

# 世界大学ランキングの恣意性

調 麻佐志 (東京工業大学 リベラルアーツ研究教育院 教授)

## はじめに

世界大学ランキングにかぎらず、あらゆる評価は多かれ少なかれ恣意的なものである。とりわけ、世界大学ランキングのように、複数の事項に対する評価をまとめた総合的な評価は恣意性の度合いが高い。なぜなら、複数の評価項目を取りまとめる際には、(数値を伴わない質的な評価であっても) なんらかの重み付けに相当する操作が必要であり、その「重み」を恣意的に定めることは避けられないからである。

とはいえ、「評価が恣意的である＝妥当でない」という理解は早急過ぎるかもしれない。そもそも評価を実施する際には何らかの目的(評価目的)を伴うはずである。その目的と照らして恣意性の影響の範囲が適切と考えられるのであれば、恣意的であっても妥当と受け止めることもできる。

それでも、その目的を離れて評価の結果である数値だけが流通する、いわゆる数値／数字の独り歩きと呼ばれる現象が社会にあまねく見られることは問題であり、その現象はおそらく数値が客観的であるとみなされることと関係する(藤垣 2013)。世界大学ランキングはまさにこの問題にかかる。

実際、没個人的な手続きによって作成され、その手続きがほぼ公開されているという意味において、現在の世界大学ランキングは概ね客観的である(調 2017)。一方で、その手続きは恣意的にしか定められないものであり、評価結果である数値・順位が客観的であるとは言い難い。そもそも客観的な評価なるものが存在するのであろうか？

本章では、数値例等を踏まえながら、世界大学ランキングの作成過程において必然的に伴う恣意性が評価結果にどう影響し得るかを検討する。

## I 恣意性とは

言語学・言語理論の領域(とりわけソシユールの周辺)で恣意性・恣意的という言葉はよく使われるが、本節で扱う恣意概念は言語理論等に直接かかわるものではない。辞書(大辞林)で恣意は以下のように説明されている。

- ①その時々気ままな思いつき。自分勝手な考え。
- ②物事の関係が偶然的であること。

このうち2つ目の意味での忖意という性質が、本報告で想定する忖意性であり、そのような性質を帯びることを忖意的であるとする。このような忖意性が発揮された代表的な事例が Times Higher Education（以下 THE）が発表する世界大学ランキングにおけるアレキサンドリア大学の躍進である（Wouters 2010）。この事例を紹介することで、大学ランキングにおける忖意性とその影響について具体的なイメージを示そう。

2010年のTHE世界大学ランキング（THEでは2010-11と表記する）においてエジプトのアレキサンドリア大学がアラブアフリカ初の上位500大学にランクインした（147位）。参考までに当時ランクインしていた国内大学を挙げると東大26位、京大57位、東工大112位、阪大130位、東北大132位といった状況であり、その他の国内大学は圏外におかれた。アレキサンドリア大と同順位にはウプサラ大、その少し下にはデルフト工科大、ジョージタウン大、アムステルダム大が位置するなど、欧米圏の錚々たる機関がアレキサンドリア大の後塵を拝した。この結果を評して、THEを作成するTHE誌の編集長はアレキサンドリア大を「ランクインした大学はみな真にワールドクラスである。（“any institution that makes it into this table is truly world class.”）」と讃えた（NYT 2010）。確かに見かけはワールドクラスである。

当該のランキングで、アレキサンドリア大は「研究の影響力」を表すと説明されている引用の項目において高く評価されていて、カリフォルニア工科大、MIT、プリンストン大に次ぐ世界第4位であった。引用以外の4項目におけるアレキサンドリア大の得点は低く、引用がアレキサンドリア大学を147位に引き上げたことは間違いない。問題は、なぜアレキサンドリア大学の「研究の影響力」が高いかである。

この「成果」は、実質的には一名の教員 Mohamed El Naschie 氏に帰着されるものであった。氏は Elsevier 社から出版されている学術誌 “Chaos, Solitons and Fractals” の創刊者兼編集者であり、同誌に323本の論文を掲載している<sup>1</sup>。しかも、氏は（当然ながら）自らの論文の中で過去の自らの論文を大量に引用していた。

THE世界大学ランキングにおいて、引用項目の評価手法（methodology）がこの年より改訂され<sup>2</sup>、従来の教員一人あたりの論文被引用数から当該大学から出版された論文一本あたりの被引用数で評価されることになった。その結果、アレキサンドリア大の引用項目は氏の成果をかなり直截に反映するものとなってしまったのである。なぜなら、アレキサンドリア大学の氏以外の研究者は、評価の基礎データを提供する学術論文データベースに収録されている雑誌<sup>3</sup>にそれほど論文を掲載することがなかったため、論文一本あたりの引用数の分母（論文数）に大きな影響を与えなかったからである。

たった一人の教員の、それも疑わしい成果によって所属大学の評価が「ワールドクラス」にまで高まるのは当然問題視されるべきであろう。翌年、アレキサンドリア大は静かにランキング表から去っていった。

どのようにTHEはアレキサンドリア大をランキングから「排除」したのであるだろうか。簡単な

<sup>1</sup> 常軌を逸した数の論文が掲載されていることについては何らかの不正があったと疑うこともできるであろう。しかし、本報告の射程外であるので検討はしない。

<sup>2</sup> この年の改訂であったが故に、アレキサンドリア大の突然のランクインがこの時に発生した。

<sup>3</sup> 詳細な議論も必要ではあるが、簡単に言えば学術的な評価の高い学術誌である。

方法である。2010年のランキングにおいては、年50本未満しかデータベースに論文が収録されていない大学は評価対象から除外されていたが、翌2011年には年200本へと除外基準が引き上げられたのである。その結果、Naschie氏以外はあまり論文が掲載されないアレキサンドリア大が基準を満たせずに除外された。

確かに極少数の教員の成果が大学全体の評価へと直結することには問題があり、改善されるべきことは間違いない。しかし、その際の除外基準として年200本か、年500本か、はたまた年201本とするのかは恣意的にしか決められない。

改善すべきという判断は適切であるものの、アレキサンドリア大は恣意的にランキング表から除外されてしまったという事実は否定できない<sup>4</sup>。たとえば、近年の一橋大学のデータベース論文収録数は年100~150本程度なのでTHE世界大学ランキングの対象から除外されることを考慮すれば、その恣意性の深刻な影響が理解されるのではないだろうか。

ドタバタはアレキサンドリア大では終わらなかった。2014年の同じTHEの世界大学ランキング(2014-15ランキング)でトルコの有力大学が大躍進した。やはり引用項目の「いたずら」が原因である。

引用項目の評価には評価年の2年前から遡って5年間に出版された論文が用いられる。そのために、この年初めて2012年に出版された論文が評価の対象となった。2013年のノーベル物理学賞受賞者の一人はPeter W. Higgs氏であり、その受賞の引き金になったのが2つの巨大国際共同実験グループ(ATLASとCMS)の実験によるヒッグス(Higgs)粒子の発見であった。実験結果は2012年にそれぞれのグループから論文として出版され、当該の二つの論文は数多くの引用を集めた。この2つの論文が大躍進の鍵である。

トルコの有力大学にはいずれかの研究グループに所属する研究者が在籍しており、とりわけCERNを拠点とするAtlas所属の研究者、同じくCMS所属の研究者の両方が在籍する3大学<sup>5</sup>(中東工科大、ボガジチ(Bogazici)大、イスタンブール工科大)は結果的に多くの引用を獲得することになった。除外基準をクリアしているとはいえ、相対的に文献データベース収録論文数の少ない各大学は、アレキサンドリア大ほどではないが、引用項目に牽引され大幅なランクアップを果たした。事実、ボガジチ大は2012年に出版した論文の獲得する被引用数の約40%をこの二つの論文で稼いでいた。

このような評価結果は決して好ましいものではなく、評価手法の改善が必要なのは間違いない。それではさらに論文数の除外基準を上げて対応したのであろうか。そうではなかった。

CERN関連の巨大グループによる論文はその重要性もあって引用を多数集めている<sup>6</sup>。しかも、トルコの有力大学だけでなく、日本や米国を含む「先進国」の大学でも被引用数の数%程度以上がヒッグス論文を含むCERN関連の巨大グループの比較的少数の論文に由来する。それゆえに、CERN関連の巨大グループに所属する研究者が在籍するか否かによって大学全体の評価が目に見えて左右されるという状況に陥っていたのであった。この結果もまた健全な評価ではな

---

<sup>4</sup> 当時はランキングの methodology ページに除外基準の改訂が明記されていたが、現在少なくともすぐに確認できる場所に methodology ページへのリンクがない。それどころか、2010年のランキングのページにはよくわかる場所に methodology へのリンクが残っている。

<sup>5</sup> 念のため付記するなら、これらの大学に問題があるわけではなく、筆者は十分優れた大学であると考えている。

<sup>6</sup> 重要性に加えて巨大グループには当然研究者が多く所属しており、それらの研究者が個人や小グループで論文を書く際にはCERN論文を引用するので「自然」と引用することの結果でもあると言える。

い。しかし、当該論文群の影響力を引き下げるために論文数による除外基準を大きく引き上げれば、ランキングの対象となる大学が激減したと考えられる。

(推測になるがおそらくそのような状況をも考慮して) THE は別の除外基準を採用したのである。具体的には翌年より著者数が 1000 名を超える論文については引用項目の評価対象から除外することとした<sup>7</sup>。同時期に、引用項目の評価手法の違いにより CERN 関連の巨大グループによる影響をそれほど受けなかった QS の世界大学ランキングにおいても引用項目の評価手法が改定され、著者が在籍する機関の数が 10 を超える論文については評価から除外された。

ここにも恣意性が見られる。1000 あるいは 10 は単に十進法でキリが良い数字なだけであり、そこをカットオフ値にする必然性はない。しかも、もともと恣意的に決まっている論文数による引き上げで対応することも可能な状況で著者数や機関数を基準としたこともまた恣意的である。さらに、このような評価を支える研究分野である科学計量学の領域では、複数著者による論文については著者数で割り算して機関に按分する分数カウントといわれる手法を採用するのが通例であり、その手法でも問題には十分かつ「エレガント」に対応できたはずである。実際、科学計量学研究者である筆者には、分数カウントを採用しないことこそが恣意的に映る<sup>8</sup>。

以上の例示で、大学ランキングに伴う恣意性およびその影響のイメージが提供できたのであろう。本報告で注目する大学ランキングに伴う恣意性は、必ずしも意図や不作為によって生じるものではない。むしろ、ランキング作成に伴う手法や基準などの選択において複数（大概是極めて多く）の選択肢が存在することが恣意性の源である。したがって、恣意性を排除することはできない。しかも、のちほど示すように、ランキング結果に与える選択の影響が無視できないほど大きな場合もあるのだ。

蛇足ながら国内にも、THE 世界大学ランキングにおける引用項目の評価による「恩恵」を大きく受けてきた大学がある。一例を上げれば、首都大学東京である。同大学が優れた中規模大学であることにおそらく異論はないだろう。しかし、中規模大学であるがゆえに CERN 関連の巨大グループによる論文の影響を強めに受け、結果として「得」をしてきたことも否定し難い。さらに、首都大学東京には CERN 論文に加えて、MEGA という「ライフサイエンス分野でよく使われる優れたソフトウェア」に関する論文があり、首都大学東京はその論文によってコンスタントに被引用数を稼ぎ、引用項目における高い評価を得てきた。

当該のソフトウェアを使って解析を実施した多くの論文によって MEGA 論文は引用される。そのような引用は当然である。しかしながら、ソフトウェアにはアップデートが付き物であり、メジャーアップデートの度に次の MEGA 論文が新たに出版され、前の論文が古くなり引用評価の対象からはずれてもその次の論文も引用されるため（論文の世代交代である）、当初の成果がいつまでも大学の評価に大きなプラスの影響を与え続けるという研究評価の観点からみて不都合が生じている。もちろん、首都大学東京が何かズルをしたわけではなく、評価手法がそのようになっているからこそ生ずる「いたずら」である。

大学の規模が相対的に小さいがゆえに MEGA 論文の評価へのインパクトは大きく、首都大学東京は THE 世界大学ランキングの引用の項目において長年国内大学 1 位であった。さらに、

---

<sup>7</sup> その後さらに改訂が加えられた。

<sup>8</sup> 分数カウントを採用することも恣意的であることにはかわりはない。しかし、少なくとも標準的な手法を採用するというのは、恣意性を極力排除するやり方としては妥当であろう。

MEGA に関わる研究者の数は限られているため、著者数や機関数による除外基準にかかる可能性もない。当該の研究者たちが在籍する間、首都大学東京はその「恩恵」を得続けるであろう。

極端な話、手っ取り早く引用項目での評価を高めた中規模以下の大学があれば、好条件を提示して MEGA の研究者を首都大学東京から引き抜くか、定年を待って受け入れれば良い。少なくとも大学全体として「研究力」を高める様々な対策よりも遥かに安価かつ確実に評価を高められる。

## II 世界大学ランキングにおける恣意性

世界大学ランキングでは、多くの選択の過程を経てその作成手法が定まり、それに沿って得点や順位の算出が行われる。現実に行われた選択の過程は明らかではないが、模式的には次のような過程が想定される。すなわち、(1) 評価目的の決定、(2) 評価対象の確定、(3) 定性評価/定量評価のいずれを取るかの選択、(4) いわゆるランキングとなるような一次元のリスト、あるいは多次元の評価軸で評価するか決定、(5) 評価項目の決定、(6) 評価内容を構成する概念（構成概念）の決定、(7) 構成概念を測定可能な概念に落とし込む（可操作化する）方法の決定、(8) 測定方法の決定、(9) 測定値の得点化方法の決定、といったプロセスを経る<sup>9</sup>。

そこで、現存の世界大学ランキングを、すなわち定量評価に基づいて複数の評価項目で測定を行い得点化した上でウェイトに基づいて総合点を算出し、順位付けすることを想定して、これらの過程にどのような選択肢があるかを簡単に検討してみよう。

### (1) 評価目的

評価目的は、当然ながらランキングを作成する目的である。「評価のための評価」という忌まわしい言葉もあるが、通常の評価であれば評価の前提として評価目的は定まっているはずである。たとえば、大学がその経営戦略を考える上で研究交流協定を結ぶべき大学を選定するため、あるいは高校生が進学先を決めるためといったことが評価の目的となる。前者であれば、目的にしたがって、自大学の研究ポートフォリオを前提として、ポートフォリオを補完する、あるいは戦略によるが逆に重なる分野で「研究力」の強い大学のランキングを作成するのが自然である。すなわち、評価する側には何らかの目的があるという意味では評価目的には必然性があるものの、評価目的は様々な目的の中から選ばれたものであるという意味では恣意的であり、その目的が決まることによってランキングの作成手法は、とりわけ評価項目として何を設定するかが全く異なってくる。

なお、既存の著名な世界大学ランキングにおいては必ずしも評価の目的が明らかではない。一例をあげると、THE 世界大学ランキングのページには以下引用するように書かれている。

The Times Higher Education World University Rankings 2016-2017 list the 980 top universities in the world, making it our biggest international league table to date. It is the only global university performance table to judge world class universities across all of their core missions – teaching,

---

<sup>9</sup> 富澤 (2004)を参考とした。

research, knowledge transfer and international outlook.

[https://www.timeshighereducation.com/world-university-rankings/2017/world-ranking#!page/0/length/25/sort\\_by/rank/sort\\_order/asc/cols/scores](https://www.timeshighereducation.com/world-university-rankings/2017/world-ranking#!page/0/length/25/sort_by/rank/sort_order/asc/cols/scores) (2017/6/8 最終確認)

この文章を読む限り、THE 世界大学ランキングは評価のための評価であるように映る。ちなみに、QS 世界大学ランキングのページには評価の目的は、少なくとも目立つところには示されていない。一方、ARWU（上海交通大学）ランキングでは、当初の目的が中国のトップ大学の世界での立ち位置を見出すことであったと記載されている<sup>10</sup>。ARWU ランキングが目的という点では多少具体的であるが、今や目指すところは同じで、世界に数多ある大学を主として一次元にリスト化すること、すなわち序列化自体が目的となっている。

## (2) 評価対象

世界の大学を序列化することが目的であるとしても、通常は対象とする大学の選別を行わなければならない。そもそも国によっては<sup>11</sup>どの機関を大学とするかから議論が必要である（自治体が運営する市民大学もある観点では立派な大学である）。

その議論をクリアしたとしても、たとえば、わが国なら大学院大学や大学校を対象とするかの選別が必要である。さらに評価は一種の比較であり、似たもの同士を比較することは容易であるが、異なるものを比較するのは難しい。たとえば、リベラルアーツカレッジとユニバーシティを総合的に比較するのは不条理な作業であるし、総合大学と単科大学や理工系大学を比較するのも困難である。したがって、評価対象とする大学の種別についても判断しなければならない。結果として、医科大学やLSEのような個性的な大学は近年まで世界大学ランキングの対象外とされていた。

加えて評価コストという現実的な理由などによっても評価対象とする大学を絞り込む必要がある。実際、国内だけで大学は779があるが（平成27年度）、THE 世界大学ランキングの対象となるのは世界全体で981大学、日本では69大学にしか過ぎない。THEの説明によるとデータ収集の対象となった大学はこれより多いものの、いずれにせよ恣意的な基準<sup>12</sup>によって取捨選択が行われている。この大学の除外基準をいくつにするかによっても、後ほど示すように評価結果はかなり変動する。

## (3) 定性評価／定量評価

### (4) 一次元／多次元

上記2つについては、既存の著名な世界大学ランキングにおいては、すでに述べたように定量的に評価し、序列化するための一次元にまとめ上げることが前提となっている。しかし、例えば定量評価によっても大学が評価できることは言うまでもない。

## (5) 評価項目

---

<sup>10</sup> <http://www.shanghai ranking.com/aboutarwu.html> (2017/6/8 最終確認)

<sup>11</sup> 公的に大学が認定されている国であればこの問題は回避できるが、

<sup>12</sup> 基準の一つは5年間にデータベースに収録された論文数が1000本以上であること。

国内の大学の評価においては、過去、教育と研究の2項目を念頭に評価が行われていたが<sup>13</sup>、とくに今世紀になってそこに社会貢献が加わり、3つの評価項目を設定して評価が実施されるようになった。このように評価項目が取捨されることによって評価結果も大きく変わる。

加えて、評価項目の括りとして何を取り上げるかによって、次の構成概念も影響される。それによって評価結果もまた左右される。たとえば、現在は上記3項目に加えて国際性が評価項目となることが通例であり、その項目の評価には留学生比率などが使われる。そこで、もし仮に(ある観点では)国際性を包括する概念である多様性(diversity)を国際性に代えるとすれば、留学生比率だけではなく、たとえば学生や教職員の男女比も評価の対象となり、評価結果に影響を与えると推測される。

世界大学ランキングでも参照データとしては学生男女比率を示しているが、評価項目に国際性はあっても多様性がないことからわかるように、スコアの算出には採用されていない。もし多様性が項目に含まれば、ランキングの対象となる国内大学のほぼすべてにおいて女子学生や女性教員の割合が国際水準よりも極度に低いため、日本の大学の順位はさらに低下するであろう。

#### (6) 構成概念

構成概念とは、評価項目を構成し、実際に測定を行う対象となる概念のことである。たとえば、研究という評価項目に対しては、研究成果の質、量、あるいは成果のインパクトなどが構成概念の候補と考えられる。仮に研究成果の量のみを構成概念とした場合には、規模の大きな大学が圧倒的に有利なランキングとなることは間違いない。

また、本報告書は(研究の)国際性についてはその質(国際共著論文比率)に加えて、その「厚み」も評価することを提案した。仮に世界大学ランキングにおいてそれが国際性の構成概念に加われば、いくつかの国の教育機関の順位は大きく変動するかもしれない。たとえば、ニュージーランドの大学から出ている論文の国際共著論文比率は高く、その国際性もまた高く評価される。しかし、隣国オーストラリアの影響は強いので、筆者は測定していないがもしかしたらニュージーランドの大学の厚みは限定されているかもしれない。このように新たな構成概念を加えることによって大学ランキングに変動が見られるはずだ。

#### (7) 可操作化の方法

構成概念が定まってもそれだけでは評価のための測定は実施できない。研究成果のインパクトを測定するには、測定可能なインパクトを表す代理変数を決めなければならない。たとえばよく使われる代理変数は論文の(学術論文による)被引用数である。しかし、インパクトとして、学界を離れ広く社会的なものを想定するのであれば、論文の(特許による)被引用数を用いても良いだろう。さらに現実的なインパクトを考えるのであれば、ライセンスの件数や特許収入などの方が適切かもしれない。一方で、医療分野に限れば診療等のガイドラインに引かれた論文の数などもインパクトの指標となり得る。

どのような可操作化を行うかによって、学術的な研究の得意な大学と産業化に近いところの研究にフォーカスしている大学で、選択の影響にも違いが生じるのは想像に難くない。

---

<sup>13</sup> 当時、公的ないし体系的な評価はそもそもなく、あくまで暗黙の評価項目である。

## (8) 測定方法

具体的な測定方法にも恣意性が認められる。研究インパクトを論文の被引用数で測定するとしても、どのように正規化するかによって結果に大きな違いが生じる。よく行われるのは、論文一本当たりの平均値で集計する被引用数の正規化であるが、QS 世界大学ランキングでは研究者一人あたりで正規化している。その違いは大きなもので、すでに例として上げたボガジチ大の2014年のケースでは、被引用数についてはTHE 世界大学ランキングでは世界43位であるのに対し、QS 世界大学ランキングでは361位であった。それが主因となって、総合的なランキングではTHEで139位であるのに対し、QSで399位と奮わなかった。正規化の方法の恣意的な選択結果がこれほど大きくランキングを動かすのである。しかも、いずれをとるかは恣意的な判断であるとはいえ、どちらの正規化の考え方にもおかしい点があるわけではない。

恣意的になってしまうのは、正規化の方法だけではない。被引用数測定の対象となる論文の出版年を過去何年のものとするか、学術分野ごとの引用のされやすさを補正するための学術分野区分をどのように設定するか、被引用数に影響を与えていると考えられる国や文化の違いを補正するか、など様々なところで恣意的な選択が求められる。とくに最後の国や文化の違いの補正については、その手法の恣意的な変更により2015年日本の大学のTHE 世界大学ランキングにおける順位は急降下したように(調 2016)、ランキングにも多大な影響を与えている。

## (9) 得点化の方法

測定方法が決定すれば、測定値が得られる。試験の結果として出てくる素点のようなものである。順位付けを一つの構成概念に基づいて行うのであれば、素点に基づいてソートし順位をつければ終わりである。しかし、現在、流通する世界大学ランキングは総合的な評価であることを標榜しており、複数の評価項目、複数の構成概念に基づき総合的な順位付けを行っている。だから、複数の素点は一次元に集約される必要がある。そのために通常は重み付き平均が用いられるので、各項に対するウェイトを決定しなければならない。当然ながら、研究と教育、あるいはその他の項目に対するウェイトを与える標準的な考え方は存在しないので、ウェイトの決定は恣意性の典型例である。ウェイトの操作はランキング結果の変動に直結する。

さらにウェイト以前に、素点を共通の基準に則った得点(スコア)に変えなければ、重み付き平均を求めることは難しい。たとえば、留学生比率0.3、論文一本あたり被引用数10という数字に対して、ウェイト付きとはいえ平均を取ることはナンセンスであることがその理由である。そのため、著名な世界大学ランキングではそれぞれのやり方で素点を得点に換算している。

THEとQSのランキングは共通してzスコアと呼ぶ得点へと素点を換算する。すなわち、対象となった全大学の素点から平均値と分散を求め、素点の分布が正規分布であることを仮定し、仮定した正規分布における素点のパーセント値を得点とするのである<sup>14</sup>。ただし、評判に関するアンケートの得票数が素点である項目に関しては、その甚だしい分布の歪みを補正するために素点の対数をとった上でzスコアに換算する<sup>15</sup>。

のちほど示すように、素点は必ずしも正規分布とみなせるわけではない。そのことを踏まえ

<sup>14</sup> 受験産業におけるいわゆる偏差値の考え方と多少似ているところがある。そして、得点分布が歪んでいるために、その意味合いが怪しいことも同じである。

<sup>15</sup> この変数変換もまた恣意的である。





ハーバード大	98.5	96.7	98.7	99.2	86.7	98.9	99.6	85.6	99.2
MIT	99.7	87.7	99.4	99.4	88.7	99.3	99.7	87.7	99.4
スタンフォード大	99.1	100.0	100.0	99.5	89.7	99.6	99.8	88.6	99.6
Caltech	98.9	98.8	99.2	99.4	88.6	99.2	99.7	87.6	99.4
東大	31.8	52.2	31.0	44.2	46.8	44.4	53.7	46.2	54.0
京大	24.5	48.5	26.5	36.1	43.5	38.4	45.4	43.0	47.7
阪大	17.4	44.3	21.3	27.5	39.8	32.0	36.1	39.3	41.2
東北大	22.1	47.2	25.3	33.3	42.4	36.4	42.5	41.8	45.8
東工大	20.8	46.5	24.3	31.7	41.7	35.4	40.8	41.2	44.8
神戸大	14.1	42.0	17.5	23.2	37.7	27.1	31.3	37.2	36.4
NUS	72.7	70.3	71.9	81.5	63.1	81.4	87.1	62.3	86.7
北京大	36.0	54.1	32.8	48.6	48.5	47.3	58.1	48.0	57.2
ソウル大	31.6	52.1	30.8	44.0	46.7	44.3	53.6	46.2	53.8
台湾大	28.5	50.6	28.5	40.6	45.4	41.9	50.1	44.8	51.1

(出典) Elsevier 社提供 Scopus 集計データを元に筆者作成。

表1のように、上位大学については対象大学数の増減や得点化方法の違いによって、得点に大きな変動は見られない<sup>16</sup>。一方、国内および得点の高いNUSを除くアジアの大学では、THE/QSの方法で得点化した場合、対象大学数が増えるにつれて得点も大きく上昇する。したがって、上位大学との得点差も目に見えて縮小する。ARWUでは小幅な下降はあるものの、上位大学も同程度下がるため<sup>17</sup>、得点差に大きな変動はない。さらに、パーセント値による場合はTHE/QSの方法と比較的似通った挙動を示すものの、東大を除く日本の大学の得点はTHE/QSよりも高めとなる傾向がある（とくに対象大学数が少ない場合）。

THE/QSの方法とパーセント値によるものとの評価結果の違いの一部は、素点であるFWCIが正規分布ではないことに由来する可能性がある。そこで、FWCIの分布について分析する。

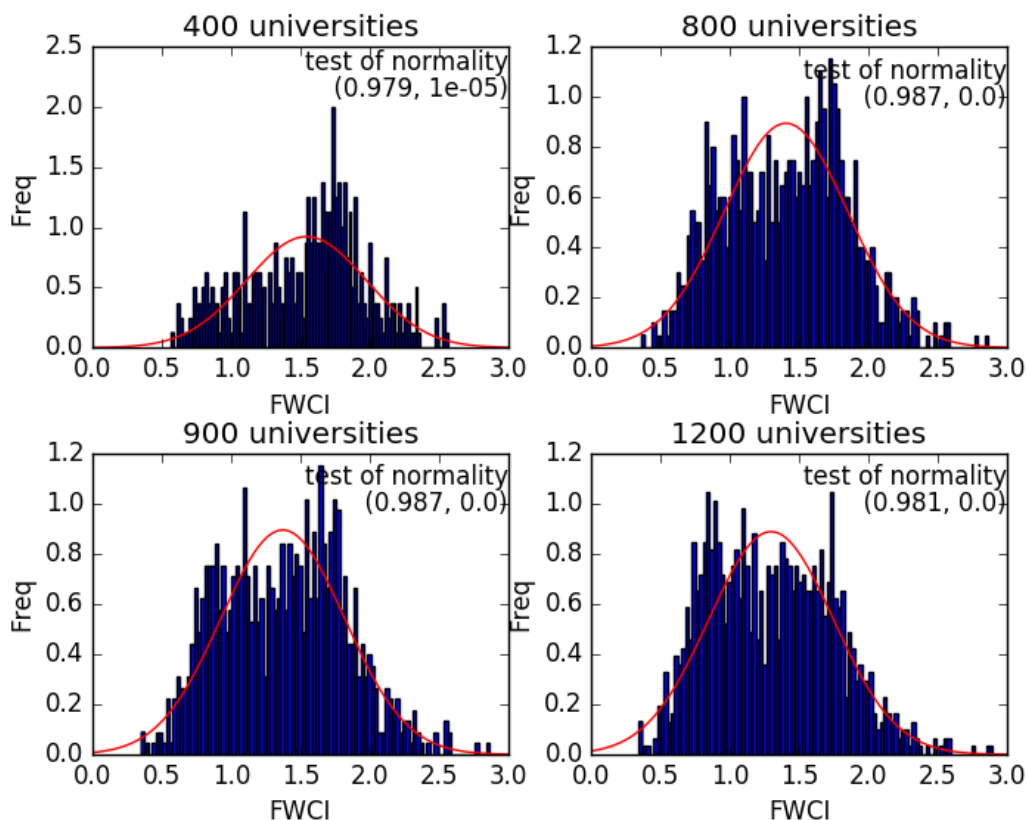
## (2) FWCIを対象としたzスコアが仮定する測定値の正規分布の仮定の検討

すでに述べたように、スコアを使った得点化の前提としては、素点（測定値）が正規分布することが暗黙に仮定されている。ここでも引用を対象として、その仮定の妥当性を検証してみよう。

データは前項と同じものを用い、5年間の論文数上位400大学、800大学、900大学（当時のTHEランキングの対象大学数）、1200大学を対象として、FWCI（の大学内平均）の分布を確認し、さらにシャピロウィルク検定によって正規性を帰無仮説とする検定を行った。分析の結果は図1の通りである。なお、青の棒はデータのヒストグラムを、赤線で示された分布はそれを正規分布でフィッティングした結果を示す。

<sup>16</sup> 正確には、ARWUランキングでは800大学に拡大した際に、新たな大学が最上位校となったため、それなりの変化が見られた。

<sup>17</sup> 対象大学を増やしたことにより、最上位校の上を行く大学が評価対象に加わったためである。



(出典) Elsevier 社提供 Scopus 集計データを基に筆者作成。

図1 FWCI の分布と正規性の検定

図1から視覚的に明らかなように、対象大学数によらず、FWCIの分布は正規分布とみなせるものではない。大学(サンプル)数が増えるにつれて正規分布に近づくこともなく、本分析で最大のサンプル数である1200大学となっても2つのピークが見られる。当然、検定結果も帰無仮説を棄却するもので、少なくともFWCIに関して当該サンプルにおいては正規分布を仮定することはすべきでない。すなわち、zスコアの適用は適切でないと考えられる。

## V まとめ

以上に論じてきたように、世界大学ランキングの作成手法を決定するすべての段階に、恣意的な選択・判断が見られる。ランキング作成に限らず評価を行う際に恣意的判断は不可避であり、恣意性があること、それ自体が妥当性の欠如を意味するわけではない。しかし、世界大学ランキングはその作成目的が明確でなく、作成手法(と結果)の妥当性を判断する上で重要な評価目的がわからないという問題がある。

数値計算により、手法や基準の恣意的な選択によって、世界大学ランキングの元となる得点(スコア)に少なからぬ影響あることが示された。しかも、世界大学ランキングではよく使われるzスコアを用いた得点化が測定値の正規性を前提としているにもかかわらず、その仮定は

少なくとも引用を対象とする場合に棄却されることが検定により確認された。このことは、当該手法を用いた世界大学ランキングの妥当性を疑う根拠となり得る。

しかしながら、世界大学ランキングという形式、すなわち一次元の序列化こそ必要としなくとも、大学の評価には社会的な要請がある。そして、あらゆる形式の大学評価にも恣意性が見られるのだから、それを排除するのではなく、多少でもマシな手法や指標を選ぶことを支援するような見識や研究が重要であるだろう。

#### 参考文献 (MS ゴシック 9pt.)

- ・藤垣 裕子 (2013) : 解題. 『数値と客観性』. みすず書房.
- ・調 麻佐志 (2017) : 世界大学ランキングと「研究力」. IDE 現代の高等教育. No. 589 pp. 45-51.
- ・調 麻佐志 (2016) : 世界大学ランキングの決まり方. nippon.com (<http://www.nippon.com/ja/in-depth/a05102/>) .
- ・富澤 宏之 (2004) : 方法論間の関係. 藤垣編『研究評価・科学論のための科学計量学入門』. 丸善. pp.110-119
- ・ARWU 2016 Methodology (<http://www.shanghairanking.com/ARWU-Methodology-2016.html>).
- ・NYT (2010): Questionable Science Behind Academic Rankings. (<http://www.nytimes.com/2010/11/15/education/15iht-educLede15.html?scp=1&sq=times%20higher%20education%20ranking&st=cse>).
- ・Paul Wouters (2010): New York Times: “Questionable science behind academic rankings”. (<https://citationculture.wordpress.com/2010/11/18/>).
- ・QS World University Rankings 2016-2017 methodology (<https://www.topuniversities.com/qs-world-university-rankings/methodology>).
- ・World University Rankings 2016-2017 methodology (<https://www.timeshighereducation.com/world-university-rankings/methodology-world-university-rankings-2016-2017>).

(しらべ まさし)