

広島大学 高等教育研究開発センター 大学論集
第44集 (2012年度) 2013年3月発行：227-242

研究評価の研究動向

西 村 君 平

研究評価の研究動向

西村 君平*

1. 研究の背景と目的

我が国では1995年の科学技術基本法制定を一つの契機として、継続的に研究制度の改革が進められている。その中で重要課題の1つとなっているのが、研究評価の改革である（内閣府，2008）。特に第4次科学技術基本計画（2011年）では、科学技術イノベーションの促進の観点から研究費制度における審査・配分機能の強化がうたわれており、今後、変化が加速していくと予測される。

このような変化は、何も日本だけに見られる現象ではない。日本を含む先進工業国では科学技術のもつ社会的経済的インパクトが高まる中で、研究評価は単に科学コミュニティ内部に閉じた問題ではなくなっている。政府、議会、企業、市民社会など様々な利害関係者は、自らの立場を研究評価に反映させることを通じて、研究制度への介入を強めようとしている。すなわち今や研究評価は研究ガバナンス¹⁾のツールという新たな役目を帯び始めているわけである（Hansson, 2006）。そしてその結果、評価視点の多様化や評価手法の複雑化が進んでいる（Frederiksen et al., 2003）。

本論考が注目するのは、このような制度的・社会的変化を受けて、新しい研究評価の研究が国内的にも国際的にも登場してきていることである。以下ではこれらの国内外の研究動向をレビューすることで、知見の体系化と今後の研究課題の明確化を行う。

2. 先行レビューと本論考の関係

従来、研究評価の研究は「研究の研究（research on research）」の一部と考えられ、研究そのものを扱うインターナルな研究（科学論に近接）、研究を支える制度的・社会的条件を扱うエクスターナルな研究（高等教育研究に近接）に区分されてきた。レフリーシステムやその背後にある価値規範構造の研究は前者に、近年の研究評価制度の変化を受けた研究は後者に含まれる（阿曾沼，2006）。

先行するレビュー論文が示したインターナル・エクスターナル分類は、研究評価研究の2つの学問潮流を浮き彫りにしており、その有用性は今日も失われていない。しかしその限界性としてレビュー範囲が国内に限定されており国際的な研究動向が軽視されている点、「研究の研究」を鳥瞰する目的意識が強いためか、研究評価の研究については部分的紹介に留まっている点が挙げられる。

ここでエクスターナルに分類されるような、ガバナンス・ツールとしての研究評価に係る国際的な研究動向を概観すると、科学技術社会論（Foltz, 2000）、科学計量学（Moed, 2005; Weingart, 2005）、

* 広島大学大学院教育学研究科教育人間科学専攻（高等教育学）／日本学術振興会特別研究員 DC

評価論 (Coryn & Scriven, 2008) といった高等教育の近接領域での知見の蓄積を見て取ることができ。またこれら知見の参照関係を見ると、いずれも科学社会学を始めとするインターナルな研究の知見に依拠していることに気付かされる。そこで更に近年のインターナルな研究に目を向けてみると、研究評価の品質について経験的研究が蓄積していることが分かる (Bornmann, 2008; Demicheli & Di Pietranton, 2008; Wessely, 1998)。

よって以下では、まず科学技術社会論、科学計量学、評価論における研究蓄積をレビューする (3節: エクスターナル)。次にその背景にある科学社会学の研究動向を、特に研究評価の品質についての経験的検討に焦点化して整理する (4節: インターナル)。最後に経験的検討の基礎にある科学観について科学論の知見に基づき考察することで、先行研究の知見の再評価と今後の研究展望を得る (5節: インターナルからエクスターナルへ)。このように先行研究を遡及的に (いわば鳥の目と言うより蟻の目で) レビューすることで、研究評価研究の中核部分に隠れた論点を明らかにできるのではないだろうか。

3. 新しい研究評価の研究

ガバナンスに係る研究評価の研究は3つに分けられる。研究対象の違いに由来する区別としてグラントピアレビュー (grant peer review) の研究とビブリオメトリクスの研究に分けられ (Coryn & Scriven, 2008), これに加えて研究評価手法や制度の良し悪しを評価するメタ評価的研究が実施されている (Weingart, 2005; Coryn & Scriven, 2008)。

(1) グラントピアレビューの研究

①定義および概念の整理: 公共性と秘教性の葛藤

グラントピアレビューとは、文字通り研究費配分を目的としたピアレビューのことで、その特徴はピアレビューとの対比で示される (Chubin & Hackett, 1990)。

20世紀初頭、近代科学が世界規模で広がりを見せ始める中で、ピアレビューは研究評価のスタンダードとして広く普及することになった。その一般的特徴は同僚研究者が評価者となること、評価の観点が学術的価値の有無に置かれること、判断の根拠は専門的判断であって、必ずしも明示化されないことなどに求められる。つまりピアレビューは科学コミュニティの内部に閉じており (Chubin & Hackett, 1990), 秘教的と言える (esoteric; 新堀, 1984)。

対してグラントピアレビューは、第二次世界大戦後、米国の NSF・NIH での実践が注目を集め世界的に普及していく (Guston, 2003)。資金配分と結びつくため、グラントピアレビューにはアカウントビリティや効率性が期待される。つまりグラントピアレビューには公共性が付与されていて、秘教的かつ公共的という葛藤を内包する性質を持つ (Chubin & Hackett, 1990)。公共性は評価者に行政官や政治家が参入したり、評価観点に社会経済的価値が入れられたり、後述するビブリオメトリクスなどを用いて判断根拠の明示化が求められることで実現すると考えられている。

②制度的・手法的な構成要素

グラントピアレビューは公共性と秘教性の葛藤ゆえに制度・手法の面で複合的で複雑である。そのため構成要素の記述整理それ自体が研究トピックとなっている (Foltz, 2000; 林, 2004; Frederiksen et al., 2003)。

先行研究が指摘する共通部分は以下の4つに集約できる。1つめはメールレビュー、申請書の査読である。メールレビューの主眼は対象の学問的価値の決定であり、評価者には専門を同じくする同僚研究者が与えられる。2つめはパネルレビューで、メールレビューの結果を受けて、委員会方式で合議する。この際には対象の学問的価値のみならず、社会経済的価値などが論じられる。委員会では当該領域のオーソリティや熟練研究者で構成される。3つめは行政裁量 (strong manager discretion) で、資金配分団体の上級経営層 (行政官・議員含む) の意思決定である。4つめはフォーミュラ (計算式) による評価結果の算出である。ビブリオメトリクスはここに援用される。

③事例研究・国際比較の知見の要約と考察

制度・手法の複雑性は他国の事例研究や国際比較、評価の実務家による実践報告の機運を高めている²⁾が、これまでその知見は必ずしも関連付けて論じられて来なかった。そこで次に公共性と秘教性の観点から知見を要約し、研究評価のガバナンス・ツール化を構造的に把握する理路を得る。

表1 事例研究・比較研究の知見の要約

公共性	秘教性
(評価者) 行政裁量の強化 (山本, 1995) ボークバレル: ・省庁・議員による利益誘導 (Rennie, et al., 1989) ・研究予算の特定目的化 (阿曾沼, 1999) ・利害関係の管理技術の開発 (Quinlan, et al., 2008) (評価の観点) 研究倫理の変容: ・PLACE (ザイマン, 1995) ・academic capitalism (Slaughter & Leslie, 1997; 上山, 2010) (評価根拠) フォーミュラの強化 (山本, 同上)	(評価者) ・オーナーシップの弱化 (米澤, 2009) ・企業経営者等の参入による素人評価 (西村, 2011) 利益相反の顕在化とその排除: ・競合研究者の評価者への採用問題 (Riggs, 1995) ・競合研究者の排除による素人評価 (Cohen, 1996) (評価観点) ・パラダイムへの隷属 (Riggs, 1995; Sowers, 1995) ・ビッグサイエンスの優遇 (Goodstein, 1995) (評価根拠) ・過剰業績主義 (Alberts & Shine, 1994; Marshall, 1995)

公共性と秘教性の観点から見れば、公共性の強化を示す「行政裁量の強化」「フォーミュラの強化」、秘教性の弱化を示す「オーナーシップの弱化」「素人評価」「過剰業績主義」は、グラントピアレビューの基礎的な方法論に関わる重要なトピックと言える。また「ボークバレル」「研究倫理の変容」は公共性の強化による研究評価の政治・経済問題化を示唆しており、政治・経済問題化は「利益相反の顕在化とその排除」にもつながっている。なお公共性と秘教性の葛藤という観点からはガバナンス・ツール化により評価観点が多様化すると予測されるが、先行研究の知見は逆に「パラダイムへの隷属」「ビッグサイエンスの優遇」といった評価観点への収斂を示唆しており興味深い。

④知見の小括と考察：事例・比較研究から構造的把握へ

研究評価制度の改革はグラントピアレビューに関する無数の研究や報告を生み、グラントピアレビューというブラックボックスに徐々に光が当たりつつある。しかし先行研究の中心は事例研究・比較研究であり、グラントピアレビューのメカニズム等に対する構造的把握は遅々として進んでい

ない。本論考で試論的に試みたが、今後は事例研究・比較研究の知見を手がかりに、グラントピアレビューの構成概念やモデルの彫琢、及びその経験的検証が求められると言えよう。

(2) ビブリオメトリクスの研究

①定義および概念の整理：研究評価への応用可能性

ビブリオメトリクスは、文献の数学的・統計的分析を行うもので、評価手法と言うよりも情報科学・科学社会学で用いられる調査手法である。論文数や論文引用数、学術誌のインパクトファクターを用いた分析がよく知られているが、近年は他にも様々な分析手法が開発されている³⁾。また分析対象となるデータベースは、もっぱらトムソンサイエンス社の Science Citation Index (SCI データベース)であったが、近年は他のデータベースの開発も進んでいる(富澤, 2004)。

我が国では特に90年代以降、ビブリオメトリクスの研究評価への応用の機運が高まった。ここでは定量的データを用いることでピアレビューの根拠資料を拡張すること(中井, 1995)、定量的データのわかりやすさを活用して国民に対する説得力を増進すること(根岸, 2000)が期待された。

他方でビブリオメトリクスの利用に対する慎重論もある(加藤, 2004)。人文・社会科学でのデータベース欠如、自然科学ですら独創性は定量化できないこと、数字のひとり歩きと過剰な業績主義の危険が指摘される。また科学者当人は概ね定量的手法に慎重であることも示唆されている。

しかし現実にビブリオメトリクスの応用は始まっており、研究システムのモニタリングツールとしての利用、個別研究・研究者の評価方法としての利用は不可逆的に進みつつある(Aksness, 2005)。

②研究システムのモニタリングツール

情報科学・科学社会学に出自をもつビブリオメトリクスは、研究システム全体の研究業績の定点観測に長けている。この特徴を活かして研究システムの国際競争力を相対評価することができる。

我が国では1980年代に基礎研究タダ乗り批判への応答、劣悪な研究環境の改善アピールなどの必要性から、日本の研究システムの国際的競争力の検討が進められた。その際に領域ごと、国公私といった設置形態ごと、大学ごとの競争力の比較検討も行われている(阿曾沼, 2006)。

また大学単位の競争力比較によるランキング開発も行われている(ライデン大学によるランキング等)。我が国でも大学ランキングにSCIを用いた論文引用ランキングが用いられている(根岸, 2009)。ビブリオメトリクス型のランキングについても、データベースの整備状況、測定誤差、データ読解能力の欠如などが指摘され、その是非について慎重な検討が勧められている(van Raan, 2005)。

③個別研究の評価

個別研究の評価はその機能に応じて更に2つに区別できる(林, 2004)。1つめは研究活動の改善・形成的機能を重視する立場で、機関の研究業績の記述や規定要因の説明が実施されている。2つめは意思決定・総括的機能を重視する立場で、ビブリオメトリクスのフォーミュラへの適用例がある。ガバナンス・ツールとしての研究評価に関心を持つ本論考では、特に後者が重要になる。

後者の例としてイギリス(米澤, 2007)とオランダの事例(林, 2004)が報告されているが、特にオーストラリアの事例は先取的で、国際的にも高い注目を集めている。

もともとオーストラリアでは、大学に対する資金配分はグラントピアレビューによる競争的資金

とフォーミュラを用いて恒常的に配分される基盤的資金の二元体制が敷かれていた。90年代に入ると、フォーミュラの改正が始まり、1996年以降は論文生産数が採用されることになった。これを受けて、1981年から2002年にかけてオーストラリアの3つの研究大学での論文発表の傾向性の変化が検証された。全体的傾向として、各大学の論文生産数は向上しているが、その発表傾向を見るとインパクトファクターの低い学会誌への発表が増加していた (Butler, 2003)。改革後のオーストラリアの研究者は、研究の質よりも量に関心があり、長期的研究業績よりも短期的業績を重視し、革新的アプローチよりも保守的なアプローチを好む傾向にあった (Marginson & Considine, 2000)。オーストラリアとの好対照事例としてスペインの事例がある。そこでは高引用論文の発表数に応じた賞与が実施され、その結果インパクトファクターの高い学会での業績が増加していた (Butler, 2003; Jimenez-Contreras et al., 2003)。こうした研究を受けて、我が国でもビブリオメトリクス応用に際しては、業績の量よりも質の向上を促進するように留意すべきと指摘されている (林・富澤, 2007)

④知見の小括と考察：評価手法としてのビブリオメトリクスの可能性と困難性

このようにビブリオメトリクスの評価手法としての有効性については慎重論も根強い。特に欧州の科学計量学では評価手法としてのビブリオメトリクスについての理論的考察も進んでおり注目に値する。そこでは「 Grantピアレビューは定性的・主観的、ビブリオメトリクスは定量的・客観的」といった通俗的の二元論にとらわれず、複眼的にビブリオメトリクスの可能性と困難性が検討されており (van Raan, 2005; Weingardt, 2005)、今後の我が国の研究・実践への示唆に富む。

(3) メタ評価的研究—研究評価手法・制度の相互比較

① Grantピアレビューの結果とビブリオメトリクスの結果の相互比較

そもそもビブリオメトリクスは研究評価に応用できるのか。事例研究が示す否定的な印象を受けて、Grantピアレビューとビブリオメトリクスの相互比較が多数実施されている (林, 2004)。そこでの焦点はGrantピアレビュー結果とビブリオメトリクス結果の相関を見ることにある⁴⁾。Grantピアレビューの結果としては評価得点や順位が用いられ、ビブリオメトリクスの指標としては論文数や引用数が用いられる。ただし分析結果にはまだ一貫性がなく、論争を呼んでいる (Oppenheim, 1997; Warner, 2000)。

②メタ評価枠組みの開発と制度の比較評価

既に述べた通り、各国の研究評価制度の異同についての比較研究は複数存在する。しかし評価制度の良し悪しを論じた研究はほとんど実施されていない。例外的に評価論の文脈で研究評価制度の国際比較評価 (Coryn, 2007) が実施されており注目に値する。

この研究では評価の品質は表2にあるような5つの規準で捉えられている。ただし規準の根拠はそれがアメリカの評価論に依

表2 コーリンらのメタ評価規準

<p>妥当性：良い研究評価は論理的に適切で論証可能な結論を下す。 信頼性：良い研究評価は、関係者が信じるだけの合理的根拠を持つ。 有用性：良い研究評価は結果が活用されやすいように設計される。 費用対効果性：良い研究評価は経済的である。 倫理性：良い研究評価モデルはバイアス管理が適切に実施される。</p>

拠することが示唆されるのみで、不透明である。その他の規準 (例えば費用対効果性や社会的正統

性など)への優越性は示されていない。また評価資料は各国の評価制度をまとめたレポートで、評価者は評価の実務経験者が担当しているが、このような測定方法の社会調査としての妥当性・信頼性に疑問が残る。なお総合得点において高い値をマークしたのは、業績ベースの資金配分と体系的評価手法を備えた国・地域(ニュージーランド等)であった。

③知見の小括と考察：研究評価の品質の多義性

メタ評価的研究はその端緒についたばかりで問題も多い。特に今後は「グラントピアレビューとビブリオメトリクスの相関は評価の何を担保する指標なのか」「メタ評価規準とその適用方法の根拠は何か」などメタ評価の仮構する「理想の研究評価像」⁵⁾が論争的になるだろう。冒頭でも述べたが、今日では研究の社会経済的効用への期待は高く、様々な利害関係者が研究評価へ介入し始めているからである。いきおい異なる「理想の研究評価像」の間で政治的葛藤が生じる可能性が非常に高い。今後のメタ評価的研究では自覚的に評価品質の多義性がわきまえられなければならない。

4. 研究評価の品質の経験的検証

科学社会学では研究評価の研究は伝統的に重視されてきたが、その中で研究評価の品質の経験的検証が進んでおり、評価論のような近接領域からも参照されている(Coryn & Scriven, 2008)。検討の焦点は信頼性、バイアス、予測妥当性に集約される(Bornmann, 2008)。そこで次に品質指標(メタ評価規準とほぼ同義)の理論的・経験的基礎に着目しながら、その知見を整理する。

(1) 信頼性

信頼性とは、評価者の判断の一貫性の程度を指す概念である。研究評価の文脈では、評価者間の判断の異同すなわち評価者間信頼性(inter-rater reliability)が特に問題となる。同じ条件(同一の判断材料、手順)で評価したにも関わらず評価結果に違いが生まれたとなると、評価者の持つ専門知識にばらつきがあったことが疑われるからである。これはピアレビューの基本的仮定である「専門家(=同僚評価者)は高水準の専門知識を共有している」への疑念に直結する(Cole, 1991)。この意味で信頼性の問題は非常に重要であり、ゆえに信頼性については研究蓄積が顕著である。

既に個別研究結果の比較分析と総合(meta-analysis)が実施されている(Cicchetti, 1991)。比較分析の結果、相関係数が0.2-0.4の範囲におさまることが明らかになっている。社会調査の場合、0.5以下は信頼性が低いと判断されることに鑑み、ピアレビューの信頼性は低いと結論付けられている。

こうした見解には反対意見もだされている。すなわちピアレビューの評価規準は多元的であるにも関わらず、多くの研究は規準ごとの得点を無視して、総合得点のみを取り扱っている。規準ごとの得点を扱えば評価者間信頼性を相対的に高まるという指摘である(Hargens & Haerting, 1990)。

(2) バイアス

研究評価の低信頼性は、専門知識以外の要因が評価結果に意図せざる偏りを生じさせている可能性、すなわちバイアスへの懸念を惹起する。具体的には性別、所属大学の威信、学閥(old boy

network) といった複数のバイアスの存在が指摘されている (Bornmann, 2008)。

しかしバイアスについては一貫した結果は得られていない。例えば我が国では評価者や科研費配分が国立大学に優先的に配分されているという根強い批判がある一方 (川野辺・前川, 2005; 竹内, 2001; 2003), そのようなバイアスは確認できないという研究結果も出ている (島, 2005)。

そもそも評価者や配分結果の偏りをしてバイアスの存在を指摘するのは不可能である。公正な評価を行ったにも関わらず、実力を反映して特定の組織 (例えば東大) に配分が偏った可能性を排除できないからである (新堀, 1974)。こうした問題は実験的手法等の、より精密な条件管理のもとで評価の過程を精査することで解決できると考えられる (Demicheli & Pietranton, 2008)。

(3) 予測妥当性

ピアレビューの目的の1つは、数ある研究の中から優れた研究を選抜し、それにリソース (紙面, 研究費等) を集中させることにある。しかし評価 (審査) の段階では対象となる研究はまだ社会に公表されていないか、あるいは開始すらしていない状態なので、その研究が今後、学術的・社会的にどの程度インパクトを持つのか不定である。このようにピアレビューは本質的に未来への予測を含む。この特徴は予測妥当性の問題として検証されている (Bornmann, 2008)。

予測妥当性の検討は、ビブリオメトリクスとピアレビューの比較と似る。まず論文の学術的価値を引用度数で操作化する。次に投稿論文を2種類に分ける。1つは最初の投稿でストレートに採択されたグループでもう1つは最初の投稿でリジェクトされ、後に小規模な学会などに再投稿し採択されたグループである。あとは2つのグループの引用分析の結果を比較すれば、最初の評価の予測妥当性が明らかになる。欧州の自然科学系の学会誌を対象に4つの事例研究が実施されているが、いずれの事例でも予測妥当性の高さを示す結果が得られている (Bornmann & Daniel, 2008)。

(4) 小括：データ解釈のための理論枠組みの不在と科学観の相違

このように科学社会学ではピアレビューの品質に関する経験的検証が着実に進められており、知見の整理総合も進んでいる。また品質指標は、科学の目的や特性といった科学観に依拠して設定されており、知識基盤も相対的に堅固である。

しかし経験的知見をどのように解釈するべきか、共通枠組みは欠如している (Bornmann, 2008)。例えば信頼性の低さをして評価者の判断のデタラメさを見て取ることもできるが、逆に科学の健全な多様性の現れとも解される (Cole, 1991)。また科研の審査の信頼性を研究すれば、超高信頼性スコアが出されるのではないかとという予測もある (慶伊・本多, 2010)。この予測は、科研の審査には過去の業績と所属大学の偏重という強固なバイアスが存在するという仮説に基づいている。この場合、信頼性の高さは判断のデタラメさを帰結する⁶⁾。

このように研究評価の品質の経験的検討を慎重に重ねてもなお、評価の品質についての解釈のばらつきという課題が残ってしまう。そこで次に科学論の知見に基づきながら、研究評価の品質に係る科学観について考察し、科学社会学における品質指標の内奥を探る。それにより評価の品質の多義性という問題を解消しうる研究展望を得ることが狙いである。

5. 研究評価の品質に係る科学観の葛藤とその解消—研究評価研究の展望—

(1) 機能主義的アプローチと構成主義的アプローチの葛藤

研究評価に係る科学社会学の理論は、機能主義と構成主義の葛藤をモチーフに描くことができる (Bornmann, 2008)。すなわち研究評価の研究において、機能主義は CUDOS 等の合理的規範・価値構造の存在を仮定して、ピアレビューの機能を説明するアプローチを指す。逆にピアレビューを通じて学術知識が構成されていく過程を記述し、それをテコに機能主義の描く規範・価値構造に書き換えを迫るアプローチを採るのが構成主義である。

(2) 科学観の相違・品質指標の共通性

機能主義と構成主義の葛藤は、それが前提にする科学観の違い⁷⁾に由来する (Cole, 1991)。機能主義は、人間の認識から独立した自然・真理の存在を仮定する。そして科学はその法則を明らかにすることを目指すと考え。この科学観に立つと研究評価の主眼は、評価対象論文が自然や真理と合致しているかを検討することに求められる。自然・真理は本質的に普遍的・斉一的と考えられるので、研究評価は本来の高い信頼性を示すと考えられる。

構成主義は人間の認識から独立した自然・真理の不在を仮定する。そして科学は情報を「自然」や「真理」にコード化する営みであると考え。この科学観に立つと、研究評価の主眼は評価対象論文と既存の「自然」や「真理」との異同を見ることに置かれる。「自然」や「真理」は本質的に文脈的・ランダム的と考えられるので、評価結果の低信頼性やバイアスこそ研究評価の本性である。

このように整理することで、機能主義と構成主義は科学観においては対立的だが、研究評価の品質指標の設定においては共通性を有していることが分かる。すなわち研究評価品質の経験的検討の「理想の研究評価像」は機能主義の理想へと一元化しており、残った問題はそれを機能主義的に擁護するか構成主義的に批判するかである (ゆえに品質の検討結果の解釈が争点になる)。よって機能主義・構成主義の葛藤というモチーフのみでは、研究評価の品質の多義性は十分に扱いきれない。

(3) 科学の逆説的性質—機能主義と構成主義の限界性—

更に敷衍すれば、研究評価への機能主義的擁護と構成主義的批判は、科学論としての同根の問題性を抱えている。1960年代以降、構成主義は研究評価が実際にどのように進められているか、エスノグラフィーの手法を用いて記述することで機能主義の限界性を示した (平川, 2002)。そこでは評価者は評価対象となる論文や申請書が自然や真理と合致しているか、追試やデータ照合することは皆無であり、せいぜい評価者の持つ専門知識に基づいて、論文や申請書の価値を論じるのみであることが示された (Knor-Cetina, 1981)。つまり機能主義の仮定は実際の科学的営みと合致していない。

逆に1990年ごろから機能主義による反論が出されている (Cole, 1991)。すなわち構成主義は本質的に文脈的でランダム的なはずの科学的知識に、限定的であれ普遍性や斉一性が存在することを説明できていないという批判である。仮に構成主義の言うように科学が全く文脈的・ランダム的であれば、なぜ科学的知識に基づく事業は成功するのか—例えばなぜ飛行機は飛ぶのか—が説明されな

い。つまり構成主義の仮定は科学的知識と合致していない。

このように単に科学の社会的側面を指摘する立場（構成主義）も科学の認知的側面を強調する立場（機能主義）も科学という行為を説明するには不十分である。むしろ「科学の逆説的性質，すなわち社会的な文脈の中におかれた科学がそれにも関わらず普遍的な真理を産むという特殊性」（伊勢田，2004，49頁）を説明しうる理論体系の構築が求められている⁸⁾。では，そのような理論体系の構築に向けて，我々はどのようにして研究評価にアプローチすればよいだろうか。

（4）研究評価への新しいアプローチ—研究評価の文脈性をキーワードに—

従来の機能主義・構成主義のアプローチは表面的に対立的だが，科学の逆説性という観点からは両者には共通性がある。それは研究評価の文脈性にネガティブな位置づけを与えることである。すなわち機能主義では研究評価における社会的文脈は科学の攪乱要因と位置づけられ，構成主義においてそれは科学の非合理的な本性と位置づけられる。こうした発想からは評価の文脈性の「有無」が非常に重視される。それが自らの科学観の最高の論拠となるからである。

しかし科学の逆説性の観点からは，研究評価の文脈性の「有無」はクリティカルな問題とならない。なぜならそれは科学の不可避で当たり前の構成要素と見なされるからである。むしろ問題は文脈性の「処理過程」に置かれる。すなわち「研究評価の過程において，いかにして文脈の諸要因を普遍的通用性のある結論へと昇華するのか」，そのメカニズムを記述・説明することが課題となるのである。またこのメカニズムが作動するために必要な特性や条件を明らかにすることで，研究評価の実態により適合的なメタ評価規準を構築できる。ひいては研究評価の品質の多義性が問題化しがちな今日の状況においても有効に作動する，研究評価の在り方を知ることができるだろう。

6. 結論：知見の要約と本論考の意義，限界性

本論考では事例研究や比較研究，実証研究を領域横断的に渉猟し，それらの研究の背後にある基本的な考え方や仮定を考察することで，研究評価に係る研究のネットワークを描き出してきた。

まず近年のエクスターナルな研究を概観することで，研究評価の多元化と複雑化は，グラントピアレビューとビブリオメトリクスの複合体の登場として記述されること，グラントピアレビューやビブリオメトリクスへの事例研究等を通して，研究評価の品質への関心が高まりメタ評価的研究が実施され始めていることを示した。今後は政策的にもメタ評価が推進されていくと予想されるが，その際には品質の多義性という問題の克服が不可欠となる。

次にこの問題の克服に向け科学社会学における品質の経験的検証の知見を整理し，その背後にある機能主義と構成主義の科学観について考察した。この考察により今後は科学の逆説性の観点が重要になること，この観点からは評価の文脈性がキーワードとなることを論じた。またこうした研究の発展の中で評価の多義性の解消につながる研究評価方法論が構築される可能性を述べた。

本考のレビューは関連領域の知見の融合や体系化に部分的にであれ寄与できたはずである。ただし，研究評価研究は想像以上の広がりや深さを持っており，本考ではその動向を追うために関連領

域の先行レビューを利用せざるを得なかった (Bornmann, 2008; Coryn & Scriven, 2008; Foltz, 2000; Hansson, 2006; Weingart, 2005)。この点に本考の限界性があることを、最後に指摘したい。

【注】

- 1) ガバナンスはやや曖昧な概念であるが、本論考ではそれを次のように理解している。すなわち従来、社会制度は中央政府によって一元的・直接的に管理される傾向が強かったが、近年、制度の自律分散化・複雑化・多様化が進み、様々な利害関係者（議会・企業・一般市民など）による多元的・間接的な管理が強化されてきている。ガバナンスは後者の統治様式を指示する概念である。
- 2) 他国・他地域の情勢についての分析やテクニカルレポート（遠藤, 2003；大場, 2000；米澤, 2009）、政府・行政機構の主導での大規模国際比較が報告されている（サルミ・ハウプトマン, 2006; OECD, 2003; ab Iorwerth, 2005; Geuna & Martin, 2003; von Tunzelmann & Mblura 2003）。また Forz (2000) では、科学者による研究評価制度への批評・反省を含む様々な報告の知見が整理されている。
- 3) ビブリオメトリクスの方法については、藤垣他 (2004) の参照。
- 4) 研究方法論的色彩が強いが、論文数、引用数、相対的引用度（論文が発表された学会誌の引用度）、領域相対引用度（相対引用度を引用度の分野平均を用いて値を統制したもの）、相対的発表戦略性（学会誌のインパクトファクターと当該領域におけるインパクトファクターの比で、相対引用度を統制したもの）を用いた指標間比較も行われている。ピアレビューの得点と最も相関が低かったのは相対的引用度数で、最も相関が高かったのは相対的発表戦略性であった (Aksnes & Taxt, 2004)。指標の定義については Vinkler (1991) に詳しい。
- 5) 研究評価に対する批判的検討が依拠する「研究評価の理想状態（公平な評価等）」を問題化する必要があるという指摘は、90年ごろには既に提起されている (Chubin & Hackett, 1990)。
- 6) ここで予測として示したものは、当該論文68-69頁における「意見」を定式化したものである。
- 7) 本節の機能主義と構成主義の背後にある科学観の相違は、実在論と反実在論に関する科学哲学の知見を参照している（特に伊勢田, 2004；戸田山, 2005）。
- 8) なおこの問いに対してコールは機能主義的回答を提出している (Cole, 1991) が、実在論的科学観を前提にしており、新規性のある議論とはなっていない。またボンマンはルーマン社会システム論による機能主義と構成主義の統合を提案している (Bornmann, 2008)。

【参考文献】

- 阿曾沼明裕 (1999) 「国立大学における研究費補助のパターン変化—特定目的化と競争化—」『高等教育研究』第2集, 135-155頁。
- 阿曾沼明裕 (2006) 「研究の研究」『大学論集』第36集, 107-126頁。

- 伊勢田哲治 (2004) 『認識論を社会化する』名古屋大学出版会。
- 上山隆大 (2010) 『アカデミック・キャピタリズムを超えて』NTT出版。
- 遠藤悟 (2003) 「米国における学術政策・学術行政評価—国立科学財団 (NSF) にかかる評価を中心に」『学術月報』第56集, 3巻, 236-255頁。
- 大場淳 (2000) 「人文社会科学の研究とその評価」『高等教育研究』第3集, 81-105頁。
- 加藤毅 (2004) 「研究成果に関する主観的評価の構造と測定可能性」『大学研究』第31集, 67-80頁。
- 川野辺裕幸・前川公志 (2005) 「高等教育政策の転換と競争的資金供給」『公共選択の研究』第45集, 5-23頁。
- 慶伊富永・本多卓也 (2000) 「自然科学系の研究その現状と大学評価」『高等教育研究』第3集, 63-79頁。
- ジョン・ザイマン (村上陽一郎・川崎勝・三宅苞訳) (1995) 『縛られたプロメテウス: 動的定常状態における科学』シュプリンガー・フェアラーク。
- ジャミル・サルミ, アーサー・ハウプトマン (丸山文裕監修) (2006) 『高等教育におけるイノベーション: 配分メカニズムの比較評価』国立大学財務経営センター。
- 島一則 (2005) 「私立大学・私立中各大学に対する科学研究費補助金の配分実態」日本私立大学協会付置私学高等教育研究所報告書『研究と研究費』同研究所発行, 25-35頁。
- 新堀通也編 (1974) 『学閥』福村出版。
- 新堀通也編 (1984) 『大学教授職の総合的研究—アカデミックプロフェッションの社会学—』多賀出版。
- 竹内淳 (2001) 「大学の科学研究費の官民格差—日本の科学界のレベルアップのための構造的課題—」『科学』6月号, 832-836頁。
- 竹内淳 (2003) 「大学の公的研究費の日米構造比較」『科学』第73集, 第2号, 137-140頁。
- 戸田山和久 (2005) 『科学哲学の冒険』日本放送協会。
- 富澤宏之 (2004) 「データベース」藤垣裕子・平川秀幸・富澤宏之・調麻佐志・林隆之・牧野淳一『研究評価・科学論のための科学計量学入門』丸善, 38-47頁。
- 内閣府 (2008) 『国の研究開発評価に関する大綱的指針』(<http://www8.cao.go.jp/cstp/kenkyu/taikou081031.pdf>) <2012年9月18日アクセス>。
- 内閣府 (2011) 『科学技術基本計画』(<http://www8.cao.go.jp/cstp/kihonkeikaku/4honbun.pdf>) <2012年9月18日アクセス>。
- 中井浩二 (1995) 『我が国における研究評価手法の総合的研究 (報告書)』(科研費総合 A)。
- 西村君平 (2011) 「21世紀 COE プログラムの評価実態調査—評価委員の属性の多様性に着目して」『大学論集』第43集, 223-238頁。
- 根岸正光 (2000) 「研究評価とビブリオメトリックス」根岸正光・山崎茂明編『研究評価』丸善, 1-18頁。
- 林隆之 (2004) 「研究評価への科学計量学の応用」藤垣裕子他, 前掲書, 148-174頁。
- 林隆之・富澤宏之 (2007) 「日本の研究パフォーマンスと研究実施構造の変遷」『大学評価・学位授与研究』第5号, 57-73頁。

- 平川秀幸 (2002) 「実験室の人類学」金森修・中嶋秀人編『科学論の現在』勁草書房, 24-63頁。
- 藤垣裕子 (2004) 「科学計量学の歴史」藤垣裕子他, 前掲書, 11-18頁。
- マイケル・ギボンズ編著 (小林信一監訳) (1997) 『現代社会と知の創造』丸善。
- 山本真一 (1995) 「外国における研究評価システムの現状—大学セクターの場合—」中井浩二 (代表) 『我が国における研究評価手法の総合的研究 (報告書)』(科研費総合 A)。
- 米澤彰純 (2009) 「研究評価と財政配分の日英比較—コンセンサス・メディアとしてのピア・レビュー—」日本教育行政学会 『学校と大学のガバナンス改革』教育開発研究所, 195-209頁。
- ab Iorwerth, A. (2005). *Methods of Evaluation University Research around the World*. Ottawa: Canada Department of Finance.
- Aksness, D. W. (2005). *Citations and their Use as Indicators in Science Policy*. Unpublished doctoral dissertation, University of Twente.
- Aksness, D. W. & Taxt, R. E. (2004). Peer Reviews and Bibliometric indicators: a comparative study at Norwegian University. *Research Evaluation*, 13(4), 33-41.
- Arberts, B. & Shine, K. (1994). Scientist and the Integrity of Research. *Science*, 266, 1660-1661.
- Bornmann, L. (2008). Scientific Peer Review: an Analysis of the Peer Review Process from the Perspective of Sociology of Science Theory. *Human Architecture: Journal of Sociology of Self-Knowledge*, IV, 2, 23-38.
- Butler L. (2003). Modifying Publication Practices in Response to Funding Formulas. *Research Evaluation*, (April), 39-46.
- Chicchetti, D. V. (1991). The reliability of Peer Review for Manuscript and Grant Submissions: a Cross-disciplinary investigation. *Behavior and Brain Science*, 14, 119-135.
- Cole, S. (1991). *Making Science*. Cambridge: Harvard University Press.
- Coryn, C. (2007). *Evaluation of researchers and their research: Toward making the implicit explicit*. Unpublished doctoral dissertation, Western Michigan University.
- Coryn, C. & Scriven, M. (2008). Editor's Note. *New Directions for Evaluation*, 118 (spring), 1-5.
- Coryn, C., Hattie, J. A. & Scriven, M. & Hartmann, D. J. (2009). Models and Mechanisms for Evaluating Government-Funded-Research. *American Journal of Evaluation*, 28(4), 437-457.
- Cohen, J. (1996). AIDS Trials take on Peer Review, *Science*, 271, 20-21.
- Cubin, D. E. & Hackett, E. J. (1990). *Peerless Science: Peer Review and U. S. Science Policy*. Albany: State University of New York Press.
- Demicheli, V. & Di Pietrantonj, C. (2008). Peer Review for Improving the Quality of Grant Applications. *The Cochrane Library*, 2, 1-15, (online).
- Foltz, F. A. (2000). The Ups and Downs of Peer Review: Making Funding Choices for Science. *Bulletin of Science, Technology & Society*, 20 (December), 427-440.
- Frederiksen, L. F., Hansson, F. & Wenneberg, S. B. (2003). The Agora and the Role of Research Evaluation. *Evaluation*, 9(2), 149-172.
- Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: an Internal Comparison.

- Minerva*, 41, 277-304.
- Goodstein, D. L. (1995). After Big Crunch. *Wilson Quarterly Science*, Summer, 53-60.
- Guston, D. H. (2003). The Expanding Role of Peer Review Processes in the United States. In Shapira, P & Kuhlmann, S. (Eds.) *Learning from Science and Technology Policy Evaluation*. Edward Elgar Publishing.
- Hansson, F. (2006). Organizational Use of Evaluations: Governance and Control in Research Evaluation. *Evaluation*, 12(2), 159-178.
- Hargens, L. L. & Herting, J. R. (1990). Anew Approach to Referee Assessments of Manuscripts. *Social Science Research*, 19, 1-16.
- Jiemenz-Contreras, E., Anegon, D. & Lopez-Cozar, E. D. (2003). The Evolution of Research Activity in Spain. *Research Policy*, 32(1), 123-142.
- Knorr-Cetina, K. D. (1981). *The Manufacture of Knowledge*. Oxford: Pergamon Press.
- Marginson, S. & Considine, M. (2000). *The Enterprise University: Power, Governance and Reinvention in Australia*. UK: Cambridge University Press.
- Marshall, E. (1995). Suit Alleges Misuse of Peer Review. *Science*, 265, 1212-1213.
- Mitroff, I. I. & Cubin, D. E. (1979). Peer Review at the NSF: a Dialectical Policy Analysis. *Social Science of Science*, 9, 199-232.
- Moed, H. (2005). *Citation Analysis in Research Evaluation*. Netherland: Springer Publisher.
- OECD (2003). *Governance of Public Research: Country Case Studies*. Paris: Authers.
- Oppenheim, C. (1997). The Correlation between Citation Counts and the 1992 Research Assessment Exercirating for British Research in Genetics, Anatomy and Archaeology. *Journal of Documentation*, 53(5), 477-487.
- Quilan, K. M., Kane, M. & Trochim, W. M. (2008). Evaluation of Large Reseach Initiatives: Outcome, Challenges and Methodological Considerations. *New Directions for Evaluation*, 118, 61-72.
- Rennie, D., Knoll, E. & Flanagan, A. (1989). The International Congress on Peer Review in Biomedical Publication, *JAMA*, 272, 91.
- Riggs, J. E. (1995). Priority, Rivalry and Peer Review. *Journal of Child Neurology*, 10(3), 255-256.
- Slaughter, S. & Leslie, L. L. (1998). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore: The Johns Hopkins University Press.
- Sowers, A. E. (1995). Funding Research with NIH Grants. *The Scientist* 9(20), 12.
- Van Raan, A. F. J. (2005). Fatal Attraction: Conceptual and Methodological Problems in the Ranking for Universities by Bibliometric methods. *Sociometrics*, 62(1), 133-143.
- Vinkler, P. (1991). Magic Triangle for three Relative Impact Indicators. *Sociometrisc*, 21(1), 143-146.
- Warner, J. (2000). A Critical Review of Application of Citation Studies to Research Assessment Exercises. *Journal of Information Science*, 26(6), 453-460.
- Wessely, S. (1998). Peer Review of Grant Applications: What Do We Know? *The Lancet*, 352(25), 301-305.
- Weingart, P. (2005). Impact of Bibliometrics upon the Science System. *Scientometrics*, 62(1), 117-131.

A Review of Studies on Research Evaluation

Kunpei NISHIMURA *

Background

Studies on Research Evaluation are developed through interdisciplinary perspectives that include Higher Education Study, STS (Science technology and society), Scientometrics, Evaluation Study, Sociology of Science, and Philosophy of Science.

Objectives

To make available findings and theory about Research Evaluation in recent years.

Search Method

Retrospective review. That includes citation search on early review (Asonuma, 2006; Bornmann, 2008; Coryn & Scriven, 2008; Foltz, 2000; Hansson, 2006; Weingart, 2005), and database search such as JSTOR, and overviews of academic journals such as *Research Evaluation*, *Bulletin of Science technology and society*, *Scientometric*, *New directions for Evaluation*, *Social Studies on Science*, and so on.

Data Collection and Analysis

Literature were read, classified and described according to their topics and findings.

Main Result

Generally, findings and theory in STS, Scientometrics, Evaluation Study are based on the Sociology of Science. And recent Sociology of Science is deeply associated with the Philosophy of Science deeply and, in particular, the theme of the “paradoxical nature of science” that is to say, the process of Science is heavily affected by social context, but as a result Science bears un-social “truth”. At the end of this retrospective review, employing the concept of “paradoxical nature of science”, this study offers some prospective discussion about research question on research evaluation.

* Doctoral Student, Graduate School of Education, Hiroshima University; Research fellow, Japan Society for the Promotion of Science