

Title	研究成果指標の多様性と標準化：概念整理と実証的分析
Author(s)	林，隆之；藤光，智香；秦，佑輔；中渡瀬，秀一
Citation	年次学術大会講演要旨集，35：258-261
Issue Date	2020-10-31
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17343
Rights	本著作物は研究・イノベーション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 Research Policy and Innovation Management.
Description	一般講演要旨

研究成果指標の多様性と標準化：概念整理と実証的分析

○林隆之（政策研究大学院大学）、藤光智香，秦佑輔（文部科学省），中渡瀬秀一（国立情報学研究所）

1. はじめに¹

研究成果の測定は、行政機関においては大学等の研究実施組織に投資した公的資金に対する成果を確認し、その後の資金配分的意思決定を行ったり説明責任を果たすために行われ、研究実施組織内では戦略策定や組織内資源配分のために行われる。研究成果の測定において留意しなければならない点は、全分野に適応可能な単一の指標は存在しないにもかかわらず、資金配分などの重要な意思決定において限られた指標群が使われやすいことであり、それによって組織や研究者の行為に悪影響が及ぶ点である（Rijcke et al., 2016）。その典型的な事例は Web of Science や Scopus 等の論文データベースによる測定を人文学・社会科学分野に対して適応し、それにより研究や成果発表の行為が変わることである。

このような課題は以前から指摘されている。しかしながら一向に具体的な解決はみえていない。本稿ではまず、このような課題が現在、具体的にいかに表出しているかを確認し、議論が不成立となる構造を検討する。その上で、人文学・社会科学分野を含めて研究成果の測定において海外等で提案・実施されている方法の日本での実現可能性と課題を、歴史学を事例とした複数の分析から検討する。

2. 国内外の状況と概念整理

資金配分の場面において、論文の引用数に基づく測定を大学等の組織単位ごとに行うことは、日本では近年みられるようになってきている。例えば、2013 年の文部科学省「研究大学強化促進事業」では「ヒアリング対象機関選定のための指標」として Top10%論文数の割合など複数の指標を用い、2014 年の「スーパースター大学創成支援事業」の「トップ型」の申請では、論文の被引用状況などの指標を示した。さらに、2019 年度より国立大学の運営費交付金配分において共通指標による配分が開始され、5つの共通指標の一つが「運営費交付金等コスト当たり Top10%論文数」とされた。これに対して、国立大学協会は委員会を設置して適切な指標についての検討を始め、教育・研究実績を 11 学系ごとに測定する指標の設定を求める提言を行った。その中では、第三期中期目標期間の国立大学法人評価における研究業績の質に関するピアレビューがまだ行われていない状況であるために、研究成果については暫定的に「教員あたり研究業績数」等の定量指標のみで構成しながらも、複数の指標を学系（分野）別に設定するように求めた。結果、Top10%論文数の指標は残りながらも、教員あたり研究業績数の指標が設定された（林 2019）。

これまでも分野ごとの研究活動や成果の多様性を尊重する必要があるという指摘は度々なされている（日本学術会議 2008,2012）。分野によって研究方法や成果の発表形態は異なり、分野内部でも様々であるため、研究の厚みを増すためには、多様な評価基準や指標が不可欠となる。一方で、上記のように何らかの形で限りある資金の配分を行わざるを得ない際には、比較が不可能なほどに雑多な指標を設定しては、結果、活用が困難となる。実際に、上記の国立大学協会の委員会においても、人文学・社会科学においては書籍が主要な成果発表の形態であることは認識されていたが、では具体的にどのような書籍を「研究業績」「1 本」として測定するべきであるか、芸術学やスポーツ学におけるパフォーマンスを含めた、それ以外の業績を集計できる形でいかに測定できるかといった問題に具体的に直面した。

このことは、多様性と標準化のトレードオフは、測定や評価の結果をどのような目的で用いるのかを明確にしない限り解決しないことを示している。その前提が無い場合に、多様性を求める学术界と、説明責任を伴う意思決定が求められる行政機関の間ですれ違った議論が展開されることになる。表 1 には大きく 2 種類の利用目的における指標設定の考え方を整理した。この上で、本稿は上記のような、ある程度の標準性を要する場面に焦点をおいて、日本では人文学・社会科学を対象に含めた測定にどのような方策が取り得るかの検討を、歴史学を対象とした分析により行う。

¹ 本研究は、文部科学省の科学技術イノベーション政策における「政策のための科学」(SciREX) 共進化プロジェクト、並びに政策研究大学院大学 SciREX センターと大学改革支援・学位授与機構との共同研究として実施した。

表1 活用目的による指標の検討方向の違い

評価の種類	評価の目的	事例	指標の検討方向	指標検討の海外動向
指標に基づく大学等の比較（レビューアーによる専門家判断を含まない）	・運営費交付金配分や組織単位の資金配分などにおいて、透明かつ客観的で、低コストに研究成果を測定する。	・学術面＝北欧諸国における運営費交付金配分（ノルウェーモデル等の学術出版の総合的把握） ・インパクト面＝大学評価でのレバンス指標としての産学連携の研究費・特許数等の少数指標（豪州 ERA）	指標の数は少数。機械的に判断が可能な状態にする必要。	質がある程度は担保された上で量を測定する指標の検討
レビューアーによる評価（ケーススタディやナラティブなどの定性的記述の中で指標を活用）	・運営費交付金配分や組織の自己・外部評価において研究内容の質的側面を含めて評価を行う。場合によっては内容的フィードバックをする。	・学術面＝日本の現況分析では多様なエビデンスを大学が記載。オーストラリア、仏国は業績説明。英国 REF、豪州 ERA、イタリアの人文社会は業績そのもの（自然科学は引用数を提供する場合も）。 ・インパクト面＝英国 REF、豪州 EI assessment はケーススタディ手法。オーストラリアでも位置評価基準として説明。ガイドラインとして示す指標について例示。	多種類の指標の例示が可能（一部はその妥当性が明らかでないような指標も利用可能）。しかし、その中でもある程度の比較可能性を担保する必要。	主要な指標についての例示。 多様なインパクトについて、記述形式の統一（英国）、インパクト概念の整理。 Altmetric の模索段階（公式の評価で指標として設定される段階ではない。）

3. 日本における研究成果の測定方法の限界とその要因

3.1. 海外における研究成果測定における前提

研究成果をいかに測定するかは、特に「研究成果量」を直接的に資金配分への一指標として用いようとする場合に重要となる。これまで、人文学・社会科学を含めた研究成果測定として一つの方法を示しているのは、ノルウェーモデルと呼ばれる北欧4ヶ国やベルギー、ポーランド、アイルランドで国全体あるいは一部大学で導入されている方法である(Sivertsen, 2018)。国により異なるが、少なくとも人文学・社会科学については民間論文データベースは使わず、国内ジャーナルや書籍を含めて業績を測定して、それらをポイント制で集計して資金配分に反映するものである。また、ジャーナルや出版社の「レベル」によってポイントに差をつけることも行われている。

これは多様な研究成果を測定可能とまでは行かないが、人文・社会科学の特性を踏まえて書籍も測定対象としており、しばしば日本でも活用可能性について言及されてきた。しかし、この仕組みにはいくつかの前提がある。

第一には、この集計はピアレビューなどの質的判断を経たチャネルによって発表された成果を「学術的出版物」と定義している。収集するデータベースには一般書等も入力できるが、学術成果の指標は、書籍においても査読が計測の最低要件となっている。

第二に、そのような学術的出版物として考え得る出版チャネル（ジャーナル、出版社）とそのレベルの判断は各国のアカデミー等の学術界によって形成されている。「研究評価に関するサンフランシスコ宣言」や「ライデンマニフェスト」など、インパクトファクター等のジャーナル単位の指標を用いることへは批判があるが、ここでは学術界自身が発表チャネルを分類している。

第三に、この方法の実現のためには国内研究業績データベースが不可欠である。たとえばノルウェーでは CRISTIN と呼ばれる共通研究業績入力システムを国レベルで形成している。このようなデータベースが形成されることにより、2016 年からは ENRESSH という 40 カ国の研究者や政策形成者らによる、人文学・社会科学の分析を行うネットワークも形成されている(Jong et al., 2020)。

ノルウェーモデルが導入可能となるには、以上のような前提となる環境の存在が必要である。

3.2. 分析1：日英における研究発表形態の差異

これ以降、上記のような研究成果の測定手法がどの程度、実現可能性があるかを人文学の歴史学を事例として検討する。日本の国立大学法人評価における「研究業績水準判定」において提出された研究成果を対象として研究成果をポイント制で計測しうるかの検討を行う。また、比較として、北欧等のデータは入手不可能であるため、評価への提出資料が公開されている英国の大学研究評価 REF (2014 年実施) との比較を行う。ただし、評価で提出する研究成果数は日本では組織構成員数の半数程度を選定するが、英国では構成員数の4倍を提出しており、構造的には英国のほうが雑多な成果を含む可能性がある。

まず、評価において提出された研究発表の形態の種類を確認すると表2のようになる。書籍（編著や章を含め）やジャーナル論文が両国とも多いことは共通であるが、英国ではそれらが合計で97%を占めるのに対して、日本では9割以下であり、それ以外の形態（報告書や史料、展示など）が1割以上と、

成果形態の多様性は日本がやや高い。

次に書籍について、日本の2回の評価の合計と英国とで、出版社の出現頻度を整理した。表3にはその結果を示している。英国は上位に大学出版社および国際的に著名な（歴史学に出版対象を限定していない）学術出版社が並び、研究成果数が多いにもかかわらず上位9社によって半分以上を占める。一方で、日本の場合は、学術書刊行に限らない大手出版社、大学出版社に加え、歴史学を中心とする人文学に特化した専門出版社が多くなっている。

例えば Oxford University Press は「全ての書籍は刊行決定前に編集スタッフと外部の査読者により評価される」とウェブサイトにも明記しているように、英国は出版前にピアレビューを伴う国際学術出版社が中心である。一方、日

本は分野に特化した多様な出版社が存在しているという特徴があるが、レビューがおこなわれているかは公表情報からは明らかではなかった。

以上のように、日本と英国では研究発表形態も若干異なり、学術書刊行の文化も異なる。日本の特徴は、比較的に多様な発表形態のもとで研究成果を生むことが可能であり、また、「学術書」が必ずしも学者のみを対象に限定しない幅広いものとなっている（佐藤ほか 2011）と肯定的に見ることが可能である一方、ノルウェーモデルで前提となる「レビューを経ている出版社の出版物を研究成果として測定する」という方針は適応しづらい。現在の日本の研究成果発表の文化を前提にすると、ノルウェーモデルの実現は困難と言わざるを得ない。

3.3. 分析2：日本における研究業績評価から得られる代替指標

上記のように研究成果の形態や出版社も多様である場合に、成果の発表チャネルによるポイント制とは別に、代替指標として考えられるものはあるだろうか。それを把握するために、上記の第二期国立大学法人評価において、ピアレビューアが研究業績判定を行った際に、提出物において記載されていた情報の何を「卓越(SS)」という最高評価を下すときの参考情報として活用したか分析を行った（第一期評価を対象とした分析結果と方法については林、土屋 2016 を参照）。

表には人文学あるいは社会科学において 5%以上の研究業績資料において用いられていた指標を示している。歴史学に限らず人文学分野全体でみると、「賞」や、「学術誌・専門誌」、「新聞・一般紙」における書評・紹介が指標として多く用いられている。一方、「出版社名」は人文学で 2%と少ない。これは、上述のように出版チャネルによる質管理が不可能であるために、

表2 歴史学における研究発表形態の日英差異

	日本 NIAD2008 (N=369)	日本 NIAD2016 (N=373)	英国 REF2014 (N=6, 431)
著書	37%	33%	21%
編著	21%	13%	5%
本の章	7%	14%	28%
ジャーナル論文	22%	26%	44%
国際会議	1%	4%	0%
報告書、ワーキングペーパー	7%	2%	1%
データベース	1%	2%	0%
史料、目録、翻訳	4%	2%	1%
展示	0%	2%	0%
ほか(学位論文、報道、ウェブサイトなど)	1%	1%	0%

表3 歴史学における研究成果を公表した出版社の分布

日本NIAD2008&2016	件数	割合	累積割合	英国REF2014	件数	割合	累積割合
岩波書店	39	10%	10%	Oxford University Press	370	11%	11%
吉川弘文館	33	9%	19%	Palgrave Macmillan	320	9%	20%
東京大学出版会	31	8%	27%	Cambridge University Press	256	7%	27%
汲古書院	17	4%	31%	Ashgate	223	6%	34%
山川出版社	12	3%	34%	Routledge	143	4%	38%
思文閣出版	11	3%	37%	Manchester University Press	142	4%	42%
名古屋大学出版会	10	3%	40%	Boydell & Brewer	135	4%	46%
京都大学学術出版会	10	3%	42%	Brill	124	4%	49%
岩田書院	9	2%	45%	Yale University Press	59	2%	51%
ミネルヴァ書房	8	2%	47%	Brepols	57	2%	53%
刀水書房	7	2%	48%	Edinburgh University Press	48	1%	54%
校倉書房	7	2%	50%	Continuum	43	1%	55%
Routledge	6	2%	52%	Pickering and Chatto	35	1%	56%
勉誠出版	6	2%	55%	Bloomsbury Academic	33	1%	57%

表4 評価における代替指標の活用状況

	人文 学	社会 科学
賞	51%	39%
新聞・一般紙での書評・紹介	25%	13%
学術誌・専門書での書評・紹介	24%	12%
著名な学術雑誌への掲載	5%	30%
インパクトファクター	0%	13%
引用	5%	9%
被引用数	0%	5%
著名な叢書の一つ	5%	1%
学会発表、芸術講演	5%	3%
招待講演・基調講演・プレナリー	11%	16%
講習会・研修会、一般講演会	5%	4%
本件の研究費	7%	1%
研究による影響（文化的貢献）	5%	0%
研究による影響（教育への影響）	1%	6%
研究による影響（審議会等委員）	0%	5%
研究による影響（政策への貢献）	1%	6%

成果発表後の「Post-publication peer-review」とよばれる、ピアによる発表成果に対する主観的判断を重視していることが現れていると考えられる。

3.4. 分析3：代替指標としての書評の活用可能性

では、このような Post-publication peer-review 情報は、資金配分等のための情報として現実的に活用可能であるか。そのために、「定量的指標」としての「書評の数」を例にとり、どの程度ピアレビュー結果と相関するのかを確認する。上記の第二期国立大学法人評価において歴史学分野に提出された研究業績のうち書籍を対象として、学術雑誌・一般雑誌の記事（論文等含む）のタイトルのデータベースである国立国会図書館雑誌記事索引データベースにおいて、タイトルに当該書籍名を含む書評・書籍紹介と考えられる記事を検索して、その数を計測し、2名のピアレビューの評価結果（各3段階評価により、合計で0～4点満点）との関係を分析した。

結果は右図である。横軸を評価結果、縦軸を書評の件数として、頻度（研究成果数）をプロットの大きさにしている。結果からは、雑誌（主に学術雑誌）に4回以上書評が掲載された書籍は、学術面あるいは社会

面のどちらかにおいて2以上の評点を示すなど、ある程度の関係はみられる。しかし、書評件数が多い書籍が最も高い評点を獲得しているとは限らない。また、書評が存在しなくとも評点が高い研究業績も多い。つまり、多くの書評が書かれることは学術的あるいは社会的に優れた研究成果であることを代替的に示す可能性はあるが、その情報によって判断ができるほどのカバー率も精度も全くない。

これらの限られた分析結果からは、評価において当該分野の研究者やレビューアーが用いた指標をもってしても、ピアレビューを代替することを可能とする測定とはならないことが示唆される。

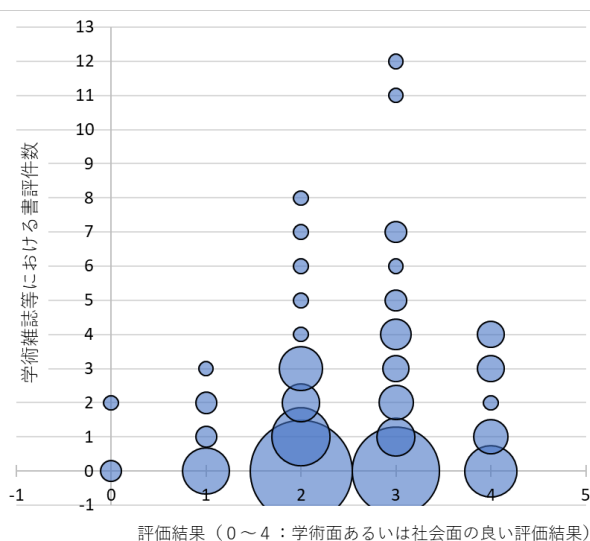


図1 評価結果と書評数の関係

4. 議論

日本では、日本学術会議の報告などにおいて研究の多様性を尊重すべき指摘はなされており、また、上記の歴史学の分析結果からは、人文学において一律な測定を行い資金配分等へ活用することができる指標を見いだすことは難しいことが示唆された。そのため、一定の質を考慮した量的測定を行うためには、多様な指標は参考にしながらもピアレビューによる総合的判断（およびその段階判定の指標化）を行うことが、日本の現状においては選択肢になるのではないか。

一方で、本稿では人文学・社会科学の（学術面を超える）社会的インパクトについては言及していない。社会的インパクトの測定についても、本稿の議論と同様に、インパクトは多様であるために普遍的な指標はなく、特定のミッションに向けたインパクト評価ツールを確立する必要があることや(Pedersen et al., 2020)、インパクトが生じるプロセス段階の連携（エンゲージメント）を含めた測定を行うなど、多様な視点を強調する議論がある一方、インパクトの多様な定性的記述について形式的な標準性（スタイルガイド）を求める方向での模索もある(Parks et al., 2018)。このように、多様な対象についてのレビューアーによる評価実施のプロセスの標準性や、人文学・社会科学を含めて研究業績や研究終了後のインパクトを追跡する研究業績システムの整備などの検討が求められよう。

参考文献

- Jong, S. de, Balaban, C., Holm, J., Spaapen, J., 2020. Redesigning research evaluation practices for the social sciences and humanities: perspectives from the European network for research evaluation in the social sciences and humanities (ENRESSH). *Deeds Days* 73, 17–35.
- Parks, S., Ioppolo, B., Stepanek, M., Gunashekar, S., 2018. Guidance for standardising quantitative indicators of impact within REF case studies, Guidance for standardising quantitative indicators of impact within REF case studies.
- Pedersen, D.B., Grønvad, J.F., Hvidtfeldt, R., 2020. Methods for mapping the impact of social sciences and humanities—A literature review. *Res. Eval.* 29, 4–21.
- Rijke, S. de, Wouters, P.F., Rushforth, A.D., Franssen, T.P., Hammarfelt, B., 2016. Evaluation practices and effects of indicator use—a literature review. *Res. Eval.* 25, 161–169.
- Sivertsen, G., 2018. The Norwegian Model in Norway. *J. Data Inf. Sci.* 3, 3–19.
- 佐藤郁哉, 芳賀学, 山田真茂留, 2011. 『本を生みだす力：学術出版の組織アイデンティティ』新曜社
- 日本学術会議, 2008. 「(対外報告) 我が国における研究評価の現状とその在り方について」
- 日本学術会議, 2012. 「(提言) 我が国の研究評価システムの在り方 ～研究者を育成・支援する評価システムへの転換～」
- 林隆之, 2019. 「国立大学運営費交付金の配分と指標・評価」『IDE 現代の高等教育』615, 64–67.
- 林隆之, 土屋俊, 2016. 「学問分野による「卓越性」指標の多様性」石川真由美編『世界大学ランキングと知の序列化』京都大学出版会, pp.325–345.