

教養教育の学習成果の測定は可能か

—2000年代のアメリカの取り組み—

吉田 文¹⁾*

1) 早稲田大学教育・総合科学学術院

1. 課題の設定

本稿は、2000年代初頭のアメリカにおける、学生の学習成果の測定をめぐる議論を、アカウンタビリティを求める産業界や政府と、学生の成長と大学の教育改善を重視する大学人との対立の構図として分析することを目的とする。

大学教育における学生の学習成果の測定は、日本でも2000年代から各種の審議会においてその重要性が訴えられるようになり、「社会人基礎力」、「学士力」と命名された、学生が卒業時に身につけるべき各種の「力」のセットが提示されてはいる。しかし、それをめぐる賛否両論の議論が、大学のステークホルダー間で継続的になされたことはほとんどなく、したがって、学習成果の具体的領域の検討や、それらを測定するための指標の開発に関しても、個別大学の範囲を越えては行われてこなかったといってい¹⁾。

他方で、アメリカでは、1980～90年代から学習成果の測定の必要性が論じられるようになり、それをめぐる議論は2000年代に入ってさらに盛んになる。加えて、個別の大学にとどまらず、広範に汎用可能性をもつ評価指標の開発に対して財団が支援をし、その妥当性が検討されて、一定の範囲での普及をみている。このアメリカにおける学習成果の評価をめぐる動向に関しては、日本においては、評価方法の紹介としては知られているが(江原 2010, 森 2010, 2012)、その背後にあるアクター間の対立や、その対立の論理にまで言及して論じられてはいない。

そこで本稿では、2000年代初期のこうした動向を、

その背景、具体的な実践とその対立点、それらの問題点や示唆に関して検討し、日本でも議論されている学習成果の評価に関して、今後考えるべき示唆を得ることを目的とする。

2. アカウンタビリティと学習成果の直接評価

徐々に縮減する州政府の財源をどのように配分するか。その配分の原理となったのが、新自由主義によるアカウンタビリティという考え方である。すなわち、投入した税金に対して、どのような成果が得られたのかを明確にするのがアカウンタビリティであり、実績にもとづく資金の配分がもっとも合理的と考えられるようになった。1980年頃からこうした考え方が論じられはじめ、90年以降には、この考え方は州政府の税金で維持されている州立大学にも及び、大学における「実績」とは何か大学外から議論されるようになる。具体的には、1998年に非営利団体として設立された全米公共政策・高等教育センター(The National Center for Public and Higher Education)が、2000年からはじめた州内の高等教育機関における、学生の大学教育に対する準備状況、進学状況、授業料や奨学金、卒業率、収益率などに関する量的な指標を作成し、州ごとの比較を行う Measuring Up というプロジェクトは、その代表的なものである。江原(2010)によれば、実績評価の対象は学士課程の教育活動であり、ほとんどの州で実績による資金配分を行っており、それが高等教育予算に占める割合は1～6%程度という。わずかな額のように見えるが、州政府が配分する予算を通じ

*) 連絡先：〒169-8050 東京都新宿区西早稲田1-6-1 早稲田大学教育・総合科学学術院 ayayoshida@waseda.jp

て大学教育に介入することになったことは大きな変化である。自律的に改革をすることを保証されている大学の自治に対する介入とする大学側の反対は強かったが、大学外からは政府だけでなく経済界も含めてこれに賛同する声は大きい。

問題は何をもち「実績」とするかである。ここで忘れてはならないのは、アメリカの地域ア krediteーション団体である。大学を大学として適格認定するのがア krediteーション団体であるが、そこでは2000年頃より学習成果を重視したア krediteーションの導入についての議論が出始めていた。そこに大きな影響力を持ったのは、2005年、当時の連邦教育省長官スプリングスの諮問に対する答申である *A Test for Leadership* (2006) である。連邦政府立のア krediteーション団体の設置、それまでのインプット重視のア krediteーションに対して、学生の学習成果を指標とすることの重要性を指摘し、卒業時における共通テストの導入を全学生に義務付けようとするなどの議論は、学習成果の評価という課題を大きく浮上させることになった。共和党内のこのスプリングス・レポートの主張は、民主党政権に交代しても消滅したわけではなかった(森 2012)。

なぜ、こうした議論が生じるかといえ、大学進学率の上昇にともない、それまでの幼稚園から高校卒業までを意味するK-12に代わって、幼稚園から大学卒業までを意味するK-16という考え方が登場したことである。教育システムのなかでドロップアウトすることなく高校卒業まで到達するよう、教育段階間の接続をスムーズにすることを課題としてきたアメリカの教育であるが、高等教育への進学者が70%になるなか、高校卒業まででは不足であり、大学卒業までを視野に含むK-16という考え方が生まれる。これによって、高校の卒業要件と大学の入学要件の擦り合わせを行い、大学教育を受けるに足る高校卒業者を大学は入学者として受け入れているかを考慮すべしとする議論が始まる。そしてさらには、高等教育段階に進学しても規定年限内での学士課程卒業率が低く、言い換えればドロップアウト率が高いことが長年の懸案事項としてあり、その改善があらためて課題となったことによる(森 2010)。

こうした各方面からの多様な議論が収斂するところは、大学における学生の学習成果を付加価値として定量的に示し、それを大学教育の実績としても用い、さらには大学教育改革の資料にもしようということであった。

3. 学習成果の評価に関する4つの方法

学生の学習成果に関する議論そのものは目新しいものではない。その動向をまとめたChunによれば、大学の質を測定し、それでもって大学を評価する方法は、大きくわけて4つに分類され、それぞれがよって立つ論理と問題点があるという(Chun 2002)。それによれば、第1は、大学に関する客観的・数量的なデータによる評価である。いくつか例を示せば、入学者の学力による大学の選抜度、中退率や卒業率、GPA、卒業生の就職状況や地位達成など学生に関する指標、教員学生比率、提供されている科目数など教育に関する指標、寄付金額、外部から獲得した研究費などの金銭的指標など、多様な領域における多様な指標を挙げることができる。これらはいずれも量的に測定が可能で、その意味において客観的であることに合理性が認められており、また、大学の質は、どの指標に関しても数値の上下に比例している、より良い質とは資源量に比例するという考え方が存在していることを特徴とする。これらのデータは、客観性という点において妥当性は高いものの、学生の学習成果を第一義的に示しているかという点において疑問が残る。

第2は、いわゆる機関のランキングである。大学教員、職員、あるいは外部の有識者などによって、大学教育のいくつかの観点に対して評価がなされ、それらを総合した機関としての評価は、他機関と相互比較可能なランキングとしての指標となる。こうした測定や評価がなされる背後には、大学関係者や有識者こそが、大学の質をもっともよく理解しているという論理が存在する。ランキングは機関や学部といった組織単位の比較が容易であるため、学生の志願行動、あるいは、ランキングを上昇させるための組織改革などには影響を与えるものの、ランキングが学生の実際の学習に影響を与えるか否かに関しては否定的な意見が多い。

第3は、学生調査である。第1と第2の方法が教員、

職員、有識者などを対象にして行われていたのに対し、この第3の方法は、学生を直接の対象として、学生生活の経験、満足度、能力の向上の程度の自己評価、教育や職業に関する将来計画などを聞くことに特色があり、学生の評価をもって大学の諸活動の改革に用いようとすることに特色がある。この学生調査は、学生の意識・意見・行動、さらには自己評価ではあるが能力の向上の程度に関しても直接明らかにできるという点に利便性が見いだされ、複数の大学で共通の調査票を用いて実施されている調査がいくつかあるほどに普及している。

しかしながら、他方で、この学生が自己申告するという点において、データの信頼性に疑義が生じるのである。というのは、たとえば4件法で同じ項目を選択したとしてもそれがどこまで同じとみなすことができるかといったことである。また、学生の学習成果としての自己評価は間接評価に過ぎず、その信頼性に関しても問題は残るとされるのである。

こうしたなかで生じてきたのが、第4の学生の学習成果を直接評価する方法である。もちろん学生の学習成果は、個々の科目ごとに多様な方法で実施され、それらを総合したGPAが成果を表すものとして長く用いられてきた。ただ、これらの方法は、学生の到達度は測定しても、向上の程度を測定しているわけではない。また、学生の履修する科目や専攻の違いを越えた共通性をもつものでもない。とはいえ、学生の学習成果を直接評価する方法の有効性は、学生の自己評価である間接評価よりもまさると言われるのである。しかしながら、具体的にどのような評価方法がよいのかとなると、それに関する合意に達することは容易ではないという点に問題がある。

学生の学習成果という観点からは、これまでも各所で採用されてきた第1から第3の方法は、測定結果の信頼性と妥当性という点で問題があり、新たに論じられるようになった第4の方法は、測定方法が定まらないという点で問題がある。しかしながら第4の学生の学習成果を直接に評価する方法、それも学習による能力の向上の程度をもって、学習成果とする方法に関しては、1990年代になるとアカウンタピリティという観点から、その重要性が指摘されるようになっていく。

州政府、連邦政府、アクレディテーション団体、大学関係者の間での新たな議論は、まさしく、学生の学習成果を直接評価するという課題であった。そのことの重要性は認識されたとしても、こうした方法に大学関係者からは反対が生じた。とりわけ、大学教育における付加価値が、安易な方法で計量され、それが故に機関間、州間での相互比較が容易になされてしまうことへの反対の声は大きかった（たとえば、Association of American College and Universities (2002), (2007b) など）。

4. 学士課程から一般教育へ

ここまでの議論は、学士課程卒業時の学習成果に関する議論である。卒業時の学力基準を定める、入学時と卒業時の付加価値をもって学習成果とするなどは、教育機関としての教育の効果やその改革の議論とも容易に結びつくものである。こうした渦中において、一般教育の学習成果をどのように考えるかという議論が登場する。学士課程卒業時の学習成果の議論が、なぜ一般教育の学習成果の議論になるのか、その議論の筋道を追ってみよう。

学士課程の学習成果を問う場合、GPAのような知識の獲得程度を用いることは一般的ではない。それは、1つには、この学習成果は、専攻にかかわらず学士課程としての共通性をもつものを想定して議論されてきたこと、もう1つには、学問的な知識の獲得だけではなく、学生は大学生活を通じて何ができるようになったのかを、市民としての社会的責任をもって価値判断ができるようになることや、学習の仕方を学習することで情報の処理能力を高め批判的思考力を涵養することこそが、学士課程の学習成果だと論じられることによる (Hersh 2002) ²。

学習成果をこのように定めると、その成果を上げるための教授法は、従来の教員中心ではなく学生中心であるべきと論じられるようになる。学習者としての学生の存在を強調し、教員はそれを補助するメンターになることで、学生は自立的学習を進められるようになり、上述の学習成果が生まれるのだと論じられる。そして、こうした学習成果と教授法は、学士課程全体を通じて達成されるものであるが、とりわけ一般教育に

においては尊重されるべきだとされるのである。なぜなら、一般教育のミッションは幅広い学習による知的なスキルの形成にあるからだというのが、その理由である (Schneider & Shoenberg 1998)。

もともと一般教育は、専攻の前段階において多様な領域の学習を目的としているが、多様な科目を提供すると、科目間の一貫性が希薄になり、学生の学習にも一貫性が欠如するという、多様性と一貫性の両立という矛盾を抱えており、大学内外から常に改革が要請されてきた (ロスブラット 1999)。1990年代からの一般教育改革が目指したのは、多様な科目の提供による多様性の確保ではなく、それらの学習を通じた学生の知的スキルの形成であった。多様な内容を学習しても、それらが、たとえば批判的思考力の涵養を目的とすれば、そこで一貫性が確保できるという考え方である (Gaff1991)。この考え方は、学生の学習成果の重視という考え方と違和感なくシンクロしていくのである。

そして、2000年頃には、それを可能にする教育方法の工夫として、協調学習、経験学習、サービス・ラーニング、探求型学習、統合学習といった手法が強調されるようになる。これらの方法により、一般教育の矛盾や問題を解決できると期待されたのである。どの科目を教えるかではなく、教授学習へのアプローチを考慮することは、教養教育 (liberal education) に根差す哲学であるとも言われたのである (Rothblatt 2003)。

アメリカにおける教養教育 (liberal education) は多義的であり、その理解は容易ではない。というのは、学士課程教育の理念を指して教養教育 (liberal education) と言われる場合、非職業的な学問分野 (liberal arts) を指して用いられる場合、また、一般教育 (general education) の課程を意味する場合もあるからである。教養教育 (liberal education) は、これらのいずれとも互換的に用いられる。それは、もともとアメリカの大学は教養教育 (liberal education) のカレッジとして始まり、大学が専門教育に特化していく過程において、4年間の教養教育 (liberal education) の課程は、2年間の一般教育の課程に縮減したが、学士課程の理念としては教養教育 (liberal education) が強調されるという歴史的経緯によるものである (ロスブラット 前掲)。したがって、学士

課程の学習成果を専攻にかかわらず共通の内容で考えようとする、一般教育改革として学習成果を求めることは、同根であるといってもよいのである。

5. 一般教育と transfer

一般教育の学習成果に対する要請には、これらとは異なるもう1つの事情があることを指摘したい。それは、アメリカの高等教育には、単位の蓄積を通じて大学間を移動する transfer (単位互換による機関の移動) という仕組みがあることである。transfer とは、移動前の機関において取得した単位を、移動先の機関における単位に読み換えることによって、機関を移動しても単位の累積加算をすることで卒業が規定年限内に可能になる仕組みである。この方式により、コミュニティ・カレッジ、総合大学、研究大学といった州内の州立大学を自由に移動できるシステムが構築されてきた。

そして、この transfer は、一般教育から専攻への移行期に多く生じるが、単位の振り替えは、科目名称の類似性とその単位数によって行われていた。それは、類似の名称をもつ科目は同じ内容が教えられ、単位数は学生が獲得した力のレベルであり、それはどの機関においても同じだという暗黙の仮定があることを意味する。

しかし、実際には、そのようなことはなく、学習成果という議論が登場した時、そのことが改めて問われるようになったのである。学士課程の学習成果を重視するならば、transfer に際しても学習成果を重視して当然ではないか、すなわち一般教育の学習成果を問わずして、学士課程の学習成果は問うことができないという議論である (Schneider & Shoenberg 前掲)。

この transfer と関わる一般教育の学習成果の議論は、2000年前後から始まる。そして、これらの議論は、州政府やア krediteーション団体の、大学に対するアカウントビリティの要請と重なって、現実のものとなる。州立大学全体にわたる一般教育の科目内容や単位の標準化、すなわち教育内容を詳述し、また学生が獲得する能力を明示することで、実質をともなった単位互換を可能にしようと、改革を求めるようになる。

これによって、学生の単位修得のロス減少し、提供する科目の精選によって増加する一方の高等教育費

の節約が可能になると州政府は論じる。効果と効率の両立が可能というわけだ。ただ、こうした州立大学全体に対する標準化の要請は、これまで大学教員個々人の管理下にあったカリキュラムへの介入であり教員の自主権の剥奪だとみる大学関係者は多い (Gaff 2001)。

実際、transferがどの程度生じていたかといえ、2003年においては、学士課程卒業者の59.2%が2機関以上に在籍していた。入学時から卒業時までを同一機関に在籍する者の方が少ないのである。大半は同一州内の機関に在籍する。また、2年制の機関から4年制の機関への移動がもっとも多いが31%に過ぎず、4年制の機関での移動19.1%、2年制の機関での移動20.7%、4年制の機関から2年制の機関への移動14.2%と実に多様な移動が生じていることがわかる (Adelman et al. 2003)。

これだけのtransferが生じていても、州内の大学に共通した一般教育の要件を、それも学生の学習成果に関してまで決定することは容易ではないようだ。2005年に実施された調査によれば、わずか10州が、州として州立高等教育機関に共通した一般教育の単位数、履修領域などのパッケージを作成し、それを利用してtransferを行うことを取り決めているが、それ以外の州では、規制力をもたないパッケージであったり、科目に付された番号がtransferに用いられるなど、機関の自由度が高いことが明らかにされている。また、一般教育によって涵養すべき能力に関していえば、ごくわずかの州で「批判的思考力」(4州)、「倫理性」(3州)、「生涯学習力」(1州)といった事項を掲げているに過ぎない (Shoenberg 2005)。

Transferというアメリカの高等教育に特徴的なシステム故に、一般教育の学習成果が問われるのであるが、州全体の足並みを揃えること、言い換えれば、1機関を超えてカリキュラムに共通性を求めることが、いかに容易ではないかをみることができる。

6. VAAIからCLAへ

これらの議論が並行するなか、次第に、どのような側面を学士課程の成果とみなすかという具体的な議論に進んでいく。興味深いのは、一般教育の学習成果と、教養教育 (liberal education) のそれとが互換的なもの

のとして議論される点である。それはすでに述べたように、教養教育 (liberal education) と一般教育との歴史的な経緯にもとづく類似性・互換性によるものであり、また、学士課程における専攻にかかわらず学士としての共通性を評価しようとする場合、それは一般教育で実施されている基礎スキルの獲得や幅広い分野の学習を評価対象にすべきと考えられるからである。

ところで、教養教育 (liberal education) は、その目的として教養人 ((liberally) educated person, cultivated person) の育成を、当初より掲げてきた。この歴史的な理念は現代でも各所で用いられ、大学のミッションに掲げられる場合も多い。この教養人 ((liberally) educated person, cultivated person) とは、円満な人格をもち公衆へ奉仕する市民³を意味し、したがって円満な人格を形成することが、教養教育に課された役割とされている (ロスブラット 前掲)。そのため、教養教育の学習成果の評価が議論となっても、必然的にその理念を受け継ぎ、知識の獲得に加えて、社会的責任に関する事項が列挙されることも特徴として指摘することができる。

教養教育の学習成果としてどのような事項を評価するか、2000年頃からいくつかのプロジェクトが始まるが、それには大きく2つのグループがあり、両者の考え方は対立する。その議論の過程を追うことで、アメリカの高等教育における学習成果の評価が孕む問題と測定の困難をみることができる。

1つは、2000年に始まったランド・コーポレーションの教育助成審議会 (Council for Aid to Education) の「付加価値評価計画 (Value Added Assessment Initiative)」⁴である。学生が高等教育機関に在学することによってどのような能力の付加価値をつけたか、それを教養教育という観点から、学生の直接評価によって測定しようとするプロジェクトである。評価する領域は、大きく分けて「知識 (knowledge)」、「判断力 (judgement)」、「知るための方法 (ways of knowledge)」の3つのカテゴリーである。「知識」とは、人文・社会・自然にわたる幅広い領域の知識、「判断力」とは、獲得した知識を用いて、決定をし、優先順位をつけ、判断を下すことをいう。また、「知るための方法」とは、情報を処理し、批判的に思考し、学

習するための方法を学習することである (Hersh and Benjamin 2002)。また、これらを、それぞれの力が社会生活を送る上で有用なものという観点から評価すること、社会生活の文脈を斟酌した評価が重要であることが強調されている。いくつかの高等教育機関での実験を通じて、項目や測定の妥当性を検証し、最終的には学習成果を測定するための標準テストを作成した。学生の受験結果は、大学の教育改善の資料とするだけでなく、州の高等教育政策の資料として利用可能にすることを計画するものであった。

その後2004年には、大学卒業時にすべての学生が獲得しているべき能力（批判的思考力、分析的論理的思考力、問題解決力、文章表現力）とその範囲をコアとして定め、それを測定する Collegiate Learning Assessment という標準化されたテストとして結実する。このテストは多肢選択のようなものではなく、実社会で生じる問題をどのように解決するか、提示された各種の資料を参考に自由記述形式で回答するものである。これを入学時と卒業時に実施し、その差分をもって付加価値とするのである。

この推進者によれば、「これらの能力は、何らかの特定のコースや専攻において教えることができるものではなく、大学生活全般において構築される集成的、蓄積的な結果である。したがってCLAは、個々人の学生や個々のコースの評価ではなく、機関としての付加価値を測定したものということになる。」(Hersh 2007) という。そして、標準化されたテストであるため、その結果の機関間で比較が可能であり、それによって、たとえば、入学者の学力レベルが同程度であっても卒業時に差が出てくるのはなぜかを問い、そして教育改革を進めることができると論じる。加えて、外部の批判を想定して、「CLAは新たなランキングのツールとして用いられることや、結果をもとに予算配分に利用されることを意図してはいない。(中略)CLAの目的は、機関の教授・学習過程の改善にある。」(Hersh op. cit.) と力説している。

CLAはその後CLA+とステップアップして現在に至り、150以上の機関で3万人の学生に利用されている。これほど短期間に多くの利用者を得るに至ったのは、アクレディテーションにおける評価項目に学習成

果が加わったことによる。スプリングスの意図した連邦立のアクレディテーション団体は設立されなかったが、アクレディテーションにおいて学習成果の評価が評価項目として付加されたため、それ用の資料とするために多くの大学が利用するようになったのである。アクレディテーションのために新たに学生の学習成果を測定するのは多くの大学にとって容易ではなく、標準化されたテストはそのための簡便な手段として用いられたのである。さらには、OECDによる大学卒業時の学力評価のプロジェクトである AHELO においても、このCLAの使用が検討されていた。

7. AAC&Uの反論

こうした方式に大きく反対したのは Association of American Colleges and Universities (AAC&U) である。これは、教養教育の推進を目的とする大学団体であり、それに関する研究プロジェクトやセミナーの開催などを行ってきた。こうした批判的思考力などを測定する標準化テスト、それによる学生の学習成果の付加価値の測定、その結果の州の高等教育政策立案における利用という方式への反対理由はいくつかある。

たとえば、「狭隘な定義のもと狭隘なテストでもって、大学における学習成果を測定すると、現在確立しようとしているエビデンス文化 (Culture of Evidence) をゆがめることになってしまう。エビデンスとして測定するものは、教育として測定しなければ、エビデンス文化は自動的に教育改善をもたらすものとはならない。スプリングス委員会は広範囲にわたる能力を標準化テストで測定しようとしているが、過去の経験と我々が現在取り組んでいる研究にもとづけば、もっと学ぶべきこと、検証すべきことが残されている。これらの能力は、学問分野における知識の獲得、理解、論理的思考の学習において獲得するものであり、また、学問分野の学習によって、Educational Testing Service (ETS) が“ソフト”と命名している、個人的、市民的、同時的、社会的、認知的知識も獲得できるのである。」(Richard J. Shavelson 2007) といった、評価すべき能力とは、学問分野の学習を通じて獲得できるものであるとして、それを欠いた標準化テストで測定することに対して反対する意見がある。

表1. AAC&Uによる教養教育に不可欠な学習成果 (2004年版)

<p>人類の文化と自然界に関する知識 (Knowledge of Human Cultures and the Natural World)</p> <ul style="list-style-type: none"> ・自然科学 ・社会科学 ・数学 ・人文科学 ・芸術 <p>認知的および実践的技能 (Intellectual and Practical Skills)</p> <ul style="list-style-type: none"> ・文章および口頭コミュニケーション ・探究, 批判的思考力, 創造的思考力 ・数量的リテラシー ・情報リテラシー ・チームワーク ・学習の統合 <p>個人的および社会的責任 (Individual and Social Responsibility)</p> <ul style="list-style-type: none"> ・市民的責任および市民としての関与 ・倫理性 ・認知的知識の獲得と行為 ・生涯学習への志向性

出典：Association of American Colleges and Universities (2005)

また、「これまでの調査研究が明らかにしているように、学生は、学習に積極的に長い時間取り組み、教員や学生と知的に相互作用をし、達成しようとする意欲を持続し、多様な人や考えに出会い交流することで、成長するのである。(中略)豚の体重を測定すれば豚が太るわけではないように、学生のテスト受験にお金をつぎ込んでも、学生は学習するようにはならない。」(Banta 2007)と、闇雲にテストをしても本題である学生が学習に取り組みことにはならないと警鐘を鳴らしている。

こうした標準化されたテストに批判的な立場をもちつつも、AAC&Uとしては教養教育を推進する立場から、教養教育の成果としてどのような側面を重視すべきか、項目設定のプロジェクトを立ち上げる。その第1版が2004年に出された表1である。そこでは3領域15項目を、専攻にかかわらずすべての学生が身につけるべきものとして設定した。

AAC&Uは教養教育の学習成果を考えるとそのものには反対しているわけではない。AAC&Uとしては、すでに2002年、2004年に出版したプロジェクトの報告書でも同様に、教養教育の成果を評価するための項目に言及しているが、それは、「これまで、学生の学習が問われることはあまりなく、それゆえに学生の達成度に関して知られていないことが多くある。よい評価とは、あらゆる学生の学習を向上すべき力を備え

ていなければならない、評価の実践は、キャンパスにおける教授・学生の取り組み・機関の支援・学習の側面において、それらが並行して改善する方向で拡大されるべきものである。」(AAC&U 2005)というスタンスに立ってのことである。すなわち、あくまでも高等教育機関内における、学生の学習を向上させるための、各種の取り組みの改善を目的としており、州の高等教育政策立案に資するといったことは評価の目的には含まれていない。そして、AAC&Uは、これらの項目が、産業界や政府からも同意を得るものであることを、項目1つ1つに対して、それらから出された各種の文書を用いて例証している(AAC&U 2005, op. cit.)。

そして、これをベースにして、2007年には改訂版が作成されている(表2)。

これは2004年版をもとに議論を重ねた結果であり、本質的には大きな変更はない。しかしながら、従前では、第2領域の1項目であった「学習の統合」が、改訂版では新たに第4の領域の「統合的、応用的学習」として設定されたこと、従前の第1の領域では習得すべき学問領域が個別に立てられていたのに対し、改訂版ではすべてが一緒になり1項目となっていること、従前の第2の領域における「探究、批判的思考力、創造的思考力」から、改訂版では「探究と分析」として独立し、従前の「チームワーク」に「チームワークと問題解決」として問題解決が付加されたことに変化を

表2. AAC&Uにおける教養教育に不可欠な学習成果（2007年版）

<p>人類の文化と物理的・自然的世界に関する知識 (Knowledge of Human Cultures and the Physical and Natural World)</p> <ul style="list-style-type: none"> ・自然科学, 数学, 社会科学, 人文科学, 歴史, 芸術の学習 <p>認知的および実践的技能 (Intellectual and Practical Skills)</p> <ul style="list-style-type: none"> ・探究と分析 ・批判的思考力, 創造的思考力 ・文章および口頭コミュニケーション ・数量的リテラシー ・情報リテラシー ・チームワークと問題解決 <p>個人的および社会的責任 (Individual and Social Responsibility)</p> <ul style="list-style-type: none"> ・市民的知識および市民としての関与—ローカルとグローバル ・認知的知識能力 ・倫理的思考力と倫理行動 ・生涯学習の基盤と技能 <p>統合的, 応用的学習 (Integrative and Applied Learning)</p> <ul style="list-style-type: none"> ・一般的な学習と専門的な学習を越えた統合と高度な学習達成

出典：Association of American Colleges and Universities (2007a)

みることができる。個別の学問分野の学習よりも、それらを通じて培われる能力、それらを統合することによる高度な学習、社会生活で求められる分析力や問題解決能力など、より応用的・実践的な能力に比重がかけられるようになったといえることができるだろう。

8. 既存の調査結果から明らかになること

これら教養教育の成果として重要な項目を作成したAAC&Uは、2004年版の項目に対応させて、学生はこれらをどの程度身に付けているのか、既存の各種調査を用い、表3に示すように教員調査（「学生がそれらの項目を向上できるように授業を構成しているか」）、学生調査（「大学での在学経験が、それらの力の向上に役立ったか」）、全国レベルの標準化されたテストの3側面から比較している⁵。教員には、学生が成果を上げるように授業構成したか、学生には大学は学生が力を伸ばせるような取り組みをしてきたかをたずねているが、それはいいかえれば、大学の実施している教育が、付加価値をつけようとしているか、実際に付加価値をつけてくれたかを評価をしているところが興味深い。

これをみると、教員は自身の授業において批判的思考力を高めたり、各種の学習を統合することに関しては当たり前のように実施しており、コミュニケーションスキルや情報リテラシーなどに関しても力を入れており、機関レベルでそれを実施する体制ができている

ことも明らかである。しかし、それらと比較すると社会的責任を教えることに関しては、重視する程度は低い。学生の授業評価をみると、力を入れている教員が多い項目では、学生も在籍機関での経験がそれらの力の向上に役立ったとしているが、力を入れている教員が少ない社会的責任に関する項目では、在籍経験がそれらの力の向上に役立ったとする比率が低くなっている。機関の目的や教員の教育意図が、学生の学習に影響を与え、それが学生の力の向上に関連しているのはいくつかと推測されている。

他方で、全国レベルのテストでは、測定されているのが如何に限られた項目であるかを明瞭に見て取れる。また、測定された数少ない項目においても、習熟度が低いレベルにとどまっており、学生が在籍経験によってそれらの力を獲得できたとする評価と大きな開きがあることも指摘したい。

AAC&Uでは、こうした分析を通じて、教育機関が何を目的としどのような教育をするかが、学生の学習に影響を与えることを示し、その多様性を無視して標準化テストで成果を測定することが教育改革につながらないとして、成果測定の議論を始める。

9. 学習成果の評価方法

標準化テストに対抗する方法として生み出されたのは、ルーブリックという測定方法であり、それはアカウンタビリティの要請によって作成されたETSの

表3. 教養教育の項目別成果

項目	教員調査 (FSSE)	学生調査 (NSSE)	全国テスト (ETS/ACTCAAP)
知識	—	・一般教育における幅広い知識86%	・ETS: 数学レベル3到達は4年生で8%, 1年生から5%上昇 ・ACT: 1年から4年で数学は低下, 自然科学はごくわずかに上昇 ・社会科学, 人文科学, 自然科学, 芸術はデータなし
文章および口頭コミュニケーション	・明晰で効果的な文章作成を重視した教育61% ・明晰で効果的な口頭表現を重視した教育49%	・文章作成77% ・口頭表現72%	・ETS: 文章作成レベル3到達は4年生で11% ・口頭表現はデータなし
探究, 批判的思考力, 創造的思考力	・批判的・分析的思考力を重視した教育93%	・批判的・分析的思考87% ・創造的思考力に関してはデータなし	・ETS: 批判的思考力熟達は4年生で6%, 77%が不十分 ・ACT: 1年から4年にかけて1標準偏差上昇 ・創造的思考力はデータなし
数量的リテラシー	・量的問題の分析を重視した教育44%	・量的問題の分析65%	—
情報リテラシー	・所属機関は学生にPCの利用を推奨91% ・情報の価値, 他者が収集・解釈した情報の検証方法を重視した教育73%	・PCや情報に関する技術77%	—
チームワーク	・他者とうまく仕事をするを重視した教育55%	・他者とうまく仕事ができる78%	—
学習の統合	・アイデア・情報・経験を統合して, 新たなより複雑な解釈や関係を見出すことを重視した教育85%	・コース終了後に多様なアイデアや概念を統合66% ・アイデアの統合が必要なプロジェクトに参加87%	—
市民的責任および市民としての関与	・コミュニティサービスやボランティアをすることを重視した教育54%	・コミュニティの福祉に寄与45% ・国政・自治体選挙の投票23%	—
倫理性	・学生が価値や倫理の獲得を重視した教育54%	・価値や倫理に関する自身の行動規範59%	—
認知的知識の獲得と行為	・学生の留学は重要45% ・自分と異なる人種や民族的背景をもつ人に対する理解度を高めることを重視した教育41% ・所属機関は多様な経済的・社会的・人種的・民族的背景をもつ学生間の交流を促進44%	・自分と異なる人種や民族的背景をもつ人に対する理解度51%	—
生涯学習への志向性	—	・自分自身で効果的な学習ができる77%	—

出典: AAC&U 2005, op. cit. より作成

注: 各項目の%は, 4件法による回答のうち, 肯定的な回答2つの合計。「—」は該当データなし。

Academic ProficiencyやCLAなどよりも, 学生の学習成果を時系列に測定する方法として妥当性が高く, 教育改善に使用できるとする。2007年版で確定された教養教育に不可欠な要素それぞれに関して16のルーブリックが構築された。ルーブリックとは, 学生が何を学習するのかを示した段階別の基準にもとづき, 学生がどの段階に到達しているのかを評価するものである。何を学習し何を獲得すべきかを学生に示すことができる, 時系列的に, たとえば入学時から卒業時まで

のいくつかの時点での測定をすることで, 学生個々人の成長の度合いを評価できるものとして, そのメリットが謳われた。それぞれの段階の違いが, 数値ではなく言語で記述されているところに特徴がある。

AAC&Uの開発したルーブリックは, 横軸に時系列を置き, すべての項目に共通するものとして初年度を基準 (Benchmark), 2年次, 3年次を途中経過 (Milestone), 卒業時 (Capstone) を最終評価として4回測定する。縦軸は, それぞれの項目で若干差異が

あるものの、その領域の力を獲得するのに必要な事柄がいくつかの次元に分けられて列挙されている。たとえば、「探求と分析」に関して言えば、縦軸には、「課題の選定」、「既存の知識と研究」、「方法論」、「分析」、「結論」、「課題と示唆」がそれに相当し、まさに研究論文の構成が評価軸となっている。また、批判的思考力では、「能力獲得」、「危険負担」、「問題解決」、「反論の包含」、「革新的な思考」、「結合・統合・変換」とやや抽象的な表現が並ぶが、ある領域の戦略や技法を獲得し、そこで生じる課題解決に潜む危険を認識し、そのうえで問題解決を行い、しかしながらそこに別な見方や考え方などへの視野の広がりをもつことで、斬新な考え方に到達し、さらに新たな課題解決策が生み出されるというプロセスを指す。

この評価軸が言語で記されるため、機関はそれぞれの特徴に合わせ評価基準を定めることができ、いくつかのルーブリックを機関の特性に合わせて統合して使用するなど、多様な利用方法があることを特色としている。さらには、それは標準化されたテストで数値として比較されるものとは異なり、学生の成長の度合いを勘案しつつ教育改善が可能になるのだという。そしてAAC&Uでは、12機関をルーブリックのリーダーシップ機関として選定し、ポートフォリオ（学生の学習成果に関するものを作品集として集積するもの）を導入し、それをもとに評価を行い、ルーブリックの有効性を検証している。さらには、パートナーシップ機関として約150機関を選定して、ルーブリックの使用とその検証を行っている。

ここで注目したいのは、15のルーブリックのうち、どの項目の使用頻度が高いかである。ルーブリック使用のリーダーシップ機関として選定された12機関が、それぞれどのルーブリックを使用しているかをみたところ、もっともよく利用されている順に示せば、「文章コミュニケーション」（9機関）、「口頭コミュニケーション」（6機関）、「批判的思考力」（6機関）、「数量的リテラシー」（5機関）、「情報リテラシー」（5機関）、「文章読解」（3機関）、「創造的思考力」（3機関）である。どの機関も選択していない項目として「チームワーク」、「異文化に関する知識と理解」、「グローバルな学習」がある⁶。

15項目のルーブリックにおいて、これだけの採用頻度の偏りが生じているのは、達成段階をいかに段階別に示すことができるか否かによるものとみることができ。スキルやリテラシーは、比較的向上段階の明示が容易であるため、ルーブリックの使用もそこに集中している。さらには、アカデミックな学習として機関内におけるレベルを構築することが容易であることも関係していよう。それに対し「チームワーク」などは、その重要性は認識されていたとしても、それをどのように段階別に評価するかをルーブリックとして示し、かつ、それを評価することが容易ではないことが、この使用頻度から明らかである。ルーブリックは、学生の成長過程を把握できる、理想的には理想的な評価方法であるのかもしれない。しかしながら、それをあらゆる領域に広げて実践に落とし込んでいくのは、それを積極的に使用している機関でさえ容易ではないようである。また、もう1つ指摘すれば、選定された12機関は小規模なりベラルアーツ・カレッジやコミュニティ・カレッジが多い。小規模で学生の顔が見えるような機関でこそ、この評価方法は有効なのかもしれない。

10. 結論と示唆

本稿では、学生の学習成果をどのように評価するかをめぐる2000年代初期のアメリカの動向を、アカウンタビリティを求める産業界や政府と、学生の成長と大学の教育改善を重視する大学人との対立の構図として分析した。その結果、どちらのグループも学士課程教育の学習成果を、専攻にかかわらず教養教育の成果として評価することには同意し、重視する観点においても大きな差異はないものの、その評価方法においては大きく異なる立場をとっていた。産業界や政府はアカウンタビリティを重視するため、一元的な基準で比較可能な評価を求め、その結果、標準化されたテストを開発した。他方で、大学人は、標準化されたテストでは、教育の多様性を損ねるとしてルーブリックを開発し、それを個々の機関の状況に合わせて利用することを推奨した。前者は機関間の比較可能な評価こそが教育改善を進めるというスタンスに立ち、後者は機関の特性の応じた評価こそが教育改善を進めるというス

タンスに立つことの差異である。そして大学業界では、前者を批判し、後者を擁護するケースが多い。

しかしながら、ルーブリックは、評価しようとするどの要素においても、評価の段階や観点が確立されているとは言えず、実際に使用されている要素はリテラシーやスキルなど、ある意味、数量的な評価が可能なものである。また、ルーブリックの評点を用いたとしてもその標準化が求められるが、それは機関内部の評価に依存せざるを得ず、標準化が容易ではない(Siefert 2012)。さらに、そもそもこうした内部評価が可能なのは、教員がその専門を超えて何が成果であるか議論して共通認識があることが前提である (Finley 2012)。

CLAを採用する機関と比較して、ルーブリック使用機関が如何に少ないかは明瞭であるが、それは日常的な評価実践においては、標準化されたテストの方がルーブリックの使用よりも簡便であることを示すものであり、それは、理念と実践、哲学とコストが容易には両立しないことを示している。

ここには同一の項目で同一の内容を測定したとしても、それを評価 (Evaluation) に用いるのか、診断 (Assessment) に用いるのか、その目的によって測定結果の意味が異なるという問題もある。Evaluationとはあくまでも外在的な基準をもとに評価するものであり、基準の明瞭性と結果の比較衡量が求められる。他方で、Assessmentは、それをを行う対象に対して、現状の認識と今後の方向性を示すものである。産業界や政府とそれに対立する大学人のどちらも Evaluation と Assessment の重要性に対して言及しているが、実はそれらを両立することは容易ではないのである。Evaluationを重視すれば、外在的な評価基準が必要であり、Assessmentを重視すれば、同一の対象に対する一定期間の変化を測定することが必要になるからである。

測定した学生の学習成果を多面的に評価する要請は世界的に強くなりつつあるが、そのなかでわれわれはどのような方法を採用していくのか。根本的な問題は、標準化テストやルーブリックという方法論の対立ではなく、EvaluationとAssessmentという両立しがたいミッションを両立させようとしていることにあるので

はないだろうか。そう考えると、このアメリカの事例はきわめて示唆的である。

注

- 1 個別大学を越えて大学間のベンチマークとして行われている数少ない事例として共同の質問による学生調査を行っているJSAAPを上げることができる。しかしながら、これは学生の学習成果の測定を第一義的な目的にしたものではなく、学生に対するアンケートから学生の成長を測定し、それに影響を与える環境要因を検討することを目的としている点で、本稿の扱う、あらかじめ学生の到達目標を決め、そのどのレベルに学生が到達したのかを測定する学習成果とは、異なるものである。
- 2 これは、教養教育のそもそもの理念が、人格形成とされたことと関連することによるものである。これに関しては、「6. 学習成果の項目」を参照。
- 3 教養教育が誕生した古典古代では、奴隷ではない「自由市民」の育成が目的であった。そのため、教養教育は、長く「支配層」の育成が目的とされてきた。第二次世界大戦後になって、「市民」の育成とされるようになった。
- 4 ランド・コーポレーションのVAAIは、2000年に初中等教育を対象にしたNo Child Left Behind Actが制定され、初中等教育の生徒の学習成果を測定し、それによって学校をランク付けし、初中等教育の質の向上を図ろうとしたことに強く影響を受けている。2000年から始まった高等教育の成果を州ごとに比較するMeasuring UPが学生の学習成果を測定していないことを問題として、より実質的な学習成果の測定を目的として始まった。さらに言えば、VAAIがもとになって、2004年には、高等教育機関に在学することによる批判的思考力の向上の程度を測定するCollegiate Learning Assessmentというテストが作成され、そのテストはOECDによる高等教育卒業時の能力測定を目的とするAHELOの議論においても使用が検討された。
- 5 教員からみた学生調査に関してはFaculty Survey of Student Engagement、学生の自己評価に関してはNational Survey of Student Engagement、全国レベルの標準化されたテストに関してはETSのAcademic ProfileとACTのCAAPを利用している。

6 ルーブック使用機関として選定された12機関において、どの項目が利用されているかを示したケーススタディ (<http://www.aacu.org/value/casestudies>) をもとに分析した。

引用文献

江原武一 (2010) 「アメリカにおける大学評価の改革動向」『立命館高等教育研究』第10号、pp. 113-124.

森 利枝 (2010) 「学習成果のアセスメントのインパクト」『学習成果アセスメントのインパクトに関する総合的研究 (中間報告書)』国立教育政策研究所、pp. 61-75.

森 利枝 (2012) 「アメリカにおける学習成果重視政策議論のインパクト」『学習成果アセスメントのインパクトに関する総合的研究』国立教育政策研究所、pp. 106-117.

ロスブラット、シェルダン、吉田 文・杉谷祐美子訳 (1999) 『教養教育の系譜』玉川大学出版部

Adelman, C., Daniel, B., and Berkovits, I. (2003) *Postsecondary attainment, attendance, curriculum, and performance: Selected results from the NELS: 88/2000 postsecondary education transcript study (PTTS), 2000*, Washington D. C.: National Center for Education Statistics.

Association of American College and Universities (2001) *The Status of General Education in the Year 2000: Summary of A National Survey*, Washington D. C.: Association of American College and Universities.

Association of American College and Universities (2002) *Peer Review, Value Added Assessment of Liberal Education*, Vol. 4, No. 2/3, Washington D. C.: Association of American College and Universities.

Association of American College and Universities (2005) *Liberal Education Outcomes: A Preliminary Report on Student Achievement in College*, Washington D. C.: Association of American Colleges and Universities.

Association of American College and Universities (2007a) *College learning for the new global century: A report from the National Leadership Council for Liberal Education and American's Promise*, Washington D.

C.: Association of American Colleges and Universities.

Association of American College and Universities (2007b) *Peer Review, Assessing Student Learning*, Vol. 9, No. 2, Washington D. C.: Association of American College and Universities.

Banta, Trudy W. (2007) "Can Assessment for Accountability Complement Assessment for Improvement?" *Peer Review*, Vol. 9, No. 2, pp.9-12, Washington D. C.: Association of American College and Universities.

Chun, Marc (2002) "Looking Where the Light Is Better: A Review of the Literature on Assessing Higher Education Quality," *Peer Review*, Vol. 4, No. 2/3, pp.16-25, Washington D. C.: Association of American College and Universities.

Finley, Ashley (2012) "How Reliable Are the VALUE Rubrics?" *Peer Review*, Vol. 13, No. 4/ Vol. 14, No. 1, pp.31-33, Washington D. C.: Association of American College and Universities.

Gaff Jerry G. (1991) *New Life for the College Curriculum: Assessing Achievements and Furthering Progress in the Reform of General Education*, San Francisco: Jossey Bass.

Gaff Jerry G. (2001) *The Status of General Education in the Year 2000: Summary of a National Survey*, Washington D. C.: Association of American College and Universities.

Hersh, Richard H. and Benjamin, Roger (2002) "Assessing Selected Liberal Education Outcomes: A new Approach," *Peer Review*, Vol. 4, No. 2/3, pp.11-15, Washington D. C.: Association of American College and Universities.

Hersh, Richard H. (2007) "Going Naked" *Peer Review*, Vol. 9, No. 2, pp.4-8, Washington D. C.: Association of American College and Universities.

Rothblatt, Sheldon (2003) *The living arts: Comparative and historical reflections on liberal education*, Washington D. C. : Association of American College and Universities.

- Schneider, Carol Geary & Shoenberg, Robert (1998)
Contemporary Understandings of Liberal EDUCATION, Washington D. C.: Association of American College and Universities.
- Shavelson, Richard J. (2007) *A Brief History of Student learning Assessment: How We Got Where We Are and a Proposal for Where to Go Next*, Washington D. C.: Association of American College and Universities.
- Shoenberg, Robert (2005) "Greater Expectations for Student Transfer: Seeking Intentionality and the Coherent Curriculum," *General Education and Student Transfer: Fostering Intentionality and Coherence in State Systems*, pp. 1-23, Washington D. C.: Association of American College and Universities.
- Siefert, Linda (2012) "Assessing General Education Learning Outcomes," *Peer Review*, Vol. 13, No. 4/ Vol. 14, No. 1, pp.9-11, Washington D. C.: Association of American College and Universities.