高等教育の質保証における学習成果の位置 : 国際比較と学習成果の可視化

著者	塚原 修一, 濱名 篤, 山口 アンナ真美
雑誌名	教育総合研究叢書 = Studies on education
号	10
ページ	177-189
発行年	2017-03-31
URL	http://id.nii.ac.jp/1084/00000503/

高等教育の質保証における学習成果の位置
－国際比較と学習成果の可視化－

Quality Assurance in Higher Education: An International Comparison of Assessment Models and Visualization of Learning Outcomes

塚原 修一*　　濱名 篤**　　山口 アンナ真美***
Shuichi TSUKAHARA　Atsushi HAMANA　Ana Mami YAMAGUCHI

抄　録

高等教育の質保証における学習成果の位置づけについて、学習成果の評価を高等教育機関の内部質保証にゆだねる欧州のチューニング、全国的な統一試験によって学習成果の直接評価を実施するブラジル、共通の判断基準表（ルーブリック）を用いて高等教育機関が学習成果を評価する米国を比較した。制度の概要とともに実施上の課題を調査した結果として、進学率が高く、高等教育機関の多様性が大きい日本が参照してよいのは米国の方式であり、その方向にそって学習成果の可視化を推進することが課題であることが示唆された。

Ⅰ　はじめに

質保証は高等教育の評価における目的概念のひとつであり、その卓越性とともに、平均的ないしすべてが到達するべき水準に注目するものである。評価の対象は投入要因（教育条件、教員資格など）や教育の過程（カリキュラム、教授学習過程など）から産出（学習成果、成績など）におよび、高等教育機関は内部質保証の仕組みを構築しなければならない（川嶋 2013、8-9）。そのさい、質保証において学習成果をどのように位置づけるか、学習成果という情報をいかに活用するか、さらには高等教育の質保証とそれ以外の目的概念（卓越性、進学機会とその平等性など）の調整をどうするかといった課題が浮上することになろう。

高等教育の評価に関する先行研究については、国内の総説に、米澤（2006）、濱名（2007）、三好（2013）、村澤（2014）がある。それらによれば、認証評価制度の導入

---

*　関西国際大学客員教授 教育総合研究所共同研究員
** 関西国際大学人間学部 教育総合研究所学内研究員
***北海道教育大学非常勤講師

（2004年）に先立つ2000年前後には外国調査がなされたが、関心は大学評価の実施にともなう諸問題に移り、学習成果の評価は注目されなかった。学術誌の特集（『IDE 現代の高等教育』８月、2011年、『大学評価研究』11号、2012年）にも学習成果は扱われていない。山内・原（2012、2013）はその例外である。諸外国の動向は各節で述べる。

学習成果の評価は初等中等教育分野には多くの研究があるが、高等教育分野では2008年に提唱されたOECDのAHELO（Assessment of Higher Education Learning Outcomes）プロジェクトが契機のひとつとなった。ニュッシュ（2008）は学習成果の測定手段（試験）について７か国18事例を紹介した。AHELOの成果はTremblay et al. (2012)などにまとめられた。深堀（2015b、295-296、302-303）は10か国の事例を、大学教育の範囲と水準に緩やかな標準性をもたせる型（欧州諸国）と、出口段階での学習成果の評価を志向する型（米国、韓国、メキシコ）に整理した。

筆者（塚原）はAHELOの国内調査の責任者をつとめ、深堀（2015b）にも関与した。学習成果を評価する端的な方法は試験である。学校の試験は学習成果の形成的な評価を目的とすることが多いが、総括的な評価である質保証では妥当性・信頼性の高い学習成果測定が大きな負担をともなったり、学生・教員・大学に緊張関係をもたらしたりしかねない（深堀 2015b、23-25）。結果が重大な意味をもつ試験には、教育や学習を試験対策にかたむける副作用もある（斉藤 1995、Koretz 2008、Reese 2013）。とはいえ評価に関する選択肢は限られ、1.質保証の重点を教育の投入要因や過程において学習成果の扱いを軽くする、2.上記の懸念はともかく学習成果の測定を実施する、3.懸念が生じにくい評価測定のあり方を考案するなどであろうか。

これらを作業仮説として世界の代表的な事例を調査した。作業仮説１の対象は欧州で2000年に発案されて世界に普及したチューニングである。作業仮説２はブラジルの全国学生学力試験であり、最終学年の学生を対象に2004年からつづく世界にまれな事例である。作業仮説３は米国のルーブリックを用いた評価であり、主に2007年以降に展開された。以下では、学習成果を評価する制度の説明とともに、制度が導入された理由や制度を円滑に運用するための工夫に注目し、社会状況と制度の対応関係を検討するなかで日本にふさわしい制度のあり方を考察する。

Ⅱ　欧州のチューニング

１．活動の概要

深堀（2015a）にそって紹介する。欧州のボローニャ・プロセスとは、各国の学位を３段階（学士・修士・博士）に統合し、単位制度の枠組みを共通化して、欧州高等教育圏の確立をめざす政府主導の取り組みである。学士課程は３年を標準とし、平均的な学生が25～30時間で達成可能な学習成果を１単位（ECTS、欧州共通単位）として、年間に60単位、３年間で180単位の取得を学士の要件とする。チューニングとは、その

実質化にむけて、大学の自律性と多様性を尊重しつつ教育の質に一定の標準性を導入する取り組みである。2000年に発案され、大学の自発的な参加によって推進された。

チューニングでは学習成果とコンピテンスが区別される。学習成果とは科目の履修により習得が期待される知識や能力であり、単位取得の条件であって、測定可能でなければならない。コンピテンスとは知識、理解、能力などが有機的に結合したもので、その涵養が学位プログラムの目的となる[1)]。標準性の導入はコンピテンスの段階で行われ、コンピテンスの枠組みが大学と社会（卒業生とその雇用主）の協議によって作成される。2015年までに、各専門分野（と一般的技能）について45の参照基準が定義された。各大学は、参照基準から重点的に育成するコンピテンスを選択して学位プロフィールを定義し、これを達成する学位プログラムを設計して必要な科目を配置する。そのうえで、学生の学習成果を評価して単位を認定するとともに、科目と学位プログラムを評価して内部質保証を推進する。

２．欧州からの考察

ボローニャ・プロセスの趣旨は、欧州の大学制度を統一することではなく、各国の差異を認めて比較可能とすることにある。欧州の大学は各国の制度により質保証されているが、水準や内容が同じではなかった。そこで、授業科目や学習成果の上位に包括的なコンピテンスの概念をおいて、欧州各国の学位プログラムを比較可能にした。大学（各専門分野の専門家）と社会（卒業生とその雇用主）の協議によって作成されたコンピテンスの枠組みには、職業的レリバンス（職業に必要とされる能力）が反映される。これらにおいてチューニングは先進的である。地域（欧州）における高等教育圏の確立という目標は大学にとって受け入れやすく、それを通して大学制度を国際的に比較可能とする途を開いたことが、世界に普及した理由のひとつであろう。

チューニングでは、学習成果の評価が各大学の内部質保証にゆだねられた。欧州の大学は、各国の制度によって「大学教育の範囲と水準に緩やかな標準性をもたせる」方式で質保証されているので、それに対応した措置といえる。さらに、国公立大学や政府が経費を支出している大学が欧州には多く、大学間格差が存在しないことを理念とする国もある。それゆえ、チューニングの結果として学位プログラムを修正するとしても、大学にはそれを遂行する余裕があり、必要なら政府の支援が期待されよう。

Ⅲ　ブラジルの全国学生学力試験

１．試験の概要

ブラジルでは1990年代から教育を市場化・民営化する政策が導入された。大学の設置基準が緩和されて規模の拡大と変化がすすみ、1991年に51万人であった入学枠が2011年には445万人に膨張した。私学が増加して７割以上をしめたが、経営基盤が脆弱

なものもあった。政府は大学の管理と評価を強化し、2004年に全国学力評価制度を導入した。その中核が全国学生学力試験（ENADE）である。その成績を重視して大学は５段階に格付けされて結果が公表された。格付けは５と４が良好、３が平均的、２と１は不充分である。この試験の特徴は、学生に受験を義務づけるが、試験の成績は大学の格付けのみに使われ、学生個人の評価や卒業判定には影響しないことにある。

学士課程の専門分野は約40に区分され、それぞれに「教育課程の国家指針」が制定された。これらはチューニングでいうコンピテンスに相当する。試験は学年末に実施される。専門分野を３群にわけて、各分野は３年ごとに試験の対象となる。試験は、当初は第１学年と最終学年の学生の各半数を対象としたが、のちに最終学年の学生全員を対象とするものになった。試験は筆記による。出題は一般教養（全分野に共通）が選択式８問と論述式２問で配点は25％、専門分野は選択式27問と論述式３問で配点は75％である。そのほか学生の意識調査９項目があり、試験時間は全体で４時間である。複数の専門家の判定によれば、教育学（教員養成課程）の試験問題は日本の大学院入試と同水準の良問である。受験者は毎年約50万人である（山口・塚原 2015)。

試験の実施主体であるブラジルの国立教育研究所には厳重な安全管理がほどこされ、試験問題の漏洩などの防止がはかられた。試験の成績と学生の評価を分離したのは、学生が試験対策や得点競争にはしる弊害をさけ、卒業判定などへの政府の関与を回避する措置とみられるが、別の問題を生み出した。すなわち、試験には出席するが解答を拒否した白紙の答案が頻出し、それが零点となって上位大学が低く格付けされる事例が生じた。これに対応して、試験の意義を学生に広報して解答をうながすとともに、解答拒否をなくすための措置として、2016年から個人の成績を卒業証書に記載することとしたが（塚原・山口 2015)、先送りされた。

２．大学の対応事例

上位大学３校と下位大学２校の事例を調査した。上位大学３校は、いずれも試験対策はしていないと回答した。ある上位大学では、教育課程に不備があることが試験の結果からわかり、その部分を修正した。評価が低かった上位大学では、その原因が教育課程の不備ではなく、受験意欲の低さにあることがわかった。そこで、試験の目的と役割について学生に向けた説明会を開催したところ、次回の試験から高い評価が得られた。上位大学の２校では、試験の専門分野分類が自校の学科構成と整合していないとして、修正を求めて認められた。上位大学の見解によれば、試験は必要であるが、管理と規制を強化するもので、地域の特徴や必要を反映した教育課程が実施できなくなった。一方、下位大学２校では、特別講義や模擬試験などの試験準備が学習成果を向上させるとして、試験は質保証に有効としていた（山口・塚原 2016)。

３．ブラジルからの考察

ブラジルの課題は、経営基盤が脆弱で内部質保証に期待しがたい私学の存在であった。これに対応して政府は外部評価制度を導入し、学習成果を重視した大学の格付けを公表した。その結果、学習成果が不充分（格付けが１）とされた大学の割合はいくらか減少し、大学進学者は格付けの低い大学を回避するようになった。また、大学の新増設に必要な条件を実質的に高めて、経営基盤が弱体な大学の増加を抑制したといえる。制度の運用では、試験を学習成果の評価として正常に機能させるために多くの努力がはらわれたが、試験の成績と学生の評価を分離する方式が定着したとはいいがたい。試験の影響は、上位大学では教育の硬直化による多様な学習や創造性の萎縮が懸念されていたが、下位大学では試験の準備が学習成果を高めると歓迎されていた。すなわち、この制度は質の保証には貢献している。試験の内容は日本の大学院入試と同水準と判定されたが、これは難易度が高いことを意味しよう。こうした出題が可能であるのは、ブラジルの純進学率が14.7％（2011年）と高くないためと考えられる。

Ⅳ　米国のルーブリックを用いた評価

１．これまでの経緯

ルーブリック（rubric）は絶対評価の判断基準表をさす教育用語として1980年代に使われはじめ（田中 2004)、高等教育分野では2000年代に注目された。スペリングス報告（2006年）を契機に学習成果の客観的測定法の開発がはじまるなかで、Banta (2007) や Braun and Wainer (2007) は付加価値（大学が付与した学力）の測定が困難とし、Arum and Roksa (2011) はテストと分析の工夫で測定可能とした。全米カレッジ大学協会（AAC&U）のVALUE（Valid Assessment Learning in Undergraduate Education）ルーブリック（2007年）は、テストなどにかわって学習成果を可視化する手段として開発された。学士課程教育を対象に、知的・実践スキル、個人的・社会的責任感、学習の統合という３領域に関する15種類の汎用的な規準表からなり（吉田 2013)、形成的評価としても活用できる総括的評価の方法といえる。

森（2015、236-238、245-248）によれば、米国のアクレディテーション制度に対しては、大学への投入要因を重視しすぎるとの批判が従来からあったが、20世紀末から学生の学習成果を明示することへの要求が高まった。スペリングス報告と高等教育法の改正（2008年）にいたる過程では、国立の評価機関の創設や、大学卒業者を対象とした統一試験の義務化も議論された。一部の評価団体は学習成果によって教育の効率を評価しはじめ、卒業率、就職率、資格試験の合格率などの証拠を大学に求めた。こうした単純な指標や標準テストに対して、学生の学習に関する現実を提示する手段としてVALUEルーブリックは開発された。

２．米国への訪問調査

ルーブリックを活用した評価の好事例を把握するため、以下の点に配慮して米国への訪問調査を実施した（塚原 2016)。第１に、教育目標が抽象的に設定されやすい領域に注目して、留学などの評価事例をとりあげた。第２に、機関研究の枠組みのなかで、学習成果の評価が教育改善に結びついている事例を探索した。第３に、日本と同じく米国の大学には多様性がみられるが、絶対評価と多様性をいかに関係づけているかに留意した。結果の概要はつぎのようである。

ある大学では、1 学年500名の学生の文章作成をルーブリックによって評価していた。ルーブリックとしては単純なものであるが、文章評価の経験がない教員には前年度に提出された文章を題材として評価の実習が事前になされた。評価の信頼性を確保するために、複数の教員がそれぞれ評価を行い、結果を対比して整合性を確認していた。

ルーブリックをもちいた評価によって学習成果を可視化するときには、評価データをどのような大学と比較して自らの特徴や改善すべき点を明らかにするかが重要である。そのさい、学生調査を実施して結果を分析すれば理解を深めることができる。このような活動を支援し、個別大学をこえた評価の信頼性を確保する場として大学コンソーシアムが機能していた。これは志を同じくする大学の組織体であり、そうした大学集団を参照基準とすることで大学の多様性にも対応していた。

留学の目的は、かつては外国語の知識・技能の修得など単一であったが、1990年代以降は、国際性の涵養、異文化理解などがくわわって目的が複合化した。留学の成果を評価するルーブリックもあるが、留学から期待される成果を学生にあらかじめ考えさせ、帰国後に省察的学習をうながすことが重要である。さらに、留学の成果を帰国後の学習にいかに結びつけるかを、留学に先立って学生に計画させる大学もあった。

ルーブリックによる評価を学習成果の改善にむすびつける方策として、ポートフォリオによって学生を個人単位で追跡している事例がみられた。学生の学習に対する指導・支援を充実させるとともに、学生調査の結果を統計的に分析するにとどまらず、回収した調査票を手元において回答した学生と面談して、(たとえば何かに不満であると）回答した理由をたずねて理解を深めていた。

３．米国からの考察

前述の森（2015）にある、実現しなかった２つの提案の方向は正反対である。国立の評価機関の創設という提案は現行制度への批判にねざすものであろう。米国の大学は入学しやすいが卒業しにくいといわれる。すなわち、学習成果は高い水準で厳格に評価されているが、それゆえ卒業率は高まらず、教育の効率が低い状態にある。この提案は教育や学習支援におけるいっそうの努力を大学にもとめるもので、伝統的な大学が対象であろう。一方、統一試験の義務化は学習成果の向上をはかるものある。学

習成果が不充分な学生を卒業させているような新興大学が主な対象であろう[2)]。

VALUEルーブリックは、チューニングでいうコンピテンスの段階で構成された抽象的・汎用的なものである。ルーブリックを活用した評価には心理測定学にもとづく標準テストほどの厳密性は期待できないが、より高次の（あるいは統合的な）能力を形成的に評価できる可能性がある（松下 2012、松下ほか 2013）。さらに、VALUEルーブリックは導入のさいに大学の文脈などにそくして書き換える必要がある（吉田 2013）。それにより、チューニングでいう学位プロフィールの定義（重点的に育成するコンピテンスの選択）と学位プログラムの設計（コンピテンスから科目と学習成果への具体化）がなされ、緩やかな標準性の確保と個別大学の多様性への対応を実現している。ルーブリックを採用する大学は増えている（Hart Research Associates 2016）。

Ⅴ　日本への示唆

１．学習成果試験

試験は周到に設計されれば厳密な測定が期待できるが、前述した懸念事項もあり、最善の選択とはいえない。学習成果試験が同一の出題に全員が解答する形式であるとすれば、高等教育システムが国公立と私立大学から構成される日本や米国などでは、その多様性に対応しきれない可能性もある。そのうえで、学習成果試験があえて採用される大きな理由は２つある（塚原 2015）。ひとつは到達度を重視する、職業に直結した教育などの場合である。厚生労働系の国家試験や工業系の資格試験がその例であるが、知識能力が不充分な者を就業させないことで社会の危険を未然に防ぐ措置といえる。もうひとつは教育制度が不備な場合である。明治初期の日本における小学校の試験進級方式がこれにあたるが、ブラジルの状況認識もそれに近いのではないか。もっとも、日本では教育制度の整備と教育目的の変更にともない、明治後期には自動進級方式に変更された（斉藤 2003）。日本の高等教育には課題があるとしても、制度が不備とまではいえないであろう。試験を学習成果の評価として正常に機能させるためにブラジルで多大な努力がはらわれたことを考慮すれば、高等教育機関の内部質保証に期待する方向が望ましいようにみえる。

２．チューニング

チューニングでは、学習成果の上位にコンピテンスの概念を導入して、欧州各国の学位プログラムを比較可能とした。一般にこうした緩やかな標準性は、一国の内部では大学設置の段階で確保されていると考えられるが、大学設置基準の大綱化（1991年）によって日本では学部・学位の名称が500種以上にのぼるほど多様化している。これらに緩やかな標準性を導入したり、職業的レリバンスを教育課程に反映させるさいにはチューニングの枠組みが有効である。チューニングでは学習成果の評価を各大学の内

部質保証にゆだね、欧州諸国の大学評価も同様な方式をとる。これは国内の大学間格差が小さいことを前提とした措置であり、日本における質保証が、この方式によって充分に推進できるかどうかには議論があろう。

コンピテンスと学習成果の関係は、個別の知識や能力は習得できているが有機的に結合されていない、という文脈でたいていは議論されよう。ところが、日本には異なる状況もあり、大学院教育が研究室の内部で完結する方式でなされているとして、コースワークの充実など大学院教育の実質化が提言された（中央教育審議会 2005)。研究室における教育指導は卒業研究や修士・博士論文の完成をめざすもので、その範囲内では知識・能力が有機的に結合されているとしても、その幅が狭く基礎的素養が弱いというのである。これは、学習成果そのもの（とそれに直結したコンピテンス）の重要性を示唆するものである。

３．ルーブリック

本稿の３事例のうち、日本がもっとも参照してよいと思われるものは米国のルーブリックを用いた評価である。その理由は、コンピテンスの段階で記述できるとともに、評価の判断基準表としての標準性と、大学の文脈にそくした運用による多様性をあわせもつことにある。この点はチューニングに重なる部分もあるが、標準性がより明確であることに注目した。ルーブリックには到達度の判断基準が示されるが、その文言だけで疑問の余地のない評定が常にくだせるとは言いがたく、判断基準の運用を詳細に具体化して、評価の結果が適切で一貫したものとなる必要がある。標準性の確保には評価者による偏差を回避することも重要で、評価の実施における事前の準備と工夫がもとめられる。多様性を前提とすれば、個別大学をこえた評価の信頼性の確保も期待される。なお、ルーブリックは評価の判断基準表にすぎず、学生への指導・支援を充実させて学習をうながすことが質保証の本質であることはいうまでもない。

４．進学率の影響

学習成果のあり方は、大学進学率の高さによっても変化しよう。荒井（2008、24-27）によれば、入学者選抜の役割は進学率によって異なる。進学率が低いときには成績優秀者の発見が関心の対象となるが、上昇につれて多様な志願者をさまざまな大学の教育課程に振り分けることが要請され、さらには大学教育に必要な学力の到達度の確認が目標となる。同様な変化は学習成果にも生じよう。これに関連して矢野（2015、33）は、学生が学力の順番に進学していれば、進学率が50％をこえても入学者の質に大きなちがいはないという。マーチン・トロウがマス段階とユニバーサル段階を区分した50％という就学率は、学生層の変化よりも、政治的な意思決定を左右する区分点として設定された可能性がある。学力の分布が左右対象の釣鐘型であれば、マス段階とユ

ニバーサル段階の前半部分には共通性があり、平均値の周囲をかこむ大集団を一括することに利点があろう。このことは、進学率が50％の状態で実現できる質保証は、進学率が60～70％に上昇しても維持できる可能性が高いことを示唆する。

５．専門教育の分野分類

ブラジルの全国学生学力試験とチューニングには専門分野の分類が組み込まれている。特定の専門分類は、ある大学に有利に、また別の大学には不利にはたらく可能性があり、事実、ブラジルではいくつかの大学が専門分類の変更をもとめた。ブラジルや欧州はともかく、これまでの日本では、いったん設定された専門分類は固定化する傾向があり、その硬直化が教育研究活動の展開を抑制してきた歴史がある。ユニバーサル化した学士課程教育において、専門分類をどの程度の細かさにするべきか、分類の柔軟性（定期的な見直しなど）をどれほどはかるのかは検討を要しよう。

Ⅵ　学習成果の可視化

学習（学修）成果の可視化は、教育再生実行会議（2013、6）に盛り込まれた。第三次提言のうち、大学の教育機能の強化をもとめる項目のなかで、全学的教学マネジメントの改善と厳格な成績評価に関する事項のひとつとされる。前者は評価の枠組みに、後者は学生の評価にかかわる。可視化とは一目でわかる説明を意味するが、その大切さを否定する者はなかろう。学習成果を可視化する方向に以下のものがあろう。

１．評定の可視化

評定とは学生個人の成績をさす。学生の多様性・個別性をふまえ、成績の内容への理解を深めて、さらなる学習につなげることが重要である。可視化された評定に対する社会の適切な理解も求められる。学生の就職活動のさいに大学の成績表は参考にしないとする企業もあるが、そうであれば質保証は高等教育機関の外部への影響が乏しい活動となる。大学などの成績や席次に生涯にわたる影響力があった方がよいとは思わないが、適度な関心はもたれるべきであろう。

２．評価の可視化

評価においては、評価の規準（評価の尺度）や基準（尺度上の目盛り）の妥当性がしばしば論議になる。それらの選択が個別大学の判断によるというだけではなく、大学の集団ないし総体の判断としてなされることが望ましい。その方向には、たとえば以下の３つが考えられよう。

・統一性：学習成果の評価をひとつの枠組みで処理し、それを高等教育界の総意と位置づけ、多様性を何らかの大学集団の判断と位置づけることをさす。米国のVALUE

ルーブリックや欧州のチューニングにはそれを実現する仕組みがある。
・国際性：国際的な比較可能性をさす。学習成果の評価は専門分野別に行われる国／地域が多い。チューニングやブラジルの学習成果試験がそうであり、米国ではアクレディテーション団体の一部が専門分野別に組織されている。日本の学位の国際的な通用性を維持するには、専門分野別の評価の枠組みを整備する必要がある。
・適時性：学習成果の評価が、適切な頻度ですみやかになされることをさす。学習成果の評価を介した教育課程や教育科目の改善を推進し、社会変化や進学率の変動に対応した高等教育機関のあり方と活動の迅速な修正に資するものである。

３．効果の可視化

学習成果の指標のひとつに、高等教育の経済的な効果をあらわす収益率がある。その値は、学校種、性別、成績、計算の方式などにより異なるが、日本では５～10％ほどである。この点は主要国と同様であり、学生（家族）、政府、社会のいずれにとっても高等教育は投資にあたいすることが示されている。ところが主要国とは異なり、日本ではこうして可視化された高等教育の効果が社会通念と政策論議の双方に受け入れられていない（矢野 2015、160-172、242-244）。このような事態の克服と、可視化された方向にそった政策などの展開が期待される。

【注】

1)一般的技能のコンピテンスの例として、分析・統合する能力、学習する能力、問題解決、知識を実践に応用する能力がある（ゴンザレス・ワーヘナール 2012、59）。
2)森（2015、242-244）によれば、2000～10年度に米国の給付奨学金予算は3.5倍に、貸与奨学金予算は2.4倍になった［最近の統計では、2000～14年度に前者は3.3倍、後者は2.2倍］。奨学金は営利大学など非伝統的・職業訓練志向の高等教育機関の学生に多く執行され、そうした教育機関の非効率性が懸念されている。

【参考文献】

荒井克弘「全米学力調査の概要とその背景―大学入学者選抜との関連において」、荒井克弘、倉元直樹編『全国学力調査―日米比較研究』金子書房、24-34、2008。
川嶋太津夫「今、大学に求められる高等教育の質保証」、濱名篤、川嶋太津夫、山田礼子、小笠原正明編『大学改革を成功に導くキーワード30―「大学冬の時代」を生き抜くために』学事出版、8-14、2013。
教育再生実行会議『これからの大学教育等の在り方について（第三次提言）』2013。
ゴンザレス，J.，ワーヘナール，R.『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、2012。
斉藤利彦『試験と競争の学校史』平凡社、1995。
斉藤泰雄「留年・中途退学問題への取り組み―日本の歴史的経験」、『国際教育協力論集』広島大学教育開発国際協力研究センター、6(1)、43-53、2003。
田中博之「新しい評価Q&A」、『CS研レポート』啓林館教科教育研究所、51、36-41、2004。
中央教育審議会『新時代の大学院教育―国際的に魅力ある大学院教育の構築に向けて―答申』2005。
中央教育審議会『学士課程教育の構築に向けて（答申)』2008。
塚原修一「日本における大学教育の質保証―学習成果アセスメントの歴史的展開」、深堀（2015b)、146-158、2015。
塚原修一「アメリカ高等教育におけるルーブリックをもちいた評価」、『教育総合研究叢書』関西国際大学教育総合研究所、9、57-67、2016。
塚原修一、山口アンナ真美「ブラジルの高等教育における質保証と学修成果試験」、『日本高等教育学会第18回大会発表要旨集録』112-113、2015。
ニュッシュ，D.『高等教育における学習成果アセスメント―特筆すべき事例の比較研究』深堀聰子訳、国立教育政策研究所、2008。
濱名篤「大学評価の研究と実践の10年」、『高等教育研究』10、129-150、2007。
濱名篤「質保証の政策評価」、『高等教育研究』18、49-67、2015。
深堀聰子「学問分野のチューニング―参照基準に基づく内部質保証」、『国立教育政策研究所紀要』144、9-14、2015a。
深堀聰子編『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、2015b。
松下佳代「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、2012。
松下佳代、小野和宏、高橋雄介「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)、107-125、2013。
三好登「大学生の学習成果に関する研究動向と今後の課題」、『大学論集』44、303-318、2013。
村澤昌崇「高等教育における評価の動向・課題」、『大学論集』46、171-189、2014。
森利枝「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（2015b)、235-250、2015。
矢野眞和『大学の条件―大衆化と市場化の経済分析』東京大学出版会、2015。
山口アンナ真美、塚原修一「ブラジルの高等教育における評価と質保証」、『大学教育学会誌』37(2)、170-178、2015。

山口アンナ真美、塚原修一「ブラジルの高等教育における評価と質保証―学生学力試験ENADEをめぐる三大学の対応事例」、『大学教育学会第38回大会発表要旨集録』92-93、2016。

山内乾史、原清治『学生の学力と高等教育の質保証』学文社、2012。

山内乾史、原清治『学生の学力と高等教育の質保証 Ⅱ』学文社、2013。

米澤彰純「高等教育の評価に関する研究の回顧と展望」、『大学論集』36、315-329、2006。

吉田武大「ルーブリック―定義、作成方法、バリュールーブリック」、濱名篤、川嶋太津夫、山田礼子、小笠原正明編『大学改革を成功に導くキーワード30―「大学冬の時代」を生き抜くために』学事出版、187-192、2013。

Arum, R. and Roksa, J. *Academically Adrift: Limited Learning on College Campuses*, University of Chicago Press, 2011.

Banta, T. (ed.) *Assessing Student Achievement in General Education*, Jossey-Bass, 2007.

Braun, H. and Wainer, H. Value-Added Modeling. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics 26*, Elsevier, 2007.

Hart Research Associates *Trends in Learning Outcomes Assessment: Key Findings from a Survey among Administrators at AAC&U Member Institutions*, 2016. (http://www.aacu.org/sites/default/files/files/LEAP/2015_Survey_Report3.pdf)

Koretz, D. *Measuring Up: What Educational Testing Really Tells Us*, Harvard University Press, 2008.

Tremblay, K., Lalancette, D. and Roseveare, D. *Assessment of Higher Education Learning Outcomes Feasibility Study Report: Volume 1 - Design and Implementation*, OECD, 2012.

Reese, W. J. *Testing Wars in the Public Schools: A Forgotten History*, Harvard University Press, 2013.

Abstract

The aim of this paper is to make an international comparison of three assessment models for quality assurance in higher education (HE). They are: the Tuning project in Europe, the National Examination for the Assessment of Student Performance (ENADE) in Brazil and the VALUE Rubrics in the United States. Our main findings are as follows. In Europe, each country has its own quality assurance system and consequently, the curriculum and the level of academic achievement vary across countries. In 2000 the Tuning project was introduced with the purpose to make the degrees compatible. To this end, it introduced generic and subject specific competences to be used as reference points for common curricula. Contrasting with Europe, in 2004 Brazil introduced ENADE, a national exam for direct assessment of learning outcomes and is currently positioned as one of the major criteria for assessing quality. It appears that ENADE has contributed to the improvement of overall quality of HE institutions. Different from these two models, in 2007 the United States introduced the VALUE Rubrics as a means of assessing student learning outcomes in a more holistic way. This model combines the competences as found in the Tuning project and the standardized assessment with the use of the scoring sheets. We found that VALUE Rubrics appear to be the most suitable model for assessing Japanese HE institutions. When HE institutions need to enhance accountability and assure quality, to develop a universal assessment framework that can facilitate the “visualization” of learning outcomes becomes increasingly important.