

# プログラムレベルの学習成果の評価 ——総和と軌跡——

松下 佳代

京都大学高等教育研究開発推進センター教授

[キーワード] 学習成果、プログラムレベルの評価、ミネソタ大学、PEPA (重要科目に埋め込まれたパフォーマンス評価)、総和と軌跡

## I. 問題と目的

学生が4 (ないし6) 年間の学士課程でどう学び、成長したかに関心のない大学教員はまれだろう。だが、どうすればそれを把握できるのだろうか。医療系であればそれほど難しくはない。ほぼ前提知識やスキルがない状態から、臨床実習に出て初歩的ながらも診療やケアに参加し、国家試験に合格する。それは学士課程教育がなければなしえなかったことだ。しかし、人社会になるとその把握はずっと難しくなる。初年次セミナーで担当した学生の卒業論文の指導・審査にあたる、あるいは研究室のゼミで長期にわたって指導するというようなことがあれば、その学生の成長を感じることもできるが、そのような機会に恵まれる教員の割合は限られている。

本稿の目的は、プログラムレベルの学習成果をどう把握し評価するかという問題について、〈総和 (sum)〉と〈軌跡 (trajectory)〉という2つの考え方を示し、それに基づいて具体的な評価方法を提案することにある。

ここでいう「プログラム」とは、学位プログラム、なかでも学士号という学位を与えるために編成されたプログラムのことであり、学士課程教育のうち正課教育の部分にほぼ対応する。2020年1月に出された「教学マネジメント指針」においても、「学位を与える課程 (学位プログラム) が、学生が必要な資質・能力を身に

付ける観点から最適化されているかという『学修者目線』で教育を捉え直す」(p. 1) ことの必要性が唱えられている。プログラムレベルの評価方法については、文科省の調査「大学における教育内容等の改革状況について」や朝日新聞・河合塾の調査「ひらく 日本の大学」でも統計がとられているが、本稿では、それを〈総和〉と〈軌跡〉という観点から検討する。〈総和〉とは、プログラムレベルの学習成果を個々の授業科目の学習成果の総和として (時間軸を捨象して) 把握するという考え方であり、一方、〈軌跡〉とは、プログラムレベルの学習成果を学生の学習の進捗にそって (時間軸を入れて) 把握するという考え方のことである。本稿の主張を先取りしていえば、現在のプログラムレベルの学習成果の評価は、〈総和〉による方法に偏っているが、そこにはいくつかの難点があり、それを〈軌跡〉による方法で補完する必要がある、ということになる。

## II. 学習成果とその評価

### 1. 「学習成果」の意味

そもそも「学習成果」<sup>(1)</sup>とは何だろうか。この用語が日本の大学教育に普及するきっかけとなった2008年の学士課程答申では、「プログラムやコースなど、一定の学習期間終了時に、学習者が知り、理解し、行い、実演できることを期待される内容を言明したもの」(用語解説)と定義している。「プログラム」とは主に学位プログラムを指し、「コース」とはそれを構成する各授業科目を指す。OECDなどでもほぼ似たような定義がなされており、この文科省の定義は、国際的な用語法をふまえたものといえる。ここで注意すべき

は「期待」という言葉である。日常用語としての「成果」は「結果」を指すが、専門用語としての「学習成果」は「目標」と「結果」の2つの意味を含むのである。英語では、〈目標としての学習成果〉はintended/expected learning outcomes、〈結果としての学習成果〉はachieved learning outcomesと表記されることもある。

さて、学習成果にこのような2つの意味があるとすると、学習成果の可視化を図る際にも、〈目標としての可視化〉と〈結果としての可視化〉の両方が求められることになる。

### (1) 目標としての可視化

〈目標としての可視化〉とは、プログラム全体・各科目のカリキュラムにおける目標を設定して、目に見えるようにすることである。例えば、AAC&U (Association of American Colleges & Universities: 全米大学・カレッジ協会) は、大学教育にとって本質的で不可欠の学習成果を Essential Learning Outcomes (ELO) と呼び、①人類の文化や自然界についての知識、②知的・実践的スキル、③個人的・社会的責任、④統合的学習の4つの柱に整理した(AAC&U, 2007)。①、②、③はそれぞれK (Knowledge)、S (Skills)、A (Attitudes) に対応しており、古典的なKSAモデルに立っているが、④に「統合」が加えられているのが特徴的である。つまり、知識、スキル、態度をそれぞれ別々に獲得させるだけでなく、これらを統合して課題に向きあい行為する能力の育成が重視されているのである。

このELOの構造や内容は文科省の「学士力」にも引き継がれている。

### (2) 結果としての可視化

〈結果としての可視化〉とは、学習の結果、身についた資質・能力を、さまざまな評価方法によって、目に見えるようにすることである。プログラムレベルの学習成果はほとんどが二値的(正か誤か)には評価しにくい資質・能力である。そこで、AAC&UはELOを評価するためのツールとして、VALUEルーブリックを開発した(AAC&U, n.d.; Rhodes & McConnell, 2017)。そこには、「探究と分析」「文章コミュニケーション」「倫理的な推論と行為」「統合的学習」など16のルーブリックが含まれている。

一方、日本の大学教育では、2012年の質的転換答申で、「学修行動調査やアセスメント・テスト(学修到達度調査)、ルーブリック、学修ポートフォリオ」といった手法が取り上げられたものの、アセスメントが第4のポリシーに加わるにはいたらず、「アセスメントプラン」としてカリキュラム・ポリシーの中に組み込まれることになった。こうして、プログラムレベルの学習成果の評価については、今なおさまざまな模索がなされている段階にある。

## 2. 学習成果の評価の多様性

学習成果の評価には多様な方法があるが、それらは、〈直接-間接〉、〈量的-質的〉、〈機関レベル-プログラムレベル-科目レベル〉といった軸で分類することができる(松下, 2017)。このうち、直接評価(direct assessment/measures)とは、学習者の知識や能力の表出を通じて「何を知り何ができるか」を学習者自身にやってみさせることで一、学習成果を直接的に評価する方法のことであり、間接評価(indirect assessment/measures)とは、学習者による学習成果についての自己報告を通じて「何を知り何ができるか」と思っているかあるいは「どのように学習を行っているか」を学習者自身に答えさせることによって一、学習成果を間接的に評価する方法のことである。例えば、人社系の成績評価でよく用いられるレポート評価は、科目レベルの質的な直接評価であり、全学的な段階評定法による質問紙調査は、機関レベルの量的な間接評価ということになる。こうした分類軸を用いることで、多様な評価方法を組み合わせながら、学習成果の多面性を把握することが可能になる。

## Ⅲ. 〈総和〉と〈軌跡〉という2つの考え方

### 1. プログラムレベルの評価の現状

プログラムレベルの評価として、従来どんな方法が用いられてきたのだろうか。文科省や朝日新聞・河合塾の調査に基づけば、主な方法として、学生調査(学修行動調査など)、標準テスト、成績評価の活用(GPAなど)、卒業論文・卒業研究、学修ポートフォリオなどを挙げることができる(松下, 2020)。

それぞれの評価方法には、メリットの一方で限界もある。例えば、学生調査は、朝日新聞・河合塾(2018)によると、全学実施の割合が2番目に高い方法となっているが(63%)、学生の自己報告による間接評価であるため、直接評価の代替にはならない。例えば、WSCUC(WASC Senior College and University Commission)のガイドブックでは、「学生の学習のエビデンスは、調査(コンピテンスや成長についての学生による自己報告)以上のものを含むべきである。[中略]学生の学習のエビデンスを提供するには、学生の学習の成果物の直接評価の結果を含む、より多くの異なるタイプのエビデンスが求められる」(WSCUC, 2015, p. 16)とされている。

## 2. 〈総和〉による方法

では、直接評価の方法にはどんなものがあるだろうか。朝日新聞・河合塾(2018)において全学実施の割合が最も高い(92%)のは、「成績評価の活用(GPAなど)」である。だが、各授業科目の成績評価を寄せ集めただけでは、プログラムレベルの評価にはならない。また、GPAは一定のプログラム期間内の成績評価の代表値ではあっても、DP(ディプロマ・ポリシー)に掲げられた複数の学習成果それぞれの達成度を示すものではない。

そこで用いられるのが、カリキュラムマップやレーダーチャートである。カリキュラムマップは、プログラムレベルの学習成果が各科目によってどう達成されるかをマトリックス形式で示した表であり、レーダーチャートはその結果を一つの図で表したものである。とりわけAP(大学教育再生加速プログラム)事業を通じて多くの大学に普及しつつある。カリキュラムマップやレーダーチャートの背後にあるのは、「個々の科目の学習成果の総和が、学習プログラム全体としての学習成果となる」(González & Wagenaar, 2008, 邦訳 p. 33)というアイデア、つまり〈総和〉の考え方である。

だが、〈総和〉による方法には次のような難点がある(松下, 2020)。

- ①カリキュラムマップで描かれるプログラムレベル

の学習成果は、DPで記述されるようなかなり抽象度の高い資質・能力である。それらは多義的で曖昧であり、科目の内容によって、また担当教員の解釈によって異なる意味をもちうる。「全く異なるさまざまなスキルや能力の緩やかな集合体に割り当てられた名称にすぎない」(Kosslyn & Nelson, 2017, p. 29)ことが多いのである。プログラムレベルの学習成果の意味が科目によって異なるのだとすれば、それを数値化して足し合わせることにどんな妥当性があるのだろうか。

②〈総和〉による方法は、時間軸を捨象している。だが、多くの学習成果は、プログラム期間内で大きく変化する。例えば、「論理的に文章を書く力」のような学習成果は、1年次の初年次教育科目と4年次の卒業論文のどちらにも含まれるが、卒業時に学生が身に付けている力は卒業論文の方でより直接に示されているはずである。だが、〈総和〉による方法では、こうした資質・能力の変化は考慮に入れられない。

評価研究者のサドラー(Sadler, 2010)は、累積的評価(cumulative assessment)の問題点を同一の達成にいたる2つの経路によって明示している(図1)。学生AとBは、T4の時点(科目終了時)で同一の能力について同一の達成を示した。だが、もし累積的評価を行って、T1、T2、T3の時点での評価も均等に足し合わせるとすれば、Bの総和はAの65%にしかならない。これは不適切だとサドラーはいう。ここで注意しておきたいのは、サドラーが批判しているのは、T1～T3をすべて最終的な評価に組み入れることに対し

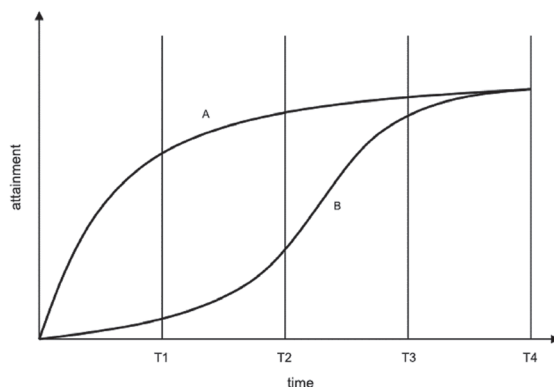


図1 2人の学生の学習経路 (Sadler, 2010, p.736)

てであって、その経路 (path) を知ること自体は意味があると考えていることである。つまり、〈総和〉による方法への批判であって〈軌跡〉による方法への批判ではないということだ。

### 3. 〈軌跡〉による方法

〈軌跡〉による方法とは具体的にどのようなものだろうか。「卒業論文・卒業研究」は、〈軌跡〉による方法の性格を内包している。卒業論文・卒業研究は、学びの軌跡を経た上で卒業時まで形成している資質・能力を総合的に反映していると考えられるからだ (図1のT4にあたる)。多くの大学・学部でDPに掲げた目標の大半が卒業論文・卒業研究に直接紐付けられているのは、そのような理由によるのだろう。また、「学修ポートフォリオ」も、根拠資料とともに学生の学びの軌跡が把握できる点で、まさしく〈軌跡〉による方法だといえる。だが、「卒業論文・卒業研究」は4年次まで把握できず、評価が主観的になりやすい。また、「学修ポートフォリオ」は、多様な根拠資料を含むため、目標としての学習成果との対応づけが適切になされないことも少なくない。

AAC&UのVALUEルーブリックは、16項目の学習成果 (ELO) について、4年間の学びの軌跡の把握を可能にするルーブリックである。各ルーブリックは、それぞれ5~6個の次元 (dimension) と5つのパフォーマンス・レベル (最も高い「キャップストーン」、2つの「マイルストーン」、「ベンチマーク」、およびそれを満たしていない場合の「ゼロ」) からなり、各次元・レベルに対応する記述語が記されている。つまり、VALUEルーブリックは、複数の次元で学習成果を分析するという意味で「分析的ルーブリック」であり、また、長期間にわたる学習成果の変化を把握するために「長期的ルーブリック」でもある。

VALUEルーブリックは、2009年の発表以来、米国での機関・プログラムレベルの評価に大きな影響を与えてきた。NILOA (National Institute for Learning Outcomes Assessment) のレポートによれば、米国において機関・プログラムレベルで学生の学習を評価するために用いられている方法は、2017年時点で、多

いものから順に、全米学生調査(76%)、ルーブリック(71%)、教室単位のパフォーマンス評価(64%)となっている。特に2009年の調査結果と比較すると、ルーブリック、教室単位のパフォーマンス評価の増加ぶりが目立っている (Jankowski et al., 2018)。

## IV. 〈総和〉と〈軌跡〉をどう組み合わせるか

### 1. ミネルヴァの場合

#### (1) 目標としての学習成果—汎用的能力の階層化—

先ほど、累積的評価に対するサドラーの批判を紹介したが、そこで対象とされていたのは科目レベルの評価であって、プログラムレベルの評価ではない。サドラー自身もいうように、学位プログラムは相対的に独自の数多くの科目からなり、科目によって目標となる知識や能力は異なるので、図1ほど単純ではない。

〈総和〉による方法の難点をふまえた上で、〈軌跡〉による方法とどう組み合わせるかを意識的に追求していると考えられるのが、ミネルヴァ大学である (Kosslyn & Nelson, 2017; 松下, 2019)。

ミネルヴァでは、抽象的で多義的・曖昧になりがちな汎用的能力を100近い<sup>(2)</sup>「知の習慣 (habits of mind)」と「基本的概念 (foundational concepts)」(まとめてHCsと呼ばれる) に分析し、4年間の正課教育や準正課活動を通じて習得できるよう、カリキュラムが体系化されている。ミネルヴァが目的に掲げるのは「実践知 (practical knowledge)」であり、それを獲得させるために、汎用的能力が次のように階層化されている。

- ・上位階層：4つのコア・コンピテンシー (批判的思考、創造的思考、効果的コミュニケーション、効果的インタラクション)
- ・中位階層：コア・コンピテンシーの具体的な側面 (例えば、「批判的思考」の場合は、A. 主張を評価する、B. 推論を分析する、C. 決定を比較考量する、D. 問題を分析する、の4つ)
- ・下位階層：個々のHCs (例えば、#audience (= 文脈やオーディエンスにあわせて口頭や文書での表現の仕方を変える)、#sourcequality (= 出典の質を決定するためにカテゴリと情報のタイプを区別する) など)

一般に「汎用的能力」(あるいは「ジェネリックスキル」)といわれるものはここでいう上位階層レベルの能力であるが、ミネルヴァではそれをHCsのレベルにまで具体化し、1年次の4つの「コーナーストーン科目」(すべて必修)でいったんすべて学ぶ。その上で、その後の専門教育や準正課活動、さまざまなプロジェクトなどにおける多様な文脈でそれらのHCsを意識的に活用しながら、適用範囲を広げ、習熟度を深めていくことが意図されている。

### (2) 学習成果の評価—1次元の長期的ルーブリック—

学習成果の評価においても中心に据えられているのはHCsの評価である。ミネルヴァでは、すべてのHCsで共通のルーブリック・テンプレート(Kosslyn & Nelson, 2017, p. 242)が使われている。このルーブリックは、1(知識の欠如)・2(浅い知識)・3(知識)・4(深い知識)・5(深遠な(profound)知識)の5段階からなる、1次元のみのルーブリックであり、また、4年間使われる長期的ルーブリックである。そのため、あらゆる成果物に適用できるような「柔軟性」と成長の度合いが把握できる「幅」をもつようデザインされている。各HCのルーブリックはこのテンプレートをカスタマイズして作成される。

先にVALUEルーブリックについて紹介したが、両者を見比べると、ミネルヴァのルーブリックは、1つのVALUEルーブリックの中の1つの次元に対応すると考えられる。例えば、文章コミュニケーションのVALUEルーブリックには、「文章作成の文脈と目的」「内容の展開」「ジャンルと学問分野の約束事」「資料(source)と根拠」「構文と技法を操ること」の5つの次元が含まれるが、「文章作成の文脈と目的」(オーディエンス・目的や文章作成課題をとりまく状況の考慮を含む)は、ミネルヴァのHCsの#audience、「資料と根拠」は#sourcequalityとほぼ重なる。つまり、AAC&Uでは、包括的な汎用的能力を目標とし、その分、分析的なルーブリックを用いるのに対し、ミネルヴァでは、目標とする汎用的能力をHCsにまで細分化・具体化した上で、1次元のみのルーブリックを用いるのである。ただし、両者とも、長期的ルーブリックによって、学生の学習の進捗(progress)の把握を行おうとし

ている、つまり〈軌跡〉による方法をとっている点で共通している。なぜ、このような長期的ルーブリックが必要になるのか。それは、ELOやHCsのような学習成果が、長期間をかけて形成され発達していくものだからである。大学・教員は、学位や単位の認定のために「一定の学習期間終了時」に評価を行うが、その形成・発達は生涯にわたって続く。

さて、ミネルヴァの1年次のコーナーストーン科目では、授業中の観察、大小さまざまな課題、学期末の最終プロジェクトなどで、このルーブリックを使いながら、HCsの評価が行われ、1年次の終わりにひとまず各科目の合否判定が行われる。2年次以降、専門教育に入ってから、科目やプロジェクトの中でHCsが活用されるたびにHCsの評価は更新され続ける。そして卒業時によろやく、すべてのHCsの得点をもとに、各コーナーストーン科目の評語が決定され、それが学士課程全体を通じて育成された学生の汎用的能力の習熟度を示すものとなる(「“time-traveling” grades」と呼ばれている)。一方、専門教育科目では、HCsだけでなく、その分野の個別特殊な知識やスキルもあわせて評価されるため、科目の成績評価は学期ごとに行われる。

### (3) ミネルヴァにおける〈総和〉と〈軌跡〉の組み合わせ方

ミネルヴァの学習成果とその評価においてまず興味をひかれるのは、〈軌跡〉による方法が実装されている点である。汎用的能力(HCs)はいったん1年次にすべて習得されるが、その後もさまざまな文脈の中で適用され、そのつど評価される(教師が目標として組み込んでいるHCsに加えて、学生自身が科目やプロジェクトに関連するとして申し出るHCsもある)。こうして卒業時までの学士課程全体を通じて、HCsがどのように広さと深さを増していくかというその進捗が可視化され共有されるのである(HC assessment dashboard上で教師も学生もいつでも見られるようになっている)。

ただし、〈軌跡〉による方法だけですべての学習成果が評価されているわけではない。まず、専門教育の科目では、HCs以外の分野固有の知識やスキルは、通

常の大学の成績評価と同じく科目ごとに評価される。また、上記のルーブリックを用いて評価された各HCの得点も、課題によって異なる重み付けを与えられ加重平均されて足し合わされる(=〈総和〉による方法)。プロジェクトのように課題が総合的・包括的になるほど、また学期が上がるほど、重み付けは増す。

ところで、こうしたHC得点の足し合わせは、前にみたサドラーの批判する「累積的評価」にあたらぬのだろうか。サドラーの批判は、科目終了時に同一の能力が同じレベルに達した場合に、その途中段階の評価結果を累積的に加算することに向けられていた。一方、多様な文脈でHCsが適用される中で獲得される知には、“Knowing that”や“Knowing how”だけでなく、“Knowing with”も含まれると考えられる(Bransford & Schwarz, 1999)。“Knowing with”とは、ある「知」を適用する文脈についての知のことである。単に「知」を適用するだけでなく、適用することを通じて「知」の方が創り変えられることもある。HCsそのものは“Knowing that”“Knowing how”にあたるが、それをさまざまな文脈へ転移していくことで“Knowing with”という新たな知が生成されるのだとすれば、サドラーのいう「同一の能力」ということではなくなり、加算することに合理的な根拠があることになる。

このように、ミネルヴァでは、何を〈軌跡〉によって、また何を〈総和〉によって把握すべきかが明確に意識され、両者の組み合わせでプログラムレベルの学習成果の評価が行われている。

## 2. PEPAの場合

ミネルヴァのカリキュラムと評価方法はよく練られているが、特定の教育理念の下でラディカルなカリキュラム改革を求めるものであり、日本の多くの大学にそのままあてはめることは難しい。日本の大学で〈総和〉と〈軌跡〉を組み合わせる方法として、「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)」(松下・小野・斎藤, 2020)を紹介しよう。ここでいう「重要科目」とは、その授業科目の目標がプログラム全体の目標に直結する科目(それまでに学んだ知識やスキルを

統合し、高次の能力を育成・発揮することを求める科目)のことである。4(ないし6)年間の学位プログラムにはそのような重要科目がいくつか含まれているはずである。そういう科目は必修になっていて、複数の教員がチームとして取り組み、学生にもテスト以外の何らかのパフォーマンスを求めていることが多い。「埋め込まれた」とは、プログラムレベルの評価が科目の評価の中で行われることを指す(Suskie, 2009)。この「埋め込み型」アプローチは、標準テストや学生調査のように、プログラムとは別にそれに追加する形で行われる「追加型(add-on)」アプローチと対比される。最後に「パフォーマンス評価」とは、学習者のパフォーマンス(作品や実演など)を手がかりに、概念理解の深さや知識・スキルなどを統合的に活用する能力を評価する方法のことである。このパフォーマンス評価では通常、評価基準としてルーブリックが用いられる。米国において、機関・プログラムレベルの評価に教室単位のパフォーマンス評価やルーブリックの利用が高まっていることを前に述べたが、PEPAもそれと同じ方向性を志向している。ただし、それは米国のトレンドに従うためではなく、このような方法が、学生の学びをリアルに把握・評価し、また育成する上で有効だと考えるからである。

PEPAは新潟大学歯学部での実践から生まれてきた考え方である。新潟大学歯学部には、歯学科(6年制)と口腔生命福祉学科(4年制)があり、それぞれに対応する学位プログラムがある。例えば、歯学科の学位プログラムは、大きく4つの時期に区分され、各期にそれぞれ「大学学習法」「PBL」「模型・シミュレーション実習」「診療参加型臨床実習」が重要科目として配置されている。これらの重要科目では、教員団で、それぞれの科目にあったパフォーマンス評価(パフォーマンス課題とルーブリック)を開発・実施している。

例えば「PBL」のパフォーマンス評価では、学生はPBLの授業と同じように、シナリオを与えられ、1週間かけてそれに対する解決策を考える。そして、その解決策を、教員を模擬患者としてロールプレイで実行する。終了後はその場ですぐに教員からフィードバックが提供される。ここでは、問題発見～解決策の提案

のプロセスと解決策の実行のプロセスのそれぞれについて、2種類のルーブリックを使った評価が行われるが、その観点(=次元)は、問題解決のVALUEルーブリックとほぼ重なる。すべての観点で4段階(レベル0～3)のうちのレベル1以上となることで合格とされる。これは他の重要科目でも同様である。

新潟大学の歯学教育プログラムでは、最も重要な学習成果を「歯科臨床能力」とし、それを歯科医療という文脈における問題解決能力と定義している。4種類の重要科目を核として、それを低学年から高学年に向けて、より専門性・総合性・真実性が高まるように配列し、それぞれで直接評価を行うことで、卒業生の質を担保しようとしている。一方、重要科目以外の個々の科目の評価は担当教員に委ねられる。ミネルヴァの場合は、すべての科目でパフォーマンス評価が行われているが、そうすると評価負担が大きくなりすぎるため、それを重要科目だけに限定することで、実行可能性の維持を図っている。

PEPAの方法はかなり複雑に見えるかもしれないが、中核部分は比較的シンプルである。学位プログラムの節目に配置された重要科目において、学生の間断的な学習成果を直接評価し、それを系列化することで、学生の学びの軌跡を把握するということである。

PEPAは、もともと歯学教育分野で開発されたので、医療系のような目標が明確な分野にしか適用できないのではないかという批判を受けることがある。これについては、現在、理工系総合大学である東京都市大学でその有効性を検討中である。都市大では、持続的社會への変革を志向する「SD PBL (Sustainable Development Project organized Problem Based Learning)」を全学科で1～3年次に1科目ずつ配置し、卒業研究につなぐことで、学びの軌跡を可視化し、高いレベルでの学習成果の達成を実現しようとしている(伊藤他, 2021)。

## V. これからの学習成果の評価

*The Tyranny of Metrics* (邦題『測りすぎ』)<sup>(3)</sup> という本の中で、ミュラー (Muller, J. Z.) は、教育、医療、警察、ビジネスなどあらゆる領域で、「経験に基づく

個人的判断の代わりに標準化された測定を使おうとする」ことによって、「意図せぬ好ましくない結果」が生じているとし、問題は、「測定ではなく、過剰な測定や不適切な測定」「測定基準ではなく、測定基準への執着」(邦訳p. 5)にあると指摘している。大学もその例外ではない。それどころか、ミュラーのケーススタディの最初に取り上げられているのは大学である。

本稿で扱ってきたプログラムレベルの学習成果の評価においては、標準テストや学生調査といった「標準化された測定」を使って説明責任を果たそうとする誘惑が強く働く。また、抽象的で多義的なプログラムレベルの学習成果を、各科目の中のそれに関連しそうな部分の総和によって算出しようとするのも「測定基準への執着」ではないだろうか。

本稿では、「不適切な測定」に陥らず、測定・評価を、大学教育や学生の学習にとって意味のあるものとするための一つの視点として、〈総和〉と〈軌跡〉という対概念を打ち出した。現在の大学教育では、〈総和〉による方法が主流だが、その問題点を指摘し、〈軌跡〉による方法によってどう補完するかを論じてきた。また、その具体的な方法として、ミネルヴァやPEPAの例をみてきた。ただし、これらについても「過剰な測定」になっていないかは検討の余地がある。過剰かそうでないかの判断は、不適切かどうかの判断以上に難しい。

本稿が、〈総和〉の方法が主流となっている現在の学習成果の評価の議論に対して、一石を投じるものになることを願っている。

## 【注】

- (1) 2012年の質的転換答申以来、高等教育の行政文書では「学修」「学修成果」が頻用されるようになったが、本稿では、引用以外はすべて「学習」「学習成果」で統一している。
- (2) 2017年時点のHCs(Kosslyn & Nelson, 2017)については訳出し、松下(2019)の資料に挙げた。ミネルヴァ学生(2021年2月現在、2年生)へのインタビューによると、HCsはたえず修正されており、現在79個になっているとのことである。
- (3) 邦訳には、「なぜパフォーマンス評価は失敗する

のか？」という副題がついているが、本稿で述べたようなパフォーマンス評価については論じられていない。

## 【文献】

- 朝日新聞・河合塾 (2018) 「学修成果の把握・可視化—ひらく 日本の大学調査より—」『Guideline』2018年11月号, 71-79.
- Association of American Colleges & Universities. (2007). *College learning for the new global century: A report from the National Leadership Council for Liberal Education & America's Promise*. AAC&U.
- Association of American Colleges & Universities. (n.d.). VALUE. (<https://www.aacu.org/value>) (2019年8月27日アクセス)
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61-100.
- González, J., & Wagenaar, R. (Eds.). (2008). *Tuning Educational Structures in Europe, Universities' contribution to the Bologna Process: An introduction* (2nd ed.). Publicaciones de la Universidad de Deusto.
- ゴンザレス, J.・ワーヘナール, R. (2012) 『欧州教育制度のチューニング—ボローニャ・プロセスへの大学の貢献—』(深堀聰子・竹中亨訳) 明石書店.
- 伊藤通子・松下佳代・斎藤有吾・中島英博 (2021) 「学習システム・パラダイムへの転換における PEPA の有効性—東京都市大学のケーススタディから—」『大学教育学会誌』43巻1号, 79-83.
- Jankowski, N. A., & Marshall, D. W. (2017). *Degrees that matter: Moving higher education to a learning systems paradigm*. Stylus. (Kindle 版)
- Kosslyn, S. M., & Nelson, B. (Eds.). (2017). *Building the intentional university: Minerva and the future of higher education*. The MIT Press.
- 松下佳代 (2017) 「学習成果とその可視化」『高等教育研究のニューフロンティア (高等教育研究 第20集)』玉川大学出版部, pp. 93-112.
- 松下佳代 (2019) 「汎用的能力を再考する—汎用性の4つのタイプとミネルヴァ・モデル—」『京都大学高等教育研究』25号, 67-90.
- 松下佳代 (2020) 「プログラムレベルと科目レベルの評価をつなぐ—PEPAの理論と課題—」『大学教育学会誌』42巻1号, 77-81.
- 松下佳代・小野和宏・斎藤有吾 (2020) 「重要科目での埋め込み型パフォーマンス評価を通して科目レベルとプログラムレベルの評価をつなぐ」『京都大学高等教育研究』26号, 51-64.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.
- ミュラー, J. Z. (2019) 『測りすぎ—なぜパフォーマンス評価は失敗するのか?—』(松本裕訳) みすず書房.
- Rhodes, T. L., & McConnell, K. D. (2017). *On solid ground*. AAC&U.
- Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, 35(6), 727-743.
- Suskie, R. (2009). *Assessing student learning: A common sense guide* (2nd ed.). Jossey-Bass.
- サスキー, L. (2015) 『学生の学びを測る—アセスメント・ガイドブック—』(斎藤聖子訳) 玉川大学出版部.
- WSCUC. (2015). *Using evidence in the WSCUC accreditation process: A guide for institutions* (2nd ed.). (<https://wascsenior.box.com/shared/static/5fum7qzstgwx0h74kae2b9zbs22gyaj.pdf>) (2019年11月12日アクセス)



---

# Program-level assessment of learning outcomes : Sum and trajectory

---

※ Kayo MATSUSHITA

**[Key Words]**

learning outcomes, program-level assessment, Minerva Schools at KGI (Minerva University), PEPA (Pivotal Embedded Performance Assessment), sum and trajectory

**[Abstract]**

The purpose of this paper is to present two concepts of assessing program-level learning outcomes, that is, learning outcomes acquired through degree programs, and to examine and propose assessment methods considering the difference of these two concepts. The concept of “sum” means to grasp the learning outcomes at the program level as the sum of individual courses while disregarding the time aspect. On the other hand, “trajectory” means to grasp learning outcomes along with the progress of student learning and with the time aspect included. The current trends in assessment of program-level learning outcomes in Japanese universities are biased toward the sum-based methods, which have the problem of adding up multiple, ambiguous, and substantially different competences, and do not take into account changes in these competences during undergraduate education. In this paper, we pointed out that it is necessary to supplement the sum-based methods with trajectory-based ones. Furthermore, as concrete illustrations that combine both methods, we examined the case of Minerva Schools at KGI and introduced an original method called Pivotal Embedded Performance Assessment (PEPA) and its practical applications.

---

※ Professor, Center for the Promotion of Excellence in Higher Education, Kyoto University