

大学評価 第2号 平成14年12月 (研究ノート・資料)

[大学評価・学位授与機構 研究紀要]

欧州における研究評価の教育コースの現状

A Report on Educational Courses on Research Evaluation in Europe

林 隆之

HAYASHI Takayuki

Research in University Evaluation, No. 2 (December, 2002) [the essay/material]

The Journal of University Evaluation of National Institution for Academic Degrees

1. はじめに	37
2. 研究評価の教育コース設立の背景	40
3. 教育コースの概要	41
3.1 教育コースの形式的特徴	41
3.2 教育コースの内容	41
4. おわりに ～ 大学の研究評価を支える人材に必要な知識・スキルとは	48
4.1 教育コースから見る研究評価の最近の傾向	48
4.2 大学評価の特殊性	49
5. 参考文献	51
[ABSTRACT]	53

欧州における研究評価の教育コースの現状

林 隆之¹

1. はじめに²

日本では大学評価に限らず研究評価や政策評価といった様々な種類の評価が1990年代末から急速に制度化されてきた。すなわち、大学評価については2000年から大学評価・学位授与機構による第三者評価が行われており、大学以外のセクターも含めた公的資金による研究開発の評価については1997年に「国の研究開発全般に共通する評価の実施方法の在り方についての大綱的指針」が決定されて本格的に行われるようになった。また一般の政策評価についても2002年4月には「行政機関が行う政策の評価に関する法律」が施行されている。もちろんこれ以前にも多種の評価が部分的、個別的、断続的には行われてきた。しかしながら国レベルでの共通のガイドラインのもとに評価が体系的に行われるようになったのは近年のことと言える。このように評価が制度化されて実際に行われ始めている中で、明らかになりつつある問題の一つが、評価に関わる人材の不足である。これは一つには十分な評価経験を有した「評価者」の不足が挙げられる。だがそれだけでなく、評価に必要なスキルを有して評価業務を運営・推進できる「評価運営者」の不足、さらには評価についての専門的知識を有してそれを支援する研究者やコンサルタントといった「評価支援者」の不足も指摘されている³。

このような評価人材の不足の問題は、評価対象が大学やその他セクターが行う研究活動（あるいは研究開発活動）である場合に一層顕著となる。なぜならば研究評価では上述の評価人材に対して、さらにその内部で多種の人材が求められるためである。まず評価者については、研究という活動はそれ自体が極めて高度な専門的知識を必要とするものであるために、それら専門的内容を理解しうる人間が評価者に入ることが不可欠である。これは同一学問分野の研究者による評価である「ピアレビュー」方式として古くから制度化されてきた。しかしながら学術雑誌の論文掲載の採否を決める場合のピアレビューとは異なり、研究機関や研究プロジェクト・プログラムの評価においては、研究内容自体を理解する能力だけでなく、研究を実施するマネジメント方法（組織構成や実施方策など）についての理解や、将来に発展する研究を見定める能力も評価者には同時に求められる。また研究活動はそれ自体で完結するものだけでなく、様々な効果を生みうるものである。科学技術的効果についても応用分野といった当該学問分野以外への波及効果が存在することは多い。そのため評価者は他分野に対していかなる効果が生じた(生じるか)かを判断できることも求められる。さらに近年は研究により生じる社会・

¹ 大学評価・学位授与機構 評価研究部 助手

² この調査は科学研究費補助金「国際的通用力を持つ大学評価システムの形成に関する日、欧、米の国際共同研究」および海外研究開発動向調査「新産業創出に資する研究評価手法の開発」により行った。

³ 2001年に改訂された研究評価の大綱的指針においても「研究経験のある人材の確保と研修等を通じた評価人材の養成等」の必要性が指摘されている。

経済的効果についても一層の期待が寄せられるようになってきている。そのため、評価者は社会・経済的ニーズについても知識を持ち、当該研究活動からの効果を判断できることも必要である。実際にはこのような多様な能力を小数の当該分野の専門家（ピア）に期待するのは困難であり、評価の対象にあわせて各種の知識を有する人材により評価者のパネルを構成しなければならない。

一方で、このような評価者が研究実施者側からの報告書や自己評価書を見るだけで質の高い評価をできると安易に期待することはできない。評価者が質の高い評価をするためには、実際にどのような効果が生じており、それが他の要素との関係や他の評価対象との比較の中でいかに位置づけられるかといった参照情報が不可欠である。そのような情報を作成する役割を担うのは評価運営者および評価支援者である。特に評価支援者は、時には評価パネルにも参加しながら、評価に関する専門知識を基に評価対象を実際に調査・分析して参照情報を作成することが求められる。しかし、たとえ研究評価に限っても対象の種類によって調査の手法は書誌計量的手法、関係者へのサーベイ調査、コスト・ベネフィット分析、歴史的な跡付けなど様々なものがあり、実際には評価支援者も複数の人間によるチームとして関与せざるを得ない。

さらには、このように想定される多種類の評価者や評価支援者について、実際の評価対象に即してどのような人材が本当に必要かを判断して選択し、具体的な実務作業のフレームワークを形成し調整することができる評価運営者が求められる。これらの評価人材の関係を図示すると図1のようになる。この図からも明らかなように、研究評価とは種類の異なる人間による極めて集団的な営みであると言える。

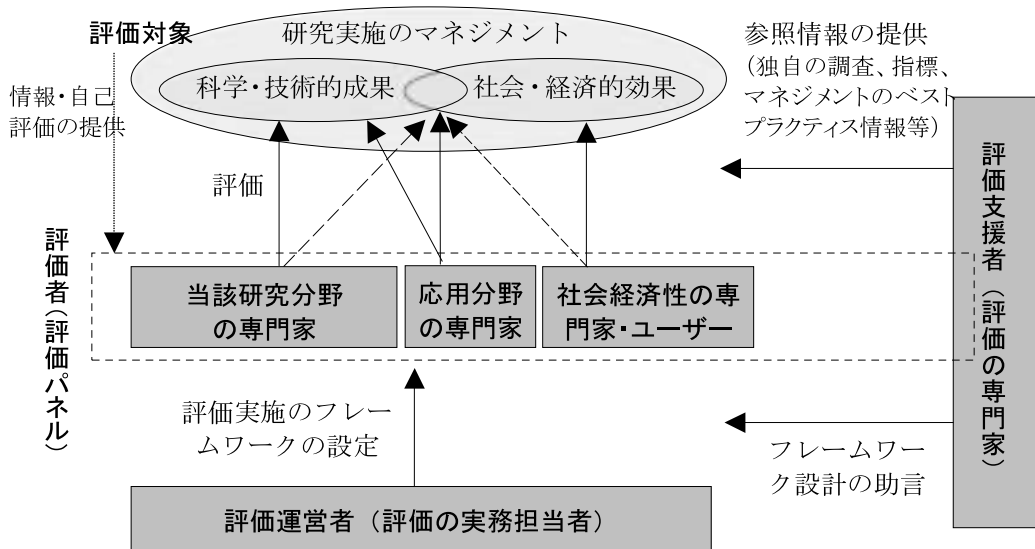


図1 一般的な研究評価システムを構成する人材

では、評価の質を高めるためには、このような各種の評価人材はどのような知識やスキルを有していることが求められるのであろうか。三種類の評価人材のうち評価者については、各自が有する専門知識をもとに研究の水準や発展性を判定することのみにその役割が限定される限りにおいては、評価という活動自体の深い見識を要求する必要性は強くない。評価者に求められることは当該評価のフレームワーク（評価の目的や方法）を十分理解しそれに沿って評価を行い得ること、ならびに、透明性や公平性のような評価の一般的要求事項を踏まえることである。ただし、評価運営者や評価支援者が提出する資料を理解する能力は求められる。

評価者がこのように当該評価のフレームワークに即することが第一に求められるとするのであれば、より重要となるのは、そのようなフレームワークを設計する評価運営者、および、それを支援する評価支援者の能力である。しかしながら、この二者について日本の現状を鑑みると、評価運営者に関しては、研究評価を専門に担当する部署や機関ができたのは最近であり、特に行政機関では数年で担当者が異動するために専門知識を有する人間が集積しないという弊害がある。同時に評価支援者についても、これまで評価を行うニーズが強くなかったために、大学や民間の調査機関においても研究評価の専門家と呼ばれるべき人材は少数しか存在しない。

このような日本の状況に対して、欧米諸国に目を向けると、研究評価の専門家と呼ばれる人材が大学や民間企業（コンサルタント）に存在しており、省庁や資金配分機関の内部にも評価実務を専門的に行う人材が存在している。欧米では科学技術政策や研究開発マネジメントの専門研究機関が存在しており、そこに所属する研究者やそこからの卒業生、あるいは大学を離れた教員が設立したコンサルタント企業などが1980年代以降に評価に関わる業務や調査研究を展開してきた。特に英国ではこの傾向が強く、専門家を中心とする評価システムは「英国モデル」とも称される（Barker 1997）。このような評価人材の層の厚さは、さらに、評価人材を再生産するプロセスである研究評価の教育コースが開講されていることにより維持されており、評価の知識の教授・流布を可能としているのである。

本稿では、そのような研究評価の教育コースを開講している英国マンチェスター大学のPREST（工学科学技術政策研究所：Policy Research in Engineering, Science and Technology）およびオランダのトゥエンテ大学の教育内容の調査を行い⁴、研究評価システムを構成する人材に求められる知識やスキルとはどのようなものであるかを検討する。ただし、これら教育コースは研究評価全般を扱うものであるため、大学の研究評価に限ったものではなく、公的研究開発プロジェクトやプログラムを評価対象とする内容も多く含まれている。そのため、本稿ではそれらを含めて教育コースの全体構成を示し、最後に、特に大学という機関の研究評価において必要な知識やスキルを議論する。

⁴ マンチェスター大学 PREST の教育内容については、2001年11月に Luke Georghiou 教授および Dr. Mark Boden にヒアリング調査を行うとともに、2002年1月に開講された教育コースに実際に参加して調査を行った。トゥエンテ大学については2001年11月に Dr. Barend van der Meulen にヒアリング調査を行った。

2. 研究評価の教育コース設立の背景

欧米では1980年代半ばから研究評価の制度化に焦点が置かれるようになった。それ以前にも学術雑誌の論文採択や学術賞の授与、教員の昇進などの形で研究者共同体内部での研究評価は行われていた。また各省庁からの研究プロジェクトへの資金配分においても事前評価は行われていた。しかしながら、機関レベルの評価や政策・プログラムレベルの評価、およびプロジェクトレベルの事後・追跡評価は制度的には行われている状況とは言えなかった。このような状況から研究評価が制度化されていく背景には、経済不況の中で研究活動に対する公的資金の増加が抑制されたために、資金提供する研究（あるいは研究実施者）を選択する必要性が増したことが一要因として挙げられる。特に英国では「Value for money（資金に見合う価値）」と言われるように、資金提供により得られる価値や効果を明らかにすることが求められるようになった。しかしながら、より本質的には、そもそも研究活動の総体としての規模自体が拡大してきたために、総花的に多数の機関やプロジェクトに資金配分を行うことが難しくなったことが挙げられる（Gibbons and Georghiou 1987, Evered and Harnett 1989）。この顕著な例が英国の Research Assessment Exercise による資金の傾斜配分であると言える。

このように評価が求められる一方で、その中で用いられる評価方法自体も複雑なものが要求されるようになってきた。一つには、厳しい資金制約のもとでは一分野内部だけでなく分野を越えた評価も必要とされる場合が多く、そこでは伝統的な分野ごとのピアレビュー方式を採用するのは困難となる。さらに、新たに「戦略研究」概念の導入や、産学連携の推進、社会経済ニーズへの貢献といった研究活動の変化を背景として、これまでのように基礎研究をピアレビューで評価し、開発研究を計画目標の達成で評価するという安易な二分法は適応し難くなったのである。

このような中で、研究評価の対象である各種の研究開発活動や研究政策を研究している科学技術政策研究者や研究開発マネジメントの研究者が評価へと関与していく。英国では Alvey プログラムのモニタリング評価がマンチェスター大学 PREST やサセックス大学 SPRU によって行われ（Georghiou 1987, Cameron and Georghiou 1997）、同時期には OECD や米国で研究評価手法の調査も行われている（e.g. Gibbons and Georghiou 1987, Office of Technology Assessment 1986）。さらに欧州では EU の第二次フレームワークプログラム（1989～1993年）以降、研究評価を研究するプロジェクトへの資金提供がなされ、その中で共同研究や国際会議を繰り返し評価研究者の間のネットワークが形成されている。このように実際の研究評価への関与と研究評価研究の推進により、研究評価の知識集積がなされ、研究評価を専門的研究テーマとする研究者が増加してきた。このような背景のもとに、英国マンチェスター大学 PREST は1980年代半ばから研究評価に関する教育をカリキュラムに組み込み、オランダのトゥエンテ大学では1989年から学外者向けの教育コースを開始した。この他にも米国 AAAS や各国の省庁内部において職員が評価部署へと異動になる際の転換教育が行われている場合はあるが、外部にも開かれた形で教育コースを開講しているのは現在（2002年1月）のところこの2校である。

3. 教育コースの概要

3.1 教育コースの形式的特徴

マンチェスター大学 PREST で行われている教育コースは5日間の集中講義の形式となっている。この教育コースはそもそもは PREST の1年間の修士課程(MSc)の中の授業の一つとして位置づけられており、それを外部にも公開したものである。そのため大半の受講者は PREST の修士課程の学生である。しかしそれら学生には各国の科学技術行政の省庁からの留学生が多く、実務に実務に適用しうる内容が教えられている⁵。一方、オランダのトゥエンテ大学では3日半という短い日数で教育コースが開かれているが、各日は朝9時から夜9時までという時間的に密なスケジュールになっている。このコースの受講者は研究評価の専門知識を身につけたい社会科学研究者、政府組織等で評価を担当している実務者、評価パネルに参加する科学者や技術者とされている。講師は、PREST の場合には主に PREST に在籍する教員により授業がなされ、外部の行政機関(英国 DTI や EU) やコンサルタント会社からのゲスト講師が事例を紹介する。トゥエンテ大学の場合は同大学の2人の教員の他に、英国(PREST)、ドイツ、米国(2002年の場合)から講師を迎えており、その他にオランダ VSNU(大学協会)や EU からのゲスト講師が事例を紹介する(2001年の場合)。

3.2 教育コースの内容

両機関で行われている教育コースの内容には若干の違いはあるものの、大筋では類似したものになっている。これは前述のように欧州では1980年代から情報交流がさかんに行われてきたことによると考えられる。両方で教えられる内容は、それぞれ表1および表2となっている。その構成を概括すれば以下のようにまとめられる。

- ・研究評価が必要となった社会的背景
- ・評価システムのデザイン
- ・評価に用いる手法
- ・評価書の作成と意思決定への利用

両コースではこれら内容を教えることで、各受講者が自分の関与する評価をテラーメイドに形成・実施する能力を身につけることが目的とされる。そのため、これらコースでは共に多くの実習を含んで教えられることに特徴がある。

以下では、主に PREST の内容を中心に紹介しながらも、トゥエンテ大学において異なっている点について補足して説明する。

⁵ PREST ではこの教育コースとは別に、英国リサーチカウンシルや DTI(通商産業省)に出向いて評価の教育・研修も時折行っている。

表1 マンチェスター大学PREST の教育コースの内容 (2002年1月)

1日目	イントロダクション (研究評価の概要) 評価の理論的基礎 ロジックチャート (講義) ロジックチャート (演習)
2日目	ロジックチャート (演習) つづき ロジックチャート (講評) 経済的効果の評価
3日目	ピアレビュー ビブリオメトリクス (講義) ビブリオメトリクス (演習, 講評)
4日目	ケーススタディ: Research Assessment Exercise 質問紙調査 (講義, 演習) ゲスト講師: 多レベルの評価 (コンサルタント) 質問紙調査 (講評) 社会的効果の評価
5日目	評価のインパクトと利用 ゲスト講師: EUにおける研究開発プログラム評価 ゲスト講師: 英国DTIにおける評価と政策形成

(1) 研究評価が必要となった社会的背景

両コースでは、まず初日にイントロダクションの意味を込めて、研究評価が必要となってきた社会的背景についての概要説明が行われる。すなわち、上述のように公的研究投資の抑制により研究の質、価値、効果に焦点が置かれるようになり、資金配分のための競争が強くなった。このような流れは1990年代のニュー・パブリック・マネジメントの導入により一段と強化されるとともに、アカウンタビリティの必要性が指摘されるようになった (Georghiou and Roesner 2000)。このように研究評価が制度化されるようになった背景をはじめに説明することで、「何のための研究評価であるのか (評価のための評価ではない)」といったコースを学ぶ上での動機付けを受講者に行う。

(2) 評価システムのデザイン

次に「評価システムのデザイン」が説明される。この内容は大きく二つに分けることができる。一つは評価を実施・運営するシステムの構成、もう一つは、具体的に評価作業を展開するための評価対象の構造化の方法である。

前者については、評価システムを構成する一般的要素の概略が説明される。これは、評価運営者が具体的に評価のシステムを形成し実施していく上で、どのような要素について考慮する必要があるかを考えるための基礎となる概念化であると言える。

PREST の教育コースにおいては、Gibbons and Georghiou (1987)などを参照しながら、評価の目的、評価のスコープ (評価対象の研究の種類、評価対象のレベル、評価の時期)、評価

表2 トウエンテ大学の教育コースの内容（2000年10月）

1日目	評価の社会的文脈	イントロダクション 評価の社会的文脈 コース概要
	評価のデザイン	評価のデザイン 演習1：目的からデザインへ（ロジックチャートの演習と講評）
	方法のカタログ	事例：オランダの大学評価（VSNU）におけるピアレビュー
2日目	方法のカタログ （つづき）	方法のカタログ（概要） 中間的インパクトとしてのネットワーク 「社会的質」の評価 経済的インパクトの評価事例：EUの研究評価 演習2：評価の方法の総合的利用（演習と講評）
3日目	政策のための評価	評価データを用いた研究政策の改善 事例：政策的介入の方法としての評価 因果関係の理解 演習3：評価レポートの作成
4日目	政策のための評価 （つづき）	演習3（講評） コースのレビュー 将来の研究評価

基準、評価に関わる人材といった4つの要素により評価システムが構成されることが説明される。特に、評価の目的は被評価者・評価の顧客・評価実施者の三者の関係を基に決まるものであり、評価を行うにはその目的を明確にする必要があることが強調される。次に、研究評価の主要な焦点である研究成果（結果や効果）について、そのカテゴリーには、暗黙的なものと形式的なもの、アウトプットとインパクト⁶、短期的成果と長期的成果、予想された成果と予想外の成果、経済的成果と社会的成果、経済的成果と構造的変化などがあり（Airaghi et al. 1999）、それら区分を意識して評価を行わなければ特定の種類の成果のみに焦点が置かれて評価が局限されてしまうことが説明される。

続いて、評価の一般的方法が説明される。これは、科学・技術的效果や経済・社会的効果といった各種の成果の測定方法を教える以前に必要な、より基盤的な方法である。評価では対象を何らかの基準と照らして価値判断することが求められるが、その方法もロジカルフレームワーク（目的・目標から期待される内容と実際に起きたことの比較）、時間的に前後の比較、コントロール・グループとの比較、反事実仮説との比較などがあり、評価目的に即してその方法を選択しなければいけない。さらに、データ収集の一般的方法（インタビュー、サーベイ、

⁶ 「アウトプット」、「アウトカム」、「インパクト」といった語が示す内容は文脈により異なるが、研究評価においては一般的に、「アウトプット」とは研究活動により生じた一次的な結果それ自体（論文、特許、ノウハウなど）を示し、「アウトカム」とはその結果により資金提供者や共同研究相手企業などの関与者へもたらされた効果を示し、「インパクト」とは研究活動には直接関与していない研究者共同体や産業界など一般への影響を示す。

人類学的観察、統計、文書)やデータ分析の一般的方法(ケーススタディ、データ分析、指標、経済学的モデリング、コスト・ベネフィット分析)、パフォーマンス指標の利用の長所と短所が説明される。

また、PRESTの講義では、一般の政策評価やプログラム評価における「評価理論」の先行研究の紹介もされる。そもそも「評価」という営み自体をいかに定義するかによって、評価が基礎を置く存在論的、認識論的、方法論的パラダイムも異なってくる。また、実施の成果に焦点を置くアウトカム・アプローチと、より実施プロセスへの焦点を強く持つプロセス・アプローチの違いも生じる。これらの講義では特定のパラダイムを推奨するものではなく、評価という営み自体が、「評価対象についての客観的な情報を得て、その成果の価値を判断する」といった単純な行為には縮減できないことを認識すべきであることを説明するものであると考えられる。なお、このような基礎概念の教授においても、資料を読んでまとめるなどの実習形式をとった双方向的な授業が行われた。

次に、「評価システムのデザイン」のもう一つの要素である、評価対象の構造化の方法が説明される。両方の教育コースでは「ロジックチャート」(あるいは「ロジックダイアグラム」と呼ばれるPRESTが用いる方法が説明される。ロジックチャートは、特に研究プロジェクトやプログラムについて、階層構造上に上位レベルの目的、下位レベルの目的(目標)、活動、期待されるアウトプット、アウトカム、インパクトの順にそれぞれの要素を並べて、それらの間の論理的関連性をリンクとして示す方法である。このようにプログラムの論理構造を目的から階層的に図示する方法は、研究評価に限らずにこれまでも提案されていることである。しかし、研究評価におけるこのロジックチャートでは目的と期待されるアウトカムを別の階層とすることにより、目的には明示的には示されにくい効果(例えば研究者の能力の向上や大学院生の育成効果)や負の効果(環境や社会・経済への悪影響)、スピルオーバー的な効果(社会経済的効果や他の学問分野への波及効果)についても示しうるようになっている。このような方法は、プロジェクトやプログラムだけでなく、機関評価における目的・目標に基づく評価にも用いることが可能なものと言える。このように論理構造を「ロジックチャート」として図示することの利点は、次の二点から説明される。一つには具体的な実施行為やそれにより期待される結果・効果を目的・目標との論理的整合性から評価することが可能となることである。また一つには実際に行為が行われた後の評価で注目すべきアウトプットやアウトカムの種類を同定することが可能となることである。

ロジックチャートの概略の説明のあとには、両大学のコースともに、EUの国際共同研究プログラムであるEUREKAを事例として、ロジックチャート作成の実習がグループごとに行われる。これはプログラムの背景や目的などが書かれた政策文書から、各階層に入れるべき要素を抽出し、どのような論理関係が示せるかを検討して、チャートを作成するものである。たとえば同一の素材をもとにしても、各グループでは異なったチャートが形成され、講師がそれらについて講評を加えていく。

(3) 評価手法

次に評価に用いる方法論が説明される。評価で用いる方法論については、対象に関する情報や関係者の主観的判断を収集する基礎的な方法と、科学技術的效果や経済・社会効果といった特定種類の成果・効果の測定を行う方法とに分けることができる。教育コースではこれらの双方が教えられる。前者については PREST ではピアレビューおよび質問調査の手法が時間枠をとって教えられる。トゥエンテ大学ではオランダ VSNU (大学協会) の大学評価を事例としてピアレビューの実際が説明される。

ピアレビューに関しては、その種類として「直接ピアレビュー」や「改善されたピアレビュー (メリットレビュー)⁷」などが説明されるとともに、どのような場合にピアレビューが有効であるか、および、一般的問題点が説明される。すなわち、ピアレビューが有効となる場合は、評価により配分される資源の制約が少ない場合や、ピアの総数が多く利害関係のないピアを選べる場合、分野の定義が明確な場合、科学的な質のみが評価基準である場合であり、これら条件を満たさない場合には信頼性あるピアレビューを行うことは困難になる。また、ピアレビューの問題点としては保守性、学際分野へのバイアス、評点のインフレーション現象、分野間の比較の困難性、ピアは部分的・不完全な知識のもとで評価をすること、社会・経済効果をピアが評価するのは困難なこと、膨大なコストがかかることなどが指摘される。この他に、ピアレビューアーの選択方法の種類が説明される。

また、サーベイについては社会調査法の基礎として教えられるような、対象者の選択、調査方法、選択肢の作り方が評価の枠組みの中で教えられる。特に評価においては、人は「自分が評価される対象である」と考える場合と、「機関や研究開発プログラムが評価対象であって自分はその対象の構成者や関与者として情報提供を求められている」と考える場合とでは、対応の仕方が異なることを意識して調査を行う必要が指摘される。この授業においても、いくつかの評価トピックスを提示して、誰にどのような項目の調査を行うかについて実習を行う。

次に各種の研究の成果を分析する方法については、科学・技術的效果、経済的效果、社会的効果とに分けて教えられる。

科学技術的效果の評価では、上記のピアレビューの他にビブリオメトリクス (書誌計量学的手法) が説明される。ビブリオメトリクスは論文や特許について、その産出数や引用数および共引用関係や共著関係などを、国別や組織別、あるいは、特定のプロジェクトやプログラムに関わるものを抽出して分析を行う方法である。実際に大学の評価でもオランダ VSNU (大学協会) では、分野によってはピアレビューへの参照情報の一つとしてライデン大学にビブリオメトリクス分析を委託している。講義では、ビブリオメトリクスはピアレビューの欠点をサポートするものであり、利点として、当該分野の専門家でなくとも測定が可能なこと、比較が可能

⁷ 改善されたピアレビュー (modified peer review) とは、通常の狭義のピアレビューでは当該分野における学問的価値 (サイエンティフィック・メリット) が主に評価されていたのに対し、評価項目に社会的・経済的效果やその可能性をも入れて評価を行うことを指す。ただし、レビューアーの中に当該学問分野以外の人間をも入れるか否かは事例により異なる。

なことなどが説明される。一方で分析に頻繁に用いられる論文データベースである *Science Citation Index* 自体に、英語論文や米国雑誌の偏重や、レフェリーのない雑誌が収録されないなどの問題があることが説明される。PREST では、講義の後にコンピュータールームにおいて実際にデータベースを用いて引用数を検索したり、データベースの問題点を確認するなどの簡単な実習が行われた。一方、トゥエンテ大学では数年前まではビブリオメトリクスの独立した時間枠をとっていたが、多くの受講者がビブリオメトリクスについては既に知っている場合が多いため、現在は VSNU の大学評価の事例紹介などの他の項目の中を含めている。

次に、研究の経済的効果および社会的効果について説明がされる。これらは、主には産業向けの研究開発プロジェクトやプログラムの評価で中心となるものであるが、大学についても工学の研究評価や産学連携活動の評価では重要な項目となりうる。講義では、まず経済効果のアウトプットおよびアウトカム・インパクトの種類として、前者には製品、製造プロセスの改善・確立、サービス、技術標準の確立、知識、スキルなどがあり、後者には競争力、雇用、組織の改変、生活の質 (Quality of Life)、地域発展、インフラ開発、エネルギーの製造・合理的利用、産業発展、規制変化などがあることが説明される。また、大学などの基礎研究による効果では、1) 価値ある知識のストックの増加、2) 能力ある卒業生、3) 科学的装置・方法の開発、4) ネットワークの形成と社会的インタラクションの誘因、5) 科学的・技術的問題の解決能力の向上、6) 新たな企業の創造などが挙げられる (Salter and Martin 2001)。

これらを踏まえた後で分析方法が説明される。公的研究開発プログラムの評価については投資効果の金銭的評価の方法 (BETA 法 (Bach et al. 1997)) なども説明されるが、それよりも経済効果の評価の背景にあるモデルを教える事に時間が割かれる。すなわち、測定された特定の経済効果について、その要因として挙げられるものは評価対象の研究活動以外にも多数の要素が絡み合っていることが通常であるし、逆に評価対象の研究活動は長年のうちに多様な種類の効果へと連鎖していくものである。そのため、インプットとアウトプットを研究活動が関与した範囲で過不足なく特定することは困難なことであり、その困難性を無視して計測を行っても妥当な結果が得られるとは限らない。それよりも各事例を詳細に跡づけることによって評価を行っていく必要性が指摘される。その上で、公的資金を民間研究開発に用いる際の正当性を分析するための、スピルオーバー概念やアディショナリティ概念 (Lukkonen 1998, 2000) なども説明される。また、経済的アウトカムが顕在化するには長期間かかり、それを数年のタイムスパンで評価することは極めて困難であるために、近年ではより研究実施プロセスへと視点を移して、その中での連携 (共同研究やユーザーとの交流による知識の移転と利用・集積) の評価へと焦点がおかれる傾向がある (Callon et al. 1997, Bozeman 1999)。トゥエンテ大学ではこの内容を「中間物としてのネットワーク」という時間枠をとって別途強調して教えている。評価において研究活動の顧客が誰であり、その顧客とのいかなる相互関係を有しているかを把握するためにもネットワーク概念が必要となり、そのようなプロセスへの視点をおくことによってこそ今後のマネジメントがどうあるべきかの示唆を得ることが可能となる。

一方で、経済的効果からは抜け落ちるような社会的効果の分析方法も説明される。この対象

には環境、健康、安全、社会参加、犯罪抑制、文化継承、科学技術の公衆理解、防衛、雇用（プラス、マイナス両面）などが入る。この内容では特に、社会的効果を受益する人間を、評価における新しいアクターとして組み込む必要が指摘される。

トゥエンテ大学ではこれら方法の講義を全て行った後に、公的研究センターの評価などを事例として実習を行う。実習では、評価目的の設定と、それに合わせた分析方法の選択が学習される。

（４）評価レポートの作成と評価結果の意思決定への利用

次に、被評価者やその監督者の意思決定へと連結しやすい評価レポートの作成の仕方について説明される。評価が受け入れられて意思決定へとつながるためには、まずは誰が評価のオーディエンスであるかを考えた上で、レポートの様式、提言の様式、評価のタイミング、評価結果の公表・普及の仕方を定めて受容性（absorbability）を高めるべきでことが強調される。例えば、提言のレベルは細か過ぎるものでも一般的すぎるものでもなく、適切なレベルを選ぶことが必要となる。さらに、評価結果が受け入れられるためには、その評価の信頼性と質が重要であることも指摘される。また、評価は意思決定を行う材料の中の一つでしかないため、それら他の要素との関係における操縦性（steerability）が求められる。トゥエンテ大学では先のEUREKAを事例として評価レポート作成の実習が行われる。

（５）事例

これら講義の間に、いくつかの評価事例が外部講師を含めて紹介される。両コースそれぞれで、大学の研究評価は事例の一つとして説明される。PRESTの場合にはHEFCE等が行うResearch Assessment Exercise(RAE)についてであり、トゥエンテ大学の場合はオランダVSNUによる大学の研究評価である。RAEについては、PRESTがHEFCEからの委託で行ったRAEによる大学へのインパクト調査（PREST 2000）の結果が紹介され、評価を行うことによってどのような効果と問題が実際に生じたかの説明がなされる。

これ以外にもEUのフレームワークプログラムの評価、英国DTI（通商産業省）における評価、研究評価を専門とするコンサルタント会社のTechnopolis社による評価、ドイツISIによる研究センター評価などが外部からの講師により紹介される。

4. おわりに ～ 大学の研究評価を支える人材に必要な知識・スキルとは

既に述べたように、この二つの教育コースは大学の研究活動の評価のみを対象としたものではなく、公的研究開発プロジェクトやプログラムも含んだ研究評価一般を扱うものである。そのため、評価の一般理論などの共通的内容もある一方で、経済・社会的効果の評価方法ではプロジェクト評価のみに適した内容も多く含まれている。逆に大学のような機関の評価については焦点が置かれずに欠けている内容もある。そのため、以下では2つの教育コースから大学評価にも共通して指摘できる最近の研究評価の傾向をまとめ、最後にこのコースでは欠けていると考えられる大学評価の特殊性について議論する。

4.1 教育コースから見る研究評価の最近の傾向

教育コースの内容からは研究評価の最近の二つの傾向を見て取れる。一つは、研究活動のアウトプットやアウトカムについて多種類のものを評価しようとする方向である。もう一つは、アウトカムだけでなく研究を実施するプロセスへも焦点を当てる方向である。

前者については、そもそも研究評価が制度化された背景自体が「value for money」に代表されるような公的資金投入による効果と納税者へのアカウントビリティが求められたことに起因する。また、研究活動自体も「戦略研究」に代表されるような社会経済的ニーズを念頭においた研究が様々な資金源のもとで行われるようになった。このような中では、学問的質を評価するのみでは、評価を行うこと自体の目的に対しても、研究者自身が研究を行っている目的に対しても、十分でない場合が生じることになる。そのため、社会経済的効果の評価は、研究プロジェクトやプログラムの評価だけでなく、大学や公的研究所の評価にも求められるようになっている。実際、オランダではVSNUの大学評価における自己評価作業の一環として研究成果のユーザーへのサーベイ調査を行う試み (Spaapen and Wamelink 1999) も見られ、米国では大学からの起業の効果分析 (Bank Boston 1997) や NSF からの資金提供による産学共同研究センターである Engineering Research Center の評価 (Ailes, Roessner and Feller 1997) の試みもある。しかしながら、このような社会・経済的効果の評価方法は依然として確立された段階とは言えない。これら进行评估するための容易な対応策としては、「改善されたピアレビュー (あるいはメリットレビュー)」として、ピアレビューにおいて社会・経済的効果についても同時に評価を行うという方法がとられることが多い。しかし、評価対象から実際にいかなる社会・経済効果が得られたのかという十分な情報がなければ、ピアレビューアーが独自の学問分野の専門性だけでそれら効果をも評価することは極めて困難となる。そのため、そのような情報を提供しうる評価支援者がシステムに組み込まれる必要性が増している。単一の方法のみで十分な分析を行うことは困難であることは教育コースの内容からも明らかであり、評価支援者は様々な種類の方法を並行的に用いて、それら进行评估者に提示する必要がある。教育コースで教えているものだけでも、指標 (評点法を含む)、ビブリオメトリクス分析、サーベイ調査、インタビュー調査、経済的モデリング、詳細な事例分析など多様なものが含まれる。評価支援

者はこのような教育コースを入り口として、さらに評価経験を積むことで評価方法のベストプラクティスを集積することが望まれている。

一方で、これとは別にプロセスへも重点を置く方向がある。ニュー・パブリック・マネジメントの導入により、インプットやプロセスではなくアウトプットやアウトカムに重点をおいた評価を定量的な形で行うことへの要求は研究評価においても見られる。しかしながら、それを過度に強調してアウトプットやアウトカムだけを評価したとしても、その評価結果から今後の研究のマネジメント（組織体制や実施体制・方法など）の改善に連結する含意が十分得られるとは限らない。過去に優れた成果を挙げた組織に資金を集中させれば全体としての効用が本当に向上するのか否かが明らかでない以上、いかなるマネジメントが有効であるかを見定めることこそが評価の目的であるという考え方である。このような積極的なプロセス志向の他にも、研究の効果を5年程度の短期では測定することは不可能であるために、その中間としてのプロセスに焦点を移さざるを得ないという志向も同時に存在する。このようなプロセス志向においても、目的や目標との論理的整合性を有するプロセスが展開されているかが評価されることになり、例えば米国のEngineering Research Centerの評価ではスピルオーバーを可能とする連携関係や企業ユーザとの情報交流をいかにしているかが評価の対象となる（Ailes, Roessner and Feller 1997）。そのため、教育コースで示されたロジックチャートのような形で論理的整合性をより明確にする必要がある。これに加えて、中間的な各フェーズにおいて、どのような行為が次の行為へと連結しているのか、あるいはどのような障害が存在するかを各対象に即して分析することが望まれる。

このような二つの方向性によって、評価を行うシステムは極めて複雑な構成となる。評価者は当該研究分野の専門家とそれ以外の専門家がパネルを構成し、評価の測定対象も多様なプロセスやアウトプットやアウトカムが挙げられ、さらに、その方法も定量的な分析方法と定性的な方法とが必要となる。このように、複雑な評価システムを形成し、各要素を取捨選択する立場にある評価の専門人材の役割はより重要となりつつあると言える。

4.2 大学評価の特殊性

上記のような共通の傾向を持ちながらも、大学という組織の評価には研究プロジェクトやプログラムの評価とは異なる知識やスキルが必要となる。そもそも大学のような機関の研究活動の評価は、研究プロジェクトやプログラムの評価とは次の点で異なっている。すなわち、研究プロジェクトの評価では研究という「行為」に焦点がおかれ、研究プログラムの評価でも複数の研究活動の集合による政策的な「行為」に焦点がおかれる。これに対して機関の評価では、行為（あるいはその支援）を行う「行為者（アクター）」自体により重点がおかれる。これは前者の行為の評価においてはインプット→プロセス→アウトプットというフローのモデルを基礎にしようが、機関の評価ではそのようなフローの変換を行う内部構造自体により注目する必要があることを意味する。具体的には、大学という機関の評価では次のような点でプロジェクトやプログラムの評価とは異なる。

第一には、組織としてのソフト・ハード面の研究支援体制が評価において重要な位置を占めることである。研究プロジェクトは小規模で短期の時限的組織で行われるのに対し、大学は大規模で恒常的な組織である。そのため、組織の戦略性、資金配分方策や研究支援体制、内部人材の構成・流動性・育成といった個別の研究活動を支える組織マネジメントに関する要素が評価においてより重要となる。同時に、プロジェクト資金では提供されないようなハード面の基本的なインフラストラクチャー（施設、機器、設備）が適切であるかも評価において重要な項目となる。

第二には、研究の具体的な活動の評価についても、その集合単位（aggregation level）の問題が生じる。研究プロジェクトは研究内容に関連性を有する少数の研究グループにより構成されるが、機関では内部に数百・千の研究グループが存在しており、特に大学においてはそれらの研究内容は極めて多様で分散的である。本来、組織の評価では、組織全体としての結果の総体やその内部構造を評価することが重要となるが、その反面、このような多様な組織では統一的な評価方法は馴染まない。また、集合単位が大きくなればなるほど、組織としての目的設定はより抽象的で時限が設定されないものとなりやすく、達成状況の評価が困難となる。

第三には、研究以外の機関の別機能（教育、社会サービス）との関連における制約（組織構成、資源）があり、機関全体の機能群の関係の中で評価を行う必要が生じることがある。

このように大学という組織は異なる専門的能力を有する構成員が複数の機能（研究、教育、社会サービス）を行う、複雑な内部構造を有する対象である。具体的に評価を行う際には、当該評価で対象とする範囲を特定の機能・組織・分野に限定することによって、便宜的に複雑さを縮減させることは行われる。だが、そうではあっても、大学の評価に関わる専門の人材には、評価システムを形成する上でこのような複雑な組織のマネジメント自体についての知識も基礎知識として求められることになり、これは狭義の「研究活動の評価」に関する知識を超えるものとならざるを得ない。その一方で、このような組織のマネジメントとしてはどのようなものが望ましいかは必ずしも明らかではなく、例えば「任期制の導入」が組織に対していかなる効果を実際に及ぼすかは明確ではない。そのため、より質の高い評価を行うためには、まずは多くの事例から「何がグッドプラクティスであるのか」を集積するところから始めることが必要となる。欧米の研究評価は、既に述べたように、これまで研究プロジェクトやプログラムの評価への関与が中心となって展開してきたという歴史的背景があるために、依然として機関評価については蓄積が多くない。そのため、大学などの機関評価に真に必要なスキルや知識は教育コースとして形式的に教えられるほどには必ずしも明確になっていない。プロジェクトやプログラム評価については数十年の経験を基にして現在の教育コースが確立されるまでに至ったように、今後の試行錯誤の中で、いかなる方法が特定の評価の目的を満たすために十分であるかは、長期的に検討していく必要があるであろう。

5. 参考文献

- Ailes, C.P., Roessner D., and Feller, I. (1997), *The Impact on Industry of Interaction with Engineering Research Center - Final Report*, SRI International
- Airaghi A. et al. (1999), *Option and Limits for Assessing the Socio-economic Impact of European RTD Programmes*
- Bach, L. et al. (1997), “Indirect effects of large-scale technological development programmes” in Callon M., Laredo, P. and Mustar P. (Eds.), *The Strategic Management of Research and Technology*, Economica
- Bank Boston (1997), *MIT : The Impact of Innovation*
- Barker K. (1997), “The ‘British Model’ — evaluation by professionals”, in Callon, M., Laredo P. and Mustar P. eds.
- Buisseret, T.J., Cameron, H.M. and Georghiou, L. (1995), “What difference does it make? Additivity in the public support of R&D in large firms”, *International Journal of Technology Management*, Vol.10 pp.587-600
- Callon M., Laredo, P. and Mustar P. eds. (1997), “Techno-economic networks and the analysis of structural effects” in Callon M., Laredo, P. and Mustar P. eds.
- Cameron H. and Georghiou L., (1997) “Managerial performance — the process evaluation”, in Callon, M. et al. eds.
- Cozzens, S.E. et al. eds. (1990), *The Research System in Transition*, Kluwer Academic Publishers
- Evered D. and Harnett S. ed. (1989), *The Evaluation of Scientific Research*, John Wiley and Sons
- Georghiou L. and Roessner D. (2000), “Evaluating technology programmes : tools and methods”, *Research Policy* Vol.29 pp.657-678
- Georghiou L. (1989), “Organization of evaluation”, in Evered D. and Harnett S. ed., pp.16-31
- Gibbons, M. and Georghiou L. (1987), *Evaluation of Research — a selection of current practices*, OECD
- Lukkonen (2000), “Additivity of EU framework programmes”, *Research Policy*, Vol.29 pp. 711-724
- OECD (1997), *The Evaluation of Scientific Research : Selected Experiences*, OECD
- Office of Technology Assessment (1986), *Research Funding as an Investment : Can We Measure Returns?*
- PREST (2000), *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape*
- Rip A. (2000), “Social Challenges for R&D Evaluation”, in Shapila P. and Kuhlmann S. eds.,

Proceedings from the 2000 U.S.-European Workshop on Learning from Science and Technology Policy Evaluation, Section 2 pp.18-41

Salter A.J. and Martin B.R. (2001), “The economic benefits of publicly funded basic research : a critical review”, *Research Policy*, Vol.30, pp.509-532

Spaapen J. and Wamelink F. (1999), *The Evaluation of University Research — A method for the incorporation of the societal value of research*, sci_Quest

[ABSTRACT]

A Report on Educational Courses on Research Evaluation in Europe

HAYASHI Takayuki*

New research evaluation systems established in several areas in Japan in the 1990s are now revealing a shortage of human resources for research evaluation. This means not only a shortage of experienced evaluators (e.g. peer-reviewers) but also professional managers in agencies and experts in universities and consulting firms for evaluation. Therefore, it is necessary to grasp what kind of knowledge and skills are required for the development of these kinds of human resources.

This report examines two educational courses on research evaluation in Europe. They are held in PREST of Manchester University in the U.K. and Twente University in the Netherlands. Both courses consist of the following contents : background of research evaluation, design of systems for research evaluation, methods for research evaluation, and the writing of evaluation reports and their use for decision making. These contents reflects the recent trends of research evaluation. One trend is to focus on several kinds of outputs and outcomes in evaluation. To measure these, peer review is not sufficient and other kinds of methods are also required. The other trend is to focus on the research process as well as output/outcome. One reason for this is that outputs and outcomes cannot be measured in short terms and the other is that the evaluation of process can provide implications for good practices of research management. In particular, to evaluate university research, the knowledge about the management of institutes beyond the research activities is also required. Therefore it is necessary to accumulate good practices in management through several case-studies in order to evaluate the institutes.

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