

Title	大学の研究成果による社会・経済・文化的インパクトの評価手法の標準化
Author(s)	林, 隆之
Citation	年次学術大会講演要旨集, 26: 815-818
Issue Date	2011-10-15
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10240
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

大学の研究成果による社会・経済・文化的インパクトの評価手法の標準化

○林 隆之（大学評価・学位授与機構）

1. はじめに

大学が実施する研究活動にも、学術的な知識の増進だけでなく、イノベーションや社会・経済的課題の解決が明示的に期待されるようになってきている。しかし、大学が主に行う基礎研究が、社会・経済・文化面にもたらした効果やインパクトを測定することは容易ではない。その理由は、インパクトの発現までに長期かかることや、一つの研究成果から複数の種類のインパクトがもたらされ、一つのインパクトに影響する要因も当該研究以外に多数あるなど、研究とインパクトに複雑な関係が存在すること等があげられる。そのため、研究開発プログラムの追跡評価などでは、しばしば、詳細なケーススタディを実施する方法がとられ、イノベーションの発現メカニズムの解明と、過去の事実を踏まえたプログラムマネジメントの改善が行われる。

一方、機関や組織を対象とする大学評価においても、社会・経済・文化面への効果やインパクトを評価することへの要請は国際的に高まっている傾向にあるが、大学評価やその結果に基づくブロックグラントの傾斜配分においては、詳細なケーススタディ手法はとりにくい。なぜならば、第一には、ケーススタディのコストが高く、国内の全ての大学（あるいは国立大学に限定しても）の研究活動を対象とする評価において実施するのは費用が高額になる。第二に、大学を一斉に評価する場合には、ある程度の比較可能性を担保することにより、学部・学科などの組織レベルでの特性を明確にし、その結果を組織のマネジメントや組織文化の改善や外部からの資金配分に用いることが求められる。そのため、個別的なケーススタディでは機能しにくい。

このような理由から、詳細なケーススタディ手法ではなく、様々な研究活動を対象としうる包括的かつライトタッチな標準的な評価方法を模索する必要がある。しかし、実際には人文学・社会科学から医学・工学などの多様な分野が対象となる中で、いかなる手法や指標を設定すべきかは不明であり、各国で様々な試行が開始され、その長短所を分析しなければならない状況にある。

本発表では、海外の大学評価におけるインパクト評価の現状を概観したうえで、日本の国立大学法人評価の中で行われた研究業績判定を対象にして、特に、ピアレビューアーがどのような根拠データ・指標を評価において重視したかに焦点をおいた分析を行う。

2. 大学評価における基礎研究のインパクト把握の国際的状況

大学の研究評価は、各国で近年見直しが行われてきた状況にある（林 2009）。その一つの視点は、学術的な質だけでなく、社会・経済・文化的なインパクトを評価基準にいかに取り込むか、であった。

英国では、20 年以上にわたって行われてきた大学の研究評価である RAE を廃止し、新たな REF を 2014 年から実施するための設計が最終段階にある。その見直しは「RAE が大学で行われている多種多様な研究を把握していることに失敗している」という指摘から始まり、その後に様々な議論を経て、新たな設計では「全ての種類の優れた研究と、優れた研究から発生する影響を認定すること」を評価の基本枠組みとしている。評価は、①研究のアウトプットの質、②インパクト、③研究環境の 3 項目から行われる。インパクトは大学の各 UoA（学科相当）から提出される簡素な「ケーススタディ」書類をピアレビューすることにより実施される計画であり、パイロット調査が行われた。ケーススタディには、研究内容とその質の証拠、インパクトを生んでいる方法、インパクトの性質・範囲、それを確認できる情報ソース、を数頁で記述するものである。これらを踏まえて、評価者は「reach」と「significance」の二つの観点から 4 段階の判定を行う。

また、豪州では、前政権のもと 2004 年に提案された「RQF」において、インパクトの評価が新たに強調されたが、新政権のもとで「設計が悪く、インパクトの定義が不明」とされ実現に至らなかった。ただし、その提案は英国 REF に反映されているとみられる。RQF では、研究の質とインパクトの 2 つを評価し、後者は大学から提出された「インパクトステートメント」をもとに 5 段階で判定を行う方法であった。新政権のもとで現在実施されている ERA ではこのような方法はとられず、各種の指標の中

に「応用」として特許や商業化収入が入っているのみである。

ケースを簡素に記述させる方法に対して、オランダは研究プログラム（研究グループに相当）の評価基準を①質、②生産性、③レレバンス、④活力と可能性、としており、全ての研究プログラムが4・5頁の自己評価書を記述し、ピアレビューアーが5段階で評価する方法がとられている。また、評価手法については KNAW の ERiC プロジェクトにおいて、各種の指標による特性分析、ステークホルダー分析などの手法が開発されている段階にある。

以上のように大学評価におけるインパクトの評価は、個々の研究課題による学術面以外のインパクトもピアレビューで簡素に段階判定する方法（modified peer review）と、組織を単位にその研究活動の特性を指標により分析する方法とに大きく分かれる。特に前者では、レビューアーが学術面でなくインパクト面についても適切に評価を行えるような客観的な根拠データや判断基準の設定が課題となる。

3. 日本の大学評価における研究のインパクトの評価：現在の方法と課題

日本の大学評価も、ピアレビューアーにインパクト面の評価も委ねる modified peer review である。日本で研究評価に該当するものは、認証評価の「選択的事項」と国立大学法人評価がある。両者において、研究成果の評価は学術的価値と社会・経済・文化的価値の二項目から行われる。以下では、全ての国立大学を対象とした国立大学法人評価（H20 年度実施分）での研究評価を分析の対象とする。

国立大学法人評価は、中期目標・計画の達成の確認によるアカウントビリティ確保、次期の計画策定等への活用、運営費交付金のうちのごく少額部分の再配分などを主な目的に行われた。評価では、大学単位の中期目標・計画の達成状況の評価とは別に、学部・研究科ごとに、研究と教育の現況分析（水準判定）が行われた。そのうちの研究の水準判定では、学部・研究科単位の「研究成果の状況」の評価を行うための一材料という位置づけで、研究業績判定が行われた。

研究業績判定では、各学部・研究科が組織を代表する優れた業績を、教員数の半分以上を上限に提出する。提出された業績は、科学研究費補助金の 278 の「細目」ごとにわけられ、一つの業績を 2 名のピアレビューアーが判定する。大学は、各業績について「学術的意義」「社会、経済、文化的意義」のどちらかの側面からの評価を望むかを申請する¹。ピアレビューでは、論文を改めて読むような一次的な評価は行わず、第三者による評価結果や客観的指標等の根拠を大学側が記述し、それを基に評価を行う、二次評価とも言えるライトタッチの方法がとられた。

判定は、「卓越(SS)」「優秀(S)」「それ以外（A（良好）、B（相応）、C（不十分）に相当）」の 3 段階で行われた。組織の全ての業績の質を評価するのではなく、優れた業績がどれほど得られているかに焦点を絞る方法である。判定基準は「卓越」「優秀」などの表現に留まり、その具体的説明は十分に形成されておらず、レビューアーに委ねられている。実態的には過去の「試行的評価」での SS、S の割合（それぞれ教員数の 5%、10%にあたる業績数）を提示することで、目安とされた。

図 1 には、実施方法の妥当性を検討するために、評価実施後に評価者へアンケート調査を実施した結果を示している。学術的意義については、評価者の 38%は、大学から提出された業績説明書に根拠が示されていたと認識しており、55%は判断をそれほど迷わずに行えたと回答している。他方、社会・経済・文化的意義については、13%のみが根拠が示されていたと答え、41%のみが迷わず判断できたとしている。すなわち、社会・経済・文化的意義は、大学から示された根拠も不十分で、判断しづらかったという結果である。なお、学問分野によって有意な差は見られない²。

同様の結果は大学からも得られている。これらの結果は、社会・経済・

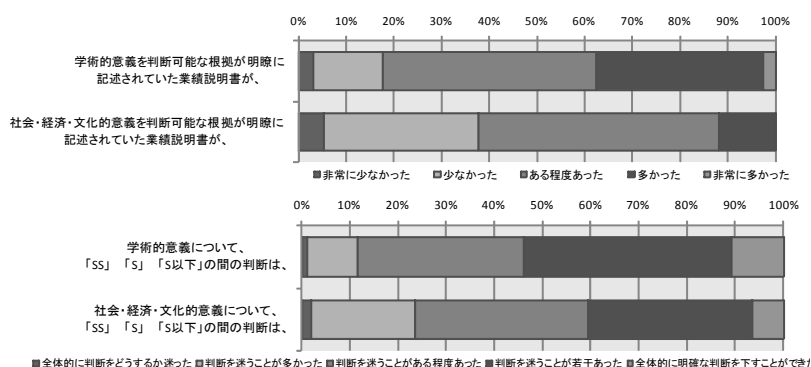


図 1 評価後のアンケートに対する評価者からの回答

¹ パスツールの 4 象限の議論(Stokes 1997)に代表されるように、学術的価値と社会・経済・文化的価値を独立した項目として区分することは、医学や工学の実態には適さない。一方で、両者を選択可能としていた「試行的評価」では、学術的価値の主張に、未だ実現していない期待される将来的な社会・経済・文化的価値が主張される場合が多かったため、方法論的に独立した項目とした経緯がある。

² 科研費の「分野」を因子とした一元配置分散分析の結果から。

文化的インパクトを示す根拠データ・指標とはどのようなものであるか、それに対する SS,S 等の判断基準はどのようなものかについての共通理解が研究者コミュニティの間に潜在的なものとしても存在していない状況を示している。そのため、行われた評価において、ピアレビューアーがどのような根拠や指標をもとに提出業績が優れていると判定したか、その判断を蓄積し分析することにより、求める根拠やその判断基準をある程度の標準性を持って形成していく必要がある。

4. ピアレビューによる研究業績判定結果の分析

(1) 社会・経済・文化的意義の提出量と判定

まずは上記のような研究業績判定の方法によって、実際にどの程度の業績が提出され、適切に判定が行われたのかを概観する。表 1 には提出数を示す。社会・経済・文化的意義への提出は提出数の 14.6% を占め、高い割合ではない。科学研究費補助金の「分野」で細分すると、申請割合が最も高いのは人文科学、社会科学であった。社会・経済的効果の通常の議論では、工学や医学などが典型的な事例として取り上げられるが、それとは異なる結果である。文化的意義を含むことで、後述するように一般書による啓蒙や芸術分野での公演などの貢献が多く提出される結果となっている。

表 1 社会・経済・文化的意義の提出量

総提出数	学術的意義	社会・経済・文化的意義
20,246	17,297 (85.4%)	2,949 (14.6%)

一方、ピアレビューの結果としては、学術的意義では提出された業績の 8 割弱が SS ないし S に認められたが、社会・経済・文化的意義では 7 割弱と下がり、提出する側（大学）と評価者との間での認識の一致度がより低い。分野別では、提出数が少なかった数物系で 8 割以上が SS,S として認められており高いが、それ以外は提出数が多い医歯薬学分野を除き、65～75%になっている。一方、評価者側の判定の整合性という点では、SS、S、S 未満の 3 段階について、同じ判定結果となったのが学術的意義で 59.0%、社会・経済・文化的意義で 56.3%であり、大きな差異はない。この値は「試行的評価」の理学において、論文自体を評価者に配布し、より細かい判定基準を設定して実施した場合の評価結果(一致が 61.5%)とも大きく変わらない（林 2003）。結果的には社会・経済・文化的意義でも判定が大きく揺れたという状況ではない。

表 2 研究業績判定作業における判定結果

	SS あるいは S と認められた割合	2 人の評価者による段階判定の差異		
		同じ	1 段階異なる	2 段階異なる
学術的意義	78.3%	59.0%	38.9%	2.0%
社会・経済・文化的意義	67.6%	56.3%	41.0%	2.7%

(なお、学部・研究科の「研究成果の水準」はこの業績判定結果のリストを一つの材料としながらも、その他の情報を含め、総合的に評価される)

(2) インパクトを実証する根拠の種類

社会・経済・文化的意義として提出された 2,949 件のうちで、2 名のピアレビューアーの双方が「卓越(SS)」と判定した業績 208 件（ただし、複数学部から同一の研究業績が提出されているため、ユニークには 174 件）を対象に、その判断に用いた根拠を分析する。2 名が統一して SS と判断したことで、ある程度の確度をもって優れた業績であると扱うことができる。評価者は大学の提出書類の文章に対して、SS と認める材料とした根拠データをマーカーで示すプロセスをとっており、その内容を分析した。

示された根拠は様々なものがあり、学問分野にも依存するが、図 2 のように一般化してその種類を区分できる。

図 2 は研究成果のユーザーへの到達と効果の発現のプロセスをフェーズ化したものであり、①計画や実施における共同から、②成果産出、③情報伝達、④利用、⑤それによる効果発現、のそれぞれに根拠データが考えられる。後ろのフェーズほどインパクトが実際に発現しており、高く評価できる。また、このような直線的な関係だけでなく、⑥次の活動への展開への円環的な移行を示す根拠も提示されている。さらに、これらを俯瞰する形でなされた⑦外部からの評価結果が根拠として示されている。

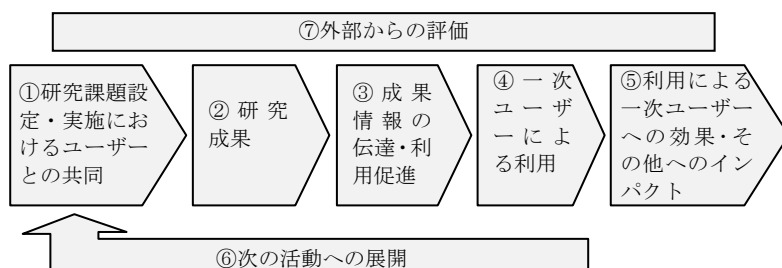


図 2 大学から提出されたインパクト根拠資料の分類

表3 社会・経済・文化面で「卓越（SS）」と判断された根拠（全分野）

	評価者が判断に用いた内容
①研究課題設定・実施におけるユーザーとの共同	・企業との共同研究実施・体制構築（例：地域クラスター形成） ・要請に基づく活動実施（例：芸術公演）
②研究成果の産出	・学術的業績数（例：国際論文の数、著名ジャーナル、インパクトファクター）* ・特許取得* ・技術内容・特性（例：新機能、新手法（診断方法、生産方法等）、長寿命化、長期継続調査、学際的活動の成果*など）、
③研究成果情報の伝達・利用促進	・情報提供活動の実施（例：ウェブサイトでの効果、研修会実施、展示会、公演、社会実験） ・開発活動への進展（例：外部資金の獲得、共同研究の申し込み） ・普及活動への展開（例：普及のための外部資金の獲得、地方自治体との共同、エージェント契約、書籍の翻訳、テレビ番組の作成） ・一般市民へのメディア報道（例：新聞、テレビ） ・政策形成への情報提供（例：審議会委員として発言） ・技術移転・特許許諾・ライセンス収入 ・ベンチャー設立・起業 ・標準化
④一次ユーザーによる利用	・多数の個人による利用（例：ソフトウェア・データベースのアクセス数、研修会への参加数） ・政府・地方自治体での活用（例：設備導入、ガイドライン採用、立法への影響） ・病院・医療介護施設での利用（例：新薬認可、治療手法の採用、薬剤の利用、電子システムの採用） ・民間企業による製品化 ・民間企業による直接利用（例：設備導入） ・一般書籍の出版部数 ・教育での利用（例：初中等教科書への収録、学習指導要領への影響、大学教育テキストへの採用、辞典での引用） ・学術的な利用（例：菌株の分譲数、論文引用数*、学術雑誌等のニュースでの取り上げ数*）
⑤利用による一次ユーザーへの効果、その他へのインパクト	・技術移転を受けた企業・ベンチャー企業の売上高、市場規模 ・開発された研究基盤を用いた更なる研究開発成果（例：ソフトウェアを用いた材料開発による特許登録、菌株の分譲による研究成果） ・使用者によるエネルギー効率の上昇
⑥次の研究活動への展開	・基金寄贈 ・企業・政府との共同研究への展開
⑦外部からの評価	・受賞（例：各種の大臣賞・地域自治体の賞、産学連携賞、ものづくり賞、発明賞、建築賞、デザイン賞、環境賞、芸術・音楽賞、出版・文化賞、学芸賞、雑誌・マスコミによる賞、科学教育・理解増進に係る賞、学会賞*など） ・書評（例：新聞、学会誌*など。評者名も併記）、公演の批評 ・招待講演・基調講演 ・外部評価結果（例：機関評価、プロジェクト評価）

注：件数は本稿においては省いている。なお1件の業績に対して複数の根拠があることをもってSSと判断されている場合も多い。

*左記記号の項目は通常は学術的価値と考えられる場合も多いものである。

表2にはこの分類のもとで、実際にどのような根拠があったか示している。実際に①から⑦に渡って様々な根拠が評価者により採用されているが、件数として最も多いのは⑦の各種の受賞(n=90)であり、書評(11)、招待講演(17)など、直接的なユーザーではない第三者により活動の評価された結果が多くを占めている。二次評価という手法による当然の結果でもあるが、その他の根拠を大学が示しにくい、評価者が自信を持って判断しにくい状況の現れとも見える。それに続いては、③の報道(41)、④の製品化(18)が多い。一方で、⑤の利用による効果やインパクトは数が少なく、最も多くて売上高(6)が根拠として採用されている。研究の社会的効果として期待されるような、具体的な社会課題の解決の実績（たとえば環境汚染物質の低減、疾病率の減少など）に係るデータは殆ど示されていない。

これらの結果は、第一には、将来可能性ではなく実際に顕在化した効果を根拠を基礎にして評価する場合には、⑦のような外部評価に加えて、③や④のような中間フェーズの指標が中心とならざるを得ないことを示している。それらの中でもどのようなデータ・指標が根拠となりうるかの共通理解を大学・評価者が構築することが今後必要となる。第二に、⑤のようなユーザーへの効果やその波及的インパクトは、研究者個人が調査をして示すことは難しく、大学や資金配分機関が詳細な調査を行うことでその結果を積極的に活用することや、実質的な効果が想定される領域（たとえば、環境、衛生、政策形成）を、顕在化した効果とは別に明確にすることにより、基礎研究の意義を別途示す必要性が示唆される。

5. おわりに

本稿ではインパクトを実証する根拠データ・指標にはどのようなものがあるかを、実際の大学評価のデータから概念化して整理した。一方で、各データ・指標がどのような状態であれば「卓越」したものと言えるのかの判断基準は、図2のような「効果発現への到達度」に加えて、「使用するユーザーの多さ」、個々のユーザーにとっての「重要性」、の3要素が評価に影響する要因と考えられる。今後、優れた業績とは認められなかった業績との比較を行うことで、それらの精緻化が別途必要となる。