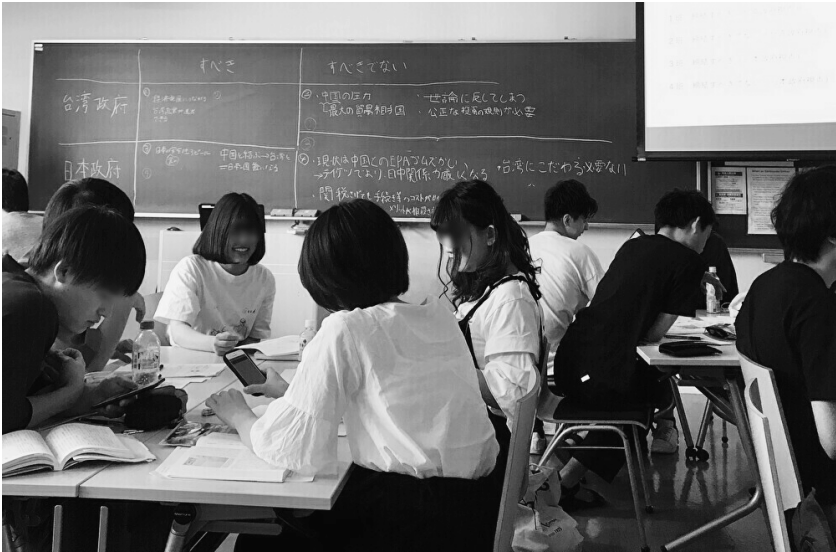


序 章

東アジアにおける人文・社会科学の研究評価制度 へのアプローチ

佐藤 幸人



日本の大学のゼミの様子（福田円氏撮影）

はじめに

近年、国の研究水準への関心が高まっている。たとえば世界大学ランキングは国の知識創造のレベルを表す指標として、多くの耳目を集めるようになった（石川 2016；ヘイゼルコーン 2018）。毎年、順位が報道されると、日本の大学の浮き沈みに、関係者ばかりか多くの人が一喜一憂している。2016年に発表された THE 社のランキングでは、日本の大学が大きく順位を落とし、ちょっとした騒ぎになったことは記憶に新しい。

世界大学ランキングの作成には、構成要素のひとつとして大学の研究に対する評価が用いられているが、その順位を研究水準を表すものとして過度に重視することには疑問や批判が投げかけられている。確かに日本の大学の研究における地位の下降は憂うべきことではあるが、上述の順位的大幅な低下についていえば、算定方法の変更によるところが大きく、それを知らずに振り回されれば滑稽なことになる。

このように研究評価には検討すべき多くの問題があり、議論が重ねられてきた。そのような問題のひとつが人文・社会科学における研究評価である。自然科学が強い普遍性を持ち、通常、特定の社会の文脈からは切り離されているのに対し、人文・社会科学は大学・研究機関および教員・研究者が帰属する社会と強く結び付いている。それゆえ、人文・社会科学の研究成果の発表ではそれぞれの社会の言語が使われることも多く、英語による発表が一般化している自然科学とは異なっている。また、後者では通常、研究成果が論文の形態で発表さるが、前者では図書として発表されることも少なくない。このような特性をもつ人文・社会科学を、グローバルに統一された尺度で評価することに対する疑問、あるいは批判が提起されているのである。

世界大学ランキングはグローバルな評価の仕組みだが、研究に対する評価を主として行っているのは各国の政府やその関連機関である。高等教育や科学技術の発展は多くの国にとって、現在、最も重要な課題のひとつであり、

研究評価はその発展を図る政策にとって不可欠の構成要素となっている。各国の人文・社会科学の評価においても、往々にして自然科学で用いられる普遍性をもつ数量的な指標が持ち込まれ、世界大学ランキングと同様の疑問や批判を招いている。

人文・社会科学の研究はどのように評価されることが適当なのだろうか。この問題は各国で検討されているが、相互の交流をとおしてそれぞれの議論や経験を共有することは必ずしも活発ではない。日本においても、他国の取組みが広く知られているわけではない。本書はこのような状況をふまえ、日本のほか、韓国、台湾、香港、中国といった東アジア各国・地域における人文・社会科学の研究評価制度とそれを構築する過程をみながら、この問題を考えてみようという試みである。わたしたちの隣の国や地域では、どのような制度によって人文・社会科学の研究が評価されているのか、それはどのような歴史的な背景をもち、どのような過程を経て現在に至ったのか、こういった論点を本書では議論している。

なぜ、東アジアなのかといえば、本書の問題意識の根底には日本に対する関心があるからである。日本と近隣の東アジアの各国・地域との間には共通性がある。第1に、長くイギリスに統治され、人口1000万人に満たない都市である香港はやや性格が異なるが、それ以外では英語以外の母語をもち、かつ一定の規模を有する学术界が形成され、その内部においても、社会の間でも、母語を使ってコミュニケーションをしている。この点はとくに人文・社会科学において顕著であり、評価制度のあり方を考えるうえでも重要である。第2に、現在の経済水準にちがいはあるとはいえ、キャッチアップ段階から卒業し、自前のイノベーションを強化しなくてはならないという課題をとともに抱えている。第3に、高等教育や科学技術をめぐるグローバルな競争が激しさを増しているが、東アジア各国・地域は日本にとって身近にいる直接的なライバルでもある。

日本はこのように東アジアの国・地域と共通の背景をもつ一方、その制度には日本にしかない特異な面を少なからずもっている。本書をとおして他の

東アジアの国・地域の制度を知ることは、日本の制度の特徴と課題に対する理解を深めることに役立つだろうと期待している。

なお、本書ではパースペクティブを広げる試みとして、東アジアのほかにイギリスとオーストラリア、オランダ、インドネシア、インドの事情を紹介するコラムも挿入している。イギリスは研究評価制度を先駆的に構築し、オーストラリアはそれを導入し、さらに独自の改良を加えている。コラムではそれを簡潔に紹介している。一方、オランダは研究評価においてプログラムベースという異なる方法を採用している。コラムではそれを説明するとともに、その社会科学に対する作用のケースも紹介している。インドネシアとインドのコラムでは、研究の量に偏重した評価制度の導入によって負の副作用が発生していることを指摘している。それによれば、インドネシアでは著者が600人あまりに及ぶ論文が生まれ、インドの社会科学では粗悪なジャーナルが氾濫する一方、東アジア同様、図書の刊行の衰退がみられる。

序章の残りの部分では、読者が第1章以降の国・地域別の各論に進む前に、その準備としてつぎの3つのことを行う。ここまで一口に評価と述べてきたが、それにはさまざまなものがある。第1節では、第1章以降を読む際の混乱を避けるため、どのような評価があるのかを整理する。第2節では、各国・地域の研究評価を理解するバックグラウンドとして、研究評価に対してどのような批判的な議論があるのかを提示する。第3節では第1章から第5章で述べられている東アジア各国・地域の評価制度の歴史や現状について、論点ごとにまとめる。これは各論を読み進むためのガイドマップであるが、各論を読んだ後に議論の整理に役立てていただいてもよいかもしれない。

第1節 評価の種類

研究に対する評価は、その成果である論文の査読から始まったが、今では多種多様な広がりを見せている（林2017）。本書ではそのうち、大学・研究

機関とその教員・研究員の評価に焦点を当てる。

第1章以降の各国・地域の研究評価をみれば明らかとなおり、大学・研究機関と教員・研究員の評価にもさまざまな種類がある。相違点としては、評価の主体と客体、すなわち誰が誰を評価しているのか、評価の目的、すなわち何のために評価を行うのか、評価の項目や方法、すなわち何をどのように評価するのかなどがある。

このようなポイントに着目して、本書でとりあげた東アジア各国・地域の経験を中心に各種の評価制度を整理したものが表序-1である。もちろん、表に示されているのは、多様な評価制度のすべてではない。以下ではこの表をもとに、どのような評価があるのかをより詳細に説明していきたい。評価の方法については、とくに数量的指標の使われ方に注目する。数量的指標とは論文の本数、論文が引用された件数、引用データなどからつくられるビブリオメトリクス（計量書誌学）指標などのことである。

まず、大学の認証とは、政府あるいは政府が認定した機関が、大学が大学としての要件を満たしているかどうかを評価することである。これは大学自身による自主的な点検と組み合わせられている。大学の主たる役割である教育と研究、それに社会サービスが評価される。研究評価の一部として、数量的指標が使われることもある。重要なことは認証の目的は大学の品質保証にあることである。したがって、すべての大学が合格することが目標であり、また、合格した大学の間でさらに序列化が行われるようなことはない。

同じような認証のための評価が学部・学科に対して行われることもある。その場合、それぞれの専門分野に即して評価が行われる。

つぎに政府が設けた大学および学部・学科向けの競争的資金を、どのように配分するのかを決めるために行う評価がある。競争的資金は有望な大学に傾斜的に配分されることから、資金をどこに配分するかを選定するため、評価によって大学は序列化される。各国が行っている世界のトップレベルの大学や学部・学科をめざすプログラムは、このような政策の典型である。ここで世界のトップレベルという場合、研究面に重きがおかれる場合が多く、研

表序-1 さまざまな評価

研究評価を 伴う制度	主体	客体	目的	項目	研究評価における 数量的指標の 使われ方
大学の認証	政府あるいはその 関連機関	大学	品質の保証	教育、研究、サー ビスに関する施 設や実績	使われることも ある
学部・学科の認 証	政府あるいはその 関連機関	学部・学科	品質の保証	教育、研究、サー ビスに関する施 設や実績	使われることも ある
大学および学 部・学科向けの 競争的資金	政府あるいはその 関連機関	大学あるいは学 部・学科	序列化して選別	研究に重点が置 かれている	非常に重視され ている
教員・研究員向 けの競争的資金	政府あるいはその 関連機関	教員および研究 員	序列化して選別	研究の実績と計 画内容	実績の評価に使 われることもある
政府の研究奨励 金	政府あるいはその 関連機関	教員および研究 員	序列化して選別	研究成果	重視されている こともある
教員・研究員の 評価	大学および学 部・学科	教員および研究 員	採用、契約の更 新、昇進などの 審査	教育、研究、サー ビスなど	重視されている
大学および学 部・学科の研究 奨励金	大学および学 部・学科	教員および研究 員	序列化して選別	研究成果	重視されている
国内の大学ラン キング	ランキング機 関。多くは民間	大学および学 部・学科	序列化し、主に 学生の選択の参 考にされる	教育面中心。ほ かに就職状況な ど	直接的には重視 されていないこ とが多い
世界大学ランキ ング	ランキング機 関。多くは民間	主に大学。最近 は学部・学科も	序列化。留學生 の選択の参考に されるなど多様	研究面に偏りが ち	重視されている

(出所) 筆者作成。

究の評価では通常、数量的指標が重視される。

表にはない方式として、大学が自主的に設定した目標の達成度を評価するという制度もある。これも政府の資金配分にかかわるが、大学間の競争性は弱くなる。

政府は大学の教員や研究機関の研究員が応募できる競争的資金も提供している。これも評価によって資金を供与する教員・研究員を選別することが必要となる。評価では研究計画の良しあしとともに、応募者の実績が問われることも多い。応募者の実績は数量的指標によって評価されることもある。また、政府が研究を奨励するため、優れた研究成果を表彰し、金銭を授与する

制度を設けている場合がある。これも報奨に値する研究成果を選び出すことが必要となる。何が優れた研究かを判断する基準として、数量的指標が用いられることがある。

大学とその学部・学科および研究機関は、上で述べたように政府やその関連機関によって評価されるが、同時にその内部においては所属する教員や研究員を評価する主体でもある。大学や研究機関は教員や研究員の採用、雇用の継続、昇進などを審査する。さらにいえば、そこで合格の基準を設定することによって、教員や研究員がその基準を満たすことを促す。

大学の教員評価の場合、主な評価項目は教育、研究、社会サービスである。ここで問題となるのは項目間のバランス、とくに教育と研究のバランスである。大学内部の教員評価は大学に対する認証評価や競争的資金の審査と連動している場合が少なくない。大学に対する評価で研究が重視され、研究評価で数量的指標に重点がおかれると、教員の評価でも研究の偏重、数量的指標の偏重が生まれやすい。なお、研究のみをおもな役割とする研究機関では、このような複数の役割の間の緊張は生じにくい。

大学や研究機関はより直接的に奨励金というインセンティブを教員・研究員に供与する場合がある。どのような研究を奨励するのかという基準として、数量的指標が重視されることは少なくない。

本章の冒頭でも述べたように、ランキングも評価を伴う。もともと、各国では学生が進学先を選ぶ際に参考とする国内の大学のランキングが作成されていた。このようなランキングでは必ずしも大学の研究面は重要視されない。むしろ卒業生の就職状況などが重要となる。

世界大学ランキングの歴史は存外に新しい。はじめての世界大学ランキングは2003年に上海交通大学が発表した「世界大学学術ランキング」である。翌2004年、THE社とQS社が共同で作成したランキングを発表した。2010年からはTHE社とQS社は別々にランキングを作成するようになった。その後、さまざまな世界大学ランキングがつくられている。

世界大学ランキングでは研究面が重視され、かつその評価では数量的指標

が重要な役割を果たしている。とくに論文の被引用件数は重要な指標となっていて、総数か教員 1 人当たりかというちがいがあったり、分野間の調整があったり、なかったりするが、大体のところ、評価の 2～3 割を占めている。このように数量的指標のウェイトが大きい世界大学ランキングが、各国の政策で重視されるようになると、その評価制度でも研究偏重、数量的指標偏重の傾向が生まれ、さらにそれが大学や研究機関における教員や研究員の評価にも反映されていくことになる。

第 2 節 適当な研究評価の模索

研究がどのように評価されるべきなのかについては多くの議論があり、今なお模索が続いている。とくに焦点となるのが研究評価における数量的指標の用い方である。本節では以下、まず数量的指標の長所と短所を示す。つぎに、あるべき研究評価をめぐる議論の国際的な組織化とその主張を紹介する。

2-1. 数量的指標の使い勝手のよさというメリット

研究を評価する主たる方法には数量的指標とピアレビューがある。ピアレビューとは同じ分野の研究者が研究成果である論文や図書に目をとおり、その質を評価することである。

一方、数量的指標とは、たとえば大学や研究機関が、あるいは教員や研究員が何本の論文を発表したか、そのうち何本が著名なジャーナルに掲載されたか、論文が他の論文などに引用された件数はいくつかといった、数量的に表される指標のことである。今述べたような書誌情報を用いたビブリオメトリクス指標のほかに、ソーシャルメディアの情報から作成されるオルトメトリクス指標がある（孫 2017）。数量的指標のなかでも、SSCI や A&HCI などのデータベースの収録ジャーナルに掲載された論文の本数、あるいはそのようなジャーナルのなかでも高いインパクトファクター（後述）をもつジャー

ナルに掲載された本数という指標が用いられることが多い。

数量的指標とピアレビューの長短は表裏の関係にある。ピアレビューは研究の内容を吟味するので、質の評価においては優れている。それに対して、数量的指標のメリットは以下で述べるような使い勝手のよさにある。

数量的指標の第1のメリットは一定の客観性である。それゆえに公正ともいえる。また、数値で表されることからわかりやすい。この点はピアレビューと対比すると明らかである。ピアレビューは大なり小なりレビュアーの主観が入ることが避けられない。そのために公正さを欠く恐れがある。また、レビュアーの評価は専門家以外からはわかりにくい。

数量的指標の第2のメリットは、数値化されていることから容易に研究を序列化できることである。ピアレビューが行うのは各分野の文脈に照らした絶対評価であり、個々の研究の間の優劣を明確に示すことには適していない。

第3のメリットは、数量的指標の方がピアレビューよりもコストが低いことである。データベースがあれば、数量的指標は簡単に算出できる。一方、ピアレビューは複数の研究者が時間をかけて行わなければならない。

第4に、多くの数量的指標は英語による発表に基づいているので、それぞれの国の学術的なコミュニティの規模を問わない。それに対して、ピアレビューは学術的なコミュニティが一定以上の規模をもっている必要がある。そもそも小さいコミュニティではレビュアーが見つからないかもしれない。

また、小さいコミュニティではピアレビューの公正性を保つことが難しい。ピアレビューの前提は評価する側と評価される側のあいだの匿名性にある。可能であれば、レビュアーには誰の研究かは知らされず、被評価者には誰が評価しているかは知らされないという、二重の匿名制が望ましい。被評価者を匿名にできない場合でも、できるかぎり利害関係のない人物を、レビュアーに選ぶ必要がある。しかしながら、小さなコミュニティではこの前提が容易に崩れてしまう。そうすると、人間関係上の好悪、しがらみ、パワーバランスや、限られたリソースやポストをめぐるストラテジックな配慮——たとえば教員評価のレビュアーが被評価者を低く評価することでそのポストにいら

れなくすれば、自分の親しい人間にチャンスを与えられると発想すること——が評価に持ち込まれ、公正性が損なわれてしまう。数量的指標にはこのような要素が入り込む余地がない。

さらに、たとえば上述のような SSCI 収録ジャーナル、なかでもインパクトファクターの高いジャーナルに掲載された本数が評価の指標として使われれば、そうしたジャーナルへの投稿を促進できる。SSCI 収録ジャーナルのほとんどは英語ジャーナルであり、インパクトファクターの高いジャーナルは国際的に著名であるから、このような指標を使って評価することによって国際化を進めることができる。

2-2. 数量的指標のデメリット

(1) 便利さが招く過度の依存

数量的指標のデメリットのひとつは、そのメリット——客観性、公正性、わかりやすさ、序列化機能、低コスト——ゆえに生じる。数量的指標は、それが的確に評価できるからということではなく、便利だからというテクニカルな理由によって、評価制度のなかで過度に重視されやすい。その結果、数量的な指標が適用しにくい研究以外の項目が軽視され、制度のバランスを歪めてしまうことである。

とくに大学の場合、評価における教育と研究のバランスという課題を抱えている。研究の評価には数量的指標を適用できるが、教育には適用しにくい。そのため、便利な数量的指標にたよることは研究偏重、教育軽視という問題を生じさせる。

それはまた、個々の大学や教員の個性を軽んじ、画一化に向かわせるということでもある。大学においても、個々の教員においても、教育と研究の間の比重のおき方には多様な選択があるが、評価において研究に偏重した数量的指標が用いられれば、教育面で高い評判を得ていた大学や教員も、研究に重点をおきなおさざるを得なくなる。

(2) 量の偏重と質の軽視

研究評価において数量的指標が重視されると、研究活動が質よりも量に傾斜するという問題が生じやすい。数量的指標でも研究の質を測ろうとはするが、すぐ後で述べるように限界があり、結果として評価制度では容易に測定できる量に比重がおかれがちになる。評価制度に対応して研究成果の量が重視されるようになると、その分、質が低下することになりかねない。

さらに、量の偏重は研究倫理に抵触しかねない行動も誘発する。本書のコラムにあるような、インドネシアにおける共著者が数百人に及ぶ論文や、インドにおける粗悪なジャーナルの氾濫は、そのような行動の典型である。研究倫理上、最も忌むべき行為である他の研究者の成果の剽窃やデータの捏造も、成果の量を求める結果として生まれることが多々ある。

(3) 引用件数とインパクトファクターの限界

数量的指標による評価の本質的な限界として、研究成果の質の測定という問題がある。数量的指標の場合、引用件数によって質を測ろうとする。すなわち、クオリティの高い研究成果は後続の多くの研究から引用されるはずであると想定している。このような想定に基づいて引用件数から論文等のクオリティを測定することは、まったくの間違いとはいえないが、引用には多様な目的があるため、引用の多いことが研究成果の重要性を示すとはかぎらない。たとえば先達への敬意を表すための引用や、否定や誤りを指摘するための引用は、引用した研究が重要であることを示すわけではない（逸村・安井 2006）。

さらに問題となるのが、研究の質の評価にインパクトファクターを用いることである。インパクトファクターはユージーン・ガーフィールドによって1972年に開発された、ジャーナルの影響力を測るビブリオメトリクス指標である。あるジャーナルに過去2年間、掲載された論文の被引用件数を、過去2年間にそのジャーナルに掲載された論文の本数で除したものである。WoSのデータを使って算出され、JCRに掲載されている。SSCIやA&HCIはWoSの一部である。

WoSに対抗するデータベースとしては、エルゼビア社によって開発され

たスコopus (Scopus) が有力である。WoS やスコopus は有料だが、近年は無料でアクセスできるグーグル・スカラー (Google Scholar) もプレゼンスを増している (孫 2012)。

インパクトファクターはジャーナルを評価する指標であり、ガーフィールド自身は個別の研究や研究者を評価することに用いてはならないと警告を発していた (Garfield 2005)。計量書誌学ではつぎのような問題が指摘されている (逸村・安井 2006; 逸村・池内 2013)。第 1 に、インパクトファクターはジャーナルの性格によって異なる。レビュー誌や総合誌は高くなりやすい。また、自誌からの引用件数はジャーナルによって異なる。第 2 に、分野によって引用のしかたが異なり、その結果、インパクトファクターも異なる。たとえば短期間に集中的に引用が行われる分野と、長期間、継続的に引用が行われる分野がある。また、基礎研究よりも応用研究、開発研究の方が引用件数、被引用件数とも多い。したがって、インパクトファクターを使って、異なる分野の研究の比較はできない。第 3 に、インパクトファクターは必ずしも安定的ではない。一時的な要因によって大きく変動することもあるし、長期的に変化することもある。

そして何よりも重要なことは、引用された件数の多い論文が必ずしもインパクトファクターの高いジャーナルに掲載されているわけではないということである。一方、インパクトファクターの高いジャーナルでも、多くの場合、被引用件数は一部の論文に集中している。つまり、インパクトファクターの高いジャーナルに掲載されている論文の多くは限られた数の引用しかされていない。このように、ジャーナルのインパクトファクターをもとに、掲載された論文のクオリティを評価することは、インパクトファクターの誤用なのである。そもそも中身をみることなしに論文の評価をしようとするところに無理がある。しかしながら、考案者のガーフィールドの警鐘をよそに、インパクトファクターの誤用は続いているばかりか、「インパクトファクター至上主義」(孫 2017) ともいえるような風潮の広がりがみられる。

2-3. 人文・社会科学の研究評価における数量的指標のデメリット

客観性やわかりやすさといったメリットが最も顕著に発揮される数量的指標は、ジャーナル論文の本数、なかでも SSCI や A&HCI などのデータベースの収録ジャーナルに掲載された論文本数であり、多くの評価制度で重用されている。しかしながら、このような指標が重視されることは、人文・社会科学に対して深刻な影響を及ぼす恐れがある。

第 1 に、ジャーナル論文中心の評価制度では、通常、限られた時間のうちに一定数以上の論文をジャーナル上に発表することが求められるため、長い時間を要する研究やリスクの高い研究を行うことが難しくなる。研究成果の発表が間に合わなければ、あるいは期待した結果が得られず、成果を発表できなければ、評価が低くなり、大学や研究機関ならば研究資金などの配分を減らされ、教員や研究員ならば、最悪の場合、職を失うことになるからである。ジャーナル論文中心の評価制度のもとで合理的な行動は、たとえ研究の価値は見劣りしても、所定の期間内に確実に成果を得られる研究課題を選ぶことである。人文・社会科学の研究には時間のかかる研究や不確実性の高い研究が少なくないが、そのような研究はたとえ高い価値があったとしても避けられることになる。

第 2 に、人文・社会科学の多くの分野においては、図書はジャーナル論文と並ぶ、あるいはそれ以上に重要な研究成果の発表形態であるが、ジャーナル論文中心の評価制度においては図書を出版することが難しくなる。ひとつには、研究成果が図書として結実するには長い時間を要するが、第 1 点で述べたように、ジャーナル論文中心の評価制度では長期に及ぶ研究が許容されにくいからである。また、一般に図書の査読体制はジャーナルほど整備されていないことから、平均をみた場合、研究成果としての図書のクオリティはジャーナル論文よりも劣ると考えられている。そのため、図書に対する評価は、その内容の良しあしにかかわらず、相対的に低く設定され、多くの場合、投入される時間や心血に見合うものではない。

第 3 に、評価制度が SSCI や A&HCI といったデータベースを用いる場合、

人文・社会科学がもつそれぞれの社会とのつながりとの間に矛盾が生じる。学界と社会のインタラクションにはそれぞれの言葉の方が適しているが、SSCIやA&HCIに収録されているのは、大部分が欧米で発行されている英語ジャーナルである。また、欧米のジャーナルの関心と、各国の社会にとって重要な研究課題は一致しないかもしれない。したがって、評価制度においてSSCIやA&HCIの収録ジャーナルを重視することになれば、欧米のジャーナルの興味に合うような研究課題が優先的に選ばれ、成果は英語で書かれる一方、それぞれの社会にとって喫緊の問題を研究し、その成果を社会にフィードバックすることは二の次になるかもしれない。

このような数量的指標のデメリットは事前には十分に理解されにくい。あるいは、被評価者である教員や研究者が敏感に察知しても、評価制度の構築の主導権をもつ評価する側はそれほど深刻には考えないことが多い。その結果、往々にしてメリットばかりが着目され、数量的指標が導入されることになる。

2-4. サンフランシスコ宣言とライデン声明

新世紀に入って、研究評価は普及し、数量的指標が広く用いられるようになる一方、それに対する疑問や批判も増大していった。2010年代になると、批判的な声が国際的に組織化されるようになった。代表的な試みとして、2012年の「研究評価に関するサンフランシスコ宣言（San Francisco Declaration on Research Assessment: Putting Science into the Assessment of Research）」（以下、「サンフランシスコ宣言」）¹⁾と、2015年の「研究計量に関するライデン声明（The Leiden Manifesto for Research Metrics）」（以下、「ライデン声明」）²⁾をみてみたい。サンフランシスコ宣言の数量的指標に対する批判は科学全般に当てはまるの

1) <https://sfdora.org/>（2019年2月20日アクセス）。

2) <http://www.leidenmanifesto.org/>（2019年2月20日アクセス）。ウェブサイトには和訳も掲載されている。なお、小野寺・伊神（2016）ではライデン声明の誕生の経緯も説明している。

に対し、ライデン声明のなかには人文・社会科学にとってとくに重要な指摘が含まれている。

ASCB および学術ジャーナルの出版社と編集者のグループは、2012 年 12 月 16 日、ASCB の年次総会の際に集まり、サンフランシスコ宣言を起草し、支持の呼びかけを始めた。2016 年 3 月 12 日の時点で支持を表明している個人は 13716 人、団体は 1173 である。

サンフランシスコ宣言はまずその主旨として、研究のアウトプットが多様であること、そのすべてが適正に評価される必要があると訴えている。アウトプットの多様性とは、アウトプットには論文だけではなく、データ、試薬、ソフトウェア、知的財産、さらには若手研究者までが含まれるという意味である。つぎに、インパクトファクターに対する批判を行っている。

続いてサンフランシスコ宣言は勧告を行っている。その中核的なポイントは、①インパクトファクターなどジャーナル・ベースの数量的指標を、研究に対する助成、研究者の採用や昇進の検討に用いることを排除する必要がある、②研究を掲載されたジャーナルによって評価するのではなく、研究自体を評価する必要がある、③オンライン出版がもたらす機会を利用する必要があるという 3 点である。サンフランシスコ宣言はこの 3 つの基本的なポイントに基づいて、一般的な勧告および研究助成機関、学術機関、出版社、数量的指標を提供する機関、研究者それぞれに対する勧告を行っている。

ライデン声明は 2014 年の科学技術指標に関する国際コンファレンスでのダイアナ・ヒックスの基調講演を起点として策定が始まり、2015 年に『ネイチャー (Nature)』に発表された。それは「研究評価における計量データの利用についてのベストプラクティスや注意点」(小野寺・伊神 2016) である。

ライデン声明は 10 の原則から構成されている。そのうち、本書にとって最も重要なのは「原則 3 優れた地域的研究を保護せよ」である。この原則は人文・社会科学におけるそれぞれの社会とのインタラクションの重要性を指摘し、研究評価でインパクトファクターを重視する結果、研究者を英語ジャーナルへの投稿に駆り立て、社会とのつながりを弱めることを戒めている。

このほか、本章ですでに述べてきたこととも重なるが、「原則 1 定量的評価は、専門家による定性的評定の支援に用いるべきである」、「原則 6 分野により発表と引用の慣行は異なることに留意せよ」、「原則 9 評定と指標のシステム全体への効果を認識せよ」も重要である。原則 1 は評価においてはピアレビューを主、数量的指標を従とすべきだとしている。原則 6 は評価がインパクトファクターに偏ることで、研究分野の特性を蔑ろにし、図書など論文以外の発表形態を軽視してはならないとしている。原則 9 は、評価制度の数量的指標の偏重によって、粗悪なジャーナルが横行したり、共著者が数百人に及ぶ論文が生まれたりすることがありえることに注意を喚起している。

第 3 節 東アジアの評価制度および日本の特徴と課題

本書の第 1 章から第 5 章では、それぞれ韓国、台湾、香港、中国、日本の研究評価制度を検討している。本節ではそれをふまえて東アジア各国・地域の共通点と相違点を、研究評価制度の導入の背景、研究評価制度とその作用、引用索引データベースの構築という 3 点に分けて提示する。さらに最後の項では、日本の特徴を改めてまとめ、そこから引き出される課題を提示する。本節は第 1 章以降を読み進む前に読めば、事前のガイダンスになるであろうし、事後に読めば議論の整理に役立つだろう。

3-1. 研究評価制度の導入の背景

東アジア各国・地域において、研究評価を含む大学・研究機関の評価が始められた背景として、2つの要因がある。ひとつは大学の増加である。これは韓国、台湾、香港、中国で明瞭に観察される。このうち前の3つの国・地域において、大学の増加は民主化と関連していると考えられる。それが最も明確なのは台湾である。台湾では1980年代後半に民主化がスタートし、1990年代以降、教育政策が変わり、大学数が増加した。香港では植民地統

治下にあったが、1984年に中国への返還が決まると一定の民主主義的な改革が行われ、そのなかで高等教育の拡大も進められた。韓国においては1980年代、民主化にやや先行しながら大学の大衆化が進行した。これは民主化と同じ社会の力が大学の大衆化にも働いたとみることができる。

大学の増加は大学の平均的な質の低下を招いたり、大学として必要な水準を満たせない大学を生み出したりする可能性がある。それを防ぐため、大学に対する認証評価が導入されることになった。韓国、台湾、中国の場合、大学の品質を保証する認証制度がつくられた。香港におけるUGCによるRAEという研究評価の実施も、大学の増加による質の低下に対する懸念が背景にある。

もうひとつの要因は国家間で科学技術をめぐる競争が激しくなり、その担い手である大学や研究機関を強化しようという世界的な潮流である。このような潮流のもと、東アジアでは世界でトップレベルの大学や研究機関を育成し、国の科学技術の水準を引き上げようとする政策を実施するようになった。韓国の「BK21」、台湾の「一流大学とトップレベルの研究センターを育成する計画（5年500億元計画）」、中国の「211プロジェクト」、「985プロジェクト」、「双一流（一流大学・一流学科）」政策、日本の「21世紀COEプログラム」や「グローバルCOEプログラム」がそれに当たる。香港は比較的、このような潮流の外に在るものの、そのRAEには「世界水準の研究を促進するため」という目的が加えられていることから、まったく無関心というわけではないようである。この種の政策では少数の大学や研究機関に集中的に資金を投入するため、その選別を行う評価が必要となったのである。

3-2. 研究評価制度とその作用

研究評価制度、とくに教員・研究員の研究評価の制度に注目するならば、ジャーナル論文が重視されているケースが多い。その分、図書は低く評価される。どこでも図書はよくて論文2本分にしかカウントされない。真面目に本を出版する場合、このような評価は費やされる時間や労力に見合うもので

はない。

SSCI 等のデータベースあるいはそのデータを使って算出されるインパクトファクターなどの指標を評価の基準として重視する傾向は、韓国、台湾、日本にみられる。韓国や台湾では SSCI 等に収録されたジャーナルに掲載された論文は、それぞれの国内のジャーナルに掲載された論文よりも高く評価される。

興味深いのは中国のケースである。中国においても SSCI 等に収録されたジャーナルに掲載された論文は高い評点が与えられる傾向があるが、『中国社会科学』のような中国国内のジャーナルに掲載された論文が、それ以上に高く評価されている場合がある。前述のように、SSCI などを過度に重視すると、人文・社会科学がそれぞれの社会から遊離してしまうという問題がある。それを防ごうという試みなのかもしれない。

では、東アジア各国・地域において研究評価制度はどのような作用を及ぼしてきたのか。一面では表序-2 に示すように、1990 年代から 2010 年代にかけて、韓国、台湾、中国の WoS 収録のジャーナルに掲載される社会科学の論文数が顕著に増加し、順位を上げ、シェアを拡大している。一方、日本の論文数は増え、シェアは拡大しているものの、順位は下降している。表にはない香港については、人文・社会科学にかぎった長期的な趨勢は不明だが、全般的な傾向からすると、研究成果は 1990 年代に増加したが、2000 年代に入ると横ばいになったとみられる。

表序-2 SSCI 収録ジャーナルに掲載された論文の本数・シェア・順位

A 経済学・経営学

	1994～1996 年 (平均)	2004～2006 年 (平均)	2014～2016 年 (平均)
論文数	136	265	565
日本 シェア (%)	1.3	1.9	2.1
順位	10	11	15
論文数	57	185	734
韓国 シェア (%)	0.5	1.3	2.7
順位	20	16	12
論文数	50	202	754
台湾 シェア (%)	0.5	1.5	2.7
順位	25	14	11
論文数	113	354	2,229
中国 シェア (%)	1.1	2.6	8.4
順位	12	9	9

B 社会科学・一般

	1994～1996 年 (平均)	2004～2006 年 (平均)	2014～2016 年 (平均)
論文数	188	343	868
日本 シェア (%)	0.6	0.9	1.0
順位	14	15	24
論文数		216	1,372
韓国 シェア (%)	圏外	0.5	1.5
順位		25	15
論文数		280	1,184
台湾 シェア (%)	圏外	0.7	1.3
順位		21	20
論文数	206	543	3,503
中国 シェア (%)	0.6	1.4	4.0
順位	11	9	7

(出所) 科学技術・学術政策研究所 (2018) をもとに作成。

もちろん、韓国、台湾、中国の WoS 収録ジャーナルに掲載される社会科学の論文数の増加がどの程度、評価制度によるものかどうかは特定できない。とはいえ、ひとつの要因であると推測することはあながち的外れではないだろう。

しかしながら、すでに述べたように、数量的指標に依存した評価制度の諸

問題は、このような研究成果の増加と表裏一体である。数量的指標偏重の副作用のうち、比較的、明瞭に現れているのは図書の出版の停滞あるいは減少である。台湾、香港、中国で観察されている。一方、韓国、台湾、中国では数量的指標への過度の依存を反省し、ピアレビューを重視しようという動きがある。しかし、いったん成立した制度は強い粘性をもつので、改めることは容易ではない。

3-3. 引用索引データベースの構築

韓国、台湾、中国では研究評価制度と並行して、引用索引データベースの構築が進められてきた。時間的にもほぼ同時である。韓国では1998年に「学術ジャーナル登録制度」がつくられ、登録された文献の情報がKCIに収められている。台湾では1990年代後半に、TSSCIやTHCIの公開が始まった。中国では1998年に、南京大学によってCSSCIが開発された。

これら引用索引データベースは、意図した結果か、意図せざる結果かはさまざまであるが、評価制度に組み込まれていった。それぞれの研究評価において、自国で発行されるジャーナルについてSSCIなどの役割を担わされることになったのである。言い換えれば、これら引用索引データベースのおかげで、それぞれの評価制度は自国で発行されるジャーナルをカバーすることが可能になったのである。

したがって、これら引用索引データベースはSSCIなどを重視する評価制度の英語偏重を是正する効果をもっていた。しかしながら、同時にデータベースに依存した評価を自国のジャーナルにも広げるという副作用も伴っていた。台湾では副作用に対する批判が強まり、TSSCIはTHCIと統合され、評価から切り離された新しいデータベースがつくられることになった。

3-4. 日本の特徴と課題

冒頭で述べたように、本書の問題意識の根底には日本への関心がある。近隣の国・地域と比べた時、日本の研究評価の特徴は何であろうか。本節の最

後に、本書から浮かび上がった日本の特異性について言及しておきたい。

第1に、表序-2に示したように、東アジアにおいて日本の社会科学のみ、国際的なプレゼンスの停滞ないし低下がみられる。原因のひとつとして考えられる相違点は、日本の人文・社会科学では教員や研究者の評価において数量的指標が相対的に強調されてこなかったことである。つまり、日本の教員や研究者はジャーナル論文の発表、とくに英語論文の発表、さらにはSSCI等のデータベース収録ジャーナルへの発表が、他の東アジア各国・地域と比べて強く求められていない。その分、日本は他の東アジアの国・地域で発生したような副作用、すなわち研究の偏重と教育の軽視、論文の偏重と図書等の軽視、英語での発表の偏重と母語を含む非英語での発表の軽視が、深刻ではないようにもみえる。

第2に、どの国・地域でも評価制度は大なり小なり複数の政策の複合体である。とはいえ、そのなかにあって日本の国立大学法人評価はやや特異にみえる。さらに探求する必要があるが、本書の各章を読むかぎり、他の東アジア各国・地域には類似の制度はみられない。また、評価に多大なリソースが投入される結果、生じる評価疲れは、他の国・地域にもみられるが、日本は国立大学法人評価によっていっそう深刻になっている恐れがある。評価疲れは大学のリソースを圧迫することで、上述のような研究上の国際的なプレゼンスの低下の原因になっているかもしれない。

第3に日本の人文・社会科学のみが引用索引データベースを構築していない。香港も独自の引用索引データベースをもたないが、規模が小さいこと、中国語文献に関しては近隣に中国および台湾があつて、それぞれデータベースを構築していることから、自前のデータベースの必要性は感じられなかったのだろう。しかし、日本の条件は香港とは異なる。日本の人文・社会科学の規模は大きく、日本以外で行われている日本語による研究活動は限られている。上述のように、教員や研究員を数量的指標によって評価しようとする傾向が弱かったため、データベースを構築する動機づけが欠けていたのかもしれない。

以上の日本の特徴は取り組むべき課題につながっている。第1に、日本の人文・社会科学のプレゼンスの低下は対策を講じる必要があるだろう。その際に数量的指標の活用が対策として浮上するかもしれない。しかしながら、本書が示すように、東アジア各国・地域の経験から、闇雲に飴と鞭を使ってSSCI等収録ジャーナルに英語論文を発表することを促したり、さらには強要したりすれば、人文・社会科学においては深刻な副作用が生じる。プレゼンスの向上をめざすにしても、バランスがとれた取組みでなくてはならない。

第2に、日本の評価制度には特異なところがあり、それが他の国・地域よりも深刻な評価疲れを生んでいる可能性がある。それは速やかに改善すべきである。制度を改めることによって大学がリソースをより有効に使えるようになれば、そのパフォーマンスを向上させることができ、第1の課題に対する対策のひとつになり得る。言い換えれば、大学のパフォーマンスをいかに引き上げるかという観点から、評価制度は構築されるべきである。

第3に、日本でも人文・社会科学の引用索引データベースを構築すべきである。引用索引データベースは研究者をはじめとする関係者に有用な研究情報を提供し、研究の発展を促すことができる。ただし、本書に述べられている東アジア各国・地域の経験が示すように、引用索引データベースは評価制度と密接な関係をもちやすく、その部分で問題を生む恐れがある。日本において引用索引データベースが構築された場合も、評価に用いられる可能性がある。そのとき、東アジアの経験を参照することで、誤った用いられ方を避けることができるだろう。

おわりに

前節で述べたように、日本の人文・社会科学において、教員や研究者に対するこれまでの評価制度では数量的指標のウェイトは限定的だったと考えられる。そのおかげで数量的指標が深刻な副作用を生み出すことも避けられて

きた。しかしながら、日本のプレゼンスを高めようと、今後は数量的指標がより重視されるようになるかもしれない。

そのとき、もし東アジア各国・地域の経験を学ぶことなく数量的指標への依存を増大させれば、同じ轍を踏むことになりかねない。編者としては、日本の人文・社会科学の評価制度がそのような愚かな道に進むことのないように、東アジアの経験から教訓を引き出し、提示することが、本書の重要な役割だと考えている。

〔参考文献〕

<日本語文献>

- 石川真由美編 2016. 『世界大学ランキングと知の序列化——大学評価と国際競争を問う——』 京都大学学術出版会.
- 逸村裕・池内有為 2013. 「インパクトファクターの功罪——科学者社会に与えた影響とそこから生まれた歪み——」『化学』 68 (12) : 32-36.
- 逸村裕・安井裕美子 2006. 「インパクトファクター——研究評価と学術雑誌——」『名古屋高等教育研究』 (6) : 131-144.
- 小野寺夏生・伊神正貫 2016. 「研究計量に関するライデン声明について」『STI Horizon』 2 (4) : 35-39.
- 科学技術・学術政策研究所 2018. 「科学技術指標 2018」 <http://www.nistep.go.jp/research/science-and-technology-indicators-and-scientometrics/indicators> (2019 年 2 月 21 日アクセス).
- 孫媛 2012. 「ビブリオメトリックスを活用した研究評価の現状と展望」第 1 回 SPARC Japan セミナー (2012 年 5 月 25 日) https://www.nii.ac.jp/sparc/event/2012/pdf/20120525_doc3.pdf (2019 年 9 月 2 日アクセス).
- 2017. 「研究評価のための指標——その現状と展望——」『情報の科学と技術』 67 (4) : 179-184.
- 林隆之 2017. 「研究評価の拡大と評価指標の多様化」『情報の科学と技術』 67 (4) : 158-163.
- ヘイゼルコーン, エレン 2018. 永田雅啓 アクセル・カーペンシュタイン訳 『グローバル・ランキングと高等教育の再構築——世界クラスの大学をめざす熾烈な競争——』 学文社.

< 英語文献 >

Garfield, Eugene 2005. "The Agony and the Ecstasy. The History and Meaning of the Journal Impact Factor." Paper presented to International Congress on Peer Review and Biomedical Publication, Chicago, 16 September.