

## まえがき

---

現代は、「測りすぎ」の時代といわれる。歴史学者ジェリー・ミュラー (Jerry Z. Muller) の著作『測りすぎ』は、大学、学校、医療、警察、軍、ビジネスと金融、慈善事業と対外援助などさまざまな領域での測りすぎ・測りまちがいを描き出して話題になった。ミュラーが大学を最初の検討対象として取り上げていることにも見てとれるように、教育の世界でも、とくに2000年代に入ってから、「エビデンスに基づく政策・実践」や「説明責任 (アカウントビリティ)」の名の下に、教育成果や学習成果の測定・評価が強く要請されるようになってきた。この時代において、評価を、教える側、学ぶ側の双方にとって意味あるものにするには果たして可能なのだろうか。そのためには何が必要なのだろうか。

本書では、大学教育を主な舞台としながら、「パフォーマンス評価」を中軸にすえて、その具体的な方法を模索し提案する。

第1章ではまず、学習評価とは何かを考える。評価はどのような構成要素によって成り立っているのか。日本語の「評価」にあたる英語の「アセスメント (assessment)」と「エバリュエーション (evaluation)」の違いは何なのか。評価は誰が行うのか。評価にはどんな原則や要件があるのか。こうした問いについて答えた上で、「学習としての評価」という考え方を、「学習の評価」「学習のための評価」と対比しながら提示する。「学習としての評価」はカナダの評価研究者ローナ・アール (Lorna M. Earl) の用語として日本では知られているが、本書ではその意味の拡張を図る。

続く第2章では、「測りすぎ」にもつながりがちな「学習成果の可視化」というテーマを取り上げよう。そして測りすぎ・測りまちがいが生じないようにするために、多様な学習成果の評価を分類し、それぞれの特徴を明らかにする枠組みを示す。

第3章では、「学習としての評価」を提案するための布石としてパフォーマンス評価とルーブリックについて論じる。こういうふうにと書くと、意外に思われる方もいるかもしれない。パフォーマンス評価やルーブリックは、「測りすぎ」の象徴のようにみなされることもあるからだ。実際、ミュラーの『測りすぎ』の副題は「なぜパフォーマンス評価は失敗するのか?」となっている。ただ、これは邦訳のみの副題であり、ミュラーの本の中で、本書でいう意味での「パフォーマンス評価」に対する批判は行われていない。本書でいうパフォーマンス評価とは、各分野での能力を最も真正な形で発揮できるパフォーマンス（作品や実演など）を手がかりに、学習者の学びや能力を評価する方法のことである（たとえば、自動車を運転する能力なら、ペーパーテストではなく実際に運転させてその能力を見るのがパフォーマンス評価である）。ルーブリックはそのパフォーマンス評価を行う際の評価基準の一つの形態であり、ルーブリックだけに焦点を当てるべきではない（その意味で、日本の教育界に広まっている「ルーブリック評価」という言い回しは誤解を招くものだ）。

「測りすぎ」の時代に多用されているのは、質問紙調査と標準テストである。結果を数値化・縮約化して、他の個人・組織・国との比較や説明責任の遂行などのために使いやすいからだ。パフォーマンス評価によって生徒・学生の学びや能力を把握しようとする試みは、主に標準テストに対抗する形で行われてきた。第4章では、私たちが大学で行ってきたパフォーマンス評価の事例を紹介し、それが確かに「学習としての評価」として機能していることを示したい。

続く第5章では、第2章で示した枠組みを用いながら、測りすぎ・測りまちがいが実際にどんな形で起きているのかを、学生調査やジェネリックスキルの標準テストの例を挙げて検討していく。

学生調査や標準テストは、学士課程全体の学びと成長を評価するために使われてきた。では、パフォーマンス評価でそれは可能なのだろうか。第6章では、4（ないし6）年間の学位プログラムの学習成果を評価する試みとして、「重要科目に埋め込まれたパフォーマンス評価（Pivotal Embedded Performance Assessment: PEPA）」の考え方を実践の具体例とともに提案する。

評価を、外から批判するのではなく、内から問い直し再構成していくという

のはなかなかの難題である。もしかすると、ミイラ取りがミイラになる危険性もある。本書がミイラにならずにすんでいるか、学習評価を意味あるものになっているか、読者の判断を仰ぎたい。