

学習成果とその可視化

松下 佳代

本稿の目的は、学習成果とその可視化について近年の動向をレビューし、論点を整理することにある。まず、学習成果を、①目標と評価対象、②直接的と間接的、③機関横断的・機関・プログラム・科目の各レベル、という3つの観点から捉えた。次に、この学習成果の多面性と関連づけながら、可視化の方法とツールを直接評価と間接評価、量的評価と質的評価の二軸で分類し、それぞれの特徴を明らかにした。さらに、こうした学習成果の可視化がもたらす第二次の可視化の例として、付加価値分析による大学教育力の可視化と、メタ分析による効果的な教育方法の可視化を挙げた。最後に、学習成果の可視化の孕む問題として、数値化可能な学習成果への切り詰め、評価から目標への浸食、多様性の喪失、評価負担の大きさの4点を提示した。

1. 問題設定

「学習成果（学修成果）」(learning outcomes)¹⁾は疑いもなく、現在の大学教育改革のキーワードの一つである。そのことは、2014年度に始まった大学教育再生加速プログラム（AP）において、「学修成果の可視化」が、「アクティブ・ラーニング」「入試改革・高大接続」と並んで初年度のテーマに選ばれたことにも表れている。

学習成果の重視は日本だけの現象ではなく世界的な潮流である。その背景としては、①知識基盤社会が到来し、大学卒業者に対するニーズが高度化し

たこと、②大学進学人口が拡大し、多様な学生が入学するようになったことにより、経済・雇用へのレリバンスの要求が高まったこと、③経済不況による国・州政府の緊縮財政や学生・保護者の消費者意識の上昇により、高等教育の質に対するアカウンタビリティの要請が強まったこと、④高等教育のグローバル化により、国境を越えた大学間の競争が始まるとともに、学位の国際通用性が求められるようになったこと、⑤MOOCなどのe-Learningの普及により、従来の単位制度のように時間によって学修を測る（例えば、1単位＝45時間の学修のように）のではなく成果によって測るという考え方が広まってきたこと、などを挙げることができよう（深堀 2015, 福留 2009, 森 2015a）。

学習成果の可視化の射程は、評価だけでなく、目標、カリキュラム、授業にも及び、正課プログラムだけでなく準正課プログラム、さらには課外活動にまで広がっている。本稿の目的は、学習成果とその可視化というテーマに対し、国内外を問わず、また、目標から評価に至るまで、視野を広げて近年の動向をレビューするとともに、今後の課題を浮き彫りにすることである。

まず、学習成果とは何かを多面的に捉え、能力（コンピテンス）との関係を明確化する（第2節）。次に、その学習成果の多面性と関連づけながら、可視化の方法とツールについて概観し（第3節）、そうした学習成果の可視化の上にさらなる可視化（いわば「第二次の可視化」）がなされていることを示す（第4節）。最後に、学習成果の可視化の孕む問題について整理し、今後の研究の展望を示す（第5節）。

2. 学習成果とは何か

(1) 「学習成果」の意味

ここまで、「学習成果」という語を、特に説明を加えずに用いてきたが、タームとしての「学習成果」は、単なる学習の結果と同義ではなく、多様な意味を内包している。ここでは3つの観点からその多面性を解剖してみたい。

(i) 目標としての学習成果、評価対象としての学習成果

まず、今日の学習成果には、学習の結果（評価対象）だけでなく、目標としての意味が含まれている。例えば、Tuning-AHELO プロジェクト²⁾では、学習成果を「学習者が、学習プロセスの終了後に、何を知り、理解して、また何をやってみせる（demonstrate）ことができると期待されてい

るかについての記述」(OECD 2009: 3, Tuning n. d.)と定義している。学士課程答申(中央教育審議会 2008)でもほぼこれと同じ定義が採用されている。この意味での学習成果には、intended, expected, desiredなどの修飾語がつく場合もある。なかでも、Biggs & Tang (2011)はintended learning outcomes (ILOs)を中心概念として学習パラダイムに立った大学教育論を展開し、大きな影響を与えてきた。

一方、これと区別して評価対象としての学習成果を指す場合には、achieved learning outcomesと表現されることもある。

(ii) 直接的な学習成果、間接的な学習成果

学習成果はまた、「学生が、授業科目、プログラム、教育課程などにおける所定の学習期間終了時に獲得し得る知識、技術、態度などの成果を指す」(大学改革支援・学位授与機構 2016: 24)と定義されることもある。この定義には、上述のTuning-AHELOや学士課程答申の定義と違って、「やってみせる」「実演する」などの言葉が含まれていない。それによって、テストやレポート、ポートフォリオのような直接指標(direct measures)で示される学習成果だけでなく、学生調査や卒業生調査のような間接指標(indirect measures)で示される学習成果も含む余地が生まれている。

一般的に、学習成果には、直接指標だけでなく間接指標によって把握されるものも含まれる。

(iii) 機関横断的な学習成果、機関レベルでの学習成果、プログラムレベルでの学習成果、科目レベルでの学習成果

Biggs & Tang (2011)は、学習成果には機関、プログラム、科目という3つの異なるレベルが存在するとしている。わが国では、2016年3月の学校教育法施行規則の改正によって、ディプロマ・ポリシー、カリキュラム・ポリシー、アドミッション・ポリシーの策定と公表が義務づけられたが、ディプロマ・ポリシーは、プログラムレベルの学習成果に関する方針である。

これらの3レベルに加えて、機関や専門分野の枠をこえて設定される学習成果がある。例えば、アメリカのAAC&U (Association of American Colleges & Universities: 全米大学・カレッジ協会)の「本質的学習成果(Essential Learning Outcomes)」(AAC&U 2007)や、「分野横断的に、我が国の学士課程教育が共通して目指す学習成果」として提案された「学士力」(中央教育審議会 2008)はその代表例である。Tuningでは、機関横断的ではあ

るが専門分野別のコンピテンス (subject-specific competences) と、さらに分野横断的でもある一般的コンピテンス (generic competences) が取りまとめられている (González & Wagenaar 2008 = 2012)。

大学教育でのタームとしての「学習成果」は、以上の (i)～(iii) の観点で分析することができる。学習成果の可視化とは、これらの学習成果を目に見える形にし、共有可能なものにすることである。

(2) 学習成果と能力 (コンピテンス)

今日の学習成果の大きな特徴は、学習成果の中身として、「知識や技能」だけではなく「能力」を掲げるということにある。例えば、Tuning では、「学習成果は、学習者によって獲得されるコンピテンスのレベルによって表現される」と明記している (Tuning n. d.)。「コンピテンス」は、ヨーロッパを中心に、まさに初等教育から高等教育、生涯学習に至るまでの広い範囲で使われる能力概念である。多少のバリエーションはあるが、基本的には、「認知的スキルとメタ認知的スキル、知識と理解、対人的・知的・実際的スキル、倫理的価値観のダイナミックな結合」(Tuning n. d.) という定義にみられるように、知識や多様なスキルや態度などの結合体として捉えられている。

深堀 (2015: 12) は、Tuning の知見をふまえて、コンピテンスとは、「学位プログラムを履修した総合的な成果として学生が獲得することが期待されている知識・技能・態度」であり、学習成果とは、「所定の学習期間内に達成可能であり、測定可能な知識・技能・態度」であるとし、抽象度の異なるアウトカムとして両者の関係を説明している。つまり、学習成果 (目標としての学習成果) は、能力の中でも、一定の学習期間での達成可能性、評価 (測定) 可能性をもつものを指すということができる。学士課程答申でも同様の捉え方がなされている。

では、能力 (あるいはその具体化としての学習成果) にはどのような内容が含まれるのだろうか。表 1 に、「本質的学習成果」と「学士力」を並置して示した。一見してわかるとおり、本質的学習成果の構成や内容は、学士力と類似性が高く、学士力の議論に大きな影響を与えたと考えられる。両者とも、基本的に、いわゆる KSA (knowledge, skills, attitudes) の 3 カテゴリーを立て、さらに、第 4 のカテゴリーとして、それらを統合・活用することによってより高度な課題を達成する能力を据えている。また、いずれの場合も、分野横断的なスキルや態度に重きが置かれていることがみてとれよう。

表 1 本質的学習成果と学士力の対応関係

本質的学習成果	学士力
人類の文化や自然界についての知識 ・科学、数学、社会科学、人文学、歴史、言語、芸術などの学習を通じて	知識・理解 ・多文化・異文化に関する知識の理解 ・人類の文化、社会と自然に関する知識の理解
知的・実際的スキル ・探究と分析 ・批判的思考・創造的思考 ・文章コミュニケーション・口頭コミュニケーション ・量的リテラシー ・情報リテラシー ・チームワークと問題解決	汎用的技能 ・コミュニケーション・スキル ・数量的スキル ・情報リテラシー ・論理的思考力 ・問題解決力
個人的・社会的責任 ・市民としての知識と関与（ローカル、グローバルに） ・異文化の知識と能力 ・倫理的な推論と行為 ・生涯学習のための基礎とスキル	態度・志向性 ・自己管理力 ・チームワーク ・リーダーシップ ・倫理観 ・市民としての社会的責任 ・生涯学習力
統合的学習 ・一般教育・専門教育での統合とより高度な達成	統合的な学習経験と創造的思考力 ・これまでに獲得した知識・技能・態度等を総合的に活用し、自らが立てた新たな課題にそれらを適用し、その課題を解決する能力

（出典）本質的学習成果は AAC&U（2007: 12）、学士力は中央教育審議会（2008: 12-3）より抜粋。

もっとも、学士力については、中央教育審議会からの依頼を受けて、日本学術会議で学問分野別の参照基準についての議論が進められ、現在（2016年12月）までに25分野で参照基準が公表されている³⁾。参照基準では、各学問分野の特性を明確化した上で、「当該学問分野を学ぶすべての学生が身に付けることを目指すべき基本的な素養」として、①基本的な知識と理解、②基本的な能力（a. 分野に固有の能力、b. ジェネリックスキル）が掲げられている。ここでのジェネリックスキルは、「分野に固有の知的訓練を通じて獲得することが可能であるが、分野に固有の知識や理解に依存せず、一般的・汎用的な有用性を持つ何かを行うことができる能力」（日本学術会議2010: 18）と定義されており、当該分野において獲得されるが、発揮されるときには分野をこえた汎用性をもつ、という形で分野固有性と汎用性が関連づけられている。

3. 可視化の方法とツール

第1節で述べたように、学習成果という概念は複数の観点から分析可能であるが、本節では、(i)で挙げた〈目標としての学習成果〉と〈評価対象としての学習成果〉の区別に焦点化して、これらの学習成果の可視化の方法とツールについて検討することにしよう。その中で(ii)と(iii)で挙げたそれぞれの学習成果との関連についてもふれることにしたい。

(1) 〈目標としての学習成果〉の可視化—学習成果の形での目標設定—

〈目標としての学習成果〉は、前述のように、「学習者が、学習プロセスの終了後に、何を知り、理解していて、また何をやってみせることができると期待されているかについての記述」であるから、「学習プロセス」をどのようなスパンで捉えるかによって、その可視化の仕方が異なってくる。学位プログラム終了後に身につけておくことが期待される学習成果は、ディプロマ・ポリシーの下、卒業までに達成すべき目標として示され、各授業科目終了後に身につけておくことが期待される学習成果は、カリキュラム・ポリシーの下、各科目の目標として書かれることになっている。

このプログラムレベルの学習成果と科目レベルの学習成果がどのような整合性をもつのかを示すために使われるツールが、コースツリーやカリキュラムマップである。現場での用語法はさまざまだが、本稿では、コースツリーを「プログラムにおける科目間の関係と履修の順次性をツリー状に表した図」、カリキュラムマップを「プログラムを構成する個々の科目とその学習成果の対応関係をマトリックス形式で示した表」(表2参照)としておく。コースツリーでは、科目とその学習成果の関係は明確に可視化されないが、カリキュラムマップでは直接的に可視化される。だが、それは、カリキュラムマッ

表2 カリキュラムマップにおける科目と学習成果（コンピテンス）の関係

科目／学習成果	コンピテンス									
	A	B	C	D	E	F	G	H	I	J
科目1		×			×					
科目2	×			×			×			
科目3		×				×			×	
科目4	×		×							×

(出典) González & Wagenaar (2008: 13) より抜粋。

プのメリットであると同時にデメリットでもある。

カリキュラムマップの背後にあるのは、「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 = 2012: 32) というアイデアであり、そこでは、コンピテンスの形成が、製品を部品に分解しそれを組み立てるのと同じように捉えられている。だが、カリキュラムは教える側の教育計画であるだけでなく、個々の学生の学びの履歴でもある（例えば、科目2と科目4のコンピテンスAの内容は、それぞれの科目が現実の時間の流れの中でどう履修されるかによって異なってくるだろう）。時間軸を捨象したこのようなツールは形式的なものになりやすい。

今日、多くの大学でカリキュラムマップが作成されているが、こうした問題が常につきまとうていることを念頭に置く必要がある。

(2) 〈評価対象としての学習成果〉の可視化—学習成果の評価—

①教育政策との関連

日本の教育政策において、〈目標としての学習成果〉を初めて明確に打ち出したのが2008年の学士課程答申だったとすれば、〈評価対象としての学習成果〉を初めて明確に打ち出したのは2012年の質的転換答申（中央教育審議会2012）であった。質的転換答申では、従来の3ポリシーに加えて、各大学でのアセスメント・ポリシーの作成の必要性が唱えられ、学習成果の具体的な測定手法として、「学修行動調査」「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」が例示された。2016年3月の3ポリシーの策定・公表の義務化では、アセスメント・ポリシーは策定・公表の義務化の対象外となったが、カリキュラム・ポリシーの中でアセスメントの方針も定めることが求められている。

②学習成果の評価を分類する2つの軸

このような教育政策にも後押しされて、近年、学習成果の評価について多様な研究・実践が進められている。松下（2016）は、それを、〈直接評価—間接評価〉、〈量的評価—質的評価〉という2軸によって整理している。

(a) 直接評価と間接評価

直接評価（direct assessment）と間接評価（indirect assessment）の違いは、その評価方法が、学習成果の直接的なエビデンスに基づくか、間接的なエビデンスに基づくかに拠っている（したがって、直接評価—間接評価とい

う対は、直接指標—間接指標、直接的エビデンス—間接的エビデンスといった対と対応している）。例えば、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——行われる評価は直接評価であり、一方、学習成果についての学習者の自己報告を通じて——「何を知り何ができると思っているか」を学習者自身に答えさせることによって——行われる評価は間接評価である（Palomba & Banta 1999, 山田 2012, 松下 2012）。この軸は、いうまでもなく、(ii) の〈直接的な学習成果〉と〈間接的な学習成果〉の区別に対応している。

アメリカの地域ア kreditation ション団体の一つであるニューイングランド協会（New England Association of Schools and Colleges: NEASC）は、「学習の直接的エビデンスをもたらす方法」（直接評価）と「学習の間接的エビデンスをもたらす方法」（間接評価）の例を表 3 のように挙げている（NEASC 2013, 森 2015b）。

間接評価だけでは、「学習者が、学習プロセスの終了後に、何を知り、理解していて、また何をやってみせることができるか」を把握することはできない。また、能力の低い人は自身の能力を過大に評価する傾向があり、逆に能力が高い人は自身の能力を控え目に評価する傾向がある——「ダニング＝クルーガー効果」（Kruger & Dunning 1999）として知られる——などの理由から、自己報告による間接評価で直接評価を代替することはできない。

だが一方、直接評価だけでは、何を学んだか（学ばなかったのか）のエビデンスは得られても、なぜ学んだのか（学ばなかったのか）を把握することは難しい。また、価値観、興味・関心などの学生自身の認知（perception）

表 3 直接評価と間接評価の例

直接評価	間接評価
・機関が作成したプログラム／機関レベルのテスト ・標準テスト ・プレテスト、ポストテスト ・ブラインドで採点される小論文テスト ・学生のプロジェクトやインターンシップについての学内・学外の審査員による評価 ・全米規模の資格試験の成績 ・学生の作品のコレクション（ポートフォリオ） ・科目ごとの評価	・卒業生・雇用主調査 ・学生調査、フォーカスグループ ・卒業時インタビュー ・インターンシップ、海外研修などのプログラムへの参加率 ・卒業生追跡調査 ・大学院進学率 ・大学院入試合格率 ・専門分野における学生の実績 ・リテンション率・編入率、就職率

（出典）NEASC（2013: 25）より作成。

表 4 量的評価と質的评价

	量的評価	質的评价
学問的基盤	心理測定学	解釈学，構成主義的学習論など
評価データ	量的データ	質的データ
評価対象	集団または個人	個人
評価目的	選抜，組織的な教育改善， アカウントビリティなど	学習や指導の改善など
評価課題	細かく分割された問題 文脈独立的	複合的な課題 文脈依存的
評価基準	客観性を重視	問主観性を重視
評価結果	数値	文章や数値
評価機能	主に総括的評価	主に形成的評価
評価方法	客観テスト・標準テスト， 質問紙調査など	パフォーマンス評価・ポートフォ リオ評価，ミニッツペーパーなど

(出典) 松下 (2012: 79) を一部改変。

や学習成果に至る学習行動などは，学生の自己報告に依拠せざるをえない。

したがって，直接評価を基本としつつ，間接評価も組み合わせて用いることが必要になる (MSCHE 2007)。

(b) 量的評価と質的评价

量的評価と質的评价の間には，評価データが量的か質的かだけでなく，それに関連して，他にもいくつかの対比がみられる (表 4 参照)。

量的評価では，何よりも測定・評価の客観性が重視される。よく使われるのは客観テストや質問紙調査などだが，これらは集団に対して実施することが容易で，比較も行いやすいので，選抜，アカウントビリティなどのために用いられる。

一方，質的评价の方は，個々の生徒・学生の学習や指導の改善のための情報を得るのに適している。質的评价は必ず主観的要素を伴うので，測定・評価の信頼性をいかに確保するかが問題になる。そこで用いられるのがルーブリックである。ただし，ルーブリックは，質を量に変換するツールでもあり，量的評価と質的评价を橋渡しする役割をもつ。

近年では，両者の中間型や混合型のようなタイプの評価も実施されている。例えば，AHELO の試行調査においてジェネリックスキルの評価に採用された CLA (Collegiate Learning Assessment) は，標準テスト型のパフォーマ

ンス評価である（松下 2012）。

③学習成果の評価の4つのタイプ

以上の2軸からなる二次元図式によって、学習成果の評価は、図1のような4つのタイプに整理できる（松下 2012, 2016）。

質的転換答申で例示された評価手法でいえば、「学修行動調査」はタイプⅡ、「アセスメント・テスト」はタイプⅢ、「ループリック」「学修ポートフォリオ」はタイプⅣに位置づけられる。例えば、AAC&Uでは、表1に示した本質的学習成果が実際にどの程度、学生に獲得されたかを評価するために、VALUE（Valid Assessment of Learning in Undergraduate Education）プロジェクトが進められてきた。このプロジェクトでは、eポートフォリオにおさめられた「学生の真正の作品（authentic student work）」を評価対象とし、それに対する評価基準として、ループリックが開発された。このVALUEループリックは、学士課程4年間をカバーする長期的ループリックであり、これまでに、問題解決、批判的思考、倫理的推論、チームワークなど、16のループリックが公開されている（Rhodes 2010, Rhodes & Finley 2013）。

④機関レベル・プログラムレベルの評価と科目レベルの評価の関係

図1には直接、表現できていないが、(iii)の観点も、学習成果の評価について検討する上で重要な観点である⁴⁾。

大学教育学会では、2013～2015年度の3年間にわたって「学士課程教育における共通教育の質保証」という課題研究が行われたが、そこで再認識されたのが、機関・プログラムレベルの学習成果の評価と科目レベルの学習成

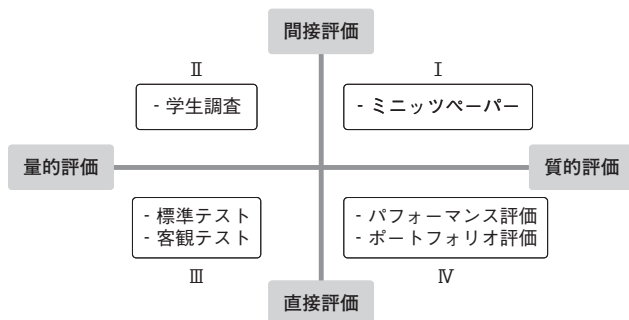


図1 学習成果の評価の4つのタイプ

（出典）松下(2012：81)を一部改変

果の評価との方法の違いであった（松下ほか 2014, 山田 2016）。例えば、わが国では、機関・プログラムレベルの評価では学生調査や業者による標準テストなどが多用され、科目レベルでは科目ごとの客観テスト・論述試験や、パフォーマンス（レポート・制作物などの作品やプレゼンテーション・演技などの実演）による評価が使われている。ポートフォリオを学生に作成させている大学は多いが、長期的ループリックなどの評価基準と組み合わせて、学生の学びや成長の評価を行い、機関・プログラムレベルの評価にあてている大学はまだ限られている（関西国際大学など）。文部科学省の調査（2015）でも、「課程を通じた学生の学修成果の把握方法」（＝プログラムレベルの評価）としてループリックが用いられている割合はわずか3.4%（2013年度）にとどまっている。

ちなみに、アメリカの状況（2013年）をみると、機関レベルの学習評価として最もよく用いられているのは、「全米規模の学生調査」（85%）、「ループリック」（69%）、「クラス単位のパフォーマンス評価〔の蓄積〕」（66%）であり、「最も価値がある（もしくは重要である）」とみなされているのもこの3つである（Kuh *et al.* 2014）。日本の状況との違いがみてとれよう。

4. 学習成果の可視化の上になされる可視化—第二次の可視化—

さて、ここまで論じてきた学習成果の可視化を「第一次の可視化」と呼ぶとすれば、そうした学習成果の可視化の上になされる可視化は「第二次の可視化」と呼ぶことができる。第二次の可視化の例として、本稿では、「大学の教育力の可視化」と「効果的な教育方法の可視化」を取り上げよう。

（1）大学の教育力の可視化

大学の教育力の可視化のために使われる方法に、付加価値分析（value-added analysis）がある（Bennett 2001, Kim & Lalancette 2013）。これは文字通り、企業の付加価値分析を援用した方法である。教育の文脈における「付加価値」とは、教育目標に対する生徒・学生の伸び（progress）に対して教育機関がなした貢献の度合いを意味する（OECD 2008）。付加価値分析では、例えば、入学時と卒業時の達成の比較を通じて、大学が在学中に学生をどれだけ成長させたか（望ましい学習成果をどれだけ伸ばしたか）を分析する。こうした付加価値分析において学習成果の可視化が前提となることは、指摘するまでもないだろう。ただし、一機関内の付加価値分析だけでは教育力が

捉えにくいため、さらに機関間で付加価値を比較することによって、大学の教育力のより明確な可視化が図られる。

大規模な学生調査（評価タイプⅡ）やアセスメント・テスト（評価タイプⅢ）では、付加価値分析が可能なように、入学時と卒業時でアセスメントを実施するのが一般的である。また、多くの学生調査やアセスメント・テストはもとも機関間比較が可能なように質問項目や調査問題が設計されている。とはいえ、入学時と卒業時の達成のスナップショット的な比較によって大学の教育力の正確な可視化が行えるかについては疑問が残る。例えば、学生の伸びは、大学のプログラムの外で得られたものかもしれないからである（例えば、大学外での語学学習や資格学習など）。

一方、ルーブリックを用いた評価（評価タイプⅣ）でも、近年のアカウンタビリティの要請に対して、付加価値分析と機関間比較が可能になる仕組みが整えられつつある。例えば、前述の VALUE プロジェクトは、eポートフォリオに学生の学習成果を継続的に蓄積させることで、スナップショット的な比較という限界を乗り越えようという試みであり、学士課程4年間をカバーする長期的ルーブリックを用いることで付加価値分析ができるようデザインされている。また、機関間比較についても、複数の州をまたいだ協働事業（Multi-State Collaborative to Advance Learning Outcomes Assessment）が進行中である⁵⁾。

大学の教育力の可視化といえば、一般によく知られているのは大学ランキングである。しかしながら、大学ランキングは、関係者の評判や容易に収集可能なデータに基づいており、しかも各指標に恣意的な重みづけがなされているという点で、「高等教育の質に対して極めて有害性を持っている」（館 2016: 20）。大学ランキングは疑似的な可視化でしかないのである。

（2）効果的な教育方法の可視化

第二次の可視化のもう一つの例は、効果的な教育方法の可視化である。なかでも、ニュージーランドの教育研究者ハッティ（Hattie, J.）の *Visible Learning*（『可視化された学習』）（Hattie 2009）は、「教授活動の聖杯を探り当てた」（Mansell 2008）と評されるほどの国際的センセーションを巻き起こした（原田・マイヤー 2015）。ハッティは、5万をこえる個別研究についての800あまりのメタ分析の研究結果を統合して、学校段階での生徒の学習成果（achievement）に関わる要因を138にまとめ（①学習者、②家庭、③

学校、④教師、⑤カリキュラム、⑥授業の6つに分類されている)、それぞれの効果を検討した。効果の高さは効果量 $d^{(6)}$ で表され、全要因について、その効果の程度が、 -0.2 から 1.2 までの目盛りをもつバロメーターで視覚的に示された ($d = 0.40$ が中位の効果)。そうして、全要因を効果の高い順に並べたランキングを提示したのである。

Hattie (2009) が研究対象としているのは初等中等教育であるが、高等教育でも、例えば、アクティブラーニングの効果についてのメタ分析を行った研究などがみられるようになってきている (Freeman *et al.* 2014)。

メタ分析や系統的レビューについては、さまざま方法論的批判もなされているが (山田・井上 2012)、エビデンス・ピラミッドの最上位に位置づけられるこうした研究手法は、エビデンスを重視する文化の中で、今後、いっそう影響力をもつことになるだろう。

5. 学習成果の可視化の孕む問題と今後の課題

以上では、学習成果の意味やその可視化の方法を整理するとともに、可視化がさらなる可視化を呼び込むことをみてきた。冒頭に述べたように、学習成果の可視化は世界的な潮流となっている。その背景は複合的であり、可視化を推し進める力が衰えることは少なくともしばらくはないだろう。

だが、学習成果の可視化には、無視できないいくつかの問題がある。そうした問題にどう対処するかが今後の課題となる。

(1) 数値化可能な学習成果への切り詰め

まず取り上げるべきは、数値化可能な学習成果への切り詰めという問題である。既に論じてきたように学習成果の評価には量的評価と質的评价があり、それぞれに異なる特徴をもつ。しかしながら、アカウンタビリティの要請が高まるなかで、次第に量的評価にウェイトが置かれるようになってきている。例えば、AP の採択校の多くが、業者の開発した標準テスト (アセスメント・テスト) と学生調査によって「学修成果の可視化」を図っている (例えば、笹川 2015 など)。

これらは統計的分析にかけやすく、組織間比較が可能であり、また、付加価値分析が比較的行いやすいという強みをもっている。しかし、こうした方法がその大学で学生が得た学習成果をどれだけ可視化しているかを私たちはあらためて問う必要がある。

AAC&U の VALUE ルーブリックは、「スプリングス・レポート」(2006 年)以降、標準テストが影響力を増す状況において、より妥当な学習評価として構想されたものであった (Rhodes 2010)。批判点として挙げられたのは、標準テストが、サンプルになった学生のみを対象としたスナップショット的な学習成果の把握の仕方であり、大学での日常的な学習内容とかけ離れていて教授・学習の質の改善に役立たないといった点であった。

もっとも、ルーブリックを用いた評価も、数値化可能な学習成果への切り詰めという問題を免れているわけではない。ルーブリックが質的評価のツールというより、単に質を数値化するためのツールとして使われるようになったときには、同様の問題が生じる。

評価において私たちは柔軟であることを求められる。評価基準から逸脱していても、そこに事前に想定していなかった新しい発想が埋もれている場合もある。評価はあくまでも学生の学習成果の質に基づいて行われるべきであり、基準はたえず修正可能性にさらされている。

(2) 評価から目標への浸食

「数値化可能な学習成果への切り詰め」が特に問題となるのは、それがさらに目標を規定するようになった場合である。すなわち、目標が設定されてそれを学習した結果が評価されるのではなく、数値化可能な学習成果にあわせて目標が設定され学習活動が行われるという「目標と結果の転倒」がみられる場合である。

例えば、一部の標準テストでは、その結果を各学生にフィードバックし、そこから自分の強み・弱みを把握させ、学習目標を設定させるということが行われている。だが、標準テストの規準と当該大学の教育目標との調整は十分行われておらず、また、テスト項目の妥当性についても（ごく一部しか公開されていないため）学問的な吟味にかけにくい。

標準テストに限らず、「学習成果の可視化」を追求する教育の下では、評価から目標への浸食が生じやすい。目標を、達成可能性、評価可能性をもつ「意図された学習成果」として規定するというそもそものアイデアが、その危険性を孕んでいることに、私たちは自覚的であるべきである。

(3) 多様性の喪失

もともと学習成果の設定において意図されていたのは、プログラム設計における柔軟性や自律性・多様性を担保することであった。例えば、Tuning

では次のように述べられている。

学習成果を用いることで、伝統的な方法で学習プログラムを設計する場合よりも、より高い柔軟性が可能になる。同等の学習成果を達成するのに、さまざまな経路がありうるということが明らかになるからである。[中略] 学習成果を用いても、他の教育機関、他の教育文化の自律性は十分に尊重される。それゆえ、学習成果を用いることで、国際的、欧州の、国内的、学内的な多様性が許容されるばかりでなく、ある学習プログラム内においても多様性が可能になるのである。(González & Wagenaar 2008 = 2012: 31)

しかしながら、学習成果が同一の基準や方法で測定されるようになると、そこですぐれた成績を挙げているところ（国、自治体、教育機関など）から学ぼうとする動きが、自発的あるいは政策誘導的に生まれる。この動きは、初等中等教育においては、2000年代に入ってから、PISA調査、全国学力・学習状況調査などを通じて強まったが（松下 2014）、近年では、大学教育でも目立ち始めている。それを促しているのが、大学ランキングであり、分野や大学のミッションの違いをこえて利用される標準テストである。

多様性と共通性の調停は、学習成果の可視化における大きな課題である。VALUE ルーブリックは、機関横断的に共有されるが、それを各大学が自大学の文脈に合わせてローカライズするというやり方を採用することで、大学をこえた共通性と大学間の多様性の両立が図られている（Rhodes & Finley 2013）。また、学生調査の代表例である NSSE（National Survey of Student Engagement）では、機関間比較をミッションやタイプが似ている大学間で行うことにより、共通性が多様性を損なわないよう運用されている⁷⁾。

多様性と共通性の調停という課題に対して単一の正解はない。私たちは対立する二項の間で納得解、最適解を探る努力を続けていくしかないのである。

（4）評価負担の大きさ

最後にやはり評価負担の大きさという問題を挙げておきたい。図1に示したように、学習成果の評価は直接―間接、量的一質的という軸によって4つのタイプに分けられ、さらに、それぞれについて機関・プログラム・科目の各レベルでの評価が存在する。既に述べたとおり、どの評価方法も固有の特色と限界をもち、組み合わせる用いることが必要だとされる。だが、そのような評価の多重使用は、複雑性を高め、コストを大きくする。

近年，わが国でも IR が急速に普及しつつあるが，とりわけ機関レベル・プログラムレベルの評価には，専門知をもつスタッフの存在が不可欠である．大学教育の最前線にいる教員の当事者意識を失わせることなく，そうしたスペシャリストと教員の間の連携をどう構築していくかということも，今後の重要な課題である．

◇注

- 1) 2012 年の質的転換答申以来，政策文書では，正課の学修を通して獲得する成果を「学修成果」と表記するようになっている．だが，learning outcomes には，単位制度の下での学修成果より広い意味が含まれており，本稿でも，正課だけでなく準正課プログラムの学びの成果にも言及することがある．そのため，以下，引用以外はすべて「学習成果」と表記することにする．
- 2) OECD-AHELO (Assessment of Higher Education Learning Outcomes) の試行調査のうち，工学と経済学分野の調査は Tuning プロジェクトとの協働で行われ，Tuning-AHELO と呼ばれている．
- 3) 日本学術会議ウェブサイト (<http://www.scj.go.jp/ja/member/iinkai/daigakuhosyo/daigakuhosyo.html>, 2016.12.31) 参照．
- 4) アメリカの中部地区アクレディテーション協会 (Middle States Commission on Higher Education) では，直接・間接と機関レベル・プログラムレベル・科目レベルという 2 つの軸で，評価方法を整理している (MSCHE 2007; 福留，2009)．
- 5) AAC&U と連携している事業を進めている SHEEO のウェブサイト (<http://www.sheeo.org/msc>, 2016.12.31)，および AAC&U (2017) 参照．
- 6) 効果量 d とは「メタ分析において，異なる研究（通常異なる測定単位が利用される）の効果を共通のものさしにのせるための，共通の指標」（山田・井上，2012，p. 281）であり，効果量 1 は，標準偏差 1 だけの違いを表す．
- 7) NSSE のウェブサイト (<http://nsse.indiana.edu/>, 2016.12.31) 参照．

◇参考文献

- 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*, Washington D.C.: AAC&U.
- Association of American Colleges & Universities, 2017, *On Solid Ground: VALUE*

- Report 2017*, Washington D.C.: AAC&U.
- Bennett, Douglas C., 2001, "Assessing Quality in Higher Education," *Liberal Education*, 87 (2).
- Biggs, John and Tang, Catherine, 2011, *Teaching for Quality Learning at University* (4th ed.), Berkshire, UK: The Society for Research into Higher Education & Open University Press.
- 中央教育審議会, 2008, 「学士課程教育の構築に向けて (答申)」.
- 中央教育審議会, 2012, 「新たな未来を築くための大学教育の質的転換に向けて—生涯学び続け、主体的に考える力を育成する大学へ (答申)」.
- 大学改革支援・学位授与機構, 2016, 「高等教育に関する質保証関係用語集」第4版.
- Freeman, Scott, Eddy, Sarah L., McDonough, Miles, Smith, Michelle K., Okoroafor, Nnadozie, Jordt, Hannah, and Wenderoth, Mary P., 2014, "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proceedings of the National Academy of Sciences*, 111 (23): 8410–5.
- 深堀聰子, 2015, 「アウトカム重視の大学教育改革—その背景と概念の整理」深堀聰子編著『アウトカムに基づく大学教育の質保証—チューニングとアセスメントにみる世界の動向』東信堂, 3–32.
- 福留東土, 2009, 「米国高等教育におけるラーニングアウトカムに関する動向」『比較教育学研究』(38): 145–58.
- González, Julia and Wagenaar, Robert, eds., 2008, *Tuning Educational Structures in Europe, Universities' Contribution to the Bologna Process: An Introduction* (2nd ed.), Bilbao: Publicaciones de la Universidad de Deusto. (= 2012, 深堀聰子・竹中亨訳, 『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』明石書店.)
- 原田信之・マイヤー, H., 2015, 『ドイツ教授学へのメタ分析研究の受容—ジョン・ハッティ「可視化された学習」のインパクト』デザインエッグ.
- Hattie, John, 2009, *Visible Learning: A Synthesis of Over 800 Meta-Analysis Relating to Achievement*, New York: Routledge.
- Kim, HoonHo and Lalancette, Diane, 2013, *Literature Review on the Value-Added Measurement in Higher Education*, Paris: OECD.
- Kruger, Justin and Dunning, David, 1999, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 77 (6): 1121–34.
- Kuh, George D., Jankowski, Natasha, Ikenberry, Stanley O., and Kinzie, Jillian, 2014, *Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in U.S. Colleges and Universities*, Champaign, IL: National Institute for Learning Outcomes Assessment.

- Mansell, Warwick, 2008, “Research Reveals Teaching’s Holy Grail,” *The Times Educational Supplement*, November 21.
- 松下佳代, 2012, 「パフォーマンス評価による学習の質の評価—学習評価の構図の分析にもとづいて」『京都大学高等教育研究』(18): 75-114.
- 松下佳代, 2014, 「PISA リテラシーを飼いならす—グローバルな機能的リテラシーとナショナルな教育内容」『教育学研究』81(2): 14-27.
- 松下佳代, 2016, 「アクティブラーニングをどう評価するか」松下佳代・石井英真編『アクティブラーニングの評価』東信堂, 3-25.
- 松下佳代・小野和宏・斎藤有吾・白川優治, 2014, 「学士課程教育における共通教育の質保証—直接評価と間接評価の開発と統合について」『大学教育学会誌』36(2): 17-21.
- Middle States Commission on Higher Education, 2007, *Student Learning Assessment: Options and Resources* (2nd ed.).
- 文部科学省, 2015, 「大学における教育内容等の改革状況について(平成 25 年度)」.
- 森利枝, 2015a, 「米国の高等教育における Competency-Based Education の展開に関する考察」『大学研究』(41): 29-40.
- 森利枝, 2015b, 「アメリカにおける学習成果重視政策議論のインパクト—ひきつづく議論のなかで」深堀聰子編著『アウトカムに基づく大学教育の質保証—チューニングとアセスメントにみる世界の動向』東信堂, 235-50.
- New England Association of Schools and Colleges, 2013, *Self-Study Guide* (Revised ed.).
- 日本学術会議, 2010, 「大学教育の分野別質保証の在り方について」.
- OECD, 2008, *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*, Paris: OECD.
- OECD, 2009, *A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering*, Tuning Association.
- Palomba, Catherine A. and Banta, Trudy W., 1999, *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*, San Francisco: Jossey Bass.
- Rhodes, Terrel, ed., 2010, *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*, Washington D.C.: AAC&U.
- Rhodes, Terrel L. and Finley, Ashley, 2013, *Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment*, Washington D.C.: AAC&U.
- 笹川篤史, 2015, 「PROG テストを利用した学生の能力伸長分析について」『長崎大学経済学部研究年報』31, 1-23.
- 館昭, 2016, 「高等教育における質のアセスメント—動向と課題」山田礼子編著『高等教育の質とその評価—日本と世界』東信堂, 19-30.

- Tuning, n. d, *Glossary of Terms*. (<http://www.unideusto.org/tuningeu/documents/glossary-of-terms.html>, 2017.1.28)
- 山田礼子, 2012, 『学士課程教育の質保証へむけて—学生調査と初年次教育からみえてきたもの』 東信堂.
- 山田礼子, 2016, 「共通教育における直接評価と間接評価における相関関係—成果と課題」『大学教育学会誌』 38(1): 42-8.
- 山田剛史・井上俊哉編著, 2012, 『メタ分析入門—心理・教育研究の系統的レビューのために』 東京大学出版会.

Making Learning Outcomes Visible**MATSUSHITA, Kayo***Kyoto University*

The purpose of this paper is to review recent research trends within the broad theme of “making learning outcomes visible” and to present the most pertinent issues in an organized form. First, we define the learning outcomes from three perspectives: (a) intended and achieved; (b) direct and indirect; (c) cross-institutional, institutional, program, and course levels. On the basis of the conceptual multiplicity of learning outcomes, we have classified the methods and tools used to make learning outcomes visible into four types grouped along two axes, direct and indirect assessments as well as quantitative and qualitative assessments, and clarify the characteristics of each type. Furthermore, by means of a value-added analysis of the quality of education and a meta-analysis of increasing the visibility of effective teaching, we give specific illustrations of how “second-order visibility” is generated by application of the concept of making learning outcomes visible. Finally, we point out four potential risks which are inseparably inherent in our study theme: the reduction of learning outcomes to quantifiable aspects only, the erosion of goals by assessments, a loss of diversity, and a heavy assessment load.