

「評価疲れ」に関する問題提起

渋井 進¹

¹正会員 大学改革支援・学位授与機構 研究開発部教授 (〒187-8587 東京都小平市学園西町1-29-1)

E-mail:shibui@niad.ac.jp

2004年の認証評価の施行および国立大学法人評価制度の導入から20年近くが経とうとしている。競争と評価によって改善を「自主的」に求められることは、今や大学にとって日常的な業務として定着した一方で、「評価疲れ」、「改善疲れ」、「徒労感」という評価に対する批判的な声も、評価制度導入時から継続的に報告されている。

しかし、「評価疲れ」は評価に対する疲労感、負担感等の不満を示す漠然とした用語であり、その原因や生じるメカニズムについても、ほとんど検討されて来なかった。本発表では、「評価疲れ」の概念について文献レビューや大学への調査を通して得られた事例等を挙げながら、「評価疲れ」の原因と想定されるいくつかの事柄について掘り下げ、どのように「評価疲れ」の克服・軽減が可能か展望する。

Key Words : Evaluation Exhaustion, University Evaluation, Measurement, Validity, Accountability

1. はじめに

本稿では、大学評価を中心として大学から報告されている「評価疲れ」の現状について解説し、問題提起を行う。文献調査、大学へのヒアリング調査を通じて得られた「評価疲れ」を引き起こす要因としての仮説を整理して提示し、どのように「評価疲れ」の克服・軽減が可能か展望する。

(1) 「評価疲れ」の用語の出現と変遷

「評価疲れ」の用語は評価制度の中でも比較的早期に導入された、行政評価制度において2000年代前半に報告されている。地方自治体の行政評価の普及は田中²⁾によると「1990年代末頃から行政評価制度を導入する自治体は顕著に増加し始め、2000年を過ぎる頃には『行政評価ブーム』」と形容されるような状況が現出した」とある。また、2002年には政策評価法が施行され、国のすべての行政機関において政策評価が実施されるようになった。

それ故、評価の検証等を通じて「評価疲れ」についても早くから報告されている。石橋³⁾においては、「『評価疲れ』という言葉に代表されるように、政策評価が自己目的化してしまう」との記載が見られる。ここでは評価結果が総花的で焦点が定まらないことから、評価に関連した書類作成の増加と、結果の活用が説明責任の点から形骸化・空洞化していることを問題として取り上げている。大学評価における「評価疲れ」についても、書類

作成の負担と結果の活用が不十分である点から見たモチベーションの低下が影響していると考えられるが、先行して実施されていた行政評価、政策評価においても既に報告されていることが読み取れる。

大学評価における「評価疲れ」については、2008年のいわゆる「学士力答申」³⁾の中で、「第三者評価制度との関連をどのように考えていくか、『評価疲れ』との批判もある中、機関別・分野別両者の効率的で実効ある評価の仕組みはどうあるべきか」と言及されており、この文脈での「評価疲れ」は、複数ある評価制度や重複から生じる大学の作業の負担を想定していると捉えられる。

また、2009年に設立され、現在まで活発に活動を続けている大学評価担当者を中心としたネットワーク組織である大学評価コンソーシアムは、その設立趣意書⁴⁾の中で、設立の背景の要因の一つとして「大学評価に係る過大なコストや『評価疲れ』が指摘される」ことへの方策の必要性を示している。ここでの「評価疲れ」も大学評価制度の導入当初に新たに大学に求められるようになった、評価作業の負担の増加を意味していると解釈できる。

茂木⁵⁾は行政評価に関する全国調査を行うことで、「評価疲れ」について扱っている。この中で、アメリカの評価学の領域においては「評価疲れ」(Evaluation Exhaustion)に関して議論されるよりも、「評価不安」(Evaluation Anxiety)について、試験や評価に対する教育心理学の知見を基に議論が進んでいると指摘している。さらに、Donaldson et al.⁶⁾を引用し、過剰な評価不安が評価制度の

有効性を損ね、職員の不満を生じさせ、徒労感としての「評価疲れ」を引き起こすという仮説を立てている。

大学の「評価疲れ」について国際的に共通に認識があるかについては、Shibui et al.⁷⁾は第1、2巡目に認証評価を受審した大学に対するアンケート調査結果を縦断的に比較し、大学の評価に対する作業負担感としての「評価疲れ」(Evaluation Exhaustion, Workload)が年々高まっていることを示してきた。これについて各国の評価機関の関係者が集まる国際会議等で発表⁸⁾され、一定の共感を得られており、実務関係者には大学の「評価疲れ」の存在は国際的に共通の認識となっているようである。

以上のように評価導入時から報告されていた「評価疲れ」であるが、ここ数年再び新聞記事やSNS等で活発に耳にするようになった。多くはいわゆる「選択と集中」に伴う会議出席や書類作成による教員の研究時間の減少を危惧する内容⁹⁾や、評価が本質的に持つ自己目的化してしまう性質¹⁰⁾を指摘するものである。

以上のレビューおよび、2022年4-5月に国立大学5校を対象に行ったヒアリング調査を踏まえ、次節では、「評価疲れ」を引き起こすと想定される4つの要因について呈示する。

2. 「評価疲れ」を引き起こすと想定される要因

(1) 評価の自己目的化

評価指標を設定することで、それが目的化することについての問題は多く報告されている。特に数値目標として設定することで、その傾向は強まることが知られている。木村¹¹⁾はJZミューラーの「測りすぎ」¹²⁾をはじめとして再燃した本問題について、「定量的な社会指標が社会的意思決定に使われれば使われるほど、腐敗の圧力に晒され、監視すべき社会的プロセスを歪めたり腐敗させたりする傾向が強くなる」というキャンベルの法則¹³⁾以来の議論を整理している。評価指標を設定することで、評価する側が意図しない悪影響や行動を結果的に引き起こす、というものが代表的と言える。

(2) 目標・計画の設定の問題

評価担当者の負担を増やす要因として、そもそもの目標と計画の設計が悪いという問題があることが、ヒアリング調査を通して報告されている。評価できないような計画を作成されてしまうと、大学の評価担当者だけではなく、評価者側も対応できず、結果として評価は機能しなくなる。

(3) 質の保証・向上(改善)とアカウントビリティ

評価においては社会に対する説明責任、アカウントビリティの遂行が、その目的の1つとして挙げられる。

大学には「大学自らの責任で説明し証明していく」とされる「内部質保証」が義務として求められており、国立大学法人運営費交付金の重点支援の評価¹⁴⁾においては、「『評価指標(KPI)』をそれぞれ主体的に設定し、PDCAサイクルの確立に努めながら自主・自律的な機能強化」を行うように求められている。

教育研究の質の保証・向上(改善)とアカウントビリティは評価における大きな2つの要素である。野田¹⁵⁾では、これらがかつては独立した概念として二項対立的な位置づけであったが、内部質保証が強調される今日では、組織が改善したことを社会に説明することがより重要視され、両者は接続的に捉えられるように変化したことを指摘している。その結果「改善したことを社会に示すために内部質保証システムを構築すること自体が自己目的化」することへの危惧も示している。

(4) 「動機づけ」の問題

動機づけは、外部要因である報酬や罰などに影響される「外発的動機づけ」と、内面から喚起される興味・関心や意欲に基づく「内発的動機づけ」に大別される。持続する実質的な改善は本来内発的動機づけに基づくはずだが、評価は外発的なものとして導入された。

外発的動機づけとして求められている評価の中で、「内部質保証」に代表されるように、自主的な改善を求められている現状は、大学の評価に対する動機づけを高めるという観点からは、困難な構造になっていると思われる。

3. 「評価疲れ」の克服へ向けた可能性

以下では、前項で述べた4つの要因ごとに、今後の評価疲れの克服・軽減に向け、どのように対処が可能かについて考察を行う。

(1) 評価の自己目的化に対して

評価が自己目的化する問題は、完全には解決できないジレンマと言える。一方で、「ランキング」¹⁶⁾の著者P. エールディはJZミューラーの「測りすぎ」における事例や結論の大部分に賛成するとしながらも、計測指標を用いることによって果たされる説明責任のメリットを挙げ、計測指標やランキングなどの定量的分析を完全に放棄すべきではないと指摘している。

目的・目標に対して指標を設定して測定する際には、その指標が測定される際に周囲にどのような結果を生むかについて、評価を設計する側はもちろんのこと、学内で指標・エビデンスを収集する際にも様々な可能性を想定する必要がある。アウトカムをKPI(Key Performance Indicator)として可能な限り数値的に定義して測定すること

がここ数年の大学評価においては望ましいとされる流れとなっているが、KPIの向上のみが目的となってしまう、他の基礎的な取組みが犠牲になるような設計は避けるようにする責任が、評価を行う側にはあるだろう。

また、複数の評価対象となる大学等に指標を用いて相対的な比較を行う場合には、その数値の定義が明確であることや、データの信頼性が担保できるものであることや、そもそも同列で比較できる性質のものであるかなどを考えることも重要である。

(2) 目標・計画の設定の問題に対して

評価が適切に行えるような目標・計画の作り方のスキルを大学が身につけることで、負担の軽減になる可能性がある。この問題に対して、J. Wholey¹⁾によって提案されている、Evaluability Assessment (評価可能性のアセスメント)の手法を大学評価に適用する取組が、田中²⁾をはじめとして行われている。評価可能性を備えた目標・計画の作り方についての教材を作成し、それらを用いて大学の評価担当者への研修を行うことで、自己評価能力の向上を図る取組である。このような研修活動による支援は、現場の評価作業に伴う「評価疲れ」の軽減へ向けて有効であると考えられる。

(3) 質の保証・向上(改善)とアカウンタビリティに対して

内部質保証に関して動機づけが低下する可能性を述べたが、内部質保証「システム」として全体が機能しているか評価機関が確認する点で、実質的な教育内容・方法等は大学に委ねる方向にあり、自主性を尊重する内容となっているとも解釈できる。筆者らが行った大学へのヒアリング調査においても、「評価疲れ」を「評価不信感」という形で述べる声があった。そこでは、「運営費交付金の重点支援の評価に関連し、KPIに関する共通指標について集計方法を決めて全大学で統一すべき。」「論文のカウンターの仕方を整理せずに、信頼できないデータをもとに数だけが問われ順位化され、予算にも反映されるという不公平感が残る。」という、データの信頼性や評価の公平性・透明性に関連した意見があった。

以上の意見は、大学は評価を拒否しているのではなく、きちんとした評価を遂行してほしい、という積極的な立場を示しているとも読み取れる。この解釈と関連したデータとして、渋井³⁾において、第2期法人評価のヒアリング調査におけるテレビ会議の利用に際し、評価者及び大学の意識を調べたものがある。そこでは、90法人の内、テレビ会議によるヒアリングを希望した大学等が半数以下の36法人に留まったデータが示されており、評価される側の大学が、予算に影響する数年に1度の評価を、直接会って行いたいと感じていたことが考察できる。また、

テレビ会議では「臨場感がない」、「真意が伝わらない」、「場の雰囲気共有できない」などの単にヒアリングにおいて情報を伝えるだけではなく、コミュニケーションにおける心理的な側面を重視する意見も得られている。

コロナ禍で訪問調査が一昨年度以来オンラインが主になっていることも、コミュニケーションを阻害する要因となっている可能性がある。権威のある評価委員がコストをかけて大学まで訪問し、きちんと丁寧に評価することは、大学の評価に対する信頼を高めると思われる。

(4) 「動機づけ」の問題に対して

そもそも多くの大学関係者、特に組織としての大学の改善等を主導する立場ではない一般の研究者にとって、評価は面倒臭いものに過ぎず、可能な限り早く終わらせたい「作業」であろう。例えば、学内の教員評価へむけた書類を作る際に、教員のモチベーションは高まらない。しかし、外部資金獲得のための書類や審査のための書類は自身の研究環境に直結するため、そこで「評価疲れ」という声はあまり耳にしない。外発的な動機づけであったとしても、プラスの報酬が得られるものであれば、行動の持続性は高まると言える。

一方で、大学人は評価という文脈でなければ、改善活動自体については本来否定的ではなく、内発的動機づけが機能していると思われる状況も見られる。例えば、FD(ファカルティ・ディベロップメント)活動への参加も近年大学には定着している。

教員個人が本来持っていた内発的動機づけに基づく改善活動へのモチベーションが、外発的に組織としてトップダウンで「内発的に」行うことを強制されることにより低下してしまう可能性もある。外発的動機を与えることにより、内発的動機づけが低下する現象は、行動経済学の分野においてアンダーマイニング効果として知られている。これを防ぐためには、すでにうまく行っている改善に向けた取組に対しては、大学の自主性に任せて無理に介入しないことも必要であり、それを適切に把握するためには大学の現場と政策立案者の間での十分な情報共有に基づく信頼関係の形成が大事であろう。

参考文献

- [1] 田中啓(2009)。「日本の自治体の行政評価」『分野別自治制度及びその運用に関する説明資料 No.14』財団法人自治体国際化協会(CLAIR), 政策研究大学院大学比較地方自治研究センター(COSLOG).
- [2] 石橋順三(2007)。「政策評価制度～制度運用の課題と展望～」『立法と調査』, 269, 30-40.
- [3] 学士課程教育の構築に向けて(答申)(平成20年12月24日 中央教育審議会)

- [4] 「大学評価コンソーシアム」設立趣意書（案）, H21. 11. 5.
<http://iir.ibaraki.ac.jp/jcache/documents/prospectus.pdf>
 (2022年9月15日アクセス)
- [5] 茂木康俊 (2015). 「行政評価ブーム後の自治体における政策評価制度の変化と課題：『評価不安』概念を手がかりにした全国調査に基づく『評価疲れ』の検討」『地方自治研究』, 30, 1-17.
- [6] Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). "Strategies for managing evaluation anxiety: Toward a psychology of program evaluation." *American Journal of Evaluation*, 23(3), 261-273.
- [7] Shibui, S., Takahashi, N., & Noda, A. (2017). "A Longitudinal Study of Effectiveness, Impact, and Challenges in the Japanese Quality Assurance System." *International Journal of Institutional Research and Management*, 1, 83-102.
- [8] Shibui, S. (2021). "A longitudinal study on the levels of awareness among universities regarding Certified Evaluation and Accreditation" INQAAHE Conference 2021, Glasgow.
- [9] 削られる大学教員の研究時間、会議など負担に 政府が対策検討. 朝日新聞デジタル. 2022-05-25.
- [10] 数値目標による評価 「測りすぎ」ていないか？ 長谷川真理子・総合研究大学院大学長. 毎日新聞, 朝刊. 2, 2022-05-29.
- [11] 木村拓也 (2022). 「キャンベルの法則について：測定評価を巡る構造的陥穽に関する考察」『九州大学教育社会学研究集録』, 23, 57-64.
- [12] Muller, J.Z. (2018). *The Tyranny of Metrics*. New Jersey: Princeton University Press. (ジェリー・Z・ミュラー. 松本裕(訳) (2019). 測りすぎ なぜパフォーマンス評価は失敗するのか, みすず書房)
- [13] Campbell, D.T. (1979). "Assessing the Impact of Planned Social Change", *Evaluation and Program Planning*, 2, 67-90.
- [14] 文部科学省高等教育局国立大学法人支援課 「令和3年度国立大学法人運営費交付金の重点支援の評価結果について」 https://www.mext.go.jp/content/20210924-mxt_hojinka-000017262_1.pdf (2022年9月15日アクセス)
- [15] 野田文香 (2020). 「認証評価の課題と可能性」『高等教育研究』, 23, 33-52.
- [16] Érdi, P. (2020). *Ranking: The Unwritten Rules of Social Game We All Play*. Oxford: Oxford University Press. (ペーテル・エールディ. 高見典和(訳) (2020). ランキング 私たちはなぜ順位が気になるのか?, 日本評論社)
- [17] Wholey, J. (2004). "Evaluability Assessment", in Newcomer, K. et. al. *Handbook of Practical Program Evaluation*, New Jersey: Wiley, 33-62.
- [18] 田中弥生 (2009). 「評価可能性のアセスメント (Evaluability Assessment) ～大学の自己評価能力向上のために～」『大学評価・学位研究』, 10, 25-44.
- [19] 渋井進 (2018). 「大学評価の面談におけるテレビ会議の活用と効果」『日本顔学会誌』, 18(1), 23.

RAISING THE ISSUE OF "EVALUATION EXHAUSTION"

Susumu SHIBUI

Nearly 20 years have passed since the implementation of accreditation evaluation in 2004 and the introduction of the national university corporation evaluation system. While the "voluntary" demand for improvement through competition and evaluation has now become a routine part of the university's work, critical voices about evaluation such as "evaluation exhaustion," "improvement exhaustion," and "feeling of fatigue" have been continuously reported since the introduction of the evaluation system.

However, "evaluation exhaustion" is a vague term that indicates dissatisfaction with evaluations, such as fatigue and a sense of burden, and its causes and mechanisms have rarely been examined. This presentation will discuss the concept of "evaluation exhaustion" through a literature review, a survey of universities, and will explore some of the possible causes of "evaluation fatigue" and how it can be overcome or alleviated.