# 学習成果としての能力とその評価

## －ルーブリックを用いた評価の可能性と課題－

松　下　佳　代

――＜要　旨＞――

後期近代社会への移行が本格化した 1990 年代以降、期待される学習成果（目標）として能力を掲げ、その学習成果を評価することが、高等教育において重視されるようになってきた。本稿の目的は、ルーブリックを用いた評価に焦点をあてて、学習成果としての能力の評価の可能性と課題を明らかにすることにある。

そのために本稿では、AAC&U（アメリカ大学・カレッジ協会）の提案した、目標としての「本質的学習成果」と、評価ツールとしての「VALUE ルーブリック」を取り上げて、それがわが国の学士課程教育に与える示唆を検討した。本質的学習成果では知識・理解と区別される重要な一般的能力（あるいはジェネリックスキル）が抽出され、VALUE ルーブリックはその評価のためのメタルーブリックを提供している。だが、その範囲は教養教育に限定されており、わが国の学士課程教育全体をカバーするものではない。本稿では、歯学教育での PBL におけるルーブリック開発の事例を取り上げて、職業教育と結びついた専門教育の中で、問題解決という一般的能力がいかに具体化され、評価されうるかを例証した。

## 1．問題と目的

### 1.1　学習成果としての能力

学生は大学で何を学ぶのか。教員は学生に何を学ばせようとするのか。この問いに対してはこれまで、「学問」「教養」「専門的な知識・技能」「市民性」など、さまざまな答えが与えられてきた。

そして、後期近代社会への移行が本格化した 1990 年代以降、この「何を」

京都大学高等教育研究開発推進センター・教授
名古屋大学高等教育研究センター・客員教授

の問いに対する答えとしてしばしば掲げられるようになったのが「能力」である。例えば、ヨーロッパでは、欧州高等教育圏の建設をめざすボローニャ・プロセスの下でチューニング・プロジェクト（Tuning Project）が実施されてきたが、このプロジェクトでは、学生に期待される学習成果を、まず「一般的能力（generic competences）」として規定し、それを「専門分野別能力（subject-specific competences）」に具体化した上で、そのような能力を学位プログラムの中で形成できるようカリキュラムや授業をデザインすることになっている（González & Wagenaar 2008、松下 2007）。このアプローチは、現在、ラテンアメリカ諸国、ロシア、米国、アフリカ諸国、オーストラリアなどの国と地域に広がっており、日本でも一橋大学がTuning Japanとしてその研究・実践を進めている。

一方、アメリカでは、アメリカ大学・カレッジ協会（Association of American Colleges and Universities: AAC&U）が、2005年から10年計画で「教養教育とアメリカの約束（Liberal Education & America's Promise: LEAP）」というプロジェクトを開始し、その中で、「本質的学習成果（Essential Learning Outcomes）」の提案を行った（AAC&U 2007）。「学習成果」は、目標と結果の両方の意味で使われる語である。目標の意味で使われるときは、「期待される（expected）」、「望まれる（desired）」、「意図された（intended）」といった形容詞がつくこともある。本質的学習成果の構成や内容は、2008年12月の中央教育審議会答申「学士課程教育の構築に向けて」で提案された「学士力」とも類似性が高く、学士力の議論に大きな影響を与えたと考えられる。

### 1.2　学習成果の評価

では、このような学習成果はどう評価されるのだろうか。実は、この問い自体が後期近代的な性格を帯びている。というのも、その背景には、1990年代以降、公的部門に民間企業の経営理論・手法を導入しようというNPM（ニュー・パブリック・マネジメント）の考え方が普及し、教育の世界にも目標管理型システムが持ちこまれたという事情があるからである。そのような状況の中で、わが国の大学教育でも、2004年度から始まった認証評価や国立大学法人評価などを通して、インプット評価、プロセス評価に加え、アウトプット評価やアウトカム評価が行われるようになってきた。学習成果の評価とは、まさにアウトカム評価なのである。いうまでもなく、教育における成果は短期間で現れるとは限らない。したがって、アウトカ

ムについては、「中間アウトカム」と「最終アウトカム」を区別する必要がある（大学基準協会 2009: 15）。学士課程教育における学習成果の評価とは、中間的なアウトカム評価という性質をもつと考えるべきだろう。

学習成果の評価の中でも本稿が対象にするのは、学習成果としての能力の評価である。例えば、チューニング・プロジェクトは、OECD-AHELO（Assessment of Higher Education Learning Outcomes：高等教育における学習成果の評価）と連携することにより（Tuning-AHELO）、AHELO の試行調査（feasibility study）において、工学と経済学の分野での学習成果を定義し、その評価を行っている（OECD 2011、2012）。

また、AAC&U では、LEAP プロジェクトの一環として、2007 年から VALUE（Valid Assessment of Learning in Undergraduate Education：学士課程教育における妥当な学習評価）プロジェクトを実施してきた。これまでに 16 この VALUE ルーブリックが開発され、多くのキャンパスで利用されている（Rhodes 2010、Rhodes & Finley 2013）。

日本では、中教審からの審議依頼を受けて、日本学術会議が分野別質保証のための参照基準の策定に着手し、現在、各分野の参照基準が次々と提案されている。さらに、2012 年 8 月の中教審答申「新たな未来を築くための大学教育の質的転換に向けて」では、従来の 3 つのポリシーに加えてアセスメント・ポリシーの作成が各大学に求められ、学習成果 [1]の具体的な評価方法として、「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。

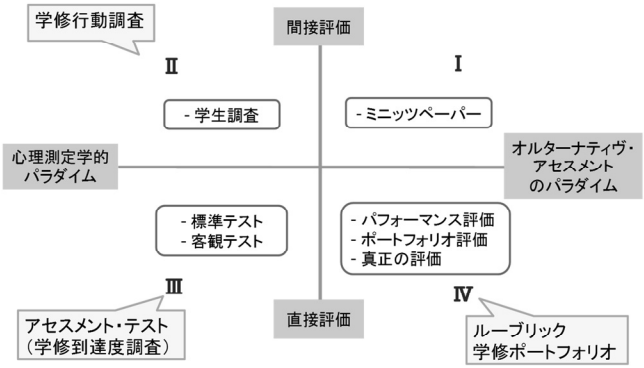
### 1.3 本稿の対象と目的

このような動きの中で、本稿では、能力の中でも、問題解決、コミュニケーションといった一般的能力に焦点をあて、また、評価の中でも、ルーブリックを用いた直接評価に的をしぼって、学習成果としての能力とその評価に関する検討を進めることにしたい。

まず、能力、なかでも一般的能力を検討対象とすることについて述べよう。問題解決、コミュニケーションなどは「汎用的技能」と呼ばれることも多いが、「汎用的技能」という言葉には、技能を実体化し、そのままでどこでも通用するというイメージがつきまとう。だが、問題解決やコミュニケーションといった能力は、抽象的なレベルで存在するわけではなく、個別特殊な文脈における具体的な行為の中で姿を現すにすぎない。多様な文脈を通してそのような具体的な行動が繰り返し行われたときに、その人

には問題解決やコミュニケーションの能力があると判断されるのである。また、「能力（competence）」には、単に個々の内的なリソース（知識、技能、態度など）をもっているというだけではなく、そうしたリソースを複雑な状況のもとで適切に結集できるという意味も含まれている（松下 2010: 20-2）。

一方、ルーブリックを用いた直接評価の検討は、図 1 に示すような学習成果の評価の構図（松下 2012）を土台としている。




出典：松下（2012: 81）を一部改変。

図 1　学習成果の評価の構図

間接評価は、学習成果を学生の自己認識の報告（何ができると思っているか）によって間接的に評価するものであり、直接評価は、学生の知識や能力の表出（何ができるか）によって直接的に評価するものである。また、心理測定学的パラダイムでは、客観性や比較可能性を重視した量的な評価方法がとられるのに対して、オルターナティヴ・アセスメントのパラダイムでは、間主観性にもとづく質的な評価方法がとられる。ルーブリックを用いた直接評価はもともとはタイプⅣに属する評価だが、ルーブリックには質を量に変換する働きがあるため、その射程はタイプⅢにも及んでいる。実際、北米では、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）のような、標準テスト型のパフォーマンス評価も普及している。AAC&U の VALUE プロジェクトは、そうした動きに対抗するものであり、標準化によらない「真正の評価（authentic assessment）」をめざしている。

このように、一般的能力に焦点をあて、ルーブリックを用いた直接評価に的をしぼって本稿のテーマに迫ろうとするときに、最も注目されるのは、AAC&U の本質的学習成果と VALUE ルーブリックである。実際、これらの研究と実践は、すでに日本でも数多く紹介され（森 2011、吉田 2011、松下 2012、深野 2013 など）、それを基盤とした学士課程教育を展開している大学も存在する。とはいえ、AAC&U の研究と実践は、アメリカの大学の教養教育を中心に行われているものであり、日本の学士課程教育にそのまま持ちこむことはできない。

そこで、本稿では、まず AAC&U の本質的学習成果と VALUE ルーブリックの特徴を明らかにした上で、そこから日本の学士課程教育にどんな示唆が得られるのかを議論する。さらに、日本の学士課程教育については、PBL（Problem-Based Learning）を検討の素材に据えることにより、学習成果としての問題解決能力とその評価の方法を考察することとする。このような作業を行いながら、ルーブリックを用いて学習成果としての能力を評価する際の可能性と課題を明らかにすることが、本稿の目的である。

## 2. AAC&U における学習成果の抽出とその評価の開発

### 2.1 本質的学習成果

AAC&U は教養教育の研究や推進などを目的として 1915 年に設立された組織であり、約 1,300 校におよぶ会員校はコミュニティ・カレッジから大規模研究大学まで多岐にわたる。本質的学習成果は、こうした会員校の多様性をこえて、アメリカの高等教育機関が教養教育を通して学生に修得させるべき本質的に重要な学習成果として抽出されたものである。

表 1　教養教育の再定義

| | 20 世紀の教養教育 | 21 世紀の教養教育 |
|---|---|---|
| 何を | ・ エリート教育のカリキュラム<br>・ 非職業的<br>・ 幸運な人間のためのオプション | ・ すべての学生にとって必要なもの<br>・ グローバル経済においても良識ある市民生活にとっても不可欠 |
| どこで | ・ リベラルアーツ・カレッジ、あるいは大規模大学での文理学部（colleges of arts and sciences）において | ・ すべての学校、コミュニティ・カレッジ、大学において、あらゆる学問分野にわたって（推奨） |
| いかに | ・ 文・理の諸分野（「専攻」）での学習を通して、あるいは（および）、下位学年での一般教育を通して | ・ 学校から大学まで、教育的な連続性をもつ学習を通して（推奨） |

出典：AAC&U（2007: 18）より訳出。

まず着目すべきは、AAC&U による教養教育のとらえ方である（表１）。従来の教養教育が「職業との結びつきをもたない全人的な教育」（藤本 2012: 63）とみなされていたのに対し、「21 世紀の教養教育」は、職業人としても市民としても不可欠なものとみなされている。また、エリートだけではなくすべての学生にとって必要であり、文理学部の専攻や一般教育を通してだけでなく学校から大学まで連続性をもって学ばれるべきものだとされている。こうしてみると、再定義された教養教育は、職業に直結する専門教育を除いて広い範囲の学士課程教育をカバーするものといえよう。

AAC&U の考える教養教育がこのように幅広いものだとすれば、「我が国の学士課程教育が共通して目指す学習成果」である学士力に本質的学習成果が援用されたとしても、さほど強引な拡張ではない。とはいえ、学士力の方は、「我が国の学士課程の多様な現実（アメリカのリベラル・アーツ型から医歯薬学教育等の職業教育まで）」をふまえてより高い「汎用性」をもつよう意図されていることには、やはり注意が必要である。

表２　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<br>・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<br>・多文化・異文化に関する知識の理解<br>・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<br>・探究と分析<br>・批判的思考・創造的思考<br>・文章コミュニケーション・口頭コミュニケーション<br>・量的リテラシー<br>・情報リテラシー<br>・チームワークと問題解決 | 汎用的技能<br>・コミュニケーション・スキル<br>・数量的スキル<br>・情報リテラシー<br>・論理的思考力<br>・問題解決力 |
| 個人的・社会的責任<br>・市民としての知識と関与（ローカル、グローバルに）<br>・異文化の知識と能力<br>・倫理的な推論と行為<br>・生涯学習のための基礎とスキル | 態度・志向性<br>・自己管理力<br>・チームワーク<br>・リーダーシップ<br>・倫理観<br>・市民としての社会的責任<br>・生涯学習力 |
| 統合的学習<br>・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 |

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。

表 2 は、本質的学習成果と学士力の対応関係を示したものである。両者を比べると、2 番目と 3 番目のカテゴリーの表現などいくつかの違いはあるものの、全体的に構成や内容は驚くほどよく似ている。共通するのは、各分野の知識・理解に加えて分野横断的な学習成果を取り出し、その中に認知的・社会的・倫理的な幅広い能力を包含していることである。

### 2.2　VALUE ルーブリック

#### 2.2.1　VALUE ルーブリックの特徴

こうした学習成果はどのように評価されるのだろうか。評価について考えるときには、評価対象と評価基準をセットで考えなければならない。VALUE プロジェクトでは、評価対象として、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）」を設定している。authentic には、〈その領域の本質的なパフォーマンスを正しく反映した〉、〈リアルな文脈で〉、〈評価のためだけの特別な課題や機会を設定しない通常の学習の中で〉といった意味が含まれており（Wiggins & McTighe 2005=2013、鈴木 2013）、客観テストやパフォーマンス評価型の標準テストに対する批判がこめられている [2)]。

一方、評価基準として開発されたのが VALUE ルーブリックである。本質的学習成果の 4 カテゴリーのうち、「人類の文化や自然界についての知識」については評価手段がすでに数多く存在するが、「知的・実際的スキル」「個人的・社会的責任」「統合的学習」については、妥当といえるような評価手段がほとんどない。そこで、正課カリキュラムと準正課カリキュラム（cocurriculum）を通じて学生が生み出した作品を対象に、ルーブリックを用いて学習の質を評価する方法を構築しようというのが VALUE プロジェクトの目的であった。まず、以下の 15 領域の VALUE ルーブリック [3)] が作られ、2013 年に「グローバルな学習」が加えられた。

- 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、読解、量的リテラシー、情報リテラシー、チームワーク、問題解決
- 個人的・社会的責任：市民参加、異文化の知識と能力、倫理的推論、生涯学習の基礎とスキル
- 統合的・応用的学習：統合的学習

一例として、文章コミュニケーションのルーブリックをあげよう（表3）。ルーブリックは一般に、規準（criterion）[4]とレベル、その中身を説明する記述語（performance descriptor、あるいは単に descriptor）によって構成される。このルーブリックは5つの規準と4つのレベルからなる（表3ではレベル2・3の記述語は省略）。1～4の数字は学年や成績を表示するものではなく、レベル1は入学してくる学生にみられるレベル、レベル4は学士号を授与される学生に期待されるレベルを表し、レベル2・3は文字どおりレベル 1 からレベル 4 に向かう道のりのマイルストーンを表している（Rhodes & Finley 2013:.6）。ルーブリックにはいくつかのタイプがあるが（松下 2012: 83）、VALUE ルーブリックは、特定の領域で一般的に適用でき複数年にまたがって使われる、一般的で長期的なルーブリックということができよう。

表３　「文章コミュニケーション」VALUE ルーブリック

| | キャップストーン<br>4 | マイルストーン<br>3 | <br>2 | ベンチマーク<br>1 |
|---|---|---|---|---|
| 文章作成の文脈と目的 | 文脈・読者・目的について完璧な理解を示し、それによって、与えられた課題に対応し、作品のあらゆる要素に焦点をあてている。 | …… | …… | 文脈・読者・目的や与えられた課題（例えば、読者としての授業者や自己の期待）に対し最低限の注意を示している。 |
| 内容の展開 | 適切で関連性があり説得力に富む内容を用いることによって、科目の習得ぶりを示すとともに、書き手の理解したことを伝え、作品全体を形づくっている。 | …… | …… | 適切で関連性のある内容を用いることによって、作品の何カ所かで、シンプルなアイデアを展開している。 |
| ジャンルと学問分野の約束事 | 特定の学問分野や文章作成課題に関連する広範な約束事（構成、内容、提示、書式、文体選択を含む）に対し、細かい注意を向けうまく遂行している。 | …… | …… | 基本的構成や提示のしかたについて一貫した体系を使おうとしている。 |
| 資料と根拠 | 当該の学問分野やジャンルにふさわしいアイデアを展開するために、質が高く、信頼でき、関連性のある資料をうまく使いこなしている。 | …… | …… | アイデアを裏づけるために、資料を使おうとしている。 |
| 構文と技法を操ること | 読み手に明確かつ流暢に意味を伝えることができる格調ある言葉遣いをしている。ほとんど全く誤りがない。 | …… | …… | 用語法に誤りがあるために、意味の伝達が妨げられるような言葉遣いをしている。 |

注　：レベル1を満たさないものはレベル0とされるので、実質的には5段階のレベルからなる。
出典：松下（2012: 103）より一部抜粋。

アメリカの大学ではもう 20 年以上にわたってルーブリックの開発･利用の経験が積み重ねられてきた。そのような経験をふまえながら、AAC&U の専門家チームによって共同開発されたのが VALUE ルーブリックである。そうして作られた VALUE ルーブリックは、今度は、個々の大学・学科・科目の文脈にあわせてローカライズされる。ここで注目すべきは、VALUE ルーブリックが、特定の個別大学のルーブリックよりもう一段抽象度が高い「メタルーブリック（metarubric）」として、あるいは、個別大学でのルーブリック開発の元になるようなプロトルーブリックとして、機能するよう意図されているということである。これによって、大学をこえた共通性と大学間の多様性の統一が図られているのである（松下 2012: 89-90）。

では、具体的に、個々の大学・学科・科目では、VALUE ルーブリックをどのようにローカライズして用いているのだろうか。

### 2.2.2　モディフィケーション（修正）とキャリブレーション（調整）

VALUE ルーブリックを用いた評価の手続きでとくに重要なのは、ルーブリックのモディフィケーション（modification：修正）とキャリブレーション（calibration：調整）である（Rhodes & Finley 2013: ch.4, ch.5）。

モディフィケーションは VALUE ルーブリックを、大学・学科・科目の文脈にあわせてローカライズするために必要な作業である。AAC&U では VALUE ルーブリックのケーススタディ[5]を行っているが､それをみると、ローカライズは次のような形で行われていることがわかる（下に行くほどローカライズによる修正の幅が大きい）。

- VALUE ルーブリックのうち必要なルーブリックだけを選択する。
- 規準やレベルはそのままで、記述語の表現を変える。
- 記述語に加えて、規準やレベルの表現も変える。
- 規準やレベルを削ったり、加えたりする。
- 複数の VALUE ルーブリックを組み合わせて、新たに 1 つのルーブリックを作る。

表 4 はこのようにしてローカライズされたルーブリックの一例である。

表４　ローカライズされた VALUE ルーブリック：
A 大学のある科目（BIO109）における採点用ルーブリック

| | 2 | 1 | 0　　得点 |
|---|---|---|---|
| 文章作成の目的<br>（文章コミュニケーション） | ･読者を考慮していること、および、研究の焦点が明確であることが、タイトルに示されていた。 | …… | …… |
| 問題の特定<br>（文脈的能力） | ･研究の主要問題が一般的な読者に関心のあるものになっていた。<br>▪ …… | …… | …… |
| エビデンス<br>（批判的思考） | ･ソース（情報源）からの情報を評価し、自分自身の解釈を生み出していた。<br>▪ …… | …… | …… |
| 問題の説明<br>（批判的思考） | ･すべての問題が記述され、明確化されていた。<br>▪ …… | …… | …… |
| 構文と技法を操ること<br>（文章コミュニケーション） | ･使用された言葉は、読者に意味を明確に伝えていた。<br>▪ …… | …… | …… |

出典：Rhodes（2013）より一部抜粋の上、訳出。

文章コミュニケーションと批判的思考という２つの VALUE ルーブリックから必要と思われる規準を選び、新たな規準を加え、レベルの数も減らしている。かなりのモディフィケーションが行われていることがわかる。

一方、キャリブレーションとは、「ルーブリックを使う人々が一堂に会して、ルーブリックがどうデザインされておりどう適用されるべきなのかについて共通理解を築くプロセス」（Rhodes & Finley 2013: 23）のことである。キャリブレーションは、ファシリテータのもと、以下のような手続きで行われる。

・ルーブリックを読み合わせ、解釈を共有する。
・学生の作品事例を用いながら、１つの規準（または１つのレベル）について各自が採点する。
・採点終了後、参加者は、自分の採点の根拠について作品の該当箇所を引きながら説明する。同じ得点を与えた参加者ごとに説明していく。
・以上のことを少なくとも２つの作品事例で行う。

Rhodes & Finley（2013）は、キャリブレーションの利点として、評価者間信頼性を高める、学問分野をこえた対話の機会をもたらす、FD の場として機能する、課題（アサインメント）のデザインについての議論につながる、という 4 点をあげている。評価論では、評価基準や評価結果を複数の評価者間で調整することを「モデレーション（moderation）」というが、キャリブレーションは評価基準のモデレーションにあたるといえよう。

### 2.3　VALUE ルーブリックの課題

以上 VALUE ルーブリックについてみてきたが、ここでその課題についてあげておきたい。1 つは比較可能性の担保という課題、もう 1 つは知識と能力の関連づけという課題である。

#### 2.3.1　比較可能性の担保

VALUE ルーブリックの目的は、とりわけ 2006 年のスペリングス報告以降強まったアカウンタビリティの要請に対して、大学が学生たちに期待される学習成果を獲得させており、その意味で教育プログラムが有効性をもちえているということを示すためのツールを作り出すことにあった。アカウンタビリティのためのツールということでいえば、すでに、直接評価としては CLA（図 1 でいえばタイプⅢ）、間接評価としては NSSE（タイプⅡ）などが用いられてきた。VALUE ルーブリックはタイプⅣに位置づき、CLA のような標準テストとは対抗的な関係に、また、NSSE とは補完的な関係にある（松下 2012: 91）。

アカウンタビリティを遂行するために用いられる方法として付加価値（value-added）の分析がある。これは、入学時と卒業時の比較を通じて、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）を分析しようとするものである。VALUE ルーブリックは長期的ルーブリックであることで付加価値分析が可能なようにデザインされている。ただし、付加価値分析だけでは、その大学の教育プログラムの有効性の程度を示す上で十分とはいえない。そこで他大学との比較が行われることになる。VALUE ルーブリックは、大学をこえた共通性をもつことによって、他大学との比較可能性を担保しようとするものであった。

しかしながら、ローカライズされたルーブリックをみると、修正の幅はかなり大きく、はたしてこれで比較可能性が担保されるのかという疑問がわく。Rhodes & Finley（2013: 21）も、「オリジナルの VALUE ルーブリ

ックのモディフィケーションが大きくなればなるほど、ローカルな結果をより広い文脈の中に位置づけることは困難になる」と述べている。

この問題に対処するために、AAC&U では現在、VALUE プロジェクトの拡張として、CAAL（Collaborative for Authentic Assessment and Learning：真正の評価と学習のための協働）プログラムが進められている[6)]。ルーブリックにそって評価された学生の作品をその得点データとあわせて収集し、リポジトリを作成するというものである。これによって、大学・学科等の多様性をこえて学生のパフォーマンスについてのベンチマーク・データ（標準となるデータ）を構築することがめざされている。

多様性を許容しながら比較可能性を担保することははたして可能なのか。AAC&U のこの活動のゆくえに注目したい。

### 2.3.2　知識と能力の関連づけ

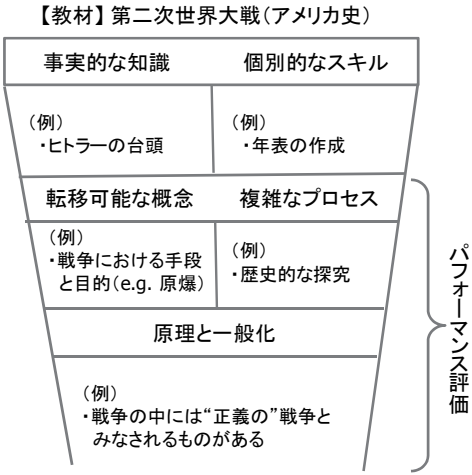
前述のように、VALUE ルーブリックは、本質的学習成果の 4 カテゴリーのうち、「人類の文化や自然界についての知識」を除く「知的・実際的スキル」「個人的・社会的責任」「統合的学習」の成果を評価するために作られた評価基準である。いいかえれば、VALUE ルーブリックは、一般的能力は評価できても、知識・理解を評価できるようには作られていない。

この問題に対しては、2 つのアプローチがありうる。1 つは、能力をみる評価と知識をみる評価を併用することである。例えば、医・歯・薬学教育などでは、学生を臨床実習に参加させる前に、基本的な臨床能力（技能・態度）を評価する OSCE（Objective Structured Clinical Examination：客観的臨床能力試験）と知識を評価する CBT（Computer-Based Testing）を行い、適格性を判断している。

もう 1 つは、能力と知識の両方をみられるような評価を行うことである。マクタイとウィギンズ（McTighe & Wiggins 2004）は、パフォーマンス評価をデザインする際に、図 2 のような「知識の構造」を想定している。

AAC&U がめざす真正の評価はパフォーマンス評価の一特殊であり、本来、「転移可能な概念」と「複雑なプロセス」、そして「原理と一般化」にあたるような深い学習が行われているかどうかをみるための評価である。だが、VALUE ルーブリックに描かれているのは主に「複雑なプロセス」に限られている。VALUE ルーブリックがメタルーブリックである以上、そこに特定の内容が含まれていないのは仕方がないが、VALUE ルーブリックを個別の文脈で用いる際には、能力形成に概念的知識や深い理解が伴

っているかを評価できるように、ルーブリックの修正や評価課題の設定を行う必要がある。




出典：McTighe & Wiggins（2005: 66）をもとに作成。

図２　「知識の構造」と評価方法の対応関係

## 3. 日本における学習成果の評価への示唆

### 3.1　日本の現状と課題

#### 3.1.1　目標としての学習成果

以上の検討結果は、日本における学習成果の評価を構想する上でどんな示唆をもたらすだろうか。

表２で示したように、AAC&U の本質的学習成果とわが国の学士力はよく似ているが、違いもある。何より大きな違いは、本質的学習成果の対象が教養教育であるのに対して、学士力の対象は、医療系や教員養成系などの職業教育も含むわが国の多様な学士課程教育全体である点である。

日本学術会議は、中教審からの審議依頼に対し、学士力の中身を構成する学問分野別の参照基準についての議論を行い、2010 年７月に「大学教育の分野別質保証の在り方について」という回答を提出した。「回答」では、参照基準の主要な構成要素として、「1. 当該学問分野の定義」「2. 当該学

問分野に固有の特性」「3. 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」「4. 学習方法及び学習成果の評価方法に関する基本的な考え方」「5. 市民性の涵養をめぐる専門教育と教養教育との関わり」を掲げており、さらに、3 の内容として、(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）をあげている。ここで、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」と定義され、専門教育と教養教育からなる学士課程教育全体を通じて形成すべきと考えられている（日本学術会議 2010: 17-20, 47）。

だが、現在までに公表されている各分野の参照基準をみる限り、ジェネリックスキルがその分野の知識・理解や専門的能力とどんな関連をもつのか、教養教育で形成されるジェネリックスキルと専門教育で形成されるジェネリックスキルがどのような相互補完関係にあるのかは判然としない。

AHELO の試行調査では、CLA を用いたジェネリックスキルの評価について、設定された文脈があまりにアメリカ的で、それぞれの国や大学の文脈が考慮されていないと指摘され、専門分野の文脈の中で測定するほうが適切ではないかと問題提起されている（深堀 2013、OECD 2013: 160-71）。それをみても、専門分野においてジェネリックスキルがどう形成・評価されるのかを明確にすることが、喫緊の課題になっているといえる。

### 3.1.2 学習成果の評価 －ルーブリックの開発

ルーブリックを用いた評価は、質的転換答申で具体的な評価方法として取り上げられたことで、今後わが国でも広がっていくと予想される。しかし、妥当性や信頼性を備えたルーブリックを開発するには、大きな労力と時間がかかる（松下ほか 2013）。また、質保証やアカウンタビリティの要請の高まりの中で、標準テストに対抗しうるような評価を行おうと思えば、一定程度の比較可能性を担保したルーブリックが望ましいだろう。

そのような状況を考えたとき、日本でも、VALUE ルーブリックと同じようなメタルーブリックがあれば、個々の大学・部局がルーブリックを開発する際の基盤となる。だが、日本で独自のメタルーブリックを作るのは困難な作業である。というのも、VALUE ルーブリックが、数多くの大学での長年にわたる評価実践をもとに共同開発されたのに対し、わが国では評価実践の蓄積が乏しいからである。したがって、わが国でのルーブリッ

クの開発は、VALUE ルーブリックなどを参考にしながら、個々の大学・部局でのルーブリック開発とメタルーブリック開発を同時並行で行うことが必要になるのである。

### 3.2　問題解決の評価事例から

#### 3.2.1　PBL について

このようなわが国の現状と課題をふまえて、以下では、問題解決という学習成果を中心に、一般的能力（ジェネリックスキル）を知識・理解と結びつけながら形成していくための学習方法と評価について考えてみよう。

評価をデザインするには、評価基準とあわせて評価課題を設定することが必要である。AAC&U では、本質的学習成果を評価するための重要な課題として、初年次セミナー、PBL（Problem-Based Learning）、キャップストーン・プロジェクトなどをあげている（AAC&U 2007: 32）。

PBL は、わが国においても、三重大学や広島大学などで全学的に導入されるなど広がりをみせているが(三重大学高等教育創造開発センター 2007, 吉田ほか 2013)、評価の観点から PBL について詳しく検討したものはあまりみあたらない。以下では、専門教育における問題解決能力の評価課題・評価基準として、私が開発に関わっている新潟大学歯学部の PBL テュートリアルにおける評価の例を紹介しよう（小野・松下 2013）。

#### 3.2.2　PBL テュートリアルの評価　－新潟大学歯学部の場合

新潟大学歯学部では、2004 年から PBL テュートリアルの取組を始め、現在、口腔生命福祉学科では第 2～第 4 学年、歯学科では第 5 学年で実施されている（小野ほか 2011）。学生 7～8 人のグループに教員のテューターが 1 名ずつ入り、シナリオ教材を用いて、次のような手順で行われる。

- ステップ 1（授業内、グループ学習）：1）事実を抽出し問題を見いだす→ 2）解決策（仮説）を立てる→ 3）学習課題を設定する
- ステップ 2（授業外、個人学習）：4）教室外で追加情報を集める
- ステップ 3（授業内、グループ学習）：5）新しく得た知識と既有の知識を統合する→ 6）仮説を検証する

ステップ３はステップ１の１週間後に設定され、その間に、学生はそれぞれに自己学習を行うことが期待されている。

新潟大学歯学部では、PBL テュートリアルの主たる目標を問題解決能力と自己学習能力の形成に置いているが、評価方法としては、知識を問うペーパーテストと卒業生に対する質問紙調査にとどまっており、求める学習成果を直接評価するものにはなっていなかった。PBL テュートリアルの直接評価の方法としては、学生に個人でステップ 1〜3 を行わせ、自分の問題解決プロセスを教員とのやりとりの中で説明させる「トリプルジャンプ」という方法があるが（Smith 1993, Mtshali & Middleton 2011）、信頼性の低さや評価負担の大きさから最近はあまり用いられなくなっている。

そこで、私たちは、「改良版トリプルジャンプ（Modified Triple Jump: MTJ）」を新たに開発した。MTJ の特徴は以下の点にある。

- 上記の３つのステップを、ステップ１（MTJ ステップ 1）とステップ 2･3（MTJ ステップ 2）に分け、これらはワークシートに記述させる。
- 新たに MTJ ステップ３として「問題解決の実行」を加え、模擬患者（教員）相手に実演させる。
- MTJ ステップ 1･2 と MTJ ステップ３のルーブリックを作成し、それにもとづいて問題解決プロセスの記述と解決策の実演を評価する。

ワークシートを用いることで教員の評価負担を軽減させ、ルーブリックを用いることで信頼性の向上を図った。

表５　PBL 評価用ルーブリックの規準（新潟大学歯学部）

| MTJステップ1・2 | | | | | | |
|---|---|---|---|---|---|---|
| 観点 | 問題発見〜最終解決策の提案 | | | | | |
| | 1. 問題発見 | 2. 解決策の着想 | 3. 学習課題の設定 | 4. 学習結果とリソース | 5. 解決策の検討 | 6. 最終解決策の提案 |

| MTJステップ3 | | | | |
|---|---|---|---|---|
| 観点 | 解決策の実行 | | | |
| | 7-1. 追加情報の収集（追加情報の収集と問題の再把握） | 7-2. 情報の統合（追加情報の統合と解決策の内容修正） | 7-3. 共感的態度（相手への共感） | 7-4. コミュニケーション（相手にあわせた解決策の表現） |

表５は、私たちの作成した２つのルーブリックの規準のみを抜粋したものである。レベルは1〜3の３段階からなる（ただし、VALUE ルーブリックと同様に、レベル１に満たないものは０とするので、実質的には４段階である）。

一方、表６は、「問題解決」VALUE ルーブリックの規準のみを抜粋したものである。

表６　「問題解決」VALUE ルーブリックの規準

| 規準 | 問題の定義 | 方略の同定 | 解決/仮説の提案 | 採りうる解決の評価 | 解決の実行 | 結果の評価 |
|---|---|---|---|---|---|---|

両者を対応させてみると、MTJ ステップ１・２は VALUE ルーブリックの「問題の定義」から「採りうる解決の評価」までを、MTJ ステップ３は「解決の実行」をカバーしていることがわかる（「結果の評価」は、ルーブリックには含まれていないが、解決策の実行の後にすぐに教員からフィードバックを行い、学生にも自己評価をさせることで行われている）。同時に、MTJ ステップ３の「解決策の実行」では、患者とのやりとりを通じて問題解決を実行することが求められる専門職の教育にふさわしい観点が、新たに設けられていることが見てとれよう。

知識と能力の関係についてはどうだろうか。PBL テュートリアルでは、問題場面を描いたシナリオを読み、解決策を立てるのに必要な知識が自分に不足していることを意識させた上で、学生たちに学習課題を立てさせ、自己学習に取り組ませる。例えば、歯周病と糖尿病を抱えながら「タバコを吸うことが生きがい」と話す患者に対してどう禁煙指導を行うかというシナリオでは、歯周病と喫煙の関係、歯周病と糖尿病の関係、効果的な保健指導の方法などの知識を得ることが期待されている。このようにして知識獲得と能力形成の両立が図られているわけである。評価においても、「4. 学習結果とリソース」や「7-2. 情報の統合」といった観点において、そうした知識が適切に獲得・活用されているかが評価されることになる。

このように、新潟大学歯学部の例は、問題解決という一般的能力（ジェネリックスキル）がいかに専門教育の中で具体化されるか、そこにおいて知識・理解と能力がどう統合されるか、そうした学習成果がどう評価されるかを例証する格好の事例といえよう。

MTJ を体験した学生は次のように述べている。「やったらやった分だけ

身になるような気がします。また、自分がどれくらい成長できるのかが分かるので、ただ学習しているだけよりも、もしかしたら手ごたえがあるかもしれません」。この学生の言葉は、適切にデザインされた評価がもつ力を雄弁に物語っている。

## 4. むすびにかえて

本稿では、AAC&U の本質的学習成果と VALUE ルーブリックの検討をふまえて、日本の大学教育への示唆を考えながら、学習成果としての能力とその評価について議論を進めてきた。AAC&U の本質的学習成果はわが国の学士力にも影響を与えたと考えられるが、それが対象としているのは教養教育であり、リベラル・アーツ教育から医療系や教員養成系などの職業教育までを含むわが国の多様な学士課程教育全体をカバーするものではない。そこで、本稿では、AAC&U が対象としていない職業教育である歯学教育を事例に、一般的能力の形成と評価の可能性を、PBL テュートリアルのための評価方法の開発を通じて探った。その結果、問題解決能力という一般的能力が専門教育の中で具体化され、ルーブリックを通じて評価されうること、またそれを通じて専門的な知識・理解との統合も可能であることが明らかになった。

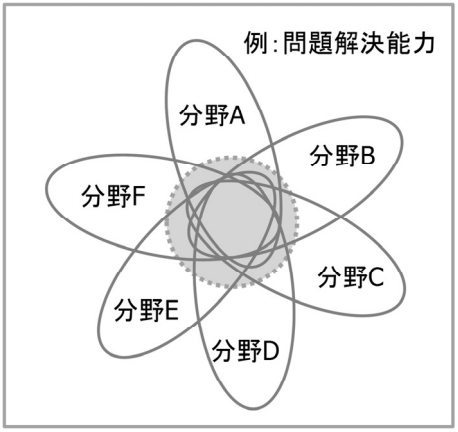



図３　一般的能力と汎用的技能

ジェネリックスキルは「汎用的技能」と訳されることが多いが、generic とは本来、「一般的な・総称的な」という意味であり、「汎用的な」という意味ではない。例えば、問題解決能力は専門分野によって異なる形をとるが、さまざまな分野の問題解決能力には共通の重なりあう部分があり（図3の点線で囲んだ部分）、それを意識的に学べば、他の分野での問題解決にも転移させやすくなる。それを私たちは「汎用的技能」と呼んでいるにすぎないのである。「問題解決能力」というのは、こうしたさまざまな形で姿を現す問題解決行為の背後にあると考えられる能力に対してつけられたラベルである。

さまざまな分野の専門家チームによって作られたVALUEルーブリックは、この「汎用的技能」の部分をかなりうまく掬いとっている。ただし、それは具体的な大学・部局・科目の文脈の中でローカライズされなければ機能しない。一方、私たちが開発したPBL評価用ルーブリックは、専門分野の中で具体化される問題解決能力を評価するためのルーブリックであり、ローカライズされたVALUEルーブリックと似たような位置にある。

わが国でも、多様な専門分野や教養・共通教育において一般的能力の形成と評価の実践を蓄積するとともに、共通の枠組みを構築していくことがいま求められている。

## 注

1) 質的転換答申から「学修成果」と表記されるようになったが、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るので、本稿では、引用以外はすべて「学習成果」を用いる。
2) パフォーマンス評価と真正の評価の概念的な区別の仕方には大きく分けて2つある。1つは、真正の評価を、パフォーマンス評価のうち特に真正性の条件を満たすものとして区別する場合、もう1つは、評価のための特別な課題や機会を設定する場合をパフォーマンス評価、通常の授業の中で行う評価を真正の評価として区別する場合である（鈴木 2013: 22-5）。
3) すべて AAC&U のウェブページ（http://www.aacu.org/about/index.cfm）からダウンロード可。日本語訳は松下（2012: 100-14）参照。
4) AAC&U では「次元（dimension）」と呼ばれることもある。その他に、「特性（trait）」や「観点」なども使われる。
5) (http://www.aacu.org/value/casestudies/, 1 October 2013.)
6) (http://aacu.org/caal/, 1 October 2013.)

## 参考文献


AAC&U, 2007, College Learning for the New Global Century: A Report from the National Leadership Council for Liberal Education & America's Promise, Washington, DC: AAC&U.

大学基準協会、2009、「内部質保証システムの構築－国内外大学の内部質保証システムの実態調査」。

深堀聰子、2013、「AHELO 国際的な学習成果アセスメント－フィージビリティ・スタディの成果と課題」『教育学術新聞』2520: 2。

深野政之、2013、「アメリカ大学カレッジ協会によるカリキュラム提言」『大学教育研究開発センター年報 2012』、51-68。

González, Julia and Wagenaar, Robert eds., 2008, Tuning Educational Structures in Europe, Universities' Contribution to the Bologna Process: An Introduction, 2nd ed.,（＝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。）

松下佳代、2007、「コンピテンス概念の大学カリキュラムへのインパクトとその問題点－Tuning Project の批判的検討」『京都大学高等教育研究』13: 101-9。

松下佳代、2010、「〈新しい能力〉概念と教育－その背景と系譜」松下佳代編『〈新しい能力〉は教育を変えるか－学力・リテラシー・コンピテンシー』ミネルヴァ書房。

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・小野和宏・高橋雄介、2013、「レポート評価におけるルーブリックの開発とその信頼性の検討」『大学教育学会誌』35(1): 107-15。

McTighe, Jay, and Wiggins, Grant, 2004, Understanding by Design: Professional Development Workbook, Alexandria, VA: ASCD.

三重大学高等教育創造開発センター編、2007、『三重大学版 Problem-based Learning 実践マニュアル－事例シナリオを用いた PBL の実践』。
(http://www.hedc.mie-u.ac.jp/pdf/pblmanual.pdf, 2013.10.1.)

森利枝、2011、「高等教育機関の特性を活かした客観的指標の模索」『文部科学教育通信』282: 16-7。

Mtshali, Ntombifikile, G., and Middleton, Lyn, 2011, "The Triple Jump Assessment". Barrett, Terry, and Moore, Sarah, eds., New Approaches to Problem-Based Learning: Revitalising Your Practice in Higher Education, New York: Routledge.

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, 2011, "A Tuning-AHELO Conceptual Framework of Expected/ Desired

Learning Outcomes in Engineering", OECD Education Working Papers, No. 60: OECD.
OECD, 2012, AHELO Feasibility Study Report, Vol. 1, OECD.
OECD, 2013, AHELO Feasibility Study Report, Vol. 2, OECD.
小野和宏・大内章嗣・前田健康、2011、「学習者主体 PBL カリキュラムの構築－新潟大学歯学部口腔生命福祉学科 7 年のあゆみ」『新潟歯学会誌』41： 1-12。
小野和宏・松下佳代、2012、「社会で活躍する歯科医療人を育てるために－専門教育と連動したジェネリックスキルの育成と評価の現状」河合塾 PROG セミナー2013。
Rhodes, Terrel, L., ed., 2010, Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics, Washington, DC: Association of American Colleges and Universities.
Rhodes, Terrel, L., June 29, 2013. Effectiveness of Liberal Education and America's Promise (LEAP) and VALUE Rubrics.（講演資料）
Rhodes, Terrel, L., and Finley, Ashley, 2013, Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment, Washington, DC: AAC&U.
Smith, Richard, M., 1993, "The Triple-Jump Examination as an Assessment Tool in the Problem-Based Medical Curriculum at the University of Hawaii", Academic Medicine, 68(5): 366-72.
鈴木秀幸、2013、『スタンダード準拠評価－「思考力・判断力」の発達に基づく評価基準』図書文化。
Wiggins, Grant, and McTighe, Jay, 2005, Understanding by Design, 2nd ed., Alexandria, VA: ASCD.（＝2012、西岡加名恵訳、『理解をもたらすカリキュラム設計－「逆向き設計」の理論と方法』日本標準。）
吉田香奈・小澤孝一郎・於保幸正・古澤修一・西堀正英・田地豪、2013、「学生の主体的学びの確立に向けた授業方法の改善－教養ゼミへの PBL の導入」『京都大学高等教育研究』19: 25-36。
吉田武大、2011、「アメリカにおけるバリュールーブリックの動向」『関西国際大学教育総合研究所研究叢書』4: 1-12。