

大学における学習成果としての能力とその評価

—— 標準化をめぐる ——

京都大学
松下 佳代

<キーワード>学習成果、能力（コンピテンス）、ジェネリックスキル、評価、標準化、標準テスト、AHELO、VALUE ルーブリック

1. 問題と目的

1990年代以降（日本ではとくに2000年代以降）、さまざまな形で提唱されるようになった能力のことを私たちは〈新しい能力〉という言葉で言い表してきた（松下編, 2010）。いいかえれば、〈新しい能力〉とは、後期近代において求められるようになった能力のことである。

〈新しい能力〉の特徴は、①多くの国々で共通に、また、初等・中等教育から高等教育・職業教育、労働政策にいたるまでの幅広い範囲で主張されていること、②目標として掲げられるだけでなく、評価の対象とされていること、③知識・技能などの認知的側面だけでなく、興味・関心などの情意的側面や対人関係能力などの社会的側面をも含む人間の能力の全体を包含していること、といった点にある（松下, 2014b, p.15）。

本稿では、大学教育を対象をしばり、〈新しい能力〉の評価について、「標準化（standardization）」の問題を中心に検討を行う。

わが国の大学教育においては、2000年代半ば以降、「就職基礎能力」「社会人基礎力」「学士力」「就業力」「基礎的・汎用的能力」などの形で〈新しい能力〉が提案されてきた。なかでも、学士課程共通の学習成果と定義される「学士力」を導入した中教審答申「学士課程教育の構築に向けて」（2008年）以来、「学習成果（learning outcomes）」としての能力という考え方が広がり、評価についてもさまざまな試みがなされるようになってきた。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」（2012年）では、学習成果の評価のための測定手法

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あげられている。

こうした学習成果の評価については、〈学習成果を、学生の自己報告によって間接的に評価するのか、学生の知識や能力の表出によって直接的に評価するのか〉、〈心理測定的パラダイムに立つのか、オルターナティブ・アセスメントのパラダイムに立つのか〉（表1）という2つの軸を立てることで、その基本的な構図を描くことができる（図1）¹。

表1 評価の2つのパラダイム

	心理測定的パラダイム	オルターナティブ・アセスメントのパラダイム
学問的基盤	心理測定学	構成主義、状況論
評価目的	説明責任、ベンチマーキング プログラム評価	教授・学習の改善 授業評価
評価機能	総括的評価	形成的評価
評価項目	分割可能性	複合性
評価場面	脱文脈性 統制された条件	文脈性 シミュレーション、真正の文脈
評価要件	信頼性を重視	妥当性を重視
評価基準	客観性	間主観性
評価データ	量的データ	質的データ
評価主体	評価専門家、政策担当者	実践者自身
評価方法	標準テスト（客観テスト）、 学生調査 など	真正の評価、ポートフォリオ評価、 パフォーマンス評価 など

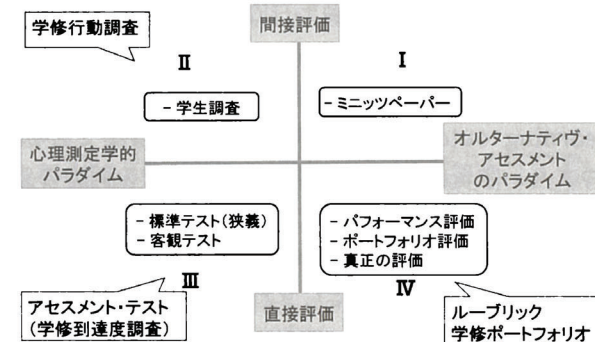


図1 学習成果の評価の構図

このように、大学教育において学習成果が多様な方法を通じて評価されようとしているなかで、とりわけ注目されるのが「標準化」をめぐる動きである。吉田（2013）は、大学教育のグローバル化のアクターとして、国家・国際機関・教育産業の三者をあげているが、同様のことは、評価の標準化にもあてはまる。文科省によって「アセスメント・テスト」が構想され、OECDによって AHELO (Assessment of Higher Education Learning Outcomes: 高等教育における学習成果の評価) のフィージビリティ・スタディ（試行調査）が実施され、業者によってジェネリックスキル (generic skills) を測定するテストが開発され教育現場に普及しつつある。これらはすべて評価の標準化を志向する動きである。しかしながら、他方で、標準化に對抗する動きもみられる。

本稿の目的は、このような大学の学習成果の評価における標準化の動きとそれをめぐる論点を検討することにある。具体的な対象としては、OECD-AHELO で用いられたジェネリックスキルのための標準テスト、およびわが国の大学教育にも影響を与えている AAC&U (Association of American Colleges and Universities: アメリカ大学・カレッジ協会) の VALUE プロジェクト (Valid Assessment of Learning in Undergraduate Education: 学士課程教育における学習の妥当な評価) を取り上げることとする。

2. 学習成果の評価の標準化

(1) 学習成果としての能力

学習成果の評価における標準化の問題に入る前に、今日、「学習成果」という言葉には独特の意味が付与されていることを指摘しておきたい。学習成果は、単なる学習の結果と同義ではない。今日の「学習成果」は、「意図された学習成果 (intended learning outcomes)」と「達成された学習成果 (achieved learning outcomes)」の両方を表わす言葉である。つまり、学習成果には、目標と評価対象の両方の意味が含まれているのである。ウィギンズら (Wiggins & McTighe, 2005) は、「成果」について、「指導によってもたらされることが意図された成果 (intended outcomes)」の略記であるとし、「意図された成果とは、求められている成果、すな

わち教育者が取り組む特定のゴールである」(邦訳 p.339) と説明しているが、これは近年新しく加わった前者の意味に焦点をあてたものになっている。

「学習成果としての能力」とは、このような意味での学習成果として、知識や技能ではなく「能力」を掲げることである。能力の代表的なものに「コンピテンス (competence)」がある。「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、生涯学習にいた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・実務的スキル、倫理的価値観のダイナミックな結合」(OECD, 2009, p.3) 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等の結合体としてとらえられている。

しかし、このような能力はきわめて包括的で、また曖昧なものであるため、このままでは目標や評価対象になりにくい。そこで、学習成果という形で具体化されることになる。つまり、学習成果とは、「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 (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, p.3) なのである。ここでのコンピテンスと学習成果の関係は、方向目標と到達目標の関係とのアナロジーでとらえることができる。

(2) 評価における「標準化」の意味

では、学習成果の評価における標準化とはどのようなことをさすのだろうか。ウィギンズら (Wiggins & McTighe, 2005) によれば、「標準化された (standardized)」評価とは、「すべての生徒について、運営上の条件と手順が統一されているテストや評価方法 (assessment)」を言い表す (邦訳 p.406)。この定義はかなり広いものであり、これにもとづけば、運転免許の路上検定のように、テストとして画一的に課されるパフォーマンス課題も「標準テスト」ということになる。

一方、西岡 (2005) は、標準テストについて、「検査実施時の受験者に対する教示のやり方、問題項目の呈示法、解答法の指示、検査時間などの検査実施法、各項目に対する受験者の反応の採点法が厳

密に規定されており、しかも受験者個人の結果は準拠集団の得点分布に基づいて作成された集団規準に照らして得点が解釈できるように作成されているテストのこと」(p.185)としている。つまり、ウィギンズらのいう①運営上の条件と手順が統一されていることに加えて、②集団規準準拠テストであることが、標準テストの条件としてあげられている。本稿では、ウィギンズらの定義を広義、西岡の定義を狭義としておこう。

評価における標準化の問題は従来、狭義の標準テストの問題として語られてきた (Ravitch, 2010)。そして、客観テスト形式での標準テストでは「真の学力」や高次の統合的な能力が評価できないという批判のもとで、そのオルターナティブとしてパフォーマンス評価や真正の評価が提案されてきたのである (FairTest²; Wiggins, 1998)。しかしながら、今日、新しいタイプの標準テストが使われるようになったことにより、この図式は変わりつつある。

(3) パフォーマンス評価型の標準テスト

—CLA のパフォーマンス評価—

新しいタイプの標準テストの典型が、米国の CAE (Council for Aid to Education: 教育支援審議会)³ によって開発され、AHELO のジェネリックスキルの評価に採用された、CLA (Collegiate Learning Assessment: 大学生学習評価) である (Benjamin et al., 2009)。

AHELO では、ジェネリックスキル、経済学、工学の 3 つの柱でテストが実施され、それぞれについて背景情報を得るための質問紙調査が学生・教員・機関に対して行われた (当初は付加価値分析も加えた 4 つの柱で計画されていたが、付加価値分析は実施にはいたらなかった) (OECD, 2012, 2013)。ジェネリックスキルのテストは 120 分で、記述式問題 (大問 2 問から 1 問を選択) と多肢選択問題 (25 問) に解答することになっている。AHELO のテスト・質問紙は、ACER (Australian Center for Educational Research: オーストラリア教育研究所) を中心とする国際的なコンソーシアムに委託されており、記述式問題は CAE、多肢選択問題は ACER が担当した。ACER は OECD-PISA においても中心的な役割を果たしている組織である。

AHELO で実際に用いられた CLA の記述式問題

のひとつをみてみよう (OECD, 2012, pp.220-233)。

Box1 課題文

あなたはミルタウン市の職員です。市長室にはこのところ、ミラクル湖で最近見つかった奇形ナマズについて市民や報道機関から多くの問い合わせが寄せられています。ミルタウン市長サリー・ビゲローは、明日の夜、ミルタウン市審議会での問題を議論しようと考えています。

明日の会議の準備をするにあたって、ビゲロー市長は、あなたに、文書ライブラリ (画面右側) に収められている文書を検討するように依頼してきました。

以下の質問に解答する際には、あなたの見解を支持するのに必要なことがらを最大限詳しく記述してください。あなたの解答は、情報の正確さだけでなく、次の点からも判断されます。

- ・考えがどのくらい明確に提示されているか
- ・考えがどのくらい効果的に組み立てられているか
- ・情報がどのくらいうまくカバーされているか

あなたの個人的な価値観や経験は重要ですが、質問はすべて、以上の情報および文書ライブラリの情報のみにもとづいて答えてください。

各質問に対する解答は下のボックスの中に記入してください。書きただけ書いてかまいません。画面のボックスの大きさに制限されることはありません。

まず、Box1 のような課題文が与えられた。「文書ライブラリ」には以下の文書が収められている。

- ・文書 1・2: 奇形ナマズについて報じた地元新聞の記事
- ・文書 3: 環境保全区域周辺地図
- ・文書 4: 工場周辺の水質調査を行った大学教授から工場をもつ会社社長への手紙
- ・文書 5: 上記の水質調査で用いられたテストについての説明 (専門書からの抜粋)
- ・文書 6: 環境保全区域の生態学調査を行った別の大学教授へのインタビュー (地元ラジオ番組のテープ起こし)
- ・文書 7: 過去 7 年間の水質の変化を示すグラフ

これらの文書は、多様な情報源、形式、内容をカバーしており、分量も A4 に換算して 12 ページにおよぶ。総合的に判断すると、近親交配説が優勢のよ

Box2 問題

文書ライブラリの資料からは、三つ目ナマズについて、少なくとも 3 通りの説明が存在することがわかります。近親交配、寄生虫、Db09 やバリジウムによる汚染の 3 つです。

問 1 a

文書ライブラリで議論されている情報、エビデンス、事実、理論のうち、近親交配説を支持しているものはどれですか。

問 1 b

近親交配説に反する議論にはどんなものがありますか。

[中略。問 2・3 では寄生虫説、汚染説について同様の質問が行われ、問 4 では、他に説明がないかが尋ねられる。]

問 5

三つ目ナマズの説明として最も正しいものはどれですか。またそれはなぜですか。あなたの解答はこの説明と対立する他の議論にどう反駁するかを論じるものでなければなりません。

問 6

市の審議会は今後、何をすべきですか。工場を閉鎖すべきか、環境保全区域を閉鎖すべきか、ミラクル湖を再生のために閉鎖すべきか、ブッシュ湖からの飲料水の取水を中止すべきか、など。あるいは何もしない方がよいのか。あなたの勧告の根拠は何ですか。

表2 OECD-AHELO におけるパフォーマンス課題（ジェネリックスキル）の評価基準

	分析的推論と評価	問題解決	効果的な文章作成
0	・測定不能。学生は質問に答えようとしていないため、評価できない。	・分析的推論と評価が0の場合は、問題解決も0となる。	・分析的推論と評価が0の場合は、効果的な文章作成も0となる。
1	・文書ライブラリで提示されている立論（もしくは分類すべき対象すべてについての重要な特徴）を支持または反駁する事実やアイデアを、全く認識していない。分析のエビデンスを示していない。 ・重要な情報を無視、もしくはひどく誤って解釈している。 ・エビデンスの質についての主張は全く示されておらず、解答は不確かな情報に基づいている。	・明解な決定や、決定に至るまでの妥当な根拠が示せていない。 *該当する場合： ・結論から論理的に導かれる手順を示していない。 ・追加調査の必要性を認識していない、あるいは、未解決の問題を扱う調査を提案していない。	・説得力のある議論を展開できていない。文章は整っておらず、不明確である。 ・事実やアイデアについての詳述は行われていない。
6	・文書ライブラリで提示されているすべての主要な立論（もしくは分類すべき対象すべてについての重要な特徴）を支持または反駁する事実やアイデアを、ほぼすべて認識している。明白でないことがらについても分析を進めている。 ・文書ライブラリから得た大部分の情報について正確に理解している。 ・情報の質について数個の正確な主張を行っている。	・決定と確固とした根拠が示せており、それらは複数の情報源から導き出された信用できるエビデンスに裏打ちされている。他の選択肢も考慮した上で、入手できるエビデンスから最善の決定を提示している。 *該当する場合： ・結論から論理的に導かれる手順を示している。 ・その意味するところも検討されている。 ・追加調査の必要性を認識し、未解決の問題の大半を扱うための具体的な調査が提案されている。	・論理的に一貫性のあるやり方で解答を構成することにより、書き手の立論を非常にわかりやすくしている。 ・それぞれの立論に関連する事実やアイデアを適切な形で包括的に詳述している。 情報源が明記されている。

（注）OECD（2012, pp.234-236）より作成。レベル2～5は省略。

うに思われるが、どの文書の内容も何らかの点で批判の余地があるものになっている。

この課題文に対して、Box2のような問題が出された。問題は全部で9問だが、問1～3は、3つの説を肯定もしくは否定する根拠を問うものであり、問5では、その中で最も適切と思われるものを選び、他の立論に反駁することを要求される。さらに、問6では、アクションのための提案を、根拠を含めて求められる。全体的にディベート的な形式になっているが、ディベートのように、肯定側・否定側の立場を役割として取るのではなく、事実論題に対して自分の見解を明らかにし、さらに政策論題にまで踏みこむことを求められている。

ウィギンズらは、パフォーマンス課題をデザインする際に考慮すべき要素として、「ゴール」「役割」「相手 (audience)」「状況」「完成作品・実演・意図」「成功を評価するスタンダードと規準」という6つの要素を抽出している (Wiggins & McTighe, 2005, 邦訳 pp.190f)。この問題には、これらの要素がすべて含まれていることが見てとれよう。ただし、ウィギンズらは、パフォーマンス課題の条件として、「内容は秘密ではないこと (not secure)」、すなわち、「課題、評価の規準とパフォーマンス・スタンダードはあらかじめ知らされており、生徒の学業を導く」ことをあげている (邦訳 p.185)。その条件も考慮に入れると、この問題は、真正のパフォーマ

ンス課題と短答式テストや小テストとの中間にある「アカデミック・プロンプト」ということになる。とはいえ、同じ OECD の PISA の問題と比べれば、AHELO のこの問題は、はるかにパフォーマンス課題的な性格を強くもったものといえる。

学生の解答は、ルーブリック式の評価基準によって評価される (表2)。規準は、「分析的推論と評価 (Analytic Reasoning & Evaluation)」「問題解決 (Problem Solving)」、「効果的な文章作成 (Writing Effectiveness)」の3つで、レベルは0～6の7段階である⁴。

このように、問題の性格からみても、評価基準からみても、CLA は、「パフォーマンス評価型の標準テスト」ということができる。

(4) 標準化をめぐる図式の変化

CLA のようなパフォーマンス評価型の標準テストでは、標準テストの意味は広義に拡張されており、表1の図式に変化が生じている。従来の標準テストの多くは、心理測定学的パラダイムに立って作られており、客観テスト形式で、評価項目は細かく区切られ、評価場面も脱文脈的であり、高次の統合的な能力を評価するには適していなかった。しかし、パフォーマンス評価型の標準テストでは、評価項目は複合的になり、評価場面も「仮想的ではあるがリアルな状況」が設定されている。また、評価基準にルー

ブリックを用いることによって、主観性をコントロールし、質的データを量的データに変換して統計的分析を行いやすいようにしている。つまり、基本的には心理測定学的パラダイムに立ちながらも、オルタナティブ・アセスメントのパラダイムを部分的に取り込むことによって、高次の統合的な能力を評価するという要請に応えようとしているといえる。

(5) 評価の標準化の抱える問題

そもそも、なぜ今日、標準化が求められるようになったのか。標準化の重要な機能は「比較可能性 (comparability)」を担保することである。「運営上の条件と手順が統一されているから」こそ、PISA では国家間での、AHELO では国境を越えた大学間での、全国学力・学習状況調査では自治体・学校間での比較が可能になる。それによって、行政機関は政策評価、大学・学校は教育評価を行い、エビデンスにもとづいて説明責任を果たすことができ、一方、生徒・学生や保護者は教育サービスについての情報を得ることができる、とされる。とくに近年では、「ベンチマーキング」の手法（同じ分野のすぐれた事例との比較を通じて、改善すべき点を探し出す手法）をとることによってすぐれた事例に学びながら教育改革を推進することが、それぞれの国、自治体、大学・学校に要請されるようになっていく⁵。そのためには、比較可能性の担保された標準テストが有効だと考えられるようになったのである。

本稿でみてきたように、標準テストの範囲は今では、パフォーマンス評価の領域にまで広がっている。だが、そこに問題はないのだろうか。

①ハイ・ステイクス化

標準テストは、それが意思決定で重要な役割を果たすことでハイ・ステイクスになりやすく、そうであればあるほど、教育や学習に弊害をもたらしやすくなると指摘されてきた。社会科学者のキャンベル (Campbell, D. T.) は、早くも70年代の半ばに、さまざまな社会システムについて、後に「キャンベルの法則 (Campbell's law)」と呼ばれることになる、次のような悲観的な法則が成り立つことを見出した。「いかなる定量的な社会指標も（ときには定性的な社会指標の場合でも）、それが社会的な意思決定の場でより多く用いられるほど、その指標は退廃への圧力を受けやすくなり、その指標がモニター

しようとする社会的プロセスを歪曲し、墮落させがちになる」(Campbell, 1976, p.49)。

実際、標準テストがハイ・ステイクスになることで、学校関係者が生徒の答案を改ざんする、テスト問題のヒントや答えを教える、テスト問題について事前に指導するなどの不正行為が、国内外で数多く報告されてきた⁶。不正行為ではないにしても、今日、わが国の学校現場では、全国学力・学習状況調査や自治体の実施する学力テストに向けた指導（過去問の訓練と解説など）が常態化していることも報告されている（寺岸, 2014）。

AHELO の場合は、まだフィージビリティ・スタディの段階であり、大学別の結果も公表されていないことからハイ・ステイクス化してはいないが、大学教育において開発されつつあるさまざまな標準テストが、将来、ハイ・ステイクス化する可能性は決して小さくはない。教育のグローバル化、サービス産業化、情報化、質保証の要請など、その前提条件は整っているからである。

②文脈的要因

ハイ・ステイクス化は、標準テストの形式に関係なく生じうる問題だが、一方、パフォーマンス評価型の標準テストならではの問題もある。

前述のように、AHELO のジェネリックスキルの記述式問題（パフォーマンス課題）では、現実世界の問題解決状況が設定されていた。しかし、AHELO の報告書では、こうした課題は、文化的な状況に埋め込まれていて、国や制度によって多様性を被りやすいこと、CLA の作った課題は国際的に使用するには過度に「アメリカ的」であったことが指摘されている (OECD, 2013, p.164, 169)。

現実世界の問題を設定しようとするれば、そこにはその文脈にどのくらい慣れ親しんでいるかという要因が必ず入ってくる。PISA の場合も同様の問題にさらされてはいるが、AHELO のジェネリックスキルの記述式問題はより高次の統合的な能力を求めるものであり、問題数が限定されるだけに、文脈的要因の影響がより大きくなったと考えられる。AHELO では、ジェネリックスキルを単独で評価するのではなく、工学や経済学のように国際的な合意を作りやすい（問題解決の文脈も共有しやすい）専門分野において、「専門分野特有のジェネリックな」コンピテンス (“discipline-specific generic”

competencies) を評価することで、この問題を回避することが検討されている (OECD, 2013, pp. 164f)。

この問題は、国際的な標準テストに限らず、国内や自治体で実施される標準テストであったとしても、問題の文脈性・真正性を高めようとするときには必ず生じる問題である。

③信頼性と妥当性、ベンチマーキングと改善のトレードオフ

従来、客観テストでは信頼性は高いが妥当性（とくに高次の統合的な能力を評価する上での妥当性）は担保しにくく、逆にパフォーマンス課題では妥当性は高いが信頼性は担保しにくいといわれてきた。AHELO でも、多肢選択問題と記述式問題⁷の間でこのトレードオフの関係がみられたと報告されている。AHELO の下した暫定的結論は、AHELO の主な目的が教授・学習の改善にあるのであれば、記述式問題を含めることで妥当性を向上させ、主な目的が国際的なベンチマークの提供にあるのであれば、多肢選択問題だけにして信頼性を高めコストを下げるべきだというものであった (OECD, 2013, p.165)。見方をかえれば、パフォーマンス評価型の標準テストでは、このトレードオフの関係を変えることはできなかった、ということになる。

このように、パフォーマンス評価型の標準テストは、高次の統合的な能力を評価対象とすることには成功したが、標準テストの陥りやすいハイ・ステイクス化のおそれは残したまま、パフォーマンス評価の困難をも呼び込むことになったと考えられる。

3. 標準化への対抗軸

(1) 標準化によらないパフォーマンス評価

—AAC&U の VALUE プロジェクト—

パフォーマンス課題型の標準テストが、標準テストの範囲を拡張しようとする動きだったのに対し、他方、CLA のような標準テストの普及を批判し、標準化によらないパフォーマンス評価の独自性を追求する動きも強まっている。

標準化への対抗軸を明確に打ち出したのが、AAC&U⁸ の VALUE プロジェクトである。AAC&U は、2005 年から LEAP (Liberal Education & America's Promise: 教養教育とア

メリカの約束) というプロジェクトを展開し、その中で、「本質的学習成果 (Essential Learning Outcomes)」の提案を行った (AAC&U, 2007)。「本質的学習成果」は、<1>人類の文化や自然界についての知識、<2>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)、<3>個人的・社会的責任 (市民としての知識と関与、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)、<4>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) という 4 つの柱とそのサブカテゴリーからなる。構成も内容も「学士力」とよく似ており、「学士力」の基盤となったと考えられるものである (松下, 2014a)。

さらに2007年からは、それに対する評価をデザインするプロジェクトとして、VALUE プロジェクトが開始された。これは、本質的学習成果として抽出された15領域 (後に16領域に拡大) のループブリック (VALUE ループブリック) を共同開発し、e ポートフォリオに収められた学生のワーク (作品や活動) について「真正の評価」を行おうとするものである。ここでの真正性には、〈その領域の本質的なパフォーマンスを正しく反映した、リアルな文脈で行われる評価〉という意味に加えて、〈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活動の中で行われる評価〉という意味が含まれており、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

表3は文章コミュニケーションの VALUE ループブリックである (Rhodes, 2010; 松下, 2012)。レベル1～4は正確に学年に対応しているわけではないが、レベル1は入学時のレベル、レベル4は学士課程修了に期待されるレベルを表している。つまり、VALUE ループブリックは、学生の成長の把握が可能な長期的ループブリックだということができる。

VALUE ループブリックはまた、各大学・部局のループブリックのリソースになるような「メタループブリック」であり、実際に各大学・部局全体の長期的ループブリックや個別の授業科目の採点用ループブリックとして使う際には、大学・部局や科目の文脈にあわせてローカライズすることになっている (Rhodes & Finley, 2013; 松下, 2014a)。VALUE

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 4	マイルストーン 3 2	ベンチマーク 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。	……	……
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。	……	……
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。	……	……
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。	……	……
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。	……	……
			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。
			適切で関連性のある内容を用いることによって、作品の何れ所かで、シンプルなアイデアを展開している。
			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。
			アイデアを裏づけるために、資料を使おうとしている。
			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）松下（2012, p.103）より一部抜粋。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。

ルーブリックをメタルーブリックとして共有しつつ、そこに大学・部局や科目の多様性に応じた修正を施すことによって、「共通性と多様性の調停」を行うことが意図されているのである。これは、標準テストによらないで標準化のもつ比較可能性という機能をとりこもうとするものだといえる。

(2) 比較可能性への対応

とはいうものの、各大学・部局で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も生じる。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だろうか。

この問題に対処するために、AAC&Uでは現在、学生のワーク（アンカー）とルーブリックによる評価結果とを大学を越えて共有することで、比較可能性を高めようというプロジェクトが行われつつある。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・データを構築することがめざされている。

4. 結論

本稿では、学習成果としての能力の評価における標準化の動きとその論点を検討してきた。標準化をめぐる対立は、今や、〈標準テスト vs. パフォーマンス評価〉という形だけではなく、パフォーマンス評価内部でもみられるようになってきている。パフォーマンス評価型の標準テストが、高次の統合的な能力を評価できるようパフォーマンス課題やルーブリッ

クをとりこむ一方で、標準化によらないパフォーマンス評価の方では、標準化のもつ比較可能性の機能を別の形で具体化するよう対応がはかられている。

このような“相互浸透”は学習成果の評価を新しい段階に導くことになるのか、それとも双方の特徴を削いだ妥協の産物をもたらすにすぎないのか。自分もその中に身を置きながら、成り行きを注視していきたい。

注

- 1 表1・図1については、松下（2012）参照。ただし、いくつかの加筆修正を行っている。
- 2 FairTest による標準テスト批判については Web ページ “What’s Wrong With Standardized Tests” 参照(<http://www.fairtest.org/facts/whatwron.htm>)
- 3 CAE は測定・評価を主な業務とする NPO であり、初等・中等教育の「全米共通教育スタンダード」にそったテスト（PARCC, SBAC）にも関わっている。
- 4 通常の CLA の採点基準は、「文章作成の技法（Writing Mechanics）」を加えた4つの規準で、1～6の6段階であるが、各セルの記述語は同じである。
- 5 PISA の場合、ベンチマーキングがどのように使われているかについては、松下（2014b）参照。
- 6 足立区の小学校で区の学力テストを情緒障害児に受けさせなかったケース（2006年）やアメリカで裁判になったアトランタ学区での大規模な不正行為事件（2009年）など。
- 7 AHELO では、工学・経済学の分野も含め、記述式

問題に「パフォーマンス課題」が用いられている。

- 8 AAC&U は、教養教育の質や公共的役割の向上を目的として1915年に設立された組織であり、1,300校あまりの加盟校で構成されている。

文献

- AAC&U (2007). *College learning for the new global century*. AAC&U.
- Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., Nemeth, A., & Shavelson, R. (2009). *Returning to learning in an age of assessment: Introducing the rationale of the Collegiate Learning Assessment*. (<http://www.collegiatelearningassessment.org/>)
- Campbell, D. T. (1976). *Assessing the impact of planned social change*. The Public Affairs Center, Dartmouth College.
- 松下佳代編 (2010). 『〈新しい能力〉は教育を変えるかー学力・リテラシー・コンピテンシーー』 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ー学習評価の構図の分析にもとづいてー」『京都大学高等教育研究』第18号, 75-114.
- 松下佳代 (2014a). 「学習成果としての能力とその評価ールーブリックを用いた評価の可能性と課題ー」『名古屋高等教育研究』第14号, 209-229.
- 松下佳代 (2014b). 「PISA リテラシーを飼いならすー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ー」『教育学研究』第81巻第2号, 14-27.
- 西岡加名恵 (2005). 「学力調査」田中耕治編『よくわかる教育評価』ミネルヴァ書房.
- OECD (2009). *A Tuning-AHELO conceptual framework of expected/desired learning outcomes in engineering*. Tuning Association.
- OECD (2012). *AHELO feasibility study report, Vol.1*. OECD Publishing.
- OECD (2013). *AHELO feasibility study report, Vol.2*. OECD Publishing.
- Ravitch, D. (2010). *The death and life of the great American school system*. Basic Books.
- Rhodes, T. (Ed.) (2010). *Assessing outcomes and improving achievement*. AAC&U.
- Rhodes, T. L., & Finley, A. (2013). *Using the*

VALUE rubrics for improvement of learning and authentic assessment. AAC&U.

- 寺岸和光 (2014). 「協同性が支える『活用する力』の育成プロセスー学校の塾化に抗して学力向上を実現するー」『日本教育学会中部地区研究会2014・報告論集』16-37.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). ASCD. ウィギンズ, G. & マクタイ, J. (2012). 『理解をもたらすカリキュラム設計ー「逆向き設計」の理論と方法ー』(西岡加名恵訳) 日本標準.
- 吉田 文 (2013). 「グローバリゼーションと大学」吉田文他『シリーズ大学1 グローバリゼーション、社会変動と大学』岩波書店.