

英語力ランキング批判
—EF-EPI・TOEFL スコア・英語教育実施状況調査—
**Criticising Rankings in English Language Education:
EF-EPI, TOEFL Scores, and English Education Implementation Survey**
キーワード: 英語教育言説、科学コミュニケーション、統計調査

責任著者: 寺沢 拓敬(関西学院大学社会学部) TerasawaTakunori@gmail.com

TERASAWA Takunori

奥住 桂(埼玉大学教育学部)

OKUZUMI Kei

浦野 研(北海学園大学経営学部)

URANO Ken

1. はじめに

本稿は、英語力を国・自治体別にランキング化する言説をとりあげ、それらの問題点を検討し、具体的な改善提案を行うことを目的とした理論的・規範的研究である。英語教育界には、学術文献・商業文献を問わず、何らかの指標を用いた英語力ランキング、および、それに基づく英語教育論が蔓延している。本稿では、こうした言説を英語力ランキング言説(あるいは、より短く、ランキング言説)と呼び、検討の俎上に乗せる。

アウトラインは次の通りである。2節で、ランキング言説として有名な国際英語力ランキングと都道府県ランキングをとりあげ、概要を述べる。3節で、これらの言説の問題点を検討する。具体的には、3.1節で方法論(とりわけ量的社会調査・統計データ分析)に関わる問題点を、3.2節で負の社会的影響を検討する。2節・3節の理論的検討をふまえ、4節・5節では具体的事例に焦点を当てる。教育界(4節)および学界(5節)においてランキング言説がいかに蔓延しており、また、どのような影響をおよぼし得るかを検討する。最後の6節で、学界・教育界はランキング言説に対してどのように向き合うべきか、具体的な提案とともに論じる。

2. ふたつのランキング言説

2.1. 国際英語力ランキング

国際英語論 (Inner, Outer, and Expanding Circles) を引くまでもなく、国によって国民の英語力に差があることは、大衆的逸話としてよく知られている。こうした序列を、逸話的な証拠ではなく、標準化英語能力テストの数値という根拠をもとに表現したのが、国際英語力ランキングである。伝統的には TOEFL スコアの国別平均値が広く用いられてきており、TOEFL ほど頻繁ではないが TOEIC が引用されることもある。また、ここ 10 年程度は、留学・語学サービス会社 Education First (EF) 社の英語能力テストを用いた EF 英語能力指標(通称 EF-EPI)が存在感を増している。

TOEFL にもとづくランキング言説は、しばらく前から、日本国内外を問わず広く流通している。これは、ETS による年次報告書 (*Data Summary*) に掲載された国別スコアを「流用」して、マスメディアや研究者が独自にランキングを作成することである。たとえば、文部科学省(2022)は、TOEFL や TOEIC の国別スコアを「流用」しながら、「各種の英語資格・検定試験において、我が国の平均スコアは諸外国の中で最下位クラス(p. 2)」であり、英語教育改革が必要だと訴えている。ここであえて「流用」と表現しているのは、運営元である ETS は、同報告書をもとにしたランキング作成は逸脱的使用であると明示的に警告しているからである(ETS, 2022, p. 19)。

対照的に、EF 社は、EF 英語能力指標ランキングを積極的に広報している。この指標は、世界中で受験可能な同社の英語オンラインテストを用いて、その受験者のスコア平均を国別にまとめたものである。同指標のウェブサイト(<https://www.efjapan.co.jp/epi/>)の冒頭ページには、集計対象国すべてが順位付きで並べられており、明らかにランキングとして表現されている。なお、本稿執筆時点における最新の結果(2023年)でいうと、全 113 国・地域中、1 位から 12 位までの地域に「英語力が非常に高い」というラベルが付されており、

以下、13–30位が「高い」、31–63位が「標準的」、64–90位が「低い」、92–113位が「非常に低い」である。なお、ここで日本など特定の国の順位について言及することは、学術的に意味が薄く、かつ、本稿の趣旨と真逆であるので控たい(理由は後述)。

このランキングが近年存在感を増しているのはその積極的な広報・メディア戦略である。同社は、例年11月にその年のランキングを発表している。おそらく留学会社としての知名度向上のためと思われるが、かなり積極的なプレスリリースも行っている。その結果、同ランキングを報じるメディアも多く、2023年ランキングで言えば、大手ではたとえば日本経済新聞(2023)や読売新聞(2024)の報道がある。さらに、ウェブメディア・個人ブログ・ソーシャルメディアでは、その何十倍～何百倍も言及されており、ウェブ上の情報としてはかなりの存在感を放っている。たとえば、グーグル検索で「日本人の英語力」と検索すると、上位の結果をEF英語力ランキング関係が独占する。私企業の独自指標(それも販促目的のもの)に過ぎない英語力ランキングが、ウェブ言論空間上では、由緒ある国際ランキング(たとえば、GDP、UNDP 人間開発指数、PISA 学力調査、FIFA ランキング)と同種のものと同様に扱われているかのようである。

ここで重要なのは、TOEFL 等の伝統的な標準化テストにせよ、比較的新興のEFテストにせよ、その最大にして唯一の目的は、個人の英語力の測定である点である。逆に、言語テストは、集団レベルの言語能力を推定するように設計されているわけではない。個人の能力測定と、大規模集団の能力平均値の推定は、根本的にアプローチが異なり、前者は測定学の枠組み、後者は社会調査論の枠組みである。つまり、個人の英語力測定のツールで、国民レベルの平均値を議論しようとする既存の英語力ランキングは、そもそも学術的・方法論的に大きな逸脱を犯している。この点は3節で詳述する。

2.2 文部科学省「英語教育実施状況調査」

もうひとつ検討の俎上に載せるのが、文部科学省「英語教育実施状況調査」である。この調査は、全国の公立小・中・高の児童生徒を対象に、毎年12月に実施されている(文部科学省, 2024)。同調査の結果が例年5月に発表され、それをマスメディア各社は「都道府県別の英語力」としてランキング形式で報道している。ただし、同調査は、英語力だけではなく様々な英語教育実践についても調査している包括的調査だが、一般的な注目度が高くメディア報道が集中するのは英語力の結果である。

ここで重要なのが、同調査の「英語力」の定義である。中3の英語力を例にとると、「生徒の英語力の状況」という調査項目における「CEFR A1 レベル相当以上の英語力を取得または有すると思われる生徒数」が英語力を持つ生徒数として定義されている。ここで注意すべきは「取得または有すると思われる」という表現であり、これは要するに「取得」と「有すると思われる」の合算値である。前者は英検等の外部試験を受験し、そのスコアがA1水準を上回っている生徒である。後者は、調査担当者(多くの場合、その学校の英語教員)がA1水準を上回っていると判断した生徒である。つまり、ここでのA1レベル以上の英語力を持つ生徒数には、調査者による主観的推定が混入している。この事実だけでも砂上の楼閣のような数値であることは明らかだが、マスメディアは、この背景まで報じることは皆無で、この調査結果がまるで揺るぎない事実かのように報道している(ただし、例外的に、*The Mainichi*(2024)は同調査の数字の裏側について詳しく報道している)。

3. 問題点

3.1. データの扱い方をめぐる問題点

本節では、上述の英語教育ランキングの問題点について、まず、方法論的な観点から2つの問題点を論じる。以下は、データ分析の初歩的な内容であるが、英語力ランキングを念頭に置いて解説している文献は乏しいので、本稿で詳述する意義はあるだろう(なお、EFランキングのイデオロギー的側面についてはSavski(2021)による検討がある)。

問題の第一が、データの代表性の乏しさである。代表性とは、手元のデータ(例、英語テスト受験者のデータ)が母集団(例、ある国の全国民)をどれだけ過不足なく反映しているかという観点である。代表性は、一般的に、サンプリングの質に左右される。最も強力なのが単純ランダムサンプリングであり、理想的な条件下であれば、非常に高い代表性が実現できる。ただ、現実の調査では種々の事情により理想的条件が担保されないことも多いので、社会統計・社会調査論の研究者は様々な工夫・改善策を提案している(廣瀬・稲垣・深

谷, 2018; 佐藤, 2015; Terasawa, 2024a)。他方、代表性が最も低いのが、「配慮ゼロのサンプリング」、つまり、自発的な調査参加者だけに頼ったり、調査遂行者にとって手近な回答者にしか調査依頼を行わないサンプリングである。そして、TOEFLもEFテストも自発的な受験者のみで成り立っている以上、その国別スコアはきわめて代表性が乏しい。この代表性の低さは、ある国の国民の英語力を過小評価あるいは過大評価するリスクが高いことを意味している。にもかかわらず、既存のランキング言説は、この問題への考慮は一切なく、不確かな点数をもとに各国民の英語力を論じているのである。

もっとも、代表性は、統計学のごく初歩的な概念であり、その重要さもよく知られている。したがって、こうした英語力ランキングの代表性の問題を理解しつつも、あくまでおおよそを理解するための参考情報として利用している研究者・著者もいるかもしれない。実際、上述の、文部科学省(2022)にも、この点の但し書きがある。しかし、結論から言うと、英語力ランキングについては、参考情報として扱えるかどうか疑わしい。なぜなら、テストスコア以外の情報がなく、データ総体としてどの程度参考になるのか／ならないのかが判断できないからである。とりわけ、EFテストについては、テスト受験者層にかかわる情報がほとんど公開されていない。同テストは、公式証明には使えないテストである。それをわざわざ受験しようとする人は、かなり特殊な集団だと思われる。反対に、高い英語力を持っていても、EFテストを受験しない人も多いだろう。たとえば、すでに英語で仕事や学業に日夜勤しんでいる人には自分の英語力を診断するインセンティブは高くないはずである。また、既に他の英語テストで自分の実力を把握できていると考えている人も受験しないだろう。つまり、英語力保持者層のうち、特定の集団が完全に脱落しているのである。問題は、この脱落がどの規模で起きているかまったくわからない点である。脱落のパターンが国によって異なれば、数値もそれに連動して変化するので、1点刻みで作られているランキングの順位は大きく変動する可能性がある。つまり、このランキングが参考情報程度にはなるのか、それとも、まったく参考にならず、むしろ、現実を歪めかねない「虚偽情報」なのかは未知数なのである。

また、代表性の問題は、経年比較にもあてはまる。各国の受験者層が不透明なのと全く同様に、各年の受験者層も不透明だからである。代表性が担保されていなければ、受験者層は年によって容易に変わり得る。したがって、たとえば、2022年と比べて2023年の日本のスコアが下がったというデータを前にしても、その原因が日本人の英語力が下がったためなのか、それとも、英語力の低い受験者が増えたためなのかはわからない。つまり、経年比較のための参考情報にもならないのである。

第二に、測定の妥当性の問題である。これは、主に英語教育実施状況調査の英語力設問に当てはまる問題である。測定の実用性とは、測りたいもの(概念)がどれだけデータに反映されているかである。逆に、測りたいものと直接関係ない要因が測定に混入すればするほどその実用性は低くなる。英語教育実施状況調査では、前述の通り、調査者(主に学校の英語教員)の主観で生徒の英語力を測定しており、測定の実用性を毀損しかねないリスクを孕んでいる。もっとも、主観が、自動的に、実用性を低下させるわけではない。仮に、調査者間で主観のブレが存在していても、そのブレが完全にランダムに生じているならば、大きな問題にはならない場合も多い。大数の法則により、個々のブレが互いに相殺し合うからである。問題は、そのブレがランダムではなく、特定の方向に偏って起きている場合、つまり、バイアスが発生している場合である。このバイアスは、まさに英語教育実施状況調査に懸念される点である。同調査には、あらかじめ設定された目標値(教育振興基本計画)の達成度合いの調査という趣旨があり、目標値を上回ろうとする強いインセンティブが介在している(黒坂, 2023)。しかも、都道府県別の結果は、毎年公開され、多くのマスメディアは都道府県ランキングで自治体間の序列を煽り立てるため、教育委員会には数値を高めに報告するインセンティブがある。ただし、このインセンティブの多寡は、自治体によっても大きな差がある。ある自治体(たとえば、文部科学省へのいわゆる「村度」度合いが高い自治体)ではきわめて大きい過大推定が起きているのに対し(黒坂, 2023)、別の自治体ではそうでもないということも容易に想像できる。このとき、前者と後者の自治体を同一次元に並べてランキング化することは、フェアな比較とは言えない。

3.2. ランキング蔓延による負の効果

以上の統計上の誤用よりもいっそう深刻だと考えられるのが、ランキング言説の引き起こす負の社会的影響である。現状、ランキングデータを引用する人々が何を目的にしているかといえば、事象の記述などではなく、

特定の意見を根拠づけるためであることは明らかである。たとえば、日本語媒体で国際英語力ランキングを引用する人は、ほぼ必ず、日本の順位の低さを強調し、日本の英語教育の遅れを訴える。都道府県ランキングも同様で、下位の都道府県の遅れを指摘する一方で、トップ自治体の取り組みを称賛する(e.g., 細田, 2024)。そもそも、英語教育実施状況調査は、その成り立ちからして目的志向が明らかである。達成目標を掲げ、それとの距離を各自治体・学校に自覚させることを目的のひとつとしているからである。さらに、明示されていないものの、ランキングで自治体を競わせることで、英語教育改革を加速させようという隠れた意図もあるだろう。

そもそも、社会調査論・学力調査論の本義からすると、調査以外の目的で調査を実施することは逸脱である(川口, 2020)。ただ、それ以上に深刻に考えるべきは、改革ツール・啓蒙ツールとして調査・データが利用(誤用)されることで引き起こされる弊害である。

弊害の第一が、誤った「エビデンス」に基づいて、不適切な教育改革が進む点である。たとえば、国レベルの英語教育課程を改善するうえで、現行教育課程で学ぶ生徒がどれくらいの英語力を平均的に持っているかは非常に貴重な情報である。しかし、もしこの数値が過大推定されていたなら、非常にまずい事態である。たとえば、実際には生徒の2割程度しかA1レベルに達していなかったにもかかわらず、「5割」という過剰推定値をもとに教育課程を構想した場合、それは多くの生徒にとって非常に負荷の高いものになりかねない。したがって、たとえ達成目標を下回ったとしても、正しい数値の報告が優先されるべきである。それは、当初の達成目標の非現実さを反省する契機にもなり、中長期的には教育の改善につながるだろう。

第二の弊害は、不適切な調査・データの存在感が増すと、真に必要な調査が行われなくなることである。生徒の英語力平均値を妥当な方法で推定することは、コストさえかければ可能である(詳細は6節)。しかし、政府の調査リソース(調査予算、調査人員、調査を組織する際の政治的コスト)には限りがある。理想的には、コストを度外視してでも必要十分の調査を行うべきだが、簡便な手法が現れると、たとえ問題含みでも、その手法で代替されかねない。こうなった場合、優良なエビデンスを提供するはずだった調査がキャンセルされるので、中長期的に英語教育への損失となる。

第三は、より抽象的な損失である。それは、学問、とくに英語教育学や応用言語学に対する信頼の毀損である。不適切な調査やデータ利用で信頼を失うのは、直接的にはそれを行った調査者・データ分析者である。しかし、不信感が積み重なると、分析者個人にとどまらず、学界全体への不信感に増幅してしまいかねない。人は、通常、ある学問を信頼しているからこそ、その分野が生み出す個々の研究知見・エビデンスも信頼する。逆に、あるエビデンスが生み出された前提条件・文脈をすべて精査した後に、信頼するかどうかを判断する人はほとんどいない。たとえば、研究者ですら、ある論文を吟味する際、引用している文献(一次文献)をすべて再検討したり、再現実験をしたり、使用されている統計手法の数学的証明を要求したり、データ捏造や利益相反がないか著者に開示を要求したりすることは稀である。この判断の背後には「この論文の著者は、知的誠実性をもって、研究を遂行したに違いない」という信頼がある。しかし、この信頼が裏切られることが続くと、容易に学問全体の不信感に転換し得る。たとえば、「この研究はどうせ業者の利権だろう」とか「文科省の施策にお墨付きを与えるためのものだろう」といった不信感が蓄積すれば、学問的・科学的・政策的コミュニケーションは機能不全に陥る。英語力ランキングの濫用にも当然そのようなリスクがある。

4. 教育現場とランキング

4.1. 英語教育実施状況調査の現場での受容

ここからは、具体的な事例をもとに、ランキング言説の問題点を検討する。本節では、英語教育実施状況調査に対する現場教員の受け止め方について検討する。著者の一人(奥住)は、2019年まで公立中学校教員として勤務し、英語科主任としてこの調査に回答する立場にあった。その経験では、少なくとも2019年頃は、この調査の回答内容や結果について教員間で話題になることはあまりなかった。話題になったとすれば、その回答に伴う負担感である。なぜなら、回答基準日の12月1日は高校受験に備えて調査書等を作成する時期と重なっていたからである。むしろ全国学力・学習状況調査(以下、全学テ)のほうが(英語は毎年実施されるわけではないが)、管理職や自治体からの「指導」を受ける機会が多いためか、話題になっていた。

ただし、現行の形式となった2019年度以降、英語教育実施状況調査の結果は自治体ランキングの形式で

報道されることも増え、現場の受け止め方も変わってきている可能性はある。自治体によっては、報告する数値について教育委員会からの「確認」があったり、結果の高い低いによって、要因を報告するレポートを課されたりするなどの、さらなる負担が増えているという声がある(黒坂, 2023)。

4.2. 現場はどう回答しているのか

では、実際の回答はどのようになされているのか。調査項目には主に (1) 児童生徒の英語力、(2) 英語担当教師の英語力・指導力、(3) 指導方法・教材などがあるが、特に悩ましいのは (1) である。前述(2.2)の通り、CEFR A1 レベル相当以上のスコアを取得している生徒数と、その水準に達していると判断できる生徒数という2種類の数値を報告する必要がある。前者については、アンケートなどで生徒に直接確認できるが、後者は担当英語教師が判断しなければならない。

「判断方法の例」として、文部科学省は「2 技能または 3 技能を測る試験におけるスコアを踏まえて判断」、「公式な記録としては認定されない試験のスコア」、「CAN-DO リストに基づくパフォーマンステストの結果等」の3つを提示している。この「試験」や「テスト」からは英検など外部試験が想起されるが、そのような明記はないので、事実上あらゆる種類のテストが該当する。たとえば、授業内でのパフォーマンステストや、いわゆる定期テスト(リスニング・ライティングを含むもの)などである。このように、定義は非常にあいまいである。

著者の一人(同上)は、中学校教員(当時)として、次のような手続きで報告していた。まず、英検の A1 対応級に合格した生徒の習熟度に注目した。そこから類推し、中学校 3 年生 2 学期の通知表評定で、4 または 5 が見込まれる生徒は、当該級に合格できたと推定した。以上の判断から、評定 4 以上が見込まれる生徒を A1 レベルと見なし、報告した。ただし、年度が変われば、この基準は変わり得るので、学校として一貫した基準で判断・報告ができていたとは言い難い状況もあった。もっとも、こうした不透明さは、教員側・学校側の問題というよりも、調査設計の問題である。調査項目の指示文がきわめて曖昧だからである。

同様の不透明さは、英語力以外の項目にも確認できる。たとえば、「授業における、生徒の英語による言語活動の割合」を調査する項目があるが、「言語活動」の定義が明確ではない(例示はあるものの、教員によっていかようにも解釈されてしまう玉虫色の定義である)。「言語活動の比率が高い地域は英語力が高い」などとまことしやかに分析がなされることもあるが、もともとの定義があいまいなため、こうした相関関係に本当に意味のあるものなのかは未知数である。たとえば、「文科省・教育委員会への忖度の程度」を交絡因子とした擬似相関の可能性もある。たとえば、文科省・教育委員会の意向に敏感な学校ほど、生徒の英語力を高めに報告しやすく、かつ、言語活動が盛んであると報告しやすいといった可能性である。

4.3. 「成果が上がっている」自治体

この調査で「成果が上がっている」と見なされている自治体の現状はどうなっているのか。2018 年度から 2023 年度まで連続で全国1位となっているさいたま市を例にとろう。2023 年、同市の A1 レベル相当以上の中 3 生は 88.4%と抜きん出ている(全国平均は 50.0%)。同市は、好成绩の要因として、(1) 独自科目「グローバル・スタディ」科の設置、(2) 学校教育法施行規則の標準よりも多い授業時数(中学校で標準 420 時間に対し 471 時間)、(3) 専科や ALT の配置などの指導体制の充実、(4) 教員研修の充実、(5) イングリッシュ・キャンプやディベート大会など英語をアウトプットする場の提供の 5 点を挙げている(さいたま市, 2023; 細田, 2024)。独自のプログラムを整備し、標準を上回る時数を確保するなど独自の努力がうかがえるが、こうした違いだけで全国平均よりも 40 ポイント近く高い数字を達成できるのかは慎重な検討を必要とするだろう。

細田(2024, p. 77)によれば、さいたま市は全中学生に、英検 IBA を用いた効果測定を実施している。英検 IBA は、いわば簡易版の英検で、聞く・読むの 2 技能を 45 分間で測定する。A1 レベル相当以上のスコアを取得する生徒数(「みなし」ではなく)について、さいたま市は 81.1%(全国平均 27.9%)と報告しており、おそらく同市の学校で統一して英検 IBA の結果をもとにした回答報告が行われていると考えられる。

ただし、さいたま市の驚異的な成果は、全学テの結果と比較するとやや印象が薄れる。2023 年全学テ・英語の平均正答率は、全国平均の 46%に対し、さいたま市は 53%であった(聞くこと・読むこと・書くことの 3 技能の平均)。もちろん、全国平均よりも 7 ポイント高く、全学テでもさいたま市が好調であることは事実だが、英語教育実施状況調査の 38.4 ポイント差という劇的な差ではない点に注意したい。

5年連続でさいたま市に次ぐ全国2位の福井県(A1レベル相当以上の中3生83.8%、2023年)についても確認しよう。同県は、好成績の要因として、(1) コミュニケーションを重視した授業改善、(2) 外部検定試験GTECの活用、(3) 外国語指導助手(ALT)の雇用、(4) 小学校における外国語教育先行実施の4点を挙げている(福井県教育委員会, 2023)。自治体自らが、外部試験の活用を成功要因として挙げているのが興味深い。福井県もさいたま市と同様にA1「取得」が82.0%と群を抜いており、外部試験の結果を効果的に「活用」していることが伺われる。全学テについても福井県は48%(全国5位)と好成績には違いないが、全国平均(46%)との差は2ポイントであり、英語教育実施状況調査における全国平均との大きな差に比べるとその差は限定的である。

以上、さいたま市・福井県の事例を検討した。他の調査結果なども鑑みると、2自治体の生徒の英語力が高いのは間違いないし、その背後に英語教師の努力があることにも疑問はない。しかし、英語教育実施状況調査の結果が正確に実態を反映しているかは別問題である。こうした問題含みの調査を根拠に、「〇〇県／〇〇市に学ぼう」といった政策論が展開されることに危惧せざるを得ない。事実、すでにそのような言論は展開されつつある(e.g., 細田, 2024)。現状では、具体的な教育施策が推進されるところまでは行っていないようだが、そうした問題のある政策過程につながる危険性を注視していくべきだろう。

5. 学界・学会とランキング

次に、学界・学会におけるランキング言説の状況を検討する。本節では、EFランキングに焦点を絞り、学術論文で同ランキングがどのように扱われているか検討する。日本国内で刊行された論文を調査するため、科学技術情報発信・流通システム(J-STAGE)を利用し、“EF English Proficiency Index”でキーワード検索を行った。検索結果から、研修報告書や学会発表要旨等を除外したところ、EFランキングを引用した13本の論文がみつかった。

各引用箇所を精査したところ、EFランキングの引用目的は次の2つに大別できることがわかった。ひとつは、特定の国のケーススタディで、その国の英語力の高さに言及するための引用である(13本中11本が該当)。たとえば、フィンランドの英語力の高さや、日本の英語力の低さを論じる文脈で、EFランキングを引用する論文が見られた。もうひとつは、複数の国の英語力を比較するために引用されるケースである(13本中2本)。前者の目的での引用が圧倒的に多いことがわかる。

次に、ランキングは論文のどの部分で引用されているだろうか。最も多いのが、序論や導入部分、つまり本題に入る前での引用である(13本中9本)。残り4本の内訳は、先行研究セクション(2本)とメソッドセクション(調査対象国を紹介する際の言及、2本)である。

以上の引用パターンからわかるのが、ランキング言説は、論文の主題ではなく、あくまで周縁的な位置づけに過ぎないことである。つまり、多くの研究者にとってランキングデータは研究の核心ではなく、あくまで特定の主張を補強するための一根拠に過ぎない。たとえば、韓国の英語力が高いとか日本の英語力が低いと主張したい研究者が、その根拠として引用する場合である。この時、引用者にとって重要なのは、自身の主張に合致するデータがあるか否かであり、データの質は検討していないだろう。データが収集・構築された文脈は考慮せず、単にその結果だけに注目しているに過ぎない。

ランキング言説が、論文において周縁的な位置しか占めていないのは、ある面では、かなり厄介な事態である。なぜなら、批判するにしても、研究への正面からの批判ではなく、むしろ重箱の隅をつつくかのようなならざるを得ないため、良心的な研究者は批判を躊躇してしまうかもしれないからである。しかし、こうしたデータ利用は、容易に伝播・増幅する。たとえば、日本の英語力の低さを支持するランキングをある論文で見かけた研究者が、そのランキングを自分の論文で引用する。そして、それを見かけた別の研究者がまた自分の論文で引用する。こうなると悪循環である。このように、多数の研究者が特定のデータを引用しているという事実が積み重なると、そのデータの質が担保されているかのように錯覚される恐れがある。各国の英語力プロフィールを正面から調査・研究しているわけでもないのに、引用事実の蓄積という非理論的・非学術的要因で、特定の主張の説得力が増すというのは、学術的コミュニケーションにとって深刻な事態である。

6. 建設的対案:結論にかえて

以上、本稿では、既存の英語力ランキングへの根本的な批判を行った。では、学界・教育界は、今後どのようにランキングと対峙していけばよいのか。本稿の締めくくりとして、建設的な対案を提示する。

対案の第一は、ランキングデータの利用をそもそもやめることである。様々なランキング・序列化が当然視される社会で暮らしているとつい忘れがちだが(Savski, 2023)、1点刻みのスコアで他国との比較を要求されるような知的場面は、実はほとんどない。日本の英語教育政策を論じる場合であれ、英語指導法を論じる場合であれ、たとえば、「日本の英語力はアジア(あるいは世界)最低だ」という情報は冗長である。最下位であろうが中程度であろうが、日本の英語教育に多くの課題があることは自明であり、日本の文脈をきちんと踏まえて改善の道を探っていけば良い。ましてや、実態を反映しているかどうかもわからない他国のスコアを準拠点にして自国の政策を論じるのは無意味であるばかりか、かなり危うい作業でもある。

それでも、数値的な国際比較が必要な場面はあるだろう。たとえば、学術的な国際比較研究である。この場合、TOEFL や EF 英語能力指数ではなく、より適切な調査データを利用すべきである。これが第二の対案である。理想的なのが、ランダムサンプリングによる国際英語力調査である(Terasawa, 2024a)。その最有力候補は、2025年にOECD学習到達度調査(Programme for International Student Assessment: PISA)の一環で行われる外国語学力調査である。国際機関主導の英語学力調査は歴史上初であり、きわめて貴重な調査になると期待されている。それだけに、日本政府が同調査への参加を見送ったのは、大きな損失と言わねばならない。日本政府・文科省には、不確かなデータ利用を自ら進んで行っている割に、信頼度の高い国際学力調査には不参加を決めるという矛盾した行動が見られる。また、PISA以外の選択肢として、英語力設問を含むランダムサンプリング調査を国際比較するというアプローチがあり得る。翻訳バイアスや自己報告バイアス(ほとんどの調査の英語力回答は現地語による自記式である)など比較可能性に課題は残るが、リサーチクエスチョンに適合すれば強力な手法である(cf. Terasawa, 2024a; 2024b)。

対案の第三が、学力調査をするならば、調査の本義から逸脱しない設計にすることである。具体的には、調査は、実態の正確な測定「のみ」を目的とし、それ以外の目的を含めないことである。この点で、現状の英語教育実施状況調査は、大きな問題がある。たとえば、調査をさせることで現場の意識改革を促す狙いや、自治体や学校を競わせるためといった狙いなど、実態解明以外の雑多な目的が混入している。別の目的が入るせいで、実態解明の精度が著しく下がり、二兎を追う者は一兎をも得ずという結果になってしまう。

幸い、同調査の問題点は明らかなので、改善への方途も明らかである。現状の全数調査を、ランダム抽出調査に変え、かつ、調査対象校の匿名性を担保すればよい。2023年調査では中3で約97万人、高3で約59万人の公立校全生徒を対象にしているが、この規模は、統計的推定にとって不必要に大きい。文科省報告書に記されている程度の示唆を得るには、この1%かそれ以下の回答者を抽出すれば、問題なく目的を果たせるだろう。また、抽出調査にすることで、全数調査で分散投入されていたリソースを集中投入できるようになり、よりきめ細かい調査・測定が可能になる。しかも、毎年調査が回ってくるわけではないとわかれば、現場の負担感も減る。また、調査対象校がランダムかつ極秘に選ばれば、自治体ランキングの意味合いは薄れるので、文科省や教育委員会に付度した回答を誘発することも抑制できるだろう。

数千人程度の抽出調査にすると、都道府県別集計値の誤差が大きくなってしまうことは事実である。しかしながら、そもそも都道府県別の集計が真に必要なのかを再考すべきである。たしかに英語教育を考えるうえで地域条件は重要な変数であるが、都道府県名よりも重要な変数は多数ある。たとえば、学校所在地の財政指数、人口規模、貧困世帯数、大学進学率等である。また、学校固有の条件(学校規模、貧困世帯出身児童の割合、英語科カリキュラム)も重要である。しかしながら、都道府県別集計は、こうした市区町村レベル・学校レベルの諸条件を曖昧にしてしまう。よりよい教育的・政策的示唆を得るうえでは、調査設計を根本的に考え直す必要があるだろう。

以上、本稿では英語力ランキング言説の問題点を、理論・方法論の面、および、具体的事例(教育現場および学界)の面から検討した。それを踏まえ、具体的な対案として、(1) ランキングを不必要に引用しない、(2) どうしても必要なら妥当性の高い調査を引用する、(3) 調査設計を再考する(具体的には、匿名性の高い抽出調査)という3点を提示した。本文でも述べたとおり、英語力ランキングは、大衆的な存在感が大きいわりに、根本的な批判が乏しい(とくに学术界からの批判は乏しい)。そうした状況に風穴を開け、よりよい学術的・政策的・教育的コミュニケーションの一助になることを願う。本稿は、紙幅の制約から、扱えなかった論

点も多い。たとえば、ランキング言説の実際の浸透度や、負の影響、あるいは実際の用例に関する批判的談話分析などの経験的分析は必要だろう。また、学界・教育界として、市民やマスメディアに向けてどのような啓蒙活動・科学コミュニケーション活動を行っていくべきかという点も重要な論点である(なお、萌芽的ではあるが、「英語業界のおかしなランキングを考える会」のような活動はある:<https://elt-ranking.jimdosite.com/>)。こうした点も含めて、本問題に広く議論が喚起されれば幸いである。

(関西学院大学・埼玉大学・北海学園大学)

引用文献

- ETS. (2022). *TOEFL iBT® test and score data summary 2021*. Retrieved from <https://www.ets.org/pdfs/toefl/toefl-ibt-test-score-data-summary-2021.pdf>
- 福井県教育委員会 (2023).「英語教育実施状況調査の結果について」. 2023 年 5 月 29 日
https://www.pref.fukui.lg.jp/doc/kyousei/kyoikuinnkai05_d/fil/1159-05.pdf
- 廣瀬雅代・稲垣佑典・深谷肇一(2018).『サンプリングって何だろう:統計を使って全体を知る方法』岩波書店.
- 細田眞由美 (2024).『世界基準の英語力:全国トップクラスのさいたま市の教育は何が違うのか』時事通信社.
- 川口俊明 (2020).『全国学力テストはなぜ失敗したのか:学力調査を科学する』岩波書店.
- 黒坂真由子 (2023).「正しい英語力測れないまりぼての『調査』『AERA』2023 年 10 月 2 日号、p.64.
- The Mainichi*. (2024). 'Unreliable': Why do Miyagi Pref. students rank last on Japan gov't English survey? July 24, 2024. Online Version.
<https://mainichi.jp/english/articles/20240724/p2a/00m/0na/013000c>
- 文部科学省 (2022).「英語教育・日本人の対外発信力の改善に向けて(アクションプラン)」2022 年 8 月 8 日. https://www.mext.go.jp/a_menu/other/mext_01982.html
- 文部科学省 (2024).「令和 5 年度「英語教育実施状況調査」の結果について」
https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1415043_00005.htm
- 日本経済新聞 (2023).「英語力、日本は過去最低の 87 位 若い世代で低下目立つ」2023 年 11 月 18 日、オンライン版. <https://www.nikkei.com/article/DGXZQOUE181BB0Y3A111C2000000/>
- Savski, K. (2023). Negotiating hegemonies in language policy: ideological synergies in media recontextualizations of audit culture. *Current Issues in Language Planning*, 24(1), 1–20.
<https://doi.org/10.1080/14664208.2021.2006945>
- さいたま市 (2023)「さいたま市中学生英語力 4 回連続全国 1 位になりました」2023 年 5 月 22 日.
<https://www.city.saitama.lg.jp/006/014/008/003/012/002/p097291.html>
- 佐藤郁哉. (2015).『社会調査の考え方(上・下)』. 東京大学出版会.
- Terasawa, T. (2024a). East Asia and English language speakers: a population estimation through existing random sampling surveys. *Asian Englishes*, 26(1), 84–105.
<https://doi.org/10.1080/13488678.2023.2191410>
- Terasawa, T. (2024b). Relationship between English proficiency and socioeconomic status in Asia: Quantitative cross-national analysis. *World Englishes, Online First*.
<https://doi.org/https://doi.org/10.1111/weng.12705>
- 読売新聞(2024).「英語能力指数 日本は87位...非英語圏113国・地域調査 最低ランク更新」2024 年 1 月 24 日. オンライン版 <https://www.yomiuri.co.jp/kyoiku/kyoiku/20240123-OYT8T50091/>