠の契機である。

このように、活動主体が日常的な自己評価を行っているということを考慮に入れるのであれば、評価の方法自体にも、工夫が加えられることとなる。それは、活動主体の自己評価能力をも評価することである。すなわち、アウトプットとしてすでに生まれているものだけではなく、その組織がどれだけ適切に自己評価できるのか、また、それを通じて自己改善をしていく潜在力がどの程度あるのか、といったことにも、評価の対象とすべきなのである^(注8)。そして、それを評価し活動主体に返していくことは、単に評価を一回限りのイベントとして終わらせるのではなく、連綿とした日常的な活動へと近づいていく重要なポイントなのである。

4. 自己評価と外部評価の相互循環

アクレディテーションを旨とする第三世代の評価では、被評価者（自己評価者）の自己評価報告書が提出されるが、それに対して、外部評価者による実地調査や聴き取り調査等を通した評価を突き合わせ、評価が下される。これに対して、被評価者（自己評価者）からは反論の機会が与えられているにしても、やはり、外部評価者は自己評価者よりも権力的に上位に位置している。このような権力的な差違を、評価のデザインに既存させる場合、外部評価者の適格性、評価の適切性に関して疑義が出ることがあり、それに対する明確な反駁証明は、原理的になしえない。

先に述べたように、評価を生成的に活かしていくためには、外部評価者の提示する見えがもたらすズレ、もしくは異化効果というべきものを、主体のリフレクションと連結させていく必要があった。外部評価者の見方は、ひとつの世界観であり、自己評価者のそれとの優劣を問題にするべきではない。むしろ、自己評価者と外部評価者とを同等の審級に置き、その世界観を交錯させることが大切であることは、本論において幾度となく強調してきた。これは、本センターの「何でも帳」において、学生と授業者との相互評価が循環し、互いの力量を高め合う営みと同じである。具体的には、自己評価者の自己評価を外部評価者が評価するばかりでなく、外部評価者の評価を自己評価者が評価し、この相互言及的な循環を複数回繰り返すというデザインが考えられるのである。

このことは、実践的な意味合いからも、活動主体の成長をもたらす。この循環の中で、自己評価者の評価の目は鍛えられ、さらに適切な自己評価ができるようになるであろう。また、外部評価者の目も鍛えられ、自己評価者（外部評価における被評価者）の活動に対して、より適切な評価ができるようになるのである。

5. メタ評価の導入

評価は、活動に対する反省である。しかしながら、その反省行為自体も、ひとつの活動である。したがって、評価

活動への評価であるメタ評価が必要であることを、前章で論じた。組織評価においても、メタ評価は、評価過程を硬直させず、生成的で発展的な評価をおこなうために不可欠のものである。組織評価において、メタ評価は、以下の3つの時期において区分されるであろう（大山他、2000）。

1）評価活動中：評価活動の形成的評価

これは、評価過程のモニタリングである。評価活動の計画は適切であるか、その計画の実現のために適切な手段がとられているか、評価活動に改善の余地はないかといった、評価作業の進行中において評価活動そのものを評価吟味し、必要があれば改善を勧告するような機能である。このメタ評価の機能は、形成的に働くこととなる。

2）評価終了時：評価活動の総括的评价

評価の終了時にメタ評価は、自身のメタ評価機能も含めて、一連の評価過程を総括的に評価する。評価活動は適切に行われたか、評価過程に改善の余地はなかったか、ということを中心に評価するものである。これは、総括的评价ではあるが、あくまでも数年のスパンの中でおこなわれる連綿とした評価活動の一契機であるという点で、形成的なニュアンスも強く持っている。メタ評価のこの機能を含めることで、当該評価は、一応の区切りとなる。

3）評価終了後：評価の実効性の評価

評価活動が終了した後も、メタ評価の活動は継続する。すなわち、評価の結果、改善が勧告された事柄、あるいは評価過程を通して得られた知見などが、組織の改善に対してどのように実効的であるかを評価していくのである。すなわち、評価結果を日常の実践的活動につなげていくインターフェイス、もしくはコンサルタントとしての役割を担うのである。

以上のようにメタ評価は、自己評価者と外部評価者との相互言及的な循環的评价へ、さらに組み込まれるリフレクションの過程である。外部評価者を自己評価者の上位に置くことは厳として避けるべきであったように、メタ評価者も自己／外部評価者の上位に置かれるべきではない。理想的には、メタ評価に関して、自己評価者、外部評価者の双方から、さらに評価やモニタリングが必要となってくるであろう。

自己評価と外部評価を循環させたり、メタ評価を不可欠な契機として置いたりする評価の在り方は、基本的には第四世代と呼ばれる参画協調型評価と同じく、ポストモダンの考え方に立っている。しかしながら、参画協調型評価が、評価活動の参加者による共通理解の醸成を重視していたのに対して、ここまで述べてきた本センターの評価においては、必ずしも、共通理解の醸成には力点を置かない。もちろん、単に主観的な見方を提出すればよいというわけではない。互いの見方にズレがあることを相互に認めただうえで、そのズレから生成的に何が生じてくるかを、評価への参画者が評価の計画も含め考察して進めていくのである。

4）組織評価の実際

以上のような発想のもと、実際に、どのような評価事業（そしてそれは同時に評価研究のフィールドでもあるのだが）が試みられ、アクションリサーチとしてどんな知見が得られたか、またどのような事が課題として残ったかを記していく。

1. 評価の実際

評価は、以下の要領で行われた。

・自己評価委員

京都大学高等教育教授システム開発センタースタッフ5名

自己点検・評価委員会の委員長には、センター長が選出された。評価期間中に、センター長の異動により、平成12年2月より、荻野文丸氏から荒木光彦氏へと交代した。

・外部評価委員

天野郁夫（国立学校財務センター・教授）

有本 章（広島大学大学教育センター・教授）

井下 理（慶応義塾大学総合政策学部・教授）
大塚雄作（文部省メディア教育開発センター・教授）
松浦良充（明治学院大学文学部・教授）
矢野裕俊（大阪市立大学文学部・助教授）（役職は平成12年3月現在）

外部評価委員の人選においては、センターのスタッフのあいだでも議論があった。「外部」ということを徹底化し、センターの活動をまったく知らない人物にまずは委嘱するのがよいのではないか、あるいは逆に、ある程度センターの活動を知っており即座に評価活動に携われる人物が良いのではないかという、相反する意見があった。議論の結果、次の1）、2）のいずれか一方、あるいは両方を満たす人物に委嘱することとした。すなわち、1）センターの日常的活動のある程度知っている人物であること、2）高等教育の専門家であり、リサーチセンターである本機関の活動に対して、研究的な視点から意見を述べるのできる人物であること、である。このような基準としたのは、「外部」の設定に難があるとは承知しつつも、できる限り組織の日常へと連結していく生成的な役割を重視したためである。この基準に従って、上記の人選をおこなった。なお、外部評価委員長には、天野郁夫氏が選出された。

・期間：平成11年5月～平成12年3月

先述したように、この評価は、自己評価と外部評価を循環させ、さらにメタ評価をそれと併行しておこなうという形である。具体的な日程としては、以下にまとめるとおりである。

表1に示したように、まずは自己評価を行い、この評価への評価も含め外部評価をおこなった。さらに、その外部評価に対して自己評価者は反論・応答し、その反論をうけたうえで外部評価者は再度外部評価をおこなうという、外部評価と自己評価を相互言及的に循環させた^(註9)。図1に評価のプロセスを模式化したものを示しておく。

表1 自己点検・評価／外部評価のスケジュール

年・月	事 項
平成11年4月	評価に関する予備的検討
5月	自己評価委員会発足、自己評価作業の開始
8月	自己評価レポートの提出
9月	第1回外部評価委員会
11月	外部評価レポートの提出
平成12年1月	自己評価者による外部評価に対する応答
2月	第2回外部評価委員会（自己評価委員と合同）
3月	メタ評価報告
3月	「自己評価／外部評価」報告書の完成

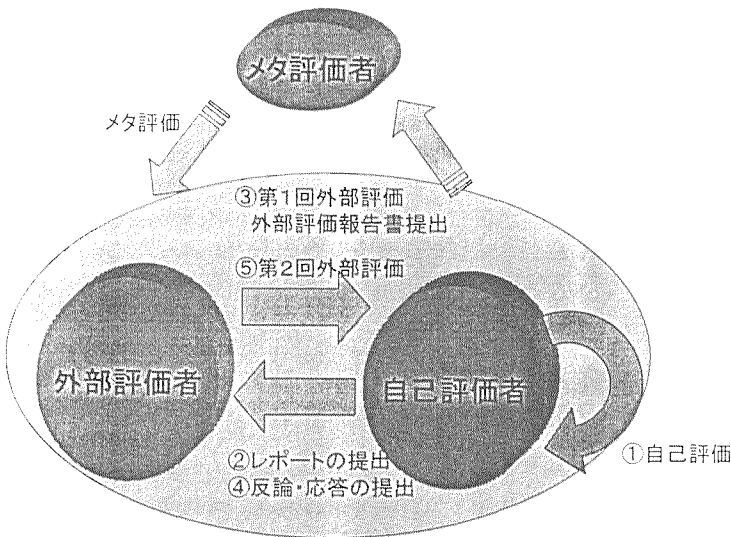


図1：本センターの評価プロセスの模式図

5）評価から何が得られたか

これまで述べてきたような工夫のもと、今回の評価が実際にどのようなものとして行われたかは、報告書（大山他、2000）を参照されたい。そこでは、評価の計画立案、自己点検評価、外部評価委員会の議事録、外部評価、自己評価者側からの応答、メタ評価などが、順を追って詳細に記述されている。

本稿では、今回のようなコンセプトに基づいた評価活動が、どのような効果をもたらしたかについて述べ、さらに、評価研究としての側面からはどのような意義があり、また、どのような問題点と課題をはらんでいるのかに関して考

察し、論を閉じたい。

1. 評価活動としての意義と問題

本センターは、当時スタッフ5人の小さなセンターである。したがって、これらすべてのスタッフが、一連の評価のプロセスに、綿密に関わることができた。これは、評価作業自体が組織改善の一助となるという、今回の評価の発想のひとつを実現するのに、きわめて適切であった。自己点検評価委員会を開くたびに、あるいは分担執筆した自己点検評価報告書を検討しあうたびに、それぞれの組織成員の共通理解が醸成されるとともに、意見の相違を発展的に統合していくことが、センター活動の理念や目的を形成していくのに役立った。実際、「フィールドワークにもとづきボトムアップ的に高等教育研究をおこなう」という現在のセンターの重要な理念は、この作業の中で確認されたものである。

さらに、外部評価者が単なる評定者ではなく、評価というプロジェクトの成員として関わることも、有益であった。例えば上述の理念は、外部評価者から提出された「センターの目的や理念の普遍の理念は何か」という問いへ応える際に確認されたものである。今回の評価のアイデアの中に、「組織の理念と目的自体の評価」があることは先述したとおりであるが、これに対してある外部評価委員からは「そのような評価をおこなうと、場当たり主義で、既存の活動を後から理由付けするようなことになり、評価の意味をなさないのではないか」といった疑義があった。また、別の外部評価委員からは「普遍の大理念と、変更可能な小理念（目的）とを区別して考えるべきだ」という指摘があった。正直なところセンターでは、「普遍の大理念」に関しては、センターの活動の雰囲気もしくは暗黙の前提として共有されており、明確に言語化され確認されたことはなかった。このことは、本センターの外部に対する説明、あるいは他大学の同種のセンターの活動との差異に関して、積極的な言明が十分にはなされてこなかったことになる。しかし、今回の評価によって、外部評価者から求められた理念の説明に応えることで、センターの活動自体が明確化されていくのにつながっていったのである。

その他、今回の評価活動自体がもたらした組織改善は、数多い。ここでは詳述はしないが、学内外研究担当教官の再編成、FD活動の推進におけるマネージメント的要素の導入、学内の教育活動改善のための積極的参加、情報共有のためのメーリングリストの開設、大学教育学というディシプリンの確立へ向けての努力など、この評価を近因あるいは遠因として、センターの活動は、より意識的かつ構成的なものとなっている。

もちろん、解決のついていない問題も多い。施設や設備の面での不備は言うまでもなく、スタッフの数が少ない割に多くのプロジェクトを抱え込み、本センターとして活動の特化した方向性が示されないままであることなどは、指摘されながらも解決されていない問題の例である。しかしながら、そのような指摘に対しても、この評価を通じて、私たちは反論する論理的根拠をしっかりと持つことができるようになった。すなわち、大学の活動を具体的なフィールドを出発点として研究していくのであるならば、そもそも大学の活動自体が広範囲にわたり複雑に絡み合っているので、大学教育研究の活動領域もある程度広がりを持たざるをえないと、言明することが可能となったのである。

2. 評価研究としての意義と問題

では、今回の評価を、評価研究としてみた場合、どのような意義と問題点が浮かび上がってくるであろうか。

積極的に評価できる意義としては、今回の評価のデザインの有効性がほぼ確かめられたことにある。この評価の作業は、決して「雑用」ではなく、組織の重要な活動として位置付けている実感が、確実にあった。この「実感」とは評価プロセスをもまさにそこで生成していくという次のような事実にもとづいている。例えば今回の評価過程は、決定された手続きでなく、評価の進行途中においても常に改善に開かれているものであった。このことは、評価の進行中に、評価作業の方法に関して、重要な多くの積極的な変更をもたらしている。例を挙げれば、外部評価者と自己評価者の応答関係を日常的に保っていくものとして、当初は電子メールを予定していたが、このメディアに関しては個人による位置づけが相当に異なるため、適切な手段ではないとの指摘がメタ評価者からあり、その結果印刷文書でのやりとりを併行することになった点などがある。また、第2回の外部評価委員会においては、外部評価委員長から、自己評価委員と外部評価委員をコンフロンテーションしてはどうかとの発案があり、急遽、合同の委員会が開催され

た。詳細は、実際の報告書を参照されたいが、このように生成的に評価を活かしていくという目的のもとで、単なる文書のやりとりで終始しない、活発な意見の交流がなされたのである。

今回の評価は、まさに動きながら考えていった評価である。したがって、反省し改善していくべき点も多い。たとえば、今回のようなある意味奇抜な評価方法をとる場合、この方法に関して、参加者の理解とコンセンサスを得るために、かなり意識的な努力が必要であるということである。もちろん、今回の評価はどのようなことを意図し、どのような手続きをとるのか、外部評価者には文書の形で伝えてはいた。しかし、それだけでは外部評価者が、自分たちがどんなプロセスに、どんな役割期待を担って参画しているのか、十分に把握するには不十分であったようである。外部評価委員、そしてメタ評価者としての井下氏は、メタ評価者の立場として「参加者がとまどうようなことでは、けっして親切的な評価とはいえない（報告書 p. 118）」と、この点を指摘した。このような事態が生じたのは、外部評価者に対する説明が不十分であったことに起因することは否めない。外部評価者は、評価の在り方に関する前理解、これまでの評価の経験などを持って参加している。本センターが実施したような特別な方法の評価を開始するにあたっては、どんなプロジェクトの中に彼らが参画しようとしているのか、事前に直接質疑応答も交えた説明の機会を設けるべきであったといえる。これは、自己評価と外部評価の循環を最大限に生成的に駆動させていくためにも、必要な前提であろう。

今回の評価の残した第二の課題は、これをさらに大きな組織に応用していくための、具体的な方策を考慮していくことである。今回の評価では、組織成員全員が自己評価者となるという、評価者と日常的活動者が完全に一致しているという恵まれた条件のもとで施行されたものである。一般的な組織では、自己評価者に組織の活動者のすべてがなるわけではなく、あくまでも一部の者にすぎない。したがって、評価者-被評価者の分裂が組織内に生じる。しかしながら、この分裂の間に相互言及のプロセスや相互評価の過程を取り入れることで、たとえば大きな組織であっても、評価を実り多きものとしていく可能性は、十分にあり得ると考えられる。この点の具体的な工夫は、今後の課題となっていくであろう。

今回の評価の第三の問題点は、メタ評価者に関するものである。メタ評価という機能自体は、評価において不可欠の契機であることは、本論で繰り返し述べてきた。しかし、それを実際に、誰がどのような形で担うのかといった、システムの現実化の段階で、解決の難しい問題点が生じてくる。これは、メタ評価者に期待される役割自体が、パラドクスと矛盾をはらむものであるからである。メタ評価者は、評価過程を俯瞰するというメタ的な視点を要請されると同時に、権力的には自己評価者・外部評価者より上位にあってはならない。しかも、メタ評価者は、評価のプロセスから一步身を引いた客観的な視点が要求されると同時に、そのプロセスの実践者たちの内部的な視点をも持っていなければならないし、プロセスの仔細に至るまで通暁していなければならない。このように、メタ評価に置いては、内属と超越の両方が要求され、それが先鋭化された形で現れてくるのである。今回の自己評価においては、このようなメタ評価が本質的にはらんでいる矛盾に関しては、自己評価者と外部評価者から一人ずつメタ評価者を立て、両者が協調しつつメタ評価を進行させていくという仕組みとし、内属と超越の問題、評価活動者との権力関係の問題の解消をはかった。しかしこれは、根本的に矛盾する問題をシステムとして解消したのではなく、メタ評価者の個人的な努力や力量に預けただけのことである。

実際、このような「鶴（ぬえ）」的な役割に関する負担感と違和感が、メタ評価者の一人から表明されるとともに、メタ評価者の役割についての明確化や選出の時期や方法に関して工夫が必要であるといった、メタ・メタ的评价ともいえるべき、積極的な提言がなされた。今後の評価研究に置いては、このような点はさらに工夫されるべきであろう。

4. 結びに代えて

わが国において評価の実施が急速な勢いで広がっていく中、評価に関する考察や研究も、同時に進行していかなければならないであろう。とりわけ、わが国では評価を支える文化が薄いぶん、なおさらである。ここまで述べてきたように、評価の本質とは、対話やインタラクションを通じて自己の反省化・相対化をおこなっていくことである。わが国の大学の授業で、グループディスカッションやディベート型の授業実践がなかなか難しいように、あるいは、組織の中で対立と葛藤を抱えるよりも波風の立たない「和」を求めてしまうように、評価を生成的に活かしてい

く文化的背景に乏しいと言えよう。これが悪いとか良いとかいう問題ではなく、評価を導入していく場合には、評価を受け入れる風土といったものにまで、本来ならば目を配るべきであるということである。たとえば、わが国の民間企業において、米国に倣い業績主義・評価主義を導入したからといって、必ずしも全体的業績が伸びるわけではない。むしろ業績悪化を引きおこし、自主的に業績主義を廃した企業の例もある。そこでは、アウトカムを強調するあまり、すぐに成果が出て評価される短期目標ばかりを追求し、結果として、いわゆるおもしろいプロジェクトや斬新な発想を殺ぐこととなったのが原因であった。米国を中心としたプログラム評価のシステムは、あくまでも、その文化的風土の中で生きるものである。また、英国の資源配分と結びつく大学評価は、あくまでも、個々の大学やカレッジが連綿と自己評価を積み重ねてきた文化の上に成り立つものであり、組織の自己評価・自己改善能力と連結して初めて機能していることは、案外知られていない (Ashcroft, K. & Palacio, D., 1996)。

ポリティカルな色合いをまといつつ、評価が猛烈な勢いで導入されつつある現在、評価自体を評価し、わが国の、あるいはそれぞれの大学に固有の評価文化とでもいうべきものを作りあげていく事が、すなわち、評価主体自身による主体的な評価への反省と実践が求められているのである。この点からも、評価を生成的に活かすための「評価理念」や評価の方法が、真剣に考えられるべきであろう。

引用・参考文献

- Ashcroft, K. & Palacio, D.: *Research into assessment and evaluation in college and universities*, Kogan Page Limited, 1996.
- 大学基準協会『大学評価の新たな地平を切り拓く (提言)』、2000年。
- Gould, S. J.: *The mismeasure of man*, W. W. Norton & Co. Ltd., 1981. 邦訳：鈴木善次、森脇靖子訳『人間の測りまちがい：差別の科学史』、河出書房新社、1989年。
- Guba, E. G. & Lincoln, Y. S.: *Fourth generation evaluation*, Sage Publications, 1989.
- 稲垣忠彦・佐藤学『授業研究入門』岩波書店、1996年。
- 石村雅雄・大山泰宏「公開実験授業検討会 ― その構造 ―」『大学授業のフィールドワーク ― 京都大学公開実験授業 ―』玉川大学出版会、2000年、142-156頁。
- 梶田叡一「学生による授業評価の意義と課題」『京都大学高等教育研究』第2号、1996年、52-69頁。
- 香取草之助 (他)「Minute Paper ― 授業改善のためのアイデア実践例」『授業をどうする！ ― カリフォルニア大学バークレー校の授業改善のためのアイデア集』所収、東海大学出版会。
- Maruyama, M.: Second Cybernetics: Deviation-Amplifying Mutual Causal Processes, *American Scientist*, 51(2), 1963, pp. 164-179.
- 大山泰宏 (他)『生成的組織評価をめざして』京都大学高等教育叢書9、2000年。
- Stufflebeam, D. L.: Empowerment evaluation: Objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, vol. 15., 1994 pp. 321-338.
- 田中毎実 (他)『平成九年度公開実験授業の記録』、京都大学高等教育叢書4、1999年。

注

- (1) 本稿は、2001年9月8日に開催された、京都大学高等教育教授システム開発センター第47回公開研究会における、筆者による同タイトルの発表の記録をもとに執筆されたものである。研究会に参加し、貴重なお意見をくださった諸氏に、この場を借りてお礼を申し上げる。
- (2) 信頼性と妥当性に関しては、本稿の2-1) でさらに詳しく言及するので、そちらも参考にされたい。
- (3) これは、あくまでも、評価という行為の過程に限定していることである。もちろん、測定評価 measurement に

においても、指標自体の開発過程、もしくは、指標の研究過程においては、信頼性ばかりでなく、妥当性も、重要な課題となってくる。

- (4) ただし、小テストによる評価結果を、具体的な生徒の学習活動の改善に結びつける方策が欠けていたので、形成的評価としての効果は乏しいものにならざるをえなかった。
- (5) この例は、シモン・ボーヴォワールがその著書『第二の性』で挙げている例を参考にしている。本来は、参画協調型評価とは文脈的に関係のない例であるが、観察者の視点によりどのように評価が変わってくるかを説明する際に適切な例であるので、あえてここで使用した。
- (6) この章での論は、拙著「教育評価の理論と実践」(『21世紀の高等教育を考える』東信堂、印刷中)と一部重なる。
- (7) meta-evaluation は、米国の評価研究者 Scriven, M が1969年に提唱し、その後 Stufflebeam をはじめとする研究者によって発展させられた概念であり、一言で言えば、評価の評価を意味している。メタ評価という言葉は、ときには、類似の調査結果や活動の結果を比較し統合をこころみる「メタ分析」の意味で使用されることもある。しかし、本論では、Stufflebeam らの用いている意味に限定して用いている。
- (8) 自己評価能力も含めて、活動主体を評価するという方法は、ポートフォリオ型評価をはじめ、学習者の評価において多く使われはじめている技法である。
- (9) なお、今回の評価では、第2回の外部評価委員会において、自己評価者と外部評価者の合同の委員会が開催された。これは、外部評価委員長の天野氏の発案によるものであるが、この意義については、3-5)で考察する。