

大学評価 第1号 平成14年10月 (論文)

[大学評価・学位授与機構 研究紀要]

高等教育における評価の諸要素とその機能
——改善志向の評価文化の形成に向けて——

Some Elements of Evaluation and their Functions
in Higher Education :

- Aiming at Formation of the Improvement-oriented Evaluation Culture -

大塚 雄作

OTSUKA Yusaku

Research in University Evaluation, No.1 (October, 2002) [the article]

The Journal of University Evaluation of National Institution for Academic Degrees

1. 大学激動時代と評価ブーム	27
評価ブームの到来	27
評価の多様性	28
2. 評価の定義と諸要素	29
「評価」の定義	29
評価に関わる諸要素	31
3. 評価の枠組と目的	32
評価の枠組	32
評価の目的	32
4. 評価に関わる人的・社会的要素	34
評価者・被評価者	34
活動主体と評価情報提供者	35
評価情報の発信源とコミュニティ形成	37
5. 評価対象のモデル化と評価手法	39
評価対象とロジックモデル	39
評価手法のいろいろ	41
評価手法の選択	43
6. 評価情報とその表現	45
評価情報	45
評価基準	46
評価結果の表示	47
7. メタ評価	48
「大学評価」のメタ評価	48
メタ評価基準のいろいろ	49
評価手法の適切性の検討	51
8. 評価のインパクトと評価文化	54
「大学評価」のインパクト	54
評価の枠組の中での「真正の評価」	55
評価文化	56
9. TQM的アプローチと大学評価	57
大学経営に対するTQMアプローチ	57
顧客満足度のとらえ方	59
リーダーシップとコミュニティ形成	61
大学の活性化と評価文化の形成	62
【文献】	63
[ABSTRACT]	66

高等教育における評価の諸要素とその機能

——改善志向の評価文化の形成に向けて——

大塚 雄作*

1. 大学激動時代と評価ブーム

評価ブームの到来

21世紀に入るや否や、大学は、未曾有の激動時代を迎えている。高度情報社会、グローバル化、18歳人口の漸減、学習指導要領の改訂、行政改革等々の社会的な動きが次々と起こっており、大学もそれに呼応して大きな改革を余儀なくされる状況にある。

例えば、国立大学は、一部では大学の統廃合が進められ、全体的にも2004年の独立法人化に向けて「国立大学法人」に関わる具体的な案が提出された。一方、2006年には総合的学習などが盛り込まれた新しい学習指導要領による学生が大学に入学してくる。また、18歳人口の漸減によって、2009年には、大学の総定員を18歳大学進学希望者が下回るいわゆる「大学全入時代」が来る。さらに、高度情報社会の中にあつて、大学でもIT（情報技術）に関わるインフラストラクチャの整備が求められ、グローバル化時代にあつて、社会に、世界に、大学を開放し、国際的に通用する大学に成長することが強く要請されてきている。

その流れの中で、1991年の大学設置基準の「大綱化」以来、各大学において「自己点検・自己評価」が義務づけられ、社会への説明責任（accountability：アカウンタビリティ）が強調されることによって、大学のさまざまな場面で「評価」が意識されるようになった。その一つの帰結が、1998年10月の大学審議会『21世紀答申』による大学評価に関わる第三者機関の設置の提案に基づいて、2000年に改組された大学評価・学位授与機構（略して機構）の誕生である。既に、全国立大学、大学共同利用機関を対象に機構による「大学評価」が開始され、2002年3月には最初の評価結果が公表されている。

また、2000年11月の大学審議会「グローバル化答申」では、大学に「国際的通用力」が求められ、それに伴って、2001年6月には、文部科学省から、各領域ごとに上位30大学（組織）に重点的に予算配当するという「トップ30（21世紀COEプログラム）」が提案された。そこでも、上位の大学、あるいは、研究拠点を選択するための「評価」が行われる。

さらに、2002年3月にまとめられた「国立大学法人」の構想の中でも、中期目標に対して6年間隔で「評価」が行われることになっている。2002年8月には、経済産業省においても、日本の産業の競争力に大学がどの程度貢献しているかについて、大学をランキング化する計画が明らかにされた。ほぼ同時期に出された中教審の答申においても、大学の設置時などの規制を緩和する代わりに、「大学の質の保証」に関して、第三者による「評価」が義務づけられ、法令違反には段階的是正措置を可能にするシステムが提案された。

* 大学評価・学位授与機構 評価研究部 教授

これらの一連の社会的動きは、いずれも、大学に対して何らかの「評価」を意識的・体系的に導入しようというものである。このように、大学激動時代を迎えて、「評価」がまさにひとつのブームとなっているのである。

評価の多様性

教育の領域で「評価」というと、学校教育の中で経験している、テストや通知票、また、入学試験などがまず連想され、点数を付けたり、合否を決めたりということが「評価」であるという一般的な通念があるように思われる。もちろん、そのような比較的手軽で、一定の客観性・信頼性が確保されていると考えられている方法による「評価」は、一つのないしは少数の価値基準の下に、多数の中から、上位、あるいは、下位の一定数を選別するというような目的をもつ場合には、効率的かつ公平な方法として広く利用されてきているものである。しかし、現代は、社会的にも価値が多様化し、それに応じて、教育的にも点数や偏差値などの一次元的な数値で評価結果を表すことに、さまざまな面で限界と問題点を生じてきており、新たな評価方法や評価観が提案されてきている（例えば、Worthen & Sanders, 1987；Harnisch & Mabry, 1991；Gipps, 1994；井上, 2001などを参照のこと）。

「大学評価」も、『21世紀答申』の副題に「競争的環境の中で個性が輝く大学」と明記されているように、大学の個性が重んじられ、それぞれの大学がその個性を伸ばさせていくことに「評価」の主眼がある。「個性」を伸ばすために、今、大学に求められることは、研究にしても、教育にしても、大学独自の理念の下に、大学の諸活動を統合的に推進していくことであろう。その意味で、非営利的な色彩をもつ大学という組織に対しても、ある種の「経営」的アプローチが意識され始めているところである。「評価」という観点からすれば、その経営的な方向性の下で、大学の諸活動を改善していくために必要な情報を的確に捉えるということが重要となる。すなわち、「大学評価」に求められることは、基本的に、個々の大学を一次元的な尺度上に載せて序列化するというのではなく、経営体としての大学のそれぞれの「個性」を反映する別個の基準に従って、個々の大学の活動状況を的確に把握することに重点が置かれるべきということである。

このように、「評価」にはいろいろな目的があって、その目的に応じて、それぞれにふさわしいさまざまな方法がある。「評価＝テスト」といった通念が、大学の活動のあらゆる場面にそのままあてはめられて「評価」が実施されることは、必ずしも効果的・効率的でないばかりか、弊害さえ生み出すことにもなりかねない。そこでまず、「評価」の多様性を視野に入れて、その上で大学における「評価」のあり方を吟味する姿勢が問われることになる。さらに、適切な方法で「評価」が行われたとしても、例えば、知能テストや学力テストなどの標準的な方法として定着した評価方法が、当初の目的や適用範囲を超えて、「IQ 神話」、「受験戦争」などと表現される社会問題を生みだしてきたように、「評価」はその目的として捉えられている枠組を超えて、教育そのものや社会に大きなインパクトをもたらすことがある点にも留意しておく必要があるだろう。すなわち、「大学評価」においても、「評価」の目的を明確にし、それにふ

さわしい方法を模索すると共に、評価の結果がその枠組を超えてインパクトを及ぼすであろう範囲についても考慮しておくことが肝要ということである。

そこで、本論では、「評価」の諸要素について概観し、それぞれの視点から、高等教育における諸活動の「改善」を志向する評価を中心に、その評価のあり方・とらえ方について論じてみることにしたい。

2. 評価の定義と諸要素

「評価」の定義

「評価」という言葉は、日常的な文脈の中で普通に用いられており、そこでは、ある対象に対して、個人的にどのような価値づけをしているかということが問題にされる。その範囲においては、国語辞典などに記されている、「価値を判じ定めること（広辞苑）」、「どれだけの価値・価格があるかを見定めること（岩波国語辞典）」などという定義で特に問題があるわけではない。

しかし、「評価」は、個人的な価値に基づく主観的な判断にとどまらず、一定の社会や組織において、組織的、体系的に行われることも多くなっており、「評価」に科学的なアプローチが期待されるニーズはそこに源があると言ってよい。そのような場合に求められる「評価」のもつべき特徴として、しばしば、「客観性」、「公平性」ということが取り上げられている。そこでイメージされる「評価」は、往々にして「評価する」という言葉がそれだけである基準の下で高い評価をすることと同義に用いられるように、評価結果がある数量もしくは順序性のある記号（ABC、優良可、……）で表現されるべきものとなっている。もちろん、ある価値を表す基準に従って判定した結果を、客観的に共有可能な数値や記号で表現することは、一つの「評価」のあり方ではある。しかし、体系的に行われる「評価」であっても、その目的によっては、数値的な表現が有効に機能するとは限らないという状況もあり得る。何に対してどのように価値を定めるのか、それを定めるのは誰か、評価の結果をどのように表現したらいいか、表現された結果を誰がどう使うのかなど、さまざまな要素がそれぞれ多様な機能をもって関連し合っていて、実は、統一的な「評価」像を具体的に浮き彫りにすることは存外に難しいことである。

例えば、教育評価の領域で、橋本・肥田野(1977)は、「教育的決定を行うための情報フィードバック」と評価を位置づけ、単に対象を数値的に表す「評定 valuation（通常は、段階評定 rating などに限定して使われる）」と、その数値をどう価値つけて教育活動に活かしていくかという意味での「評価 evaluation」を明確に区別している。これは、少なくとも、1933年からアメリカで始まったいわゆる「八年研究」の時点において、「評価」は「測定」や「テスト」などと区別される概念として意識されていたことである。そこでも、「評価」は、「測定値」等を利用して、教育のプロセスを有効に進めるための活動までもが含意され、「測定」を含むより広い概念として捉えられている。この考え方は、教育評価の領域ではその後も基本的には定着しており、最近でも、『教育評価重要用語300の基礎知識（森・秋田,2000）』で、「教育とは、

『価値』を目指す営みである。」とし、「教育評価とは、価値に照らして教育のプロセスを省察する営み」というように、活動の結果よりもむしろ「プロセス」を強調する定義がなされている。

また、経営的な視点からも、PDCA サイクル (plan→do→check→act あるいは action)、あるいは、PDSA サイクル (plan→do→study あるいは see→act) という言い方がしばしばなされ、ある活動 (do) を計画 (plan) に照らして評価 (check) した結果を次の活動 (act) に結び付けることが意識され、それがサイクルとして循環していくべきことが強調されている (例えば、Seymour, 1993)。いずれにしても、ある活動の「プロセス」の中で、それをよりよい方向へと導いていくためのものとして、「評価」が位置づけられていることになる。

大学の活動の改善に資する「大学評価」に際しても、「評定」にほぼ対応する「メジャーとしての評価」に対して、活動の改善のための「プロンプタとしての評価」という考え方が提唱されている (京都大学高等教育教授システム開発センター, 2000)。そこでは、客観的な「評定値・測定値 (measure)」よりも、主観的な世界観の交流が活動の「改善」のためのプロンプタ (prompter) としてむしろ有効であることが強調されている。

活動の「改善」に焦点が当てられた定義として、「本質的に評価は追求活動を行っている主体が行うものであり、それゆえ、それが自己評価であるということにならざるを得ない」という続 (1980) の考え方も注目される。そこでは、ある主体の活動を進めていくために行われるものが本来の「評価」であるとし、「他者評価」は、その意味で、「評定」的なものと捉えている。「自己評価」は、従来の「客観性」を重んじる立場からすれば、その点で大きな問題を抱えることになるが、活動を変えていくのは最終的にはその活動主体でしかあり得ず、その活動のプロセスにおいては、それに関連する情報はある種「客観的」であったとしても、それらをどのような価値基準の下にどう「評価」して自らの活動に結びつけていくかは確かに他者の入り込む余地のない部分であると言える。

このように、「評価」を単なる価値づけの結果としてでなく、ある活動のプロセスの中で捉えていく際には、逆に、「客観性」そのものが「評価」自体に具備されるべき要件であるかどうかという点についても問い直してみる必要がある。そこでは、最終的に「自己評価」であったとしても、活動の「プロンプタ」と評価が位置づけられるためには、活動の主体と、評価のための情報提供者との間に、ある種のコミュニケーションが生起していくことがしばしば有効となる。その点で、近年、「参加型評価 (協働参画型評価)」という言い方も各方面で見られるようになってきており、「客観性・公平性」という見地からは分離されるべき「活動主体」と「評価情報提供者」が、共に評価の作業に携わっていくことの有効性が強調されてきているところでもある (Patton, 1997; 早田, 2001)。

また、もう一つ、「評価」に関して、近年強調されてきていることに、アカウントビリティ (accountability) がある。Patton (1997) は、「評価の活用は重要な社会的課題・説明責任の新たな方向をもつ」と述べており、対象となる活動に関する評価結果を公表・公開していくことが「評価」の一環として求められてきていることが指摘されている。確かに、活動主体の活動は、

表1 教育評価に関わる諸要素とさまざまな評価

①評価の枠組と目的
枠組：社会・大学・科目・講義・教師・学生… →大学評価・教育評価・授業評価・学習評価… 目的：選別・配置・管理・改善・指導・学習・診断・研究… →総括的評価・形成的評価・診断的評価・到達度評価・判断志向型評価・改善志向型評価・ 知識志向型評価・個人間評価・個人内評価・量的評価・質的评价
②評価に関わる人的要素
評価者 vs. 被評価者：自己・他者・評価機関・相互・共同… 活動主体 vs. 評価情報提供者 →自己評価・他者評価・相互評価・外部評価・第三者評価・参加型評価・共同型評価
③評価の内容と方法
内容：学力・動機づけ・満足度・delight・教育効果・学習達成・研究成果… (活動のプロセス：theory・input・process・output・outcome・impact…) 方法：テスト・質問紙・観察・面接・記録・資料…
④評価情報
質的(定性的)情報 vs. 量的(定量的)情報 評価基準：相対評価・個人間評価・集団準拠評価・絶対評価・個人内評価・到達度評価・ 観点別評価・目標準拠評価… 評価結果の表示：数量的表示・ランキング・プロフィール・質的記述…
⑤メタ評価
副次的効果・社会的問題・測定の適切性(信頼性・妥当性・識別力・評価目的の達成度…) 評価方法の適切性…
⑥評価のインパクト
ラージスケール・アセスメント vs. オーセンティック・アセスメント アカウンタビリティと社会の支援 評価文化：評価リテラシーの普及と共有

自らの力のみで行えるのではなく、通常は、周囲の支援を必要とする。そして、それに対する応答として評価結果を公表する義務があるということが出来るし、また、そのことを通して、周囲からの更なる支援へと結びつけていくことで、活動の一層の改善がもたらされるということも少なくない。

評価に関わる諸要素

このように広範にわたる多様な「評価」のあり方を整理するために、「評価」に関わる要素をリストアップしてみたものが表1である。まず、問題とする「評価」が含まれる「枠組」と評価の「目的」(①)を定める必要がある。その目的に基づいて、「評価者」が「被評価者」(②)から、ある「内容」を対象として、ある「方法」(③)により、評価に利用するための「評価情報」(④)を収集することになる。さらに、「評価」そのものについては、とりわけ、体系的・組織的に行われている場合には、その適切性等についての「メタ評価(評価の評価)」(⑤)が求められることも少なくない。また、評価は、その枠組を超えて、さまざまな影響を他にも及

ぼすことがあり、その際の「評価のインパクト」(⑥)についても考慮しておくことが肝要である。

以下では、表1に取り上げた要素に従って、それぞれの問題点や課題について議論していくことにしたい。

3. 評価の枠組と目的

評価の枠組

「評価」のあり方を考える際には、まず、それがどのような「枠組」において行われるかということを決めておく必要がある。例えば、いわゆる「学生による授業評価」を行う場合、ある大学や学部の枠内でそれを行うのか、あるいは、一つの授業の枠内で教師が独自に行うのか、といったことがまず明らかにされるべきであるということである。「大学評価」などで、「分野別評価」において「研究」と「教育」の評価を別に行っているように、評価の対象によって「枠組」が設定される場合もあるであろう。

同じ「授業評価」であっても、枠組がどのように設定されるかによって、その目的や機能が異なってくる。一つの授業の中で閉じている枠組で行われるのであれば、授業の改善が評価の主たる目標として位置づけられるであろうし、また、より広く学部や大学という枠組での授業評価であれば、大学の授業全体の質の維持といった管理的な評価目標が含まれてくるであろう。さらに、インターネットなどを通じて、授業評価結果が社会にも開かれて公表されるような場合には、「アカウントビリティ」としての機能も果たす可能性が出てくるであろう。

このように、評価する枠組が定められることによって、そこで行われる評価の「目的」が具体的かつ明確に設定されることが可能になる。さらに、そうすることによって、「評価」に関わる人的要素、評価対象の内容や方法、評価結果の適切なフィードバックのあり方やその活用方法に関しても明確な指針を構成していくことができるようになる。

しかし、評価の枠組はしばしば開かれて、重層構造になっている場合や、いくつかの枠組が互いに影響を及ぼすということがあり得る。例えば、「授業評価」も、大学という枠組で実施されていることが多いが、評価のための調査自体は通常はクラスという枠組の中だけで行われており、基本的に授業改善に役立てるといった目的が強調されている。「大学評価」なども、ひとつの「大学」の活動の改善という目的で行われたとしても、その評価結果が社会に公表されることによって、その評価目的とは異なる望ましくない使われ方をされてしまうことが懸念されたりもしている。この点については、「評価のインパクト」という観点から、後節で再び触れることにするが、これ以降、ひとまず、評価の枠組が比較的明確に定められている状況の下での評価の諸要素について見ていくことにする。

評価の目的

「評価」の枠組が定められれば、そこで、どのような「目的」をもって評価が行われるかということのを的確に把握しておくことが望まれる。評価の「目的」にはさまざまなものがあって、

例えば、教育評価における「目的」として、橋本・肥田野（1977）は、「管理目的」、「指導目的」、「学習目的」、「研究目的」などをあげている。「管理目的」をもつ評価は、学校や社会等の広い枠組の中で、選抜・配置・処遇などに利用されることになる。「指導目的」の評価は、より範囲が限定されたクラスなどの枠組の中で、教師等の指導者側の視点から、学習者側の状態の診断や把握、それに応じた指導・授業・教材開発などに利用される。「学習目的」の評価は、学習者を中心とする枠組の中で、自らの学習の適切性の確認、動機づけの促進、学習目標の達成に向けての学習改善などに活用される。「研究目的」の評価は、広く教育という領域の中で、教師自身を含む教育を研究する者（研究者のみならず教育計画立案者等も含む）が、研究の対象となる諸要素に関わる「信頼性」かつ「妥当性」の高い情報を収集し、それらの間の関連性について検討するために利用される。ただし、この分類は、それぞれが完全に分離されるものではなく、お互いに密接に関連すべきものである点には留意しておくべきであろう。

また、選抜や処遇などが主たる目的となる場合は「判断志向型評価」、指導や学習などの改善が目的となる場合は「改善志向型評価」、調査・研究を目的として行われる場合には「知識志向型評価」などとする「評価」の分類法もある（Patton, 1997）。

さらに、活動のプロセスのいくつかの相ごとに、そこで目指される特徴に応じた評価の分類も可能である。教育活動の最初の時点では、学習者等の状況を把握する必要があるため、そのための評価を「診断的评价」と呼ぶ。また、指導や学習のプロセスの中途において、指導方法や教材、学習の方向性等が適切かどうかを確認するための評価を「形成的評価」と呼ぶ。さらに、一つの教育活動の最後に、そこで目標とされたことが全体的に達成されたかどうかを確認するための評価を「総括的评价」と呼ぶ。また、活動のプロセスあるいは最終的段階において、それぞれの段階やいくつかの観点に関して設定された比較的細かい目標が達成されたかどうかという評価を、「到達度評価」と呼ぶこともある。

「評価」は、目的に応じて、また、評価の枠組に含まれる諸要素に応じて、最適な方法も異なってくる。例えば、「管理目的」のうち、多くの学習者の中から一部の者を「選抜」という目的の下では、ある基準に従った数量的な指標によって、個人のレベルを的確に識別して序列化することが望まれることになる（個人間評価）。一方、教育の質を保証するといった目的の場合には、機関や個人の差ではなく、ある一定の水準にそれぞれが達しているかどうかを見極めるための方法が望まれることになる。このような状況では、個々の機関や個人の長短を明らかにするという意味で、個人「内」差を浮き彫りにする「個人内評価」ということが意識されることになる。

また、同じ選抜であっても、ごく少数の選抜をめざすのであれば、いずれにしても個人差を明らかにする「個人間評価」が必要とされるが、より高い精度が要求されることにもなる。そのような場合には、評価や測定にはそれなりの資源がかかることでもあり、一旦、多数からある程度の少数に効率的な「量的評価」によって絞り込み、その後、資源的にはより負担がかかるものの精度のより高い、面接などの方法によって、インテンシブに「質的评价」を行うこともあるであろう。

このように、評価の目的は多様であり、それぞれに応じて、最適な評価のあり方も異なっているのである。

4. 評価に関わる人的・社会的要素

教育の領域での評価は、教授者や学習者以外にも、学校等の経営者、学校を取り巻く社会、学習者を取り巻く家族等々、多くの立場の人々が、直接的、間接的にある種の利害関係者(stakeholder)として関与し得る。例えば、学校では、教授者が学習者の学業の状況について評価することになるが、いわゆる成績としてフィードバックされた評価結果は、学習者の親などにも関心の高いことであり、また、しばしば、内申書として、進学の際の選抜のための評価情報として、上位学校の管理者にも利用される。このような「評価」に関わる人的な要素は、「評価」そのものに大きな影響を及ぼすことになる。

評価者・被評価者

評価の「枠組」における基本的な単位として、評価する側と評価される側の区別を明らかにしておく必要がある。通常は、評価者と被評価者は同一ではなく、ある人物や組織等に関わる教育事象を他者が評価することが想定されている。例えば、学習者の成績評価は、教授者が「評価者」となり、学習者が「被評価者」となる。また、学生による授業評価では、学習者が「評価者」となり、教授者が「被評価者」となる。このような評価を「他者評価」と言う。それに対して、自らが「評価者」となると同時に「被評価者」ともなる「自己評価」という状況も、最近ではしばしば見かけられるようになってきている。大学の評価に関しても、大学設置基準の大綱化以降、各大学の自由度が広げられた引き替えとして、自らの教育研究活動を改善していくための「自己評価」が義務づけられ、それが全国的に積み重ねられてきているところである。

「評価者」と「被評価者」は、社会的に上下関係などがある場合、上位の「評価者」からの評価は、下位の「被評価者」にとってはある種の強制力として働く。特に、入学試験などのように、その「評価」が「被評価者」にとって重要な意味をもつような場合にはなおさらであり、場合によっては、その重荷が本来の教育活動自体に望ましくない影響を及ぼすこともあり得る。一方、往々にして、「自己評価」などの場合には、自らのもつ価値の歪みに引きずられ、必ずしも必要な点に対して強制力が働くとは言い難い部分もあることはしばしば指摘される場所である。事実、大学評価でも、大学審議会『21世紀答申』において、大学で行われてきている「自己評価」が実効的に機能していないと指摘され、「第三者評価」の必要性に言及されているところである。

なお、「大学評価」においては、「第三者評価」と「外部評価」という言葉は、区別して使われることが多い。いずれにしても、「他者評価」と言えるが、「外部評価」が大学の「自己評価」の一環として行われることが多いのに対して、「第三者評価」は、大学の自主的な「自己評価」活動とは異なり、大学外の評価担当機関によって定められる方法に従って、当該大学の点検等

が行われ、そこで収集された資料に基づいて、その外部組織が評価を行うことになる。

また、「自己評価」と「他者評価」の中間的な位置づけとして「相互評価」と呼ばれる評価を取り上げることができる。「相互評価」は、例えば、学習者同士、あるいは、大学同士（大学基準協会により導入されている）など、ほぼ同レベルにある立場のもの同士が、まず、一方が他方の評価者となって評価を行った後、その立場を逆転させて、評価された側が評価者、評価した側が被評価者となって、双方で評価し合う評価である。この場合は、外からの評価という意味で、ある程度の強制力をもたせることができると共に、「評価」という作業を通して、お互いにある種の学習の機会が提供される場にもなっている。ただし、両者の関係が限定されるので、馴れ合いに陥ったり、評価結果によっては双方の関係をこじらせるという可能性もないではない。

「自己評価」は、強制力ということのみならず、基本的に、自分自身を見ることができない部分もあるわけで、実際には評価自体が難しいと言える。活動を「改善」という目的の場合には、最終的に評価情報を自らの価値に基づいて改善に結び付けるのは、続（1980）の言うように自身でしかないが、評価情報自体は、「他者」から得られるものが有用である場合がむしろ多い。しかし、一方で、「他者評価」は、表面的に見える一側面に関しては適切な情報の提供が可能であるが、その背景や文脈などについて他者は知り得ないことも少なくなく、「改善」に効果的に結び付けるための評価は必ずしも容易でない。「他者」として、専門家や経験豊富な立場の人を選ぶのは、そういう理由によるところが大きい。しかし、最も理想的であるのは、「他者」からの評価情報に匹敵する質の高い情報を自ら収集でき、それを適切に処理して次の活動に結び付けられるといった「自己評価能力」をもつことである。自己評価能力の高さは、いわゆる「プロフェッショナル」と呼ばれるような、相当に高度なレベルにある専門家にしばしば観測されることであろう。彼らは、素人からすれば一見全く問題なくうまくいっているように見える場合でも、自ら不十分な点を的確に把握しており、それによって適切な調整を行って次の更に高度なパフォーマンスへと結びつけていくことができる。機構が行う「大学評価」においても、個々の大学の自己評価が前提とされているが、何よりも各大学等の自己評価能力の向上という点が、「大学評価」の第一義的なねらいとしてもっと意識されてもいいのではないかと思われる。

活動主体と評価情報提供者

「評価」の厄介なところは、例えば、教授者が「評価者」として、「被評価者」である学習者に関する成績評価を行う場合、その評価結果自体は被評価者が学習を促進させるための評価情報として活用するなど、「被評価者」が同時に、自己評価における「評価者」でもあるという循環的な構造となることがしばしば起こる点である。つまり、「評価者」も「被評価者」も、それぞれの立場で、ある種の評価活動を行っているのである。その混同を避けるために、他者等による評価情報を利用する「被評価者」を「活動主体」と呼び、活動主体の活動を促進するために評価情報を提供する「評価者」を「評価情報提供者」と、ここでは区別して呼ぶことに

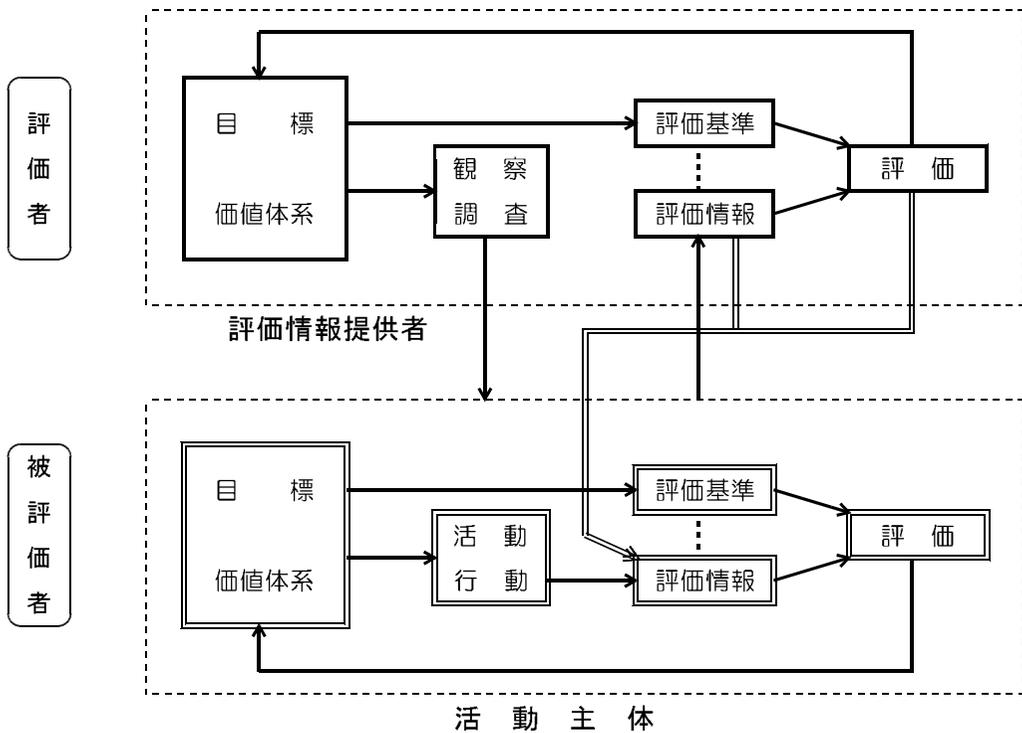


図1 評価における人的関係の模式図 (大塚, 1991を一部改訂)

したい (図1参照)。

これらの言葉を利用すると、「学習活動」という枠組においては、「活動主体」は学習者であり、その活動主体が学習を進めるために利用し得る情報の一つに教師からの成績評価が位置づけられ、この場合、教師が「評価情報提供者」ということになる。ただし、通常の初中等教育の枠組では、学習者は、自らの学習を改善するという主体的な意識をもっていないことも少なくなく、敢えて、活動主体、評価情報提供者という割り当てをする必要もない状況といえるかもしれない。

一方、学生による授業評価などで、「被評価者」である教授者が、自らの教授活動を改善するための一情報として、「評価者」となっている学生からの意見を求めている場合には、「教授活動」という枠組において、教授者が「活動主体」となり、学生は「評価情報提供者」とすることができる。この場合、教授者は、自ら意識的に「学生による授業評価」を導入するのであり、「活動主体」と位置づけることが望ましい状況であろう。また、「学生による授業評価」は、教授者の教授活動の「自己評価」に活用する評価情報として位置づけることができるであろう。

大学評価においても、「自己評価」の一環として、外部評価委員を大学自らが大学外の然るべき有識者に依頼して「外部評価」を行う場合には、大学自身は「活動主体」であり、「評価情報提供者」としてそれらの「他者」による評価が行われると見ることができる。それに対して、大学の主体的な意図とは独立に政策的に導入されるような「第三者評価」は、その評価の第一義的目的が、大学自身の活動の改善であったとしても、大学は「被評価者」としての位置

づけが強調されることになる。ただし、現行の機構による「大学評価」は、その経緯からすれば、国立大学等が自らの活動改善のために導入したわけで、その視点からすれば、「活動主体」の「自己評価」に必要な「評価情報」を提供するための「大学評価」であるという見方もできる。すなわち、大学自身の意識の持ち方によって、「第三者評価」を、自らの活動改善のための「自己評価」に必要な情報収集の手段と位置づけることができる。このように、同じ評価であっても、それがどう捉えられているかという、評価に関わる人それぞれの意識によって、その性格が異なってくることに留意しておく必要があるだろう。

同じような評価の枠組、目的、人的要素があるという場合でも、その人的要素の特徴によって、効果的な評価のあり方が変わってくることを示唆するものとして、経営的な視点からマグレガーが提唱した、X理論・Y理論（McGregor, 1960）が参考になるだろう。

例えば、経営者が多くの比較的単純な作業に関わる労働者を雇用して、その全体を経営目標の達成に向けて動かし、ある場合には、それにふさわしくない者を切り捨てることも許されるような状況では、被雇用者の活動レベルをランキングするなどして、時にはそれに応じた報償等を与えるような方式が有効である。すなわち、活動を促進すべき評価の対象者が、その活動は生活のために仕方なく行うもので、できれば他のことをしていたいというような状況では、対象者の活動水準に関する「個人間評価」が有効になるということである。これをX理論と呼んでいる。この場合は、被雇用者は「被評価者」に対応することになるが、それを「活動主体」としては位置づけにくい。ここでは、被雇用者は一個人としてと言うよりも、経営体における集団の一要素とみなされ、その集団全体のレベルの向上が目標とされている。

それに対して、被雇用者が一個人として、それぞれ主体的に必要なとされる活動に取り組むことができるのみなされる場合には、ランキングなどの個人間評価はむしろその活動を阻害する可能性が大となる。その場合には、現状の活動状況が目的達成に向けてどのようなレベルにあるかに関する的確なフィードバックやそれに基づく適切な活動指針が与えられる方が、自らを効率的・効果的な活動に導いていくことができる。すなわち、主体性をもった個人に対しては、プロセスに基づく「個人内評価」が有効になるということであり、ここでは、「被評価者」を「活動主体」と位置づけることができる。この後者の場合をY理論と呼んでいる。

評価における人的要素に関しては、単にその外面的属性のみならず、内面的・心理的特徴も十分に考慮すべきことを、X理論・Y理論は示唆してくれていると言えるだろう。

評価情報の発信源とコミュニティ形成

単なる「被評価者」である場合と、「活動主体」として評価情報を活用するという場合では、同じ「評価情報」でもその活かされ方は大きく違ってくる。ある評価の枠内における人的要素が、「活動主体」となるかどうかは、基本的には、その内面的な意識的・心理的側面の寄与するところが大きい。さらに、評価の枠組を超えたところから影響が及ぶこともある。例えば、学生による授業評価において、「授業」という評価の枠組の中では、ある意味で社会的に下位の学生からの評価情報は、それだけでは教師に授業改善に着手させる大きな強制力をもつとは

限らない。そこで、学内で授業評価結果を公表するとか、ファカルティ・ディヴェロップメントの研修会に報告し合う機会をもつなど、評価結果を評価枠組の外に開くことによって、間接的に教師の授業改善への動きを引き出ししていくことが考えられる。アカウントビリティを果たすという評価の新しい側面も、社会に公開して活動に関する理解を求めることを通して、「評価」にそういった外的なプレッシャーを付加する機能をもっていると見ることもできよう。

ただし、外的に強制力をもつ「評価の枠組」外からの影響は、必ずしも「活動主体」の本来の目的に向けた活動を促進するとは限らないという点には留意しておくべきであろう。それを、自ら目指す活動に活かすためには、そのような外的なフィードバック情報の意味を納得して受け止められるということが必要となる。その情報が活動の改善に向けて利用されるためには、それが「活動主体」の中で価値あるものとして位置づけられることが不可欠であるからである。

外からの情報に価値を置くことができる一つの要件は、評価情報の発信源が、自らと同じ方向性をもった価値観の下に評価情報を提供してくれていると認識できるということである。もし、そのような人から提供された評価情報であるならば、たとえ教師に対する学生という下位の立場からの情報であっても、教師にとっては重要な改善への引き金とすることができる。言い換えれば、そこには、「活動主体=被評価者」と「評価情報提供者=評価者」の間に、一定の人間関係が成立しており、ある種の共同体（コミュニティ community）が形成されているとみなすことができる。なお、ここで言う「コミュニティ」は、「地域社会」を指すのではなく、何らかの部分で、同じ価値、同じ目標をもった、人の集合・人のつながりというように幅広く捉えたものである。

「コミュニティ」という視点から見ると、評価情報が「活動主体」にとって意味があるということは、その情報が、自らが属している何らかのコミュニティの中で発信されたものと認識されていることを意味する。逆に、外的強制力の強い評価情報は、得てしてそのようなコミュニティという括りを見出しにくい場合が少なくなく、X理論の状況のように、活動の「主体性」が脅かされる可能性すら生じてくる。通常、活動の「改善」ということ自体、ある種のコミュニティの枠組の中で意味があるのであり、そのコミュニティ外にとっても「改善」であるかどうかは必ずしも自明のことではない。その意味で、「活動主体」が属するコミュニティで共有される価値に基づく「評価情報」が、活動の「改善」に有効に働くということである。すなわち、適切な「評価」が行われるということは、そこにある種のコミュニティが形成されている、あるいは、形成されていくことを意味するものであり、また、活動が改善されるということは、「活動主体」がそのコミュニティにおいていっそう確固たる役割を担うようになることであるとみなすことができる（大塚，2001）。そのようなコミュニティは、輻輳し合いつつ、複雑に相互作用し合っており、必ずしも容易に同定できるものではないが、評価の枠組において、どのようなコミュニティが形成されているかを見極めることは、評価のあり方を論じる際の一つの重要なポイントになると思われる。

教授・学習の場面において、教師と学習者との間にある種の「学習コミュニティ learning community」が形成されていれば、教師からの学業成績のフィードバックは、学習者の学習

を促進するためにより有効に機能するであろう。また、それが機能しているのであれば、ある種のコミュニティがそこに形成されていると見ることもできる。特に、教授者と学習者のコミュニケーションの場が保障され、学習者の長所・短所や、関心の方向性などを踏まえて、より具体的な今後の学習の方向性が提示されるなど、質的な情報の交換を密にすることは、そのような学習コミュニティを形成するための有力な手段となるであろう。それに対して、学習者にとって、教師がコミュニティの埒外の存在としてしか認識されていないのであれば、教師からのフィードバックは、学習の目的からずれる部分で学習者に影響を及ぼしてしまうことになりかねない。例えば、学業達成度をランキングとしてフィードバックした場合には、学習者の関心は、学習そのものよりも、そのランキングを上昇させるテクニックに移ってしまうことは避けられないことである。また、この場合には、そもそも、学習者の個人内評価としての情報提供も十分にその機能を果たさない可能性が大である。逆に、ある学習コミュニティにおいては、個人内評価が、その価値の下で当然のものとして受け入れられ、有効に機能するということがあり得るであろう。

最近、初中等教育では、とりわけ、総合的学習などの評価として注目されているポートフォリオ評価（井上、2001）など、学習成果を多面的に蓄積し、それに基づく教師と生徒のコミュニケーションを促進させることが意図された新しい評価法が導入されようとしていることは、まさに、「コミュニティ」の視点からすれば望むべき動向である。また、これと同様の流れの中に、「評価者」と「被評価者」が共同して「評価」を進める参加型評価など（早田、2001）を位置づけていくこともできるであろう。

しかし、なおかつ、学業成績は、教授者・学習者というコミュニティだけのものではなく、一個の学習者が何を学んできたかという情報を、より広範な社会的学習システムの中で共有するためにも必要とされる場所であり、ある種客観性があると思われる評定値の形での評価結果を表すこともまったく否定できるわけではない。それだけに、評価の枠組外からの評価情報の活用のされ方を十分に把握し、その活用目的と利用範囲を的確に共有することが望まれるであろう。

5. 評価対象のモデル化と評価手法

評価を実施する上での中心的課題は、評価の目的を達成するために、何を対象にして、それをどのように評価するかということである。それを明らかにするためには、評価の対象となる事柄に関するある種のモデル、あるいは、何らかの理論が構成され、それを踏まえてある評価方法が導き出されることになる。

評価対象とロジックモデル

「評価対象」という語は、主として評価を行う活動の「内容」を指すが、しばしば、「被評価者（組織なども含む）」を意味することもある。しかし、「被評価者」のすべてを評価することとはあり得ないことであり、「被評価者」の学習状況であるとか、あるいは、それが大

学評価の場合であれば、大学の教育活動であったり、経営状態であったりするわけで、いずれにしても「被評価者」の側面が評価されることになる。「対象」というときは、その側面を指し、「被評価者」は、「対象者」とか「対象機関」などと区別して呼ぶ方がわかりやすい。

評価対象は、当然のことながら、評価の目的に応じて変わる。例えば、「学習目的」では、学習者の「学習」状況が評価対象となるし、「指導目的」では教師の「教授」の状況が、「管理目的」では「学校経営」の状況などが、それぞれ評価の対象として取り上げられたりするであろう。さらに、「学習」状況の評価するためには、ある学習者のある学習内容に関する学習の達成度のみならず、その学習者の全体的な学力、性格、学習動機づけ、学習時間、学習後の満足度、また、その学習で利用された教材の質、教授者の力量、等々、その学習に関わるさまざまな要素が評価対象として取り上げられていくことになる。それは、「学習」に関わるどのような理論的モデルに依拠するかによって、定められていくことになる。

評価対象に関するモデル化は、基本的には、「評価」の領域の課題ではなく、そこでの対象に関わるそれぞれの領域における研究課題である。理論的に確固たるモデルが構成されていて、評価目的に沿って、そのモデルを採用することができれば、それに基づいて、適切な評価内容と方法が明確に定められることになる。また、それが可能な場合に、理論と実践が適切に結び付いていくことにもなる（大塚，1994）。ただ、「評価」の目的によっては、研究的な蓄積がそれ程なく、日常的・経験的なレベルのある種のモデルを背景に評価せざるを得ないような場合も少なくない。

例えば、政策評価などにおいては、政策自体に一定の期間をもって成果を出すことが明示される性格上、政策立案のスタートから、それが目指す最終時点まで、「直線型」の評価が行われることになる（龍・佐々木，2000）。その点は、「PDCA サイクル」などの、プロジェクト的活動を改善しつつ推進していくための「循環型」の評価とは性格を異にする部分である。政策評価は、その直線的な流れをいくつか区切って、それぞれの段階での評価が提言されている。まず、政策等に掲げられた目標を達成するために、適切に、具体的に取り組むべき課題が設定されているかという点が評価される。これを、「セオリー評価」と呼んでいる。政策などの場合には、目標とその下位目標や具体的課題等の構造とフローを図式化したものは、ロジックモデルと呼ばれているが、その背景に特に学問的な理論が意識されているとは限らず、ロジックモデルそのものの整合性を適切に点検しておくことが望まれることになる。

ロジックモデルが適切であることが確認できれば、それに基づいて、政策実行のプロセス及び成果等の評価のための評価対象を明らかにすることができる。政策に基づいて諸活動が適切に行われたかどうかを確認する評価は「プロセス評価」と呼ばれ、また、目的・目標に即した成果が得られたかどうかを確認する評価は「インパクト評価」と呼ばれている。

「プロセス評価」では、政策に基づいて行われる諸活動の記録や観察などから、その活動自体に課せられた下位の目標が達成されているか、あるいは、政策全体のより上位の目標の達成に貢献しているかどうか、といった評価が行われることになる。実際の活動のプロセスの中では、活動の改善のためのモニタリングとして、逐次、プロセス評価が行われることになるが、

その場合は、評価結果を当該活動にフィードバックする「循環型」の評価となる。

また、プロセス評価では、そこでの評価対象となる個々の活動が、当初の政策に掲げられた大目標の達成のためにブレイクダウンされて行われている場合が多い。その場合には、個々の活動を統合して、より上位の目標を達成するための調整に関わる活動主体も必要とされることになり、その上位のレベルにおいても、並行して、プロセス評価を行っていくことが望まれる。すなわち、重層構造をもつ活動の場合には、それぞれのレベルの視点から、当該活動を改善するための評価を行う責任を負う活動主体が存在することが考えられる。そのような活動主体を含む一つのまとまりを「評価単位」と呼ぶことにする。評価単位は、最終的には、その政策に関与する個人のレベルまで下りていくことが可能であるが、限りある評価作業においては、適切なレベルに「評価単位」を定めることが肝要であり、それを明確に設定することもロジックモデルの役割ということになるであろう。

「インパクト評価」では、政策で目標とされたことが達成されたかどうか、言い換えれば、ロジックモデルにおいて定められる最終的な成果を評価対象として、それを反映する情報や指標などに基づいて、総括的な評価が行われることになる。その際には、当初に投入した資源に対してどのくらいの効果があったのか（コスト・パフォーマンス評価）などの視点からの評価も求められる場合があるだろう。ただし、インパクト評価は、目標によっては、政策の最終時点での情報だけで判断できない場合もあるし、また、成果を表す指標などは、政策によってのみ影響を受けるものでない場合がほとんどであり、評価対象が明らかにされたとしても、必ずしも容易に評価が可能とは言えない部分がある。

評価対象を明確にするためのロジックモデルの構成とそれに基づく評価のあり方は、「大学評価」においてもほぼ同様にあてはめ得ることであろう。大学の理念的な「目的」から、それを実現するための課題としての諸「目標」、さらに、それらの下位目標を実現するためのさまざまな活動の関係性が、ロジックモデルとして明確にされ、それに従って、下位の「評価単位」ごとの評価が、より上位の「評価単位」の評価に反映されていくようなシステムが構築できれば、大学等の活動を改善するためにより有効な「評価」が可能となると思われる。また、ロジックモデルとして、大学の活動が明示的に図式化されることを通して、高等教育関連の研究領域での理論化もまた進化していくことが予想され、そこで提出された理論を背景に、より有効なロジックモデルが構成されていくことにもなるであろう。

評価手法のいろいろ

評価対象が明らかにされると、それに関する情報をどのように収集するか、いわゆる「評価方法」が問題とされることになる。「評価方法」というと、情報を収集する部分のみならず、その情報をどのように解釈して次の活動に結びつけていくかという情報の「活用」の部分まで含めることも可能であるが、ここでは、主として、情報収集の「手法」に焦点を絞って議論しておくことにする。なお、情報収集の部分は、先に触れた定義によれば、活動の改善に結びつけられる点が強調される広義の「評価」の全体を表すというよりも、「評定」や「測定」に関

わるものとして位置づけられるという点に留意しておく必要があるだろう。そこで、ここでは、情報を収集する手法を「評価手法」と区別して呼ぶことにしておきたい。

「評価手法」は、評価対象がどのようにモデル化されるかということに依存するが、そのモデルから唯一の手法が導かれるということではなく、なおいくつかの可能性が残される。ただし、そこで絞り込まれたいくつかの手法は、いずれも、ある利点と限界をもつ。また、評価を行うために利用できるリソースや時間的制約の問題も現実的には重要な課題となるので、評価対象のモデルに基づいて最適な手法を選択するという視点のみならず、評価を実施する状況や文脈に関わる要件も考慮しておく必要がある。その意味で、多様な評価手法の得失を踏まえて、適切な選択を行うことが求められる。

教育や心理に関する評価においては、日頃の活動の中で蓄積された文書・作品・資料・各種指標などが活用される他、テスト、質問紙調査、観察、面接等が「評価手法」として取り上げられることが多い。さらに、テストや質問紙と一口に言っても、「学力」、「学業達成」、「学習意欲」、「性格」、「興味」、「学校適応」等々、評価の目的に応じて、そこで対象とされる内容は多岐にわたり、しかも、それぞれに関する理論的モデルに基づいて多様な評価手法が開発されてきている（上里, 1993；堀, 2001；松原, 1995；大沢他, 2000などを参照のこと）。また、それぞれの対象に関する理論的モデルが整備され、利用範囲の広い場合には、一定の標準化の手続きを踏んだ標準検査等の「評価手法」が開発されることもあれば、一方、評価の「枠組」が小規模の場合には、教師等の評価者が、必要に応じてその都度作成しなければならないこともある。対象によっては、テストや質問紙を構成する質問項目のレベルから、その都度新たに開発していかねばならない場合もある。

さらに、対象が限定されたテストであっても、マークシート方式の多肢選択方式の形式で構成されるものもあれば、自由に論述される記述式のテストもあり、それぞれの手法を構成する個々の項目の形式もさまざまである。最近では、コンピュータなどの情報機器が導入される場合も少なくなく、多肢選択方式という一つの形式であっても、紙媒体のみならず、コンピュータテストなどによる手法の開発も進んでいる。ここでは、選択肢に、音声や映像が利用される、いわゆるマルチメディア・テストなど、利用される情報モードも広がりを見せつつある。一方、「観察」や「面接」などの手法においても、ビデオやオーディオ・テープに対象となる言動を収録して記録するなど、映像・音響情報の評価への活用も急速に広がりを見せているところである。

また、近年は、「活動主体」の活動の改善のための評価ということが強調されてきていることもあり、以上にあげた多様な手法を組み合わせた、多面的な評価手法が利用されることが多い。それによって、「個人内差」に関する情報が厚みをもって収集されることになる。なお、「多面的」という点では、「横断的 (cross-sectional)」に一つの時点で各種の情報が収集される場合と、「縦断的 (longitudinal)」に時間を追って情報が蓄積される場合がある。

評価手法の選択

このように評価手法は多岐にわたるが、それらをどのように組み合わせるかは、評価の目的、評価対象のモデル、可能な評価手法、人的・経済的リソースの制約などから決められていくことになる。いかに正確で詳細な評価情報が得られるからと言って、いくらでも金と時間を注ぎ込めるというわけでもなく、また、逆に、正確に測定をしようとするほど、被評価者にさまざまな影響が及んで正確な測定ができなくなるという、「不確定性原理」的な現象は、マクロの世界であっても看過することはできない。むしろ、正確とは言えなくても、ある程度の範囲で評価対象に関する情報が得られるのであれば、安価かつ有用な評価を行うことが可能になるので、そのバランスを検討していく視点が先ずは求められることになる。

例えば、授業評価では、通常、「評価者」である学生が、「どの程度理解できたと感じるのか」、「どの程度授業を面白いと感じたのか」といった、授業に関する項目を評定する質問紙調査がしばしば利用される。この種の項目は、概ね、授業において、教師や学生がどのような行動をしているのか、どのようなことを感じているのかといったことのリストから選ばれている。そのリストは、教授者の教授技術・熱意、教材・授業の適切性、学習者自身の学習態度など、いくつかの観点から分類・整理することができる。そのような分類は、授業に関する一つのモデルとして位置づけることができ、そのリストから授業評価調査が構成されていくことになるが、言うまでもなく、そのリストに挙げられ得るすべての項目を授業評価調査に含める必要はない。もちろん、すべての項目が利用できれば、正確で詳細な情報が得られる可能性もあるが、被評価者である学生の負担、調査用紙にかかる費用、調査に回答する時間の長さなど、それを許さない制約がある。そこで、評価項目の分類において、授業評価調査がカバーすべき範囲を明らかにし、それぞれの分類からバランスを保ちながら、少数の項目を抽出して、質問紙調査が作られることになる。

また、学期の最後の授業で、そのコース全体を総括的に授業評価するという場合には、20、30といった項目がバランスよく選ばれる必要があるが、毎回の授業ごとに授業評価の情報を得ようという場合は、せいぜい代表的な7～8項目に絞り込んだ簡単な質問紙が限界であろう。そのような場合には、通常、授業の感想や疑問点、授業のポイントなどを1～2分程度の時間で学生に書かせる「ミニッツ・ペーパー」などの試みと併せて行われる場合が多く、学生の負担と授業自体の時間的制約が重視されることになる。

さらに、ある授業のクラス全体の特徴を把握したい場合には、評定法式の質問項目によるクラスの評定平均値などの統計指標が有用となるが、具体的にどのような点で授業を改善すべきかを探るためには、授業で感じたことなどを自由に書かせる「自由記述」項目がむしろ有用である。ただし、「自由記述」は、ある程度の記入時間を必要とするし、書くべきことを思いつくかどうか、あるいは、それを書きたいと思うかどうかといった学生の気分的なものに左右される場合があり、網羅的な意見を吸い上げることは難しい。逆に、書かれた意見は得てしてインパクトが強く、それに引きずられることもあるが、必ずしもそれがクラス全体を代表してい

るとは限らない。また、学生数が多い場合などには、自由記述のすべてを丹念に自らの授業にフィードバックすることも困難になるという限界もある。

なお、授業評価で問題とされることに、学生に授業を評価する能力があるかどうかという点がしばしばあげられる。学生は、授業の内容については学習の途上にあるわけで、授業の論理的体系性や、全体的なカリキュラムにおける位置づけなどについて見極める力を期待することはできない。その点で授業を評価したいのであれば、少なくとも、教授者と同レベルのいわゆるピア・レビューアーに授業を公開するなどして、授業後のディスカッションなどを通してフィードバック情報を得る手法が有効であろう（京都大学高等教育教授システム開発センター、1997）。また、自らの教授技術について改善を試みたいという場合には、自分の授業をビデオなどに収録して、それを自ら振り返ってみるということもしばしば行われている（伊藤・大塚、1999）。

ただし、公開授業やビデオ収録による評価は、実施すれば、授業についてより質が高く詳細な情報を得ることが可能であるが、いわゆる「ピア」を集めるのが大変であったり、収録したビデオを分析するのに手間がかかるなど、現実的には、実施に至らずに終わっている大学や教員も少なくないであろう。一方、授業は、教授者だけのものではなく、どんなに整備された授業であっても、学習者にそれがどう受け入れられているかというところに根本的な課題があるはずで、学生が授業にどう臨み、どのように感じたか、また、何を学んだかといったことについての情報を集めることも常に求められるべきことである。その意味で、学生による授業評価調査も捨てがたい評価手法に違いはない。しかも、リソース的にそれ程多くのコストが、人的・時間的・経済的にかからないということもあり、評価情報の質は一定のレベルのものに過ぎなかったとしても、継続的に積み重ねてもいき易く、また、その価値も十分にある比較的手軽な手法であると言えるだろう。ただし、継続的な評価情報の積み重ねは「改善」には有効であることはわかって、教員個人で対応するのは相当に大変なことである。また、公開授業などの他の評価手法などが必要とされるときにも、それらはそう気軽に取り入れられるものではない。そこで、大学や学部などにある種の授業評価のサポート体制が準備されているかどうかということも、評価手法の選択にクリティカルに影響を及ぼす重要な要件となっている。

もう一点、評価手法を選択する際に、得られた評価情報をどのように捉えて価値づけるかという情報収集「後」のプロセスが、情報収集の手法の選択に影響を与えるということをあげておく必要があるだろう。正確な測定値さえ得られていれば、それを後で適当に重みづければ、思うような価値判断に結びつけられるというのは安易な考え方である。「思うような価値判断」をすることは、まさに評価の目的であって、それに応じた評価対象のモデル化がなされ、評価情報の収集が行われて、はじめてその「価値判断」に結びつけていくことができるのである。

例えば、機構の「大学評価」で問題とされていることの一つに「評価結果の一律予算配分方式への反映」ということがあるが、これは現行の評価方法ではほとんど不可能と言わざるを得ない。現行の評価では、「予算」と関係させることができるとすれば、各大学が何らかの予算要求の時に自主的に自らの評価結果を利用することぐらいであろう。少なくとも、理屈

的には、各大学ごとの「目的・目標」に即した別個の評価基準に基づいている現行の評価結果では、それがいかに的確に導かれたものであったとしても、何らかの予算配分機関で一律の予算配分方式に適用するにはほとんど情報量をもたないと言ってよい。「予算配分」に利用する評価を行うのであれば、どのような趣旨の予算に関して、どのような基準によって配分するかという具体的な見通しがあって、そのためにどのような評価情報を集めるべきかが明らかにされ、それに応じた適切な「評価手法」が選択されていく必要がある。すなわち、評価手法の選択は、その後の使われ方を十分視野に置いて判断されるべきことであり、それだけを取り出して論ずることはできないのである。このように、「評価」に関わる諸要素は、その全体の流れの中から切り離して議論することは基本的に困難であり、それぞれの相互作用に十分留意しながら検討していく姿勢が望まれる。

6. 評価情報とその表現

評価情報

本論では、評価のために利用される評価対象に関する情報を「評価情報」と呼んでいる。日常的には、評価情報を得ること自体を「評価」と呼んだりもするが、先に見たように、評価情報を得ることと、「評価」そのものは区別して捉えておくことが望ましい。評価情報がある明示的な価値基準に即した指標として得られ、なおかつその評価基準を「活動主体」がそのまま受け入れられるような場合は、「評価情報」と「評価」自体とは明確に区別できない。しかし、一般的には、評価情報は、「評価者」が採用した価値基準に基づくものであって、それを「活動主体」が別の価値基準によって重みづけて改めて解釈し直し、次の行動に結び付けていくことになる。

評価情報は、「質的情報」と「量的情報」に分類することができる。「量的情報」は、数値や順序づけ可能な記号で表現されたものであり、いずれにしても数値に変換して扱うことができる。「質的情報」は、言葉や映像などで示され、数的には表現しにくい情報である。評価という「量的情報」を提示することであるという印象が強いが、とりわけ改善に結び付ける「評価」においては、「質的情報」の方が効果的である場合も少なくない。

「評定」、「測定」などでは、多くの場合、「量的情報」が収集される。数値で表される量的情報は、多くの場合、ある価値基準を表す明示的な尺度によって測定されることから、いわゆる「客観性」の高い評価であるという印象が与えられる。数値による比較は簡便であり、ある統計的な方法によって重みづけられて総合点を求めるなど、一定の手続きによって改めて価値づけし直して再表現することも容易であることから、比較的大規模な選抜などを目的とする評価に主として使われることになる。また、活動の改善の具体的な方法が明らかにされていないような状況では、その活動に強制的に向かわせるために、量的情報が有用である場合も少なくない。

それに対して、「質的情報」は、「客観性」は必ずしも高い印象は与えないが、「活動主体」のもつ背景や文脈の下に「具体的」に評価対象に関する情報を伝え得る特徴をもっている。「客

観性」に関しては、ビデオに収録する映像や、比較的多くの評価者によって合議の上で導かれた情報など、方法によっては一定の「客観性」が確保できると考えられる場合もある。また、「具体的」であるということは、それによって、「活動主体」が自らの活動を振り返ったり (reflection)、また、他者との間のディスカッションを引き起こすのに有用であり、それを通して、次なる活動の改善に結び付けられやすくなる。それはまた、評価者、被評価者の間に一種のコミュニケーションをもたらし、先に触れた「コミュニティ」の形成にも有効に機能する可能性をもっている。

評価基準

評価情報を価値づけるためには、通常、ある種の基準が必要となるが、基準をどのように設定すべきかということも、評価の目的によって変わってくる。

例えば、選抜などを行う場合には、集団の中での相対的な位置づけが可能になるように、人と人との比較が可能になる形の基準が必要とされる。この場合は、そこで対象とされる準拠集団における相対的な位置を表す基準（ノルム norm：規準と記す場合がある）が参照されることになる。いわゆる、「相対評価」、「個人間評価」、「集団準拠評価」などと呼ばれている評価がこれに当たる。

一方、「学習目的」などの評価の場合には、ある単元の学習がどの程度のレベルにあるかが判断でき、ある基準に達することができれば、次の単元に進むといったことが想定される。この場合は、次の単元に進むことが可能である水準を示す基準が必要とされる。通常は、あるテスト等において、一定の水準が定められ、それを超えるかどうかが問われることになる。このときの到達度を表す基準（クライテリオン criterion）に基づく評価は、「絶対評価」、「到達度評価」、「目標準拠評価」などと呼ばれる。また、到達度に関して、いくつかの尺度を用意して、学習者が到達している点と弱点をプロフィールの形に表すことによって、次の学習に活かす「個人内評価」もしばしば採用されている。

なお、通常、クライテリオンを定めることは、必ずしも容易なことではなく、しばしば準拠集団の中での相対的な位置づけの情報をもって代替することがある。むしろ、絶対評価と言っても、完全に「絶対」的な評価はあり得ないと考えておいてよい。また、領域によっては、準拠集団を定めることが難しかったり、到達度を表す基準を理論的に定めることも難しいということがあり得る。そのような場合は、「よい」と思われる事例が備えている要素を調べて（ベンチマーキング benchmarking）、実際の評価対象がそれぞれの要素を達成しているかどうかをチェックしていく場合もある。この場合は、その個々の要素が基準、あるいは、観点を構成することになる。

一方、質的な評価情報の場合には、明示的な基準が定めにくい。この場合には、評価対象に関する一定のモデル等に基づいて、いくつかの観点を設定して、その観点ごとにどの程度の水準にあるかについてチェックしていく方式がとられることがある。このような評価のことを「観点別評価」と呼ぶことがある。

ただ、質的な情報の場合には、敢えて観点ごとにチェックすることが、評価の目的に照らして適当かどうかは考慮の余地があるところであり、むしろ、「評価者」、「被評価者」の主観的な「世界観」が評価情報を解釈する枠組を構成することになり、それが一つの基準に位置づけられる。自己評価等の中で取り入れられるリフレクションでは、まさに、自分の「世界観」の中に評価情報を位置づける作業が行われているとみなすことができる。また、他者とのディスカッションでは、「評価者」のもつ世界観と「被評価者」のもつ世界観とのぶつかり合いを通して、「被評価者」が自らの活動に関してある種の「気づき」がもたらされることが期待されることになる。

評価結果の表示

評価対象のモデルに従ってある種の基準が設定されると、それに即して収集された情報をどのように表示すべきかということが次の課題になる。

「量的情報」による「相対評価」では、準拠集団における位置づけを明らかにするために、順位という形で示されることが少なくないが、集団の規模の違いを勘案するために、その順位に対応してさまざまな変換が施される。最も多く採用されているのは、集団における測定値の分布が正規分布に近似されることが多いことから、標準得点、 z 得点などと呼ばれる、正規分布に従うことが想定されている得点に変換する方法がしばしば利用される。 z 得点は、平均値・標準偏差に対応する値に応じて、偏差値（平均50・標準偏差10）、偏差値IQ（平均100・標準偏差15）、SATやTOEFL等の標準得点（平均500・標準偏差100）などの値にさらに変換されて利用されることがある。

また、測定値に歪みがあったり、外れ値が含まれるなど、正規分布からかなり逸脱する場合には、 z 得点を利用するのは望ましくない。そのような場合には、測定値を順序化して、その相対順位が、正規分布を図2(a)の5つの部分に分けたときのどこに属するかということで、5段階評定する方法がしばしば利用される。さらに、「単に順位を集団の人数に対する百分率で表すパーセンタイル順位」、「パーセンタイル順位に等しい正規分布の累積確率 p に対応する標準正規分布の得点 z を対応させる正規化得点（図2(b))」、「それを偏差値化した T 得点」などが利用される場合もある。

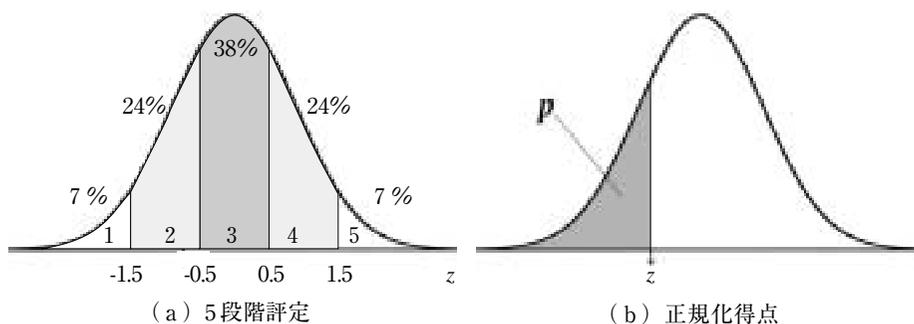


図2 5段階評定と累積確率に対応する正規化得点

質的な評価情報に関しては、一定の方式があてはめられるわけではないが、質的研究法の影響もあり、そこで取り上げられているさまざまな要素の関連図を図示したり、要素のカテゴリ分けなどが評価結果として記載されることが多い。また、基本的には、質的な評価情報は、一つのストーリーの中に位置づけて、文章の形で記載することが通常であり、その具体性を通じて活動の改善に結びつけられることになる。

なお、「量的情報」も、最近では、単一の数値で表されるということはむしろ少なく、複数の指標を並べて図示することによってプロフィールの形で表し、そのパターンごとに分類が行われて、それぞれの種類の質的な記述が試みられる場合もある。一方、「質的情報」では、その一側面を数量化して示すことによって、質的情報の具体性によってかえって導かれやすい過大解釈に陥らないような配慮も可能となる。このように、「量的情報」と「質的情報」は、相反するものではなく、それぞれの利点を活かしつつ、むしろ相補的に活用すべきものである。

7. メタ評価

「大学評価」のメタ評価

「大学評価」の最初の結果が2002年3月公表されて以降、国立大学協会からのコメントやいくつかの新聞の社説などで、機構の「評価」に関するメタ評価の必要性を主張する記事が散見された。「メタ評価」とは、評価に関する評価を指す。確かに、評価をやりっ放しにせず十分に吟味する姿勢は常に問われるべき大切なポイントである。しかし、「評価」自体、評価目的を達成するために実施されるひとつの事業であって、さまざまな要素が複雑に絡み合っていることから、評価のプロセスの中で同時進行的にメタ評価を行うことは言われるほどに簡単なことではない。「大学評価」のメタ評価を強調している主張においても、その具体的な手法についてまで提示されているわけではなく、いずれも理念的な範囲の言及に留まっている。それも当然で、現時点では、メタ評価に関しては、まだ十分な定見が固まっていない状況にあると思われる。

メタ評価は、ある「評価」を評価対象とする「評価」であり、以上に見てきた「評価」の諸要素を同様に含むものである。したがって、メタ評価の「目的」設定、それに応じた「対象」の同定、「手法」の選定、得られたメタ評価情報の解釈と活用といった、プロセスを経ることになる。

まず、メタ評価の「目的」としては、通常、「評価の目的が達成されたかどうか」をチェックするというのが、第一に取り上げられるであろう。そして、その目的を達成するために、当該「評価」において選ばれた「評価対象」、「評価手法」、「評価情報」などが適切であったかどうかという検討が行われていくことになる。例えば、「大学評価」では、「大学等の教育研究活動の改善に資する」ということと、「評価結果を公表して社会からの理解と支援を得る」という2つの目的をもって行われており、それが達成されたかどうかを検討することがまず求められることになる。平成12年度着手の全学テーマ別評価では「教育サービス面における社会貢献（略して、「教育サービス」）」が取り上げられているが、大学における「教育サービス」活

動に改善が見られたかどうか、まずはメタ評価の対象として取り上げられることになる。「教育サービス」活動の改善の有無をチェックするためには、評価直後ではなく、一定期間の後に、大学等における「教育サービス」活動の状況を調査等で確認していくことが考えられる。この種のフォローアップ調査は、メタ評価の目的からすれば、必ずしも全学テーマで対象となった全大学等を対象に調査する必要はないが、いずれにしてもそれなりのコストがかかることになる。また、「公表によって社会からの理解と支援を得る」という全学テーマの「評価目的」についても、「社会」を直接対象にして、「教育サービス」の認知度や、利用度、支援度などについて情報を収集することが望まれるが、このような調査にもかなりのコストがかかる。したがって、当該「評価目的」の達成度に関してのメタ評価は、予算の割り当て、調査への協力等、社会的な理解が得られることが前提となるであろう。

「大学評価」に関わるメタ評価への要望は、むしろ、そのプロセスの適切性の検討という点に向けられているものと思われる。「改善に資する」評価、「社会からの理解と支援を得る」評価であるためには、評価結果が、現状を的確に反映するものであり、また、適切に「水準」の判定に結びつけられていることが前提となる。しかし、「大学評価」のプロセスの適切性を評価対象とする場合には、厳密には、個々の大学の「自己評価書」及び「根拠資料」等に遡って、そこから「評価報告書」に評価結果が適切にまとめられたかどうかをチェックする必要があるが、たとえサンプルを抽出して行えばいいという部分があったとしても、それにかかる労力と費用を考慮すれば、そのメタ評価手法の実現可能性が低いことは明白であろう。したがって、「大学評価」のプロセスに関するメタ評価は、機構がそのプロセスの中で収集した、自己評価や評価結果に関する、「被評価者」である大学等からのアンケート調査、また、「評価者」として参加した「専門委員」、「評価員」などからの評価実施に関わるアンケート調査、その他の外部団体などに照会した「大学評価」に関する諸意見などを通じて、評価のプロセスの適切性を検討していくことが現実的なメタ評価的対応として位置づけられることになるであろう。

評価のプロセスに関するメタ評価を本格的に行うのであれば、評価の当初から、メタ評価の担当者を配備し、同時進行的に適切な方法によってモニターしていく体制を導入することが現実的であろう。ただし、これも、労力とコストの問題があり、必ずしも実現可能性の高い方法ではない。しかし、「大学評価」のような組織的な評価の場合には、被評価者となる大学等、あるいは、それを取り巻く社会のコンセンサスが得られさえすれば、しかるべきメタ評価のシステムを同時並行的に導入することも可能であり、またそれが望まれてもいくことになるであろう。

メタ評価基準のいろいろ

「大学評価」は、それ自体、社会的な「事業的」色彩をもっており、その意味で、事業・開発・プログラムなどの評価の適切性に関する基準は、そのまま「大学評価」のメタ評価の参考にすることができるであろう。

例えば、OECD（経済協力開発機構）のDAC（開発援助委員会）では、開発援助活動の適

切性について検討するための観点として5つの基準を挙げている (<http://www1.oecd.org/dac/Evaluation/htm/evalcrit.htm>; 龍・佐々木, 2000などを参照のこと)。ここでは、「開発援助活動」の代わりに「評価」という言葉をあてはめて、メタ評価の基準として以下に表現し直して示しておきたい。

①妥当性 (relevance) : 「評価」目的に沿って「評価」が実施され想定される結果が得られているか。

②目標達成度 (effectiveness) : 「評価」目的が達成されているか。

③効率性 (efficiency) : 「評価」にかけられた労力やコストは効率的であったか。

④インパクト (impact) : 「評価」により、直接・間接に、何らかの変化がもたらされたか。

⑤自立発展性 (sustainability) : 「評価」の後に、「評価」や「評価」からもたらされた変化が継続しているか。

「大学評価」の例では、「妥当性」、「効率性」などの観点は、評価のプロセスの中で、計画的にチェックしていくべき点であり、「目標達成度」、「インパクト」、「自立発展性」に関しては、評価数年後に改めて調査すべき観点ということになるだろう。なお、「妥当性」については、直後に触れる、評価手法としての測定の「妥当性 (validity)」とは意味が多少異なるという点に注意されたい。

また、アメリカで策定されたプログラム評価基準 (教育評価基準合同委員会, 1994; Patton, 1997) では、メタ評価に関して、以下の4つの観点と、それぞれの観点におけるチェック項目が取り上げられている。

①有用性基準 (utility standards) : 評価の想定利用者の情報ニーズに役立つこと。

- (1) 利害関係者が明確にされていること
- (2) 評価者の信頼性が確保されていること
- (3) 情報の範囲と選択が適切であること
- (4) 結果の解釈のための価値判断の基礎が明確であること
- (5) 評価報告が明瞭であること
- (6) 報告がタイムリーであり、広く普及されていること
- (7) 利害関係者をフォローして評価のインパクトが検討されていること

②実現可能性基準 (feasibility standards) : 評価が現実的であり、効率的であること。

- (1) 評価作業の進め方が実際的であること
- (2) さまざまな利益集団の異なる立場を配慮して計画され実施されていること
- (3) 効率的、かつ、資源の活用が正当化できること

③正当性基準 (propriety standards) : 評価が多くの人に影響を及ぼすことを配慮し、法的、倫理的に問題なく評価が行われること。

- (1) すべての層の参加者のニーズに応えられるように設計されていること
- (2) 評価の当事者が負う責務が明示されていること

- (3) 評価に関係している人々の権利を尊重するように設計されていること
- (4) 評価者が被評価者と接する際に、被評価者の価値を尊重していること
- (5) 完全かつ公正な評価結果が記述されていること
- (6) すべての評価結果が適当な制約と共に、評価に関係する人々に公表されていること
- (7) 利害関係が評価のプロセスや結果に影響を及ぼしていないこと
- (8) 評価にかかる会計が健全な手順に従っていること

④正確さ基準 (accuracy standards) : 評価結果が十分な情報を伝えるということ。

- (1) 評価結果が明確かつ正確に記述されていること
- (2) 評価対象の背景・状況が検討されていること
- (3) 評価の目的とそれに基づく手続きが明確に記述されていること
- (4) 評価に用いられた情報源が明らかにされていること
- (5) 有効な評価情報が収集されていること
- (6) 信頼性の高い情報収集が行われていること
- (7) 収集された情報から評価結果が一貫性をもって導かれていること
- (8) 量的情報が評価の目的に即して適切に分析されていること
- (9) 質的情報が評価の目的に即して適切に分析されていること
- (10) 評価結果が、評価の関係者に利用しやすいように明示的に正当化されていること
- (11) 評価者の個人的な感情や偏向による歪みがおさえられ、公平な評価結果となっていること
- (12) 評価自体が、形成的、総括的に評価されていること

これらのメタ評価観点は、網羅的であり、そのすべてを満たしているような「評価」自体を見出すことは存外に困難なことであろうと思われる。そもそも、メタ評価を行う「メタ評価者」としては、例えば、当該「評価」の「評価者」自身、「被評価者」、第三者の同様の「評価」を行っている個人や組織など、いくつかの立場が想定されるが、それぞれメタ評価の観点も異なってくる。また、当該「評価」の目的によってもメタ評価の観点は異なるであろう。例えば、「形成的評価」であれば、当該「評価」の対象となる活動のプロセスにおいて、どのようなタイミングで「評価」を行っていくかという「評価計画」が適切であるか、また、その活動のプロセスをモニターする手法が適切であるか、といったことがメタ評価の観点とされるであろう。それに対して、「総括的评价」であれば、「評価結果」の報告がその内容や報告先の範囲などの点で適切であるかといったことがメタ評価の観点となる。このように、「評価」の特徴によって、メタ評価の観点は適切に選択される必要があるが、そのような際に、少なくとも上記諸観点は、十分に視野に入れておくべき基礎として位置づけられるものであろう。

評価手法の適切性の検討

評価手法、とりわけ、測定 of 適切性については、教育評価、教育測定学などの領域で、理論的な整備が進んでいるところである (Linn, 1989等を参照のこと)。特に、「量的情報」は、あ

る種の「測定」値として得られていることが通常であり、測定理論の見地からその適切性について検討する理論と方法が提案されている。

例えば、測定値に誤差がどの程度含まれるか、あるいは、何度測定しても同様の結果が得られるかという、「信頼性 (reliability)」と呼ばれる観点がまず問われることになる。さらに、測定値が評価対象の意図される特性を反映しているかという点で「妥当性 (validity)」と呼ばれる観点も合わせて確認しておくことも望まれる。「信頼性」に関しては、測定論的に比較的検討しやすい定式化がなされており、一定の集団で測定値がデータとして集積されれば、統計的に検討していくことがある程度可能である。一方、「妥当性」については、測定対象に関する理論的モデルが明確であるかどうか、その対象に関する他の指標と比較できるかどうかといった点にも依存する。その点で、測定値が得られれば検討できるというものではなく、測定対象に関わる領域の理論的研究の進展に待つ部分が少なくない。

もう一点、測定値のもつ「識別力 (discrimination power)」あるいは「情報量 (information)」という点についても目を向けておくべきである。旧来の測定理論においては、「信頼性」は、ある測定方法において、測定値の値域のすべての範囲において一定の値をもつことが前提となっていたが、多くの場合、ある測定方法は、特性値のある部分においてのみ十分な情報をもたらすということがある。その場合、その特性値の付近において、「識別力」が高いという言い方をする。あるいは、その特性値における情報量が大きい、測定の精度が高いといった言い方をする場合もある。

「識別力」の考え方は、評価の手法を定めるときにも重要な視点を与えてくれる。「大学評価」などにおいても、限られた資源と労力の範囲で、あらゆるレベルのチェックをしなければならないというのはほとんど不可能であり、評価目的に従って、その目的を達成するための「評価情報」が限定的に得られるように設計できれば効率的である。例えば、「質の保証」などの評価目的がある場合には、大学としての一定の基準を満たしているかどうかという、いわば、ミニマム・リクワイアメントの部分の識別力が高くなるような評価手法を講ずればよい。また、国際的な通用力のあるような研究活動を促進していこうという評価目的があるとするれば、国際的にも卓越した研究であるかどうかというレベルの識別力が高くなるような評価手法を採用すればよい。同じ評価手法で、特性値のすべてのレベルをカバーできるというのではなく、識別力が低くなるようなレベルでは、誤差も大きくなるので、評価結果も、識別力の高い部分が強調されるような表示方法が望まれることになるであろう。その意味では、「大学評価」における「水準」判定の記述などは、いたずらに段階数を増やすよりも、ある部分の識別が明確になるような範囲で設定する工夫が望まれていくことになるであろう。

「質的情報」に関しても、それをコード化したり、カテゴリーチェックするなどの方法で分析的に検討する場合には、「信頼性」、「妥当性」、「識別性」などの観점에서検討することができるが、一般には、その検討手法はまだ十分に定式化されているとは言えない状況にある。近年、質的情報については、質的研究法の開発が進む中で、グラウンディド・セオリー・アプローチ (grounded theory approach) などの解析手法が普及しつつあるところであり、質的情報の

まとめ方に関わる評価についても議論が進められているところである (Glaser & Strauss, 1967; Flick, 1995; Holloway & Wheeler, 1996などを参照のこと)。その中で, Holloway & Wheeler (1996) では, 質的研究の評価に関してしばしば取り上げられる観点として, 「信用可能性 (credibility)」, 「移転可能性 (transferability)」, 「明解性 (dependability)」, 「確認可能性 (confirmability)」の4点が挙げられている。これは, 「研究」の評価であるが, メタ評価に置き換えてその概要を簡単に示すと以下のようなになる。

「信用可能性」は, 「評価者」が「評価対象」にどのように関わったかを明らかにすること, いくつかの方法によって「評価情報」を得ること (トライアングレーション triangulation), また, 専門家や「評価」に関わっている人 (被評価者など) に「評価情報」のチェックを得ることなどによって, 高めることができる。「大学評価」においては, 「質的情報」の形で評価結果を公表している部分が多いが, この点に関しては, 「専門委員」, 「評価員」などが公表され, 複数の評価手法に基づいて「評価結果」を導き, それに対して, 「意見の申立て」の機会が与えられるなど, 十分な取組がなされていると言えるであろう。

「移転可能性」は, 「評価結果」がどのように一般化され得るかという観点である。「質的研究」では, 比較的少数の対象に関して, 文脈に依存した結果が提示されるので, この点は厄介な問題である。少なくとも, 理論的枠組が確かであること, また, 研究対象の背景や文脈が明確に記述されることが望まれる。そのことによって, 同様な背景や文脈と思われる状況に, 研究的知見が適用できることになる。「大学評価」も, 個々の大学の事情が異なる以上, 「評価」は, それぞれの大学の背景や活動の文脈に即して行わざるを得ない。その意味では, 「大学評価」は, 事例的・臨床的な性格を根本的に有しており, 「評価結果」の記述は, 「活動主体」の大学が, 自らの状況に合った結果であると認知されて, はじめて「改善」に結び付けられていく。したがって, 「評価結果」は, 対象となる活動に関する的確なロジックモデル等に基づいて, 背景や文脈なども適切に記述されている必要がある。この点に関しては, 現行の「大学評価」は, ロジックモデルが十分に洗練されているとは言い難く, 大学等の「自己評価書」においても活動の背景や文脈が的確に記述されていない場合も少なくないという状況にあり, 改善の余地が多く残されている段階と言えるであろう。

「明解性」は, 「評価」のプロセスが他者から後追いできるようになっているということである。「信用可能性」が大きければ「明解性」も大きくなるが, そのことを明示するためには, 「評価」のプロセスが外部からモニターされることが必要となる。この点に関しては, 「大学評価」では, 評価のプロセスにおいて「評価者」, 「被評価者」対象のアンケート調査を随時行ったり, また, 評価結果に関して専門委員会及び大学評価委員会で重複的にチェックする手続きを経るなど, モニターという点ではそれなりの配慮がなされていると言える。

「確認可能性」は, 「評価情報」とそれらの出所である資料等が明確に関連づけられていて, 「評価結果」の読み手にも同じように確認が可能であるということである。この点に関しても, 「大学評価」では, 自己評価書及び根拠資料によって, 大学等の自己評価を追跡することが, 評価作業の基本となっている。また, 評価報告書においても, 評価結果を導く根拠を明示する

ように配慮されている。

しかし、以上の観点に関して、それぞれどの程度機能したかという点では、「大学評価」はまだ始まったばかりでもあり、多くの課題が残されているという段階であると思われる。一般に、質的評価、量的評価にかかわらず、メタ評価に関しては、コストや労力の点から、「評価」のプロセスに明示的に組み込まれていない場合が少なくない。「測定」という枠組みにおいては、その検討の定式化が行われており、とりわけ大規模なテストや質問紙については信頼性・妥当性に関する検討を行うことが前提となっているが、質的な評価、活動の改善のための評価に関するメタ評価のあり方については、手法の定式化も含めて、課題が山積している段階にあると言えるであろう。

8. 評価のインパクトと評価文化

「大学評価」のインパクト

「評価」は、評価の枠組を超えて、さまざまな影響力をもつ。とりわけ、大規模な評価の場合は、その影響力も広範に及ぶ。「大学評価」もその一つであり、機構の最初の「評価」結果は、2002年3月22日記者発表され、その翌日の新聞などで広く報道された。

しかし、例えば、全学テーマ別評価の「教育サービス面における社会貢献」における「項目別評価」の「水準判定」について、「取組の目的及び目標の達成への貢献」という表現を受けてのものと思われるが、ある新聞では「社会貢献度」といった言葉を見出しに使う、しかも、水準ごとに大学名を並べて大学の社会貢献度が序列化されている印象を助長するような記事が見られた。あくまで、「大学評価」は、大学個々の目的・目標に即して、それぞれの教育研究活動の改善に資する目的で行われているものであり、大学の序列化を可能とする評価ではない。また、「大学の社会貢献度」を客観的に測定するような評価は行っておらず、あくまで、各大学の「教育サービス」の目的・目標に即した項目別評価である。にもかかわらず、ある程度は予想されたこととはいえ、一部の項目ごとの水準が、その本来の意味が隠されて、評価もされていない大学全体の「社会貢献度」なるものの相対的な順序尺度として公表されてしまうということが実際に起きているのである。

また、この評価によって、大学の「教育サービス」活動が改善されていくことを願う、「評価情報提供者」の立場からすれば、「活動主体」である大学が、現時点では、本来の業務としての位置づけはまだされにくい自主的サービスを促進するためには、意欲的に取り組んでいる事項を取り上げて前向きの評価をフィードバックすることが効果的であるのは自明のことでもある。そもそも、「取組の目的及び目標の達成への貢献」に関する項目別評価では、目的・目標に関わらない「取組」は評価の対象として取りあげられないわけで、その項目別評価の水準判定の分布が高い水準に偏っているのは、むしろ妥当な評価が行われている証拠でもある。それが新聞紙上では、「お手盛り評価」と批判的に書かれ、評価目的に取り上げられてはいない「文部科学省の予算配分」のためには評価のあり方の見直しが必要であるといった論調になってしまっている。それが、社会のみならず、大学側にも伝わり、評価自体に対する疑心暗鬼を呼び

起こしてしまうなど、「大学評価」そのものにも少なからず影響が及ぶ事態となっている。

一方、評価の結果以前に、「大学評価」を実施するだけで、大学側からある種の対応策が引き出されるという効果も散見されている。例えば、「教育サービス」の項目別評価として、「改善のためのシステム」が取り入れられたことによって、ここ数年、「教育サービス」に関わる委員会が新たに設置された大学もある。これは、「教育サービス」活動を促進するという「大学評価」の目的からすれば望ましいことであり、他者の評価が入ることによるある種の「強制力」が働く証左でもある。しかし、「評価」の対象を、ある時点での「教育サービス」の状態として、その状況をできるだけありのまま表現して評価結果とすることが求められるのであれば、そのような「強制力」は、既に、大学の状態に影響を及ぼしているという点で、「ありのまま」と言えるかどうかは微妙な問題となる。評価の強制力が増すことによって、被評価者の状態が影響を受けるという、いわば「不確定性原理」的状況が生ずる可能性があるということにも留意しておく必要があるだろう。

評価の枠組の中での「真正の評価」

「評価」は、評価の「枠組」の中だけで済まされない問題を常に内包している。例えば、大学の入学試験は、大学のある種の「管理目的」のために行われる「評価」とみなすことができるが、大学という基本的な評価枠組を超えて、初中等教育に対して大きな影響を及ぼしていることはここで改めて言うまでもないことであろう。入学試験は、「大学」にとっての「評価」のみならず、受験生にとっては一生を左右する大きな意味をもち、しかも社会的に大規模に行われる「評価」である。この種の「評価」を「ラージスケール (large scale)」の評価と呼ぶことがある。ラージスケールの評価は、生徒の学習目的を、学ぶ内容等を身に付けるということから、高い評価結果を得るためへと、本来とは異なる方向に歪める危険性のあることが指摘されている (Harnisch & Mabry, 1991)。

そこで、本来の教育活動の目的を達成するために、いくつかの新たな評価方法が提案されてきている (Worthen & Sanders, 1987 ; Harnisch & Mabry, 1991 ; Gipps, 1993 ; Hodson, 1998 ; 井上, 2001など)。その中で最近強調されてきている「評価」として、「真正の評価 (authentic assessment)」と呼ばれる考え方を取り上げておく必要があるだろう。これは、「評価」が教育のプロセスの中に埋め込まれていて、「評価」として特別に浮き上がらないような「評価」である。教育のプロセスに埋め込まれるということは、テスト、テストで、学習者をたたき上げるということでは決してない。むしろ、「テスト」ということが意識されないような形で、自らの学習が適切に行われているかをその流れの中で自然に確認できて、学習が自ずと促進されていくことが期待されているものである。つまり、「真正の評価」はその活動の向上に不可欠であって、「評価情報」を得ることによって、「活動主体」は自ずとそこで求められている活動を促進させていくことができるのである。そこでは、「評価」は、既に行われたことの情報をフィードバックするという部分に力点が置かれるのではなく、これから先の活動をどうすべきかということにフィードフォワードする役割が強調されてきているのである (井上, 2001)。

最近では、欧米では、評価を evaluation ではなく assessment と呼ばれることが多くなっているのも、使い分けは一概に明瞭であるわけではないが、そのような点が意識されているということもあるであろう。

「真正の評価」の具体的な方法としては、現実の課題場面と同様な状況で実際に活動を行わせるパフォーマンス・アセスメント (performance assessment), 自らの活動記録や作品等を蓄積して (紙ばさみ portfolio に挟み込む形で), それら全体についてリフレクションするポートフォリオ評価 (portfolio assessment) などがある。特に, 新学習指導要領で注目されている総合的学習の評価方法として, ポートフォリオ評価の活用可能性がわが国でも広く試みられ始めているところである。また, 英国では, ポートフォリオ評価に類するプログレス・ファイル (progress file) を大学の標準的な学生の学習記録として導入し, それに基づいて学習法の指導も行う動きが起こってきており (<http://www.qaa.ac.uk/crntwork/progfileHE/guidelines/progfile2001.pdf> 参照), 生涯学習社会・グローバル化社会における学習促進のための学習評価手法としても注目されている (大塚, 2002)。

この種の評価は, 評価の枠組を超えて, 望ましくない影響を広く及ぼす危険性が相対的に小さく, 評価対象となっている活動に活動主体が専心できやすいという特徴をもっている。したがって, 「活動主体」の「改善」が目的となる評価では有力な評価法となるが, 一方で, 質的な評価情報が扱われることが多いので, それをどのように処理するかという難しさがある。ただ, 「評価情報提供者」と「活動主体」の間にコミュニケーションを誘導するきっかけを作ることができたり, いわゆる「コミュニティ」作りの観点からは, 今後ますます重要視されていく評価の考え方になっていくものと思われる。

評価文化

「評価」を, 「活動主体」の活動の「改善」のためのツールとして捉える考え方は, 新しいわけでは決していないが, 一般にはまだそう定着しているものとは言い難い。ましてや, 「評価」がある「コミュニティ」においてより十全に機能するという見方, また, それを「コミュニティ」を形成するためのツールとして捉える考え方は, 現時点では特にコンセンサスが得られているわけではない。「評価」は, 対象を数値化することであり, 序列化することであるといった評価観が優勢である現状では, 「評価」の多様性, 複雑性を少しでも多くの人々に共有してもらい, 「評価」の枠を広げていくことがまずは求められる第一のことであるだろう。評価の多様性も含めて, 評価に関わる基本知識を啓蒙し, 少しずつ新しい評価の意義を広げていく努力は, 「評価」がブームになっている現代においては, 緊急を要する課題になっているとも言えるだろう。

「評価」についての基礎知識に基づいて, 「評価」を適切に使いこなせる力のことを, 「評価リテラシー」と呼ぶことがある (Stiggins, 1992; 大野木, 1993; 大塚, 1994)。「評価リテラシー」は, 以上に見てきたような「評価」の諸要素のあるべき姿を把握する力, 「評価情報」の意味するところとその限界を見極める力, 評価の枠組を超えた評価の影響力を予測する力な

どを含んでいる。そのような力を有することによって、「評価」に振り回されるのではなく、それを適切なツールとして、活用していくことができる。そこで、新しい評価の考え方などに触れる機会が増やされ、「評価リテラシー」を高め、「評価」が「活動主体」の改善のために必要不可欠な、「活動主体」にとって、脅威な存在ではなく、元気の源として、万人の中に根付いていくことが望まれる。そのこと自体、新たな評価観を共有するひとつの「評価コミュニティ」とでもいうべきものが形成されることとみなすこともできるだろう。そのように、新たな「評価文化」を我々が共通の背景として持ち合わせることによって、「改善」に資する「評価」がより有効なものとして機能していくことにもなるであろう。

9. TQM 的アプローチと大学評価

大学経営に対する TQM アプローチ

「評価」について、その要素をいくつか取り上げ、それぞれの視点から「評価」について考察してきたが、最後にまとめに代えて、近年、大学に求められている経営的センスのひとつとして、TQM（総合的品質経営：Total Quality Management）というコンセプトを取り上げ、その視点から、「大学評価」への示唆を引き出してみることにしたい。

大学経営における TQM 的アプローチについては、Seymour (1993) がいくつかの視点を示している（表 2）。TQM「的」と表したのは、それらの同種の内容がしばしば別の呼び方がされているからで、Seymour はそれらを「戦略的品質経営」としてまとめている。いずれにしても、単に産み出された「品（大学の場合は、送り出す学生や研究成果）」だけの質を問題にするのではなく、それに関わる組織とすべての「人」の「質」を高めることによって、「品」の「質」も向上させようという「総合的」な経営を志すものである。

表 2 に示したように、TQM では、まず、経営に関与する「人」を取り上げる。経営の対象である「顧客」を明確にし、そのニーズを満たし、さらに、それを超える喜びを与えることを目指す。それと共に、経営に携わるすべての「人」の志気が高められるということを重視する。そのために、経営を導く「リーダーシップ」、経営を推進する「人的資源の開発」が求められる。そして、それらの「人」が自らの役割を十分に発揮するために、経営に携わる人々の「恐怖の減少」と、職務の「認知」及び「報奨」の適切な手配が有効となる。それはまた、経営に携わる人の「チームワーク」に結び付いていき、その連携の下に、より高い質の経営が実現されることになる。

経営に携わる人々の満足感を高めるためには、まずは、現状を明確に把握すること、すなわち、状況の「測定」が的確に行われることが必要とされる。そして、その結果は、最終的なゴールではなく、次の活動に常に結び付けられ「不断の改善」が行われることが望まれる。また、そこで収集される情報は、断片的なものである場合が多いが、経営や組織は「システム」として有機的に動いていくものであり、常に情報間の関係性に留意しながらそれらを解釈していく姿勢が肝要である。それを通して、個人に偏することなく、「組織としての問題解決」に結び付けられていくことになる。

表2 TQMにおける一般原理の抜粋 (Seymour, 1993より筆者が抜粋・補筆)

●TQMにおける高品質を得るための一般原理

- ①顧客の要求に適合・それを超えること…学生, 企業, 社会, 家族 (外部顧客)
- ②すべての人の職務…教員, 事務職員, 技官, 補佐員, 学生 (内部顧客)
- ③リーダーシップ…ミッション, 学長, 学部長, プロジェクト主査, 授業担任
- ④人的資源の開発…新規採用, ファカルティ・ディヴェロプメント
- ⑤恐怖の減少
- ⑥認知と報奨
- ⑦チームワーク
- ⑧測定…自己点検・自己評価, 学生による授業評価, インタビュー, アンケート
- ⑨不断の改善…PDCAサイクル
- ⑩システム論的アプローチ
- ⑪組織的問題解決

●TQMと大学経営の結びつき

- ①品質は, 教室内の教員と学生間の相互作用やアクレディテーションの基準への適合といった事柄を超えた広がりをもつ。戦略的品質経営は, この広い定義を包含する多次元の原理である。
- ②大学は学習の進展を追求する。戦略的品質経営は学習組織を創造する構造的システムである。
改善を目指すための理由・現状の記述と問題点の特定・問題点の根本原因の認識と検証・根本原因を正すための対抗策の計画と実施・対抗策の達成状況の確認・問題点とその根本原因の再発防止のための標準化・残された問題についての計画と実施プロセスの効果評価
- ③大学は孤立した個々の部品の集合として運営されている。戦略的品質経営は統合された, 目的の明確な全体性を高める, 統合力である。
戦略的品質経営の基本公理の多くは, 組織にシナジー (共同作用) を発生させる効果をもつ。

●大学文化を変えるシステム

- ①組織全体を覆う継続的で我慢強いリーダーシップがなければ品質の高い文化はありえない
- ②品質の高い文化は, 機関の新規採用と雇用のプロセスを通して大いに補強され, 拡大される
- ③品質の高い文化には, 開放的で, 正直で, 効率的で, 効果的で, 壁を破るような, 間違いのないコミュニケーションが必要である
情報を共有しようという意志・フィードバックによる改善・組織の壁を破ること…コミュニティ形成

●品質を物語ること

- ①品質に関する, 真に効果的なコミュニケーションは, 大学がすべての行為において品質を改善しようとする日々の努力に対する密接な理解を得られてはじめて可能となる。(コミュニティ)
- ②品質を伝えることは真の理解と信頼を得るチャンスである
- ③品質を物語を語ることは, 聞くことと語ること, その双方向の行為である (質的評価)
- ④品質を伝えることは, プライドと自尊心を強める, 自己妥当化のプロセスである
- ⑤品質を伝えることは顧客の満足を媒介する (アカウントビリティ)

そのような TQM の基本的な原理は、大学の経営にもあてはめられる。そこで強調されていることは、PDCA サイクルを循環させ、体系的に諸活動の改善を図ることを基本に、それを通して、大学内に一つのコミュニティを形成していくことと言い換えることができるであろう。大学にとってのアカウントビリティということも、顧客の理解を得、大学の持つ価値を共有してもらうという意味で、コミュニティに顧客を引き込む一つの工夫と見ることもできる。そして、そのような交流を産むために、「品質を物語る」ことの大切さが強調されているが、「物語る」ことはそこに関わる「人」に共有されやすい「評価結果」の表現でもあり、ある種のコミュニティの形成とは無関係ではない。そのような充実したコミュニティが形成されていくことが、ここでの TQM 的アプローチの最終的な目標と捉えておくこととしたい。

顧客満足度のとらえ方

TQM の大きな特徴の一つに、「顧客の要求に適合し、それを超えること」ということがある。例えば、大学では、教育や研究活動が行われているが、その活動の質が高ければいいというのみならず、「顧客」がそれに満足 (satisfaction) し、さらに期待を越えて喜びを感じること (delight) が問われることになる。

「顧客」は、ある種のサービスの「受け手」であり、大学の活動を考えた場合、「学生」を思い浮かべるが、学生を採用したり研究成果を活用できる企業など、大学の外部にも「顧客」は存在する。これを「外部顧客」と呼ぶことがある。また、大学内部においても、教授活動を支援するセクションがあれば、大学教員自身が教育というサービスの提供者であると同時にそのセクションからサービスを受ける「顧客」にもなり得る。これらを「内部顧客」と呼ぶことがある。このように、ある活動に関与する人々のほとんどが「顧客」として、ある種の要求をもって活動に参加しており、その要求に応えるサービスが求められることになる。

そこで、まず、対象となる活動に関与する「顧客」の範囲を明確にしてそのニーズを把握することが求められる。そして、「活動主体」の活動によって、「顧客」がどのような満足を得たかを知ることが、活動の「質」を高めるために必要とされていくことになる。この場合、ある活動の枠組が定められたときに、そこに含まれる「顧客」が「被評価者」となり、「活動主体 (その活動を管理する経営者も含む)」が「評価者」ということになる。

「活動」の評価という意味では、「評価対象」として、例えば、「顧客」のその活動に対する満足度が取り上げられることになるが、「顧客」自身は、それぞれの目的をもって当該活動に対する自らの満足度を評定することになるわけで、その点では「顧客」が当該活動の「評価者」、「評価情報提供者」となっている。したがって、「顧客」が自らの満足度を評定するときの評価の枠組を明らかにしておく必要がある。それを押える一つのポイントが、「顧客」のニーズという位置づけができるであろう。すなわち、「顧客」のニーズの違いによって、どのような満足度の差が出てくるかを把握する姿勢が望まれるということである。

また、「満足」ということ自体、いくつかの側面をもっており、それに関するモデル化も工夫される必要がある。例えば、「満足度」は、「顧客」のニーズから来る「活動」への「期待」

と、現実を得られたものの「差」によって決められるものと考えられるが、単純に、その差の単調な関数として表せるものではない。まず、満足度は、「期待」の重要度・大きさに依存する。それが大きければ、満たされた際の満足度も大きくなる。また、その重要度の質的な違いも満足度に影響を及ぼす。重要であっても、それが当たり前提供されるものであれば、「満足度」はそれ程高いものにはならないが、希少なものが提供されるほど「満足度」は高くなる。しかし、逆に、「希少」なものは、提供されなくても「不満度」は大きくならないが、「当たり前」のものが提供されないと「不満度」が高くなるということがある。また、「顧客」が「期待」していないもので、「顧客」の別の次元の満足度を高めることが提供されると、その提供されるものがそれ程大きなものでなかったとしても、「満足度（喜び delight）」を高めることもできる。このように、「満足度」自体、いくつかの側面をもっていることから、多面的に情報を収集することが望まれることになる。

「満足度」に関する「評価」に関して留意すべきことは、「満足度」を高めることだけに改善の方向性（PDCAの「act」）を定めることが逆効果になる可能性もあるということである。

「満足度」を高めるために行われる施策は、当然にあるコストがかかるわけであり、それは、経済的なもののみならず、人的な労力や時間を費やすことにもなり、その結果、別のある要素を失うという可能性を常に考慮しておく必要がある。それが、詰まるところ、「システム」的な情報の見方に関わる部分であり、「外部顧客」だけでなく、「内部顧客」も含めて、総合的に「満足度」がバランスよく高められることを目指すTQMの基本原理に強調されていることでもある。すなわち、「外部顧客」の満足度を高めるために、「内部顧客」の満足度を低めてしまつては、最終的には、「質」は保たれなくなるという点に配慮する必要があるということである。

このように、TQM的な視点からすれば、「評価対象」としては、外部の顧客のみならず、「内部顧客」という概念を持ち込んで、それらの総合的な「満足度」を、複数の手法や指標によって把握するような評価のためのモデルを採用することが望まれる。これは、「大学評価」や「授業評価」の現行の「評価」においても同様で、その結果をTQM的視点をもって活用するかしないかで、評価情報の活かされ方は大きく違うことになるであろう。

例えば、現行の授業評価調査では、さまざまな項目によって、学生の授業に対する満足度に関して多面的な情報を得ていると捉えることができる。そこでは、学生が授業を理解できているかどうかということがまずは前提となるが、それのみならず、学生が新たな学習を志向しなくなったり、何かしら新しい発見が見出されたりということが付加的にあれば、学習を大きく促進させる原動力になり得るであろう。そのように、満足度を多少なりとも多角的に捉えておくだけで、「理解度」が低いという結果が出たとしても、ただいたずらに教えるレベルを下げてわかりやすさを追求すればいいというものでもないことが示唆されるであろう。また、「満足度」を満点に近づけていくことも、必ずしも適切な目標になるわけではない。授業改善のために取り組む、教える側のコストや、教える動機づけも考慮すべきであつて、授業の「外部顧客」であるべき学生の満足度だけを「評価」するのでは、十分な授業改善をもたらすことはお

ぼつかないということを肝に銘じておくべきであろう。

リーダーシップとコミュニティ形成

「経営」という枠組においては、経営者の「リーダーシップ」が求められるのはごく当然のことである。TQMにおいても、それは、重要な役割を担っているが、「リーダー」がもつべき特質に関する評価観点を、TQM的視点、あるいは、コミュニティ形成という視点から検討しておくことも、「評価対象」を適切にモデル化するために意義のあることであろう。

組織の構成員、「内部顧客」の志気を高めるためには、ある種の「コミュニティ」の形成が前提となる。何故なら、「コミュニティ」に属するという事は、そこでの価値、すなわち、活動の目標が明確に共有されることであり、その目標達成に向けて自らの活動を方向づけることができるからである。そして、その際に、「リーダー」は、そのコミュニティの中心的な存在として、コミュニティの成員にその目標を明示的に伝え、それを浸透させることを通して、彼らを牽引することになる。また、同時に、「リーダー」は、コミュニティ構成員のある種の「モデル」になって、自身がその構成員の活動の方向性を規定することにもなる。それは、「リーダー」が構成員を叱咤して実現するものではない。そのことによって、構成員の「恐怖心」が大きくなり、十分な活動促進が果たせないということさえ起こりかねない。むしろ、「リーダー」の存在は、「コミュニティ」の構成員に自然に受け止められるべきものである。

この点は、徒弟制度の観察から導かれた正統的周辺参加論 (legitimate peripheral participation: Lave & Wenger, 1991) が参考になろう。徒弟制度では、弟子をスパルタ式に育て上げるイメージが強いが、正統的周辺参加論で強調されている点は、名人は弟子に直接的な技能指導をすることはそれ程顕著ではなく、弟子は名人の技能のみならず名人の背中を見て全人格的に育っていくということである。すなわち、名人は、そのコミュニティの中心的存在であって、それを目指して、そのコミュニティの周縁的存在であった弟子が、コミュニティにおいて確固たる位置を占めるようになっていくのである。正統的周辺参加論では、ある学習コミュニティにおいて、周縁的な存在から自らがコミュニティに定着していくプロセスそのものを「学習」と捉える新しい学習観を提供しており、ここで筆者が提示している「評価観」もまさにその考え方に基づくものに他ならない。

そのような視点から「リーダー」のもつべき特質をあげるならば、(1)「リーダー」から構成員に目標が明確に伝えられ浸透しているかどうか、(2)構成員同士が十分なコミュニケーションをとる場が準備されており機能しているかどうか、(3)チームワークが十分とれているかどうか、(4)「リーダー」を構成員が自らが目指すべき模範的存在として捉えているか、といったところになるであろう。そのような「リーダー」の下、構成員自身が成長しているのであれば、(5)構成員が順調にそのコミュニティの中で、その周縁的な存在から、ある役割を明確に分担できるような位置へと成長しているかどうか、(6)コミュニティの構成員の新陳代謝が円滑に進められているかどうか、といった観点も間接的に「評価対象」として組み込んでいくことが可能であろう。

「リーダー」の存在は、大学や授業を評価する際の観点として重要であるばかりでなく、活動の改善に評価結果を結び付けるためのある種の強制力としても重要な機能を果たすことになる。適当な評価のモデルが構成されたとしても、評価結果が次の活動に活かされないのであれば意味がないわけで、いわゆる PDCA サイクルを回すためにリーダーシップの存在が必要となるのである。その際には、組織としてのミッション・ステートメントを明示すること、それによって構成員が評価情報を自ら価値づけるように方向付けられること、さらに、構成員が得られた評価結果の片面的な解釈に陥ることのないように配慮すること、などもリーダーの役割となっていくであろう。

また、一方で、次の活動に結び付けることが前提となる評価の場合には、評価結果が構成員にただ伝わればいいと言うものではなく、構成員が改善のために行動を起こしたときに、それが比較的容易にできるような支援体制を整備しておくことも大切である。それもまた、リーダーの役割ということになるであろう。

なお、コミュニティという場合には、そのリーダーは必ずしも組織の長である必要はない。「組織」と「コミュニティ」という語を分けて使う理由は、コミュニティが目に見えない形で自然に形成されていく性格をもっているという点にある。つまり、一つの組織の中にあっても、ある役割を果たす人の周囲には、自然にあるコミュニティが形成されていき、その人を一種のリーダーとして、そこに取り込まれる構成員が共通の目標をもって、その達成のために活動を進めていくという図式が想定できる。活性的なコミュニティは、その主たる構成員が、ある種のサブ・コミュニティのリーダーとなって活動していて、そのサブ・コミュニティ同志の連携が円滑に進められ、母体となるコミュニティを活性化しているということである。そこには、「長」を明示的に名乗る必要のない、人的なつながりがあると見ることができるであろう（金子, 2002）。

大学の活性化と評価文化の形成

TQM では、評価結果をストーリーとして表現し、それを組織で共有していくことが基本理念として取り上げられている。ある種の評価情報は、もちろん、数量的に表現されていた方が、「活動主体」にとってもわかりやすく、一般的には公平感も高く感じられるが、目標に向けての活動を促進・改善するには必ずしも有効に機能しないことも以上に見てきた通りである。むしろ、数量的な情報に基づいたとしても、最終的には、一つのストーリーとして記述されていることが、評価情報を利用する「活動主体」がイメージを作りやすく、具体的な活動の改善に結び付けていきやすい。そして、そのストーリーを通して、「活動主体」は、自らの活動に「プライド」をもち、活動への自己効力感 (self-efficacy) を高めていくことができる。Seymour (1993) が「品質を伝えることは、プライドと自尊心を強める、自己妥当化のプロセス」と述べているように、その過程は、自らを「活動主体」とした循環的な評価サイクルを産み出していくことになる。

一方、そのストーリーは、まさに、大学・組織・「コミュニティ」の「文化」としても、蓄

積まれていくことになる。そのストーリーには、「活動主体」がもつミッションが明らかにされているわけで、例えば、それが、大学の学風を作り、「コミュニティ」の特徴を浮かび上げらせ、独自の教育研究の風土を形成していくことになる。そして、そういう大学や組織の文化が、次なる「評価」の枠組を構成していくことにもなる。

また、そうした循環が定着していくことによって、そこに、一つの新しい「評価文化」も形成されていくことになる。すなわち、「順位や点数で表すことが評価」という通念から、「改善のためのストーリーを作り上げる作業そのものが評価」といった見方が自然にできるようになるということである。そのような評価を通じて、「コミュニティ」がより確かなものに成長し、また、その中で、個々人が自らの役割を確かにしていくことができる。そこに、「コミュニティ」形成の原動力があるということである。明確な「コミュニティ」に我々が含まれれば、徒弟制の弟子たちが切磋琢磨してお互いに成長していくように、「コミュニティ」内に競争的な風土があったとしても、健全に「コミュニティ」自体は機能していくことにもなる。要は、それができる文化的背景をまずは共有していくことができるかどうかというところに、重要なキーが隠されているのである。

「大学評価」は始まったばかりであり、それが真に大学自身の活動の改善に結び付けられるかどうかは、ここで概観してみた TQM 的なアプローチによる積み重ねによって、新しい「大学文化」が形成され、引いては新しい「評価文化」が定着していくかどうかという点にかかっている。本拙論が、その一助につながるようであれば、望外の喜びとするところである。

【文献】

- 上里一郎（監修） 1993 心理アセスメントハンドブック 第2版 西村書店
- Flick, U. 1995 *Qualitative Forschung*. Rowohlt Taschenbuch Verlag GmbH, Reinbek bei Hamburg. 小田博志・山本則子・春日常・宮地尚子（訳）2002 質的研究入門——〈人間科学〉のための方法論 春秋社
- Gipps, C. V. 1994 *Beyond Testing: Towards a theory of educational assessment*. The Falmer Press Ltd. ギップス, C. V. 鈴木秀幸（訳）2001『新しい評価を求めて——テスト教育の終焉』 論創社
- Glaser, B. G. & Strauss, A. L. 1967 *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Co. 後藤隆・大出春江・水野節夫（訳）1996 データ対話型理論の発見——調査からいかに理論をうみだすか 新曜社
- Harnisch, D. L. & L. Mabry 1991 大塚雄作（訳）新しい評価法の開発とその評価——アメリカにおける最近の教育評価の動向から 学習評価研究, 7, 54-75.
- 橋本重治・肥田野直（監修）1977 教育評価の考え方（最新教育評価法全書1巻）図書文化
- 早田幸政 2001 評価の客観性と「主観評価」に関わる考察 大学評価研究, 1, 33-41. 財団法人大学基準協会
- Hodson, D. 1998 *Teaching and Learning Science*. Open University Press. 小川正賢（監訳）

- 2000新しい理科教授学習論——子ども一人ひとりの見方・考え方を損なわずに科学を学ばせるには 東洋館出版社
- Holloway, I. & Wheeler, S. 1996 *Qualitative Research for Nurses*. Malden: Blackwell Science Ltd. 野口美和子(監訳) 2000 ナースのための質的研究入門——研究方法から論文作成まで 医学書院
- 堀 洋道(監修) 2001 心理測定尺度集Ⅰ～Ⅲ サイエンス社
- 井上正明 2001 「教育評価学」連続講義(Ⅱ) —《エバリユエーション》から《アセスメント》へ— 福岡教育大学紀要, 50, 4, 103-129.
- 伊藤秀子・大塚雄作(編) 1999 ガイドブック・大学授業の改善 有斐閣
- 金子郁容 2002 新版・コミュニティ・ソリューション——ボランティアな問題解決に向けて 岩波書店
- 京都大学高等教育教授システム開発センター 1997 開かれた大学授業をめざして——京都大学公開実験授業の一年間 玉川大学出版部
- 京都大学高等教育教授システム開発センター 2000 生成的組織評価を目指して——自己点検・自己評価, 外部評価報告書—— 京都大学高等教育叢書9
- Lave, J. & Wenger, E. 1991 *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. レイヴ, J.・ウエンガー, E. 佐伯 胖(訳)・福島真人(解説) 1993状況に埋め込まれた学習——正統的周辺参加 産業図書
- Linn, R. L. (ed.) 1989 *Educational Measurement 3rd. Ed.* 池田央・藤田恵璽・柳井晴夫・繁榊算男(編) 1992 教育測定学・原著第3版 上巻 学習評価研究所
- 松原達也(編著) 1995 最新・心理テスト法入門 日本文化科学社
- McGregor, D. 1960 *The Human Side of Enterprise*. New York: McGraw-Hill. 高橋達男(訳) 1970 新版 企業の人間的側面 産業能率短期大学出版部
- 森敏昭・秋田喜代美(編) 2000 教育評価——重要用語300の基礎知識 明治図書
- 大野木裕明 1994 テストの心理学 ナカニシヤ出版
- 大沢武志・芝祐順・二村英幸(編著) 2000 人事アセスメントハンドブック 金子書房
- 大塚雄作 1991 放送大学に対する学生のニーズと評価——ある評価活動モデルからみた大学評価の一事例 大塚雄作(研究代表者) 大学個性化のための多元的・多面的大学評価法の研究, 25-87. 平成2年度文部省科学研究費補助金(総合研究A) 研究成果報告書 放送教育開発センター
- 大塚雄作 1994 理論と実践を結ぶ測定・評価に向けて 教育心理学年報, 33, 102-111.
- 大塚雄作 1996 電子メディア時代の理解と表現を探る 若き認知心理学者の会(編) 認知心理学者教育評価を語る, 180-191. 北大路出版
- 大塚雄作 2001 FDと学生による授業評価 図書館情報大学(編) 2001 知の銀河系七 FD(ファカルティー・ディベロプメント) 講演会の記録 図書館情報大学
- 大塚雄作 2002 生涯学習における学習継続の促進に向けて 大学教育学会誌, 24, 1, 93-99.

- Patton, M. Q. 1997 *Utilization-focused Evaluation*. Sage Publications, Inc. パットン, M. 大森彌 (監修) 長尾真文・山本泰 (編集) 2001 実用重視の事業評価入門 清水弘文堂書店
- 龍 慶昭・佐々木 亮 2000 「政策評価」の理論と技法 多賀出版
- Seymour, D. T. 1993 *On Q: Causing Quality in Higher Education*. American Council on Education and The Oryx Press. 館 昭・森利枝 (訳) 2000 大学個性化の戦略——高等教育の TQM 玉川大学出版部
- Stiggins, R. J. 1992 評価法のリテラシー 学習評価研究, 11, 12-29.
- 続 有恒 1980 教育心理学の諸課題 金子書房
- Worthen, B. R. & Sanders, J. R. 1987 *Educational Evaluation*. White Plains: Longman.

[ABSTRACT]

Some Elements of Evaluation and their Functions
in Higher Education :

—Aiming at Formation of the Improvement-oriented Evaluation Culture—

OTSUKA Yusaku*

In the 21st century, various evaluations are expected to be politically imposed on universities in Japan. In order for universities to change in a better directions, it is necessary to introduce evaluation for improvement. Evaluation has some elements as listed below. (i) framework and the purpose of evaluation; (ii) human factors in the process of evaluation; (iii) contents and methods of evaluation; (iv) evaluation information and its expression; (v) meta-evaluation; (vi) impact of evaluation. In this paper, in order to utilize evaluation pertinently, it was surveyed how those elements would function in the process of evaluation.

Consequently, the following suggestions have been obtained. (i) Very big diversity is included in evaluation based on the purpose of evaluation, and suitable evaluation needs to be performed. (ii) Although we tend to think evaluation should be its result in quantitative manner, in the evaluation for an improvement, it is common for qualitative information to be more useful. (iii) In order to use evaluation information for an improvement of activity, it is effective for the evaluation to be clarified in a certain “community”. (iv) As for evaluation, it is effective to be positioned as a tool for forming a “community”. (v) It is desired to form the evaluation culture from which the evaluation used for an improvement of activity will become natural. (vi) The “TQM (Total Quality Management)”-approach is effective also in practical use of evaluation.

*Professor, Faculty of University Evaluation and Research, National Institution for Academic Degrees