

広島大学 高等教育研究開発センター 大学論集  
第 37 集 (2005年度) 2006年 3 月発行：263—276

## 大学教育におけるプログラム評価の現状と課題

—教育成果を根拠とした形成的評価の確立を目指して—

串 本 剛



# 大学教育におけるプログラム評価の現状と課題

—教育成果を根拠とした形成的評価の確立を目指して—

串本 剛\*

## 1. 研究の目的と背景

わが国において、教育制度の一環として大学教育の自己評価が位置づけられて以来、およそ15年が経過した。その間、多くの大学がその結果を報告書として公表してきただけでなく、近年では第三者評価の基礎資料として提出する機会も増えたことにより、教育の自己評価はかなり定着した活動となっている。しかし、いまだその方法や活用に関して十分に吟味されていない側面があることも否定できない。そのひとつが、本稿が主題とする教育成果を根拠とした評価である。

教育成果による評価には、「教育成果を把握し評価する」段階と、「評価結果を教育改善に活かす」段階の2つを考えることができる。総括的評価では前者のみが関心となるが、形成的評価においては教育改善への活用を考慮した情報収集が求められるため、両者は密接に関連することになる。本論の目的は、形成的評価の観点から、教育成果を根拠とした評価に関する現状を批判的に分析し、その課題を明らかにすることである。

次節ではまず、「教育成果を把握し評価する」段階について、各大学が公表している自己評価報告書を材料に、教育成果を把握しうる方法として実際にどのようなものが用いられているかを明らかにした上で、それらの利用実態における問題点を検証する。続いて第3節では、現行の自己評価に見られる、「評価結果を教育改善に活かす」際の難点に触れ、その解決策のひとつとして、プログラム・セオリー評価の必要性を提起する。そして最後に、全体のまとめを行うと共に、残された研究課題を述べることにしたい。

なお本研究では、基本的な評価の単位を、コースや専攻あるいは学士課程といった教育プログラム、換言すれば具体的な目標を共有した4年（ないしは6年）間にわたる教育諸活動の総体、を念頭においていることを付言しておく<sup>1)</sup>。

## 2. 教育成果を把握し評価する—自己評価報告書に見る実状

### 2-1. 概要

現在日本の大学では、教育成果を把握するためにどのような方法が用いられているのか。これを明らかにするために、本節では各大学が公表している自己評価報告書を分析する。ここでいう自己評価報告書とは、教育活動について包括的に評価を行っているものを指す。従って、該当する場合には年次報告書等も含むが、「自己点検・評価報告書」と銘打たれていても、授業アンケートや卒

\* 広島大学大学院教育学研究科教育人間科学専攻（高等教育学）／日本学術振興会特別研究員

業生調査の結果をまとめただけのものは含まれていない。本研究で分析の対象としたのは、広島大学高等教育研究開発センターに保管されている自己評価報告書の内、2000年以降に発行されたものに掲載されている213大学（国立54，公立14，私立145）516学部（国立118，公立30，私立368）についての内容である。以下の考察では、学部を基本的な分析の単位としている。

516学部のうち、自己評価の項目として「教育の成果」（あるいはそれに類する「効果」、「目標の達成」など）を設定しているのは41学部を過ぎず、一割にも満たない。そこで本研究では、教育成果の把握方法を「実際に教育成果を把握している」方法だけではなく、「潜在的に教育成果を把握しうる」方法も含むものとして定義した。そして把握方法の計上に際しては、利用に言及しているだけの場合は含まず、それを通して得られた情報が記載されている場合のみを用いた。つまり、仮に「学生アンケート」を例に挙げると、「教育の成果」という項目で「学生アンケートを実施している」と記述してあるだけのときは計上していないが、「学生の生活」という項目の中で、教育成果についての設問はなくても調査結果が示されている場合には、1件として数えている。本節で示される数字の解釈においては、以上の点に留意されたい。

自己評価報告書から見出された把握方法は、全部で19種類であった。これらには、アウトプットに関わる「進級状況」、「卒業生数（率）」と、アウトカムに関するものがあり、後者の中には、半分以上の学部で利用されている「進路」から、2学部しか例のない「卒業研究」や「学会発表」まで含まれている<sup>2)</sup>。アウトカムの把握方法をその性質に従って分類すると、図表-1に示したような4つの類型を考えることができる（括弧内の数字は学部数）。ひとつめの分類軸は、収集される情報が主観的（意見）か客観的（事実）かを区別する。主観的情報としては、学生や教員に対するアンケート調査などにより収集される意見が、客観的情報としては、卒業生の進路や各種試験の成績・合格率などが挙げられる。もうひとつの軸は、教育成果をプログラム修了時あるいはその前後に一括して把握するか、プログラムを構成する各授業で分割的に把握しそれを累積して全体の評価に利用するか、という情報収集方法の別を意味している。

把握方法の利用の多寡を類型間で比較すると、一括×客観に分類される把握方法をいずれかひとつでも用いている学部が338学部（全学部の65.5%）あるのに対し、主観的情報は一括、分割の違いに関わらず15%程度の学部で利用されているに過ぎず、個別×客観つまり成績に関する情報を教育評価の根拠として示しているのは11学部、2.1%しかなかった。

図表-1 教育成果の把握方法とその類型

		情報の性格	
		客観的	主観的
収集方法	一括	一括×客観 (338) 進路 (295) / 職業 (152) 国家試験 (55) / 資格 (80) 就職先 (42) / 進学先 (13) / 受賞 (12) テスト (5) / 卒研 (2) / 学会発表 (2)	一括×主観 (80) 学生アンケート (59) 教員アンケート (24) 卒業生アンケート (12) 雇用者アンケート (4)
	分割	分割×客観 (11) 成績 (11)	分割×主観 (83) 学生による授業評価 (81) 教員による授業評価 (11)

## 2-2. 設置形態・専門分野・設置年・入試難易度による違い

前項で示した傾向は、学部の種類によって変わってくるのだろうか。設置形態、専門分野、学部設置年、入試難易度の4変数<sup>3)</sup>と、アウトプットを含めた5つの把握方法類型をクロスさせたものが、図表-2である。値は、当該の類型に含まれる把握方法を、ひとつでも用いている学部の割合である。表中斜体になっている部分は、1%水準で有意差が認められた関係であり、そこに注目してみると次の点が看取できる。

まずアウトプットの情報は、国立が他の設置形態よりも、理系の分野で文系よりも、また入試難易度では難しい場合に多くの学部が利用している。これは理系の学部が国立に多く、入試難易度も高いことが関係していると思われる。標準修業年限での卒業率は、国立大学で若干低いことから、卒業が難しいために注意が払われていることが推測される。

客観的情報によりアウトカムを把握する方法については、学部の設置形態による相違が見られる。といっても分割×客観に関しては、この類型に含まれるのが「成績」のみで、事例も少ないため、考察を加えるのが難しい。他方、一括×客観には10の把握方法が分類されている。どの把握方法においても、国立の学部が多く採用している傾向が見られるが、特に「進路」に関する情報は、7割以上の学部で自己評価報告書に記載されている。これは逆説的ではあるが、分析の対象となった自己評価報告書のうち、私立大学によるものの多くが、大学基準協会(JUAA)の大学評価に対応して作成された報告書であることが関係していると思われる。なぜならば、JUAAへの維持会員加盟申請に伴う自己評価では、卒業生の進路に関する情報を含むことが必ずしも求められていないため、それを掲載することを見送った学部が多かったと考えられるからである<sup>4)</sup>。

最後に主観的情報を利用した把握方法について見てみると、分割×主観に含まれる学生や教員による授業評価は、設置年が比較的新しい学部においてより多く利用されているのがわかる。新設学部の方がこれらの方法を実施することに抵抗が少ないためであろうか。これに対し一括×主観の把握方法は、入試難易度により利用度に差が見られ、比較的合格が易しい学部が積極的に取り組んでいるようである。この類型に含まれる把握方法別に確認すると、有意差は認められないものの、雇用者アンケートを除く全ての方法において、類型全体と同様の傾向が見出された。

以上の考察で得られた知見が、教育評価に際して、日本の学部がどのような情報の把握方法を用いているのかの一端を明らかにしていることは確かである。しかし本節の導入部分でも述べたとおり、分析の対象となった把握方法が、常に教育成果についての情報を得ることを目的として用いら

図表-2 把握方法の類型別利用割合

% (学部数)

	設置形態			専門分野		設置年		入試難易度		
	国 (118)	公 (30)	私 (368)	文系 (342)	理系 (172)	～1990 (439)	1991～ (77)	易 (139)	中 (135)	難 (146)
アウトプット	41.5	20.0	27.4	26.6	36.6	31.0	26.0	24.5	21.5	41.8
分割×客観	5.9	0.0	1.1	1.5	3.5	2.5	0.0	0.7	0.7	3.4
一括×客観	80.5	66.7	60.6	63.5	69.2	64.0	74.0	66.2	64.4	61.6
分割×主観	16.9	30.0	14.7	17.3	14.0	14.4	26.0	12.9	16.3	11.6
一括×主観	17.8	20.0	14.4	15.2	16.3	15.7	14.3	22.3	11.1	8.2

れているわけではない。以下ではこの点を考慮し、特にアウトカムの把握方法における現状の問題点を指摘し、参考となる実践例を交えながら代替案を提示する。

### 2-3. 類型ごとに見るアウトカムの把握方法の実践例

〈分割×客観〉既述のように、成績に関する情報を自己評価報告書に記載している学部は少ない。さらにその希少な事例においても、教育成果を評価するという点で十分な方法が採られているわけではない。ひとつには、そもそも成績情報が、科目間の成績評定格差を調査すること等を目的に、教育の「取組」に関する一指標として用いられている場合がある。また、仮に教育成果として情報が把握されていても、個別授業における分布に言及するのみで、それを累積し、プログラム全体の改善に活かすという視点が欠如しているのである。これらに代わる方法を現行の事例の中に探すのは難しいが、構想中のものとして広島大学工学部化学工学講座のアイデアは興味深い（滝瀧ほか2005）。同講座では、これまでに日本技術者教育認定機構（JABEE）によるプログラム認定に対応して、目標ごとの学習時間管理を実施してきたが、2006年度以降の入学生を対象に、目標別の成績レベル管理も計画している。具体的には授業毎に、成績と授業時間およびプログラムとしての目標のひとつに対するその科目の関与度の積を算出し、それを累積したものを当該目標にかかわる総学習時間で除すことにより、目標別に成績レベルを把握するという方法である。これが実行されれば、教員の側が学生の達成を各教育目標について把握し、不足している面を意識しながら教育改善ができるようになると考えられる。また成績レベルをプログラム修了時だけでなく、学年ごとの推移としても示していくことにより、学生の学習意欲を喚起するという作用も期待されている<sup>5)</sup>。

〈一括×客観〉この類型に含まれる把握方法として、比較的多くの学部で利用されているのが、学生の卒業後の進路やその職種・業種（「職業」）に関する情報の調査である。しかしこれらの情報は、医療系や教員養成系などの分野であればまだしも、専門職業人の養成を目的としない大多数の分野においては、教育成果を評価する根拠としては不十分であると言わざるを得ない。これに対し、学習の成果を直接的に測定する方法としては、資格試験や国家試験が若干利用されているものの、その他の客観テストやプログラムでの学習の総まとめである卒業研究に関する情報は例外的にしか示されていない。その例外を挙げれば、例えば三重大学人文学部の自己評価報告書には、卒業論文の成績評価内訳だけでなく、成績ごとの「審査所感」例や論文題目も記載されている（三重大学2001）。確かにこれらの情報だけでは、論文作成者がどのような能力を身につけたかを判断することは難しいという点など、いまだ改善の余地はあるものの、卒業者数（率）というアウトプットのデータに較べれば、教育成果の把握方法としてより適切であるといえるのではないだろうか。また三重県立看護大学（2001）は、「教育課程の評価」という項目の中で、クリティカル・シンキング（批判的思考力）検査の結果について検討している。同大学では、32の「推論」問題で構成された独自の客観テストを開発しており、入学時と卒業時のテスト結果の比較を行っている。日本では、批判的思考力に代表される汎用的能力の育成が奨励されている割に、欧米に比べ客観テストの開発が遅れが見られる。三重県立看護大学の取り組みは、同大学のみ小規模なものではあるが、注目に値する。

〈分割×主観〉主観的情報を用いた方法の利用に見られる問題点は、分割か一括かを問わず、アンケートの質問項目が、「成果」ではなく「取組」について尋ねるために設定されている例が多いところにある。本類型に含まれるのは各授業レベルで実施されるアンケート（授業評価）だが、100種類の学生による授業評価用の質問紙を分析した申本（2005）によれば、教育・学習成果の把握のために用いられている質問は全体の10%強に過ぎず、その内容も理解や関心の喚起を問う、具体性に欠けるものを中心となっている。このように、能力や知識の獲得について直接尋ねる項目を設けているアンケートは決して多くないが、武蔵野女子大学（現武蔵野大学）の文学部、人間関係学部、現代社会学部で平成14年度に実施された授業評価では、身に付いた力に関して次のような質問項目が設けられた（武蔵野女子大学・武蔵野女子大学短期大学部 2003）。

- |                         |                       |
|-------------------------|-----------------------|
| ① 積極的に学ぶ意欲              | ⑤ 基礎的な専門知識や技術         |
| ② 物事を論理的に考える力           | ⑥ 社会に出てから役立つ実用的な知識や技術 |
| ③ 物事をさまざまな視点から考える力      | ⑦ 人間として生きていくための教養     |
| ④ 自分の意見や考え方を分かりやすく表現する力 | ⑧ その他                 |

もちろんこの授業評価には、授業内容・方法や学生自身の受講態度など一般的に良くみられる評価項目も含まれているが、教育成果について尋ねたことにより、「自分の意見や考え方を分かりやすく表現する力」や「物事を論理的に考える力」が、他の能力に比べ十分に育成されていないという結果を得ることができている。恐らく今後、この結果にもとづき教育内容や授業方法の改善が図られていくのだろう。

〈一括×主観〉最後は、教育プログラムを通して得られた成果について、修了時に一括して意見を尋ねるという把握方法である。この場合、誰に意見を求めるかということがひとつの問題となるが、今回の分析の範囲では在学生に聞いている例が最も多い。雇用者や卒業生が対象になると、教育成果に関連した質問項目の割合が比較的多くなるが、在学生が対象の場合はいわゆる「学生実態調査」という名目で、学生生活に関する情報収集に終始してしまうことが少なくない。しかし、次のような例もある。弘前大学人文学部では、課程制カリキュラムの導入に伴い、第一期卒業生となる4年生を対象として2002年3月にアンケート調査を実施した（弘前大学人文学部 2002）。その際、所属した課程を通して以下のような成果が得られたかを尋ねている。

- |             |                |
|-------------|----------------|
| ① 基礎的な知識や技能 | ⑥ 奥深い知識        |
| ② 最先端の知識    | ⑦ 新しい発見        |
| ③ 社会で役立つ知識  | ⑧ 気付かなかった自分の発見 |
| ④ 社会で役立つ技能  | ⑨ 友達との出会い      |
| ⑤ 幅広い知識     | ⑩ 良い教師との出会い    |

友人や教師との出会いが教育成果と呼べるかは意見の分かれるところであろうが、プログラムの教育目標が期待したとおりに重点化され達成されているかを知るためには、有意義な設問群であると考えられる。生活実態調査にこうした項目を織り交ぜてみることは、ひとつの検討に値する試みかもしれない。

### 3. 評価結果を教育改善に活かす—プログラム・セオリーの明確化

教育は実践されること自体に意味があるわけではなく、何らかの成果が期待される活動である。従って改善が必要な箇所を知る上で、教育成果を評価することの重要性に疑義を挟む余地はない。しかし、単に教育成果を評価しただけでは、何を改善すべきなのかが明らかにならないことも事実であり、例えば大山（2001）はアウトカムズ評価の欠点のひとつとして、「成果にもとづき評価をおこなうので、そのような成果を規定している要因がどのようなものかつかみにくく、組織の改善には活かしにくい」（48頁）という点を指摘している。

一般に「暗箱評価」（Black box evaluation）と呼ばれるこうした事態に対して懸念が抱かれるのは、これまでの大学教育の評価の議論が、互いに独立した項目群を用いた網羅的評価を前提としてきたことを考えた場合、ある意味当然ではある。以下では、その点をまず確認したうえで、対応策のひとつとしてプログラム・セオリー評価（Evaluation based on program theory）について論及する。

#### 3-1. 日本におけるこれまでの大学教育評価論

1980年代後半以降、日本では大学教育の評価が本格的に論じられ始めたが、当時はまだ「大学評価の一領域としての教育評価」という認識が支配的であった。その中において、関（1989）は教育評価を独立的に取り上げ総合的に論じており、図表-3に示した教育評価の分類を行っている。関はアメリカの大学教育評価論者である Dressel の議論を参考に、Input 評価、Process 評価、環境評価、Output 評価の4つを区別した上で、評価の焦点<sup>6)</sup>についてさらに図表にあるような詳細な分類を行った。ただし、この分類は総合的であるがために、評価の単位が個々の学生や大学全体、あるいは教育制度といった形で多様であり、一貫性に欠けているという問題が見られる。

もうひとつの分類例は、大学基準協会（2000）によるものである。大学基準協会は、戦後半世紀以上にわたり大学評価にかかわってきた機関であり、2001年の省庁再編に伴う第三者評価機関の導入、および大学評価・学位授与機構による試行的評価の開始などを契機として政策提言を行った際、教育評価を図表-3にあるように整理している。関（1989）に類似して、評価の焦点としてインプット・プロセス・アウトプットを挙げているが、関が入学から卒業までの期間における特定の時点を根拠として分類しているのに対し、大学基準協会は、プロセスにあたる教育実践を中心に、それ以前にあらかじめ決定されているものをインプットとして、また実践の結果に関する情報はアウトプット・アウトカムとして位置づけている点に相違が見られる。この分類は比較的すっきりしているが、例えば、インプットの中に、施設・設備といったハード面の条件以外の項目も含まれており、そのため授業方法はプロセスであるのに対し、単位認定方法がインプットになるなど、分類の基準が曖昧になっているといった難点もある。

本論でこれまで教育成果と括ってきた要素は、いずれの分類にも含まれている。しかしそれは、あくまでも複数ある評価の焦点のひとつであり、「大学という組織の教育に係る部分」を網羅的に列挙したときの一要素に過ぎない。そのため、こうした分類を参考にして実施した教育評価では、評価の焦点間の相互関係に対する視点が欠落しがちであり、教育成果を評価しても、その結果を根



図表－3 従来の議論における教育評価の分類例

関 (1989)	大学基準協会 (2000)
① 教育目標 ② Input：入学（入学試験に関する諸活動・計画・支援体制・実施成果等） ③ Process：教育過程（教育課程，教授・学習過程，授業科目，課外活動等） ④ 学生団体等（自治的活動の実情等） ⑤ 教授団（採用基準や教育的成果等） ⑥ 組織運営（入試方法，意思決定システム，学長の権限・役割，教育改善システム等） ⑦ 環境（キャンパスの快適性，施設・設備の整備状況，教育支援体制等） ⑧ Output：卒業（卒業資格認定の方法と実態，卒業生の社会における活動状況等） ⑨ マクロな教育システム（教育政策，大学連合体の役割・権限・活動・計画等）	① インプット評価：カリキュラムの内容と履修方法，単位計算や単位認定のあり方，教員組織，教育用施設・設備など ② プロセス評価：履修指導，学生の学修の活性化と教員の教育指導方法改善のための措置，授業形態・授業方法など ③ アウトプット評価：単位取得率，留年率，退学率・卒業率，就職・進学等を含む卒業後の進路状況など ④ アウトカム評価：「目に見えるもの，目に見えないもの全てに亘る教育上の結果や成果が，当初の目的・目標に対応した「成果」といえるか否かが改めて検証される」

拠として何を改善すべきかがはっきりしない構造になってしまうことが避けられないのである。項目羅列型の評価が普及した背景としては，ひとつには大学教育の目標が具体化されにくいために，教育成果の評価基準が判然としないという問題が挙げられるが，もうひとつ一点目とも関連して大きな要因と考えられるのが，大学教育の評価が伝統的に，Eisner (2003) が言うところの鑑識と批評 (Connoisseurship and criticism) により実施されてきたという事実である。

鑑識と批評とは，専門家が見分ける物事の良し悪しを根拠として下される評価のことである。大学教育においては，自己評価であればプログラムの実施主体であり評価の責任者でもある教員 (団) が，外部評価であればプログラムのそれと近似した専門分野を研究している学外の同僚 (peer) が，その専門知識を根拠に教育の「取組」の是非を判断することが慣例となってきた。これは無論日本に限ったことではなく，アメリカのアクレディテーションを見ても，また欧州で行われている大学教育の評価を見ても明らかである。この方法を採用限り，評価の焦点間の関連に十分な注意が払われていないとしても，何の不思議もないことになる。

しかし，教育成果を根拠とした評価は，あらかじめ明示されている教育目標が評価基準となる，いわゆる目標標準拠評価 (Objective-based evaluation) であり，鑑識と批評とは性格を異にする。両者の優劣は一義的に決定できるものではないが，少なくとも教育成果による評価を行う場合には，従来どおりの項目羅列型評価では対応できないことになる。

### 3-2. プログラム・セオリー評価

日本の大学教育評価論において，教育プログラムが評価の単位としてあまり取り上げられてこなかったことは先にも述べたが，それはアメリカにおける大学の教育プログラム評価研究 (例えば Conrad and Wilson 1985 や Black and Kline 2002 など) とは異なり，行政による施策や非営利事業，あるいは学校評価といったプログラム評価論一般の知見や概念が援用される傾向がほとんど見られないことにも現れている。プログラム・セオリー評価は，このように日本の大学教育研究の中では

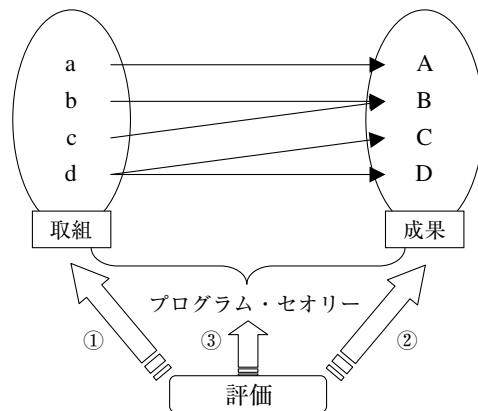
馴染みの薄いプログラム評価論の中で、1960年代から70年代にかけて提唱され始めたものである。

プログラム・セオリーとは、「活動に内在する原因を構成する手段と、結果を構成する目的との関係」（三好ほか 2003, 40頁）であり、大学教育の文脈に沿って平易に言い換えれば、「教育目標として表現される成果とそれをもたらす為の取組の関係」ということになる。プログラム・セオリー評価では、はじめにその関係を明示化した上で、それをういながらプログラムの評価を行うこととなる。

プログラム・セオリーは、初めから明示されている場合と、暗黙の前提となっている場合とがあり、後者の際にまず必要となるのが、プログラム・セオリーを顕在化させるという作業である。日本の大学教育では、各大学は、専門の学芸と幅広い教養を身につけさせ、豊かな人間性を涵養するという理念に基づいて教育諸活動を展開しており、暗黙のプログラム・セオリーは当然存在する。ただし、上述の理念は教育活動を規定するには抽象的過ぎるため、教育目標を具体化し、各目標に対応する教育活動の内容（授業科目など）を表した形のプログラム・セオリーが示されなくてはならない。これには大変な時間と労力が必要となることが予想され、現に教育目標の具体化を求めたNIADによる試行的評価事業では、多くの大学が困難に直面している（大学評価・学位授与機構 2005）。しかし第2節で広島大学の事例を挙げたように、JABEEに対応する中で、技術者教育関連の分野では多くのプログラムで実践されていることも忘れてはならないだろう。

明確に示されたプログラム・セオリーは、本題である教育を評価する際の基礎となるわけだが、そこでは大きくふたつの評価の焦点を考えることができる。ひとつは、教育の「取組」である。ここでは資源の投入やそれを用いた教育活動、つまり従前の分類にそって言い換えればインプットとプロセスが、当初意図されたとおりに実施されたかを評価する。もうひとつの焦点は教育の「成果」であり、どれだけの人数がプログラムを修了したかというアウトプットと、卒業生が意図されていた通りの能力を身につけたかというアウトカムが評価の対象となる。前節で扱った教育成果の様々な把握方法は、ここで活用されることになる。以上の内容を図示したものが、図表-4である。「取組」と「成果」の対応は一对一の場合もあれば、複数の「取組」がひとつの「成果」に、ある

図表-4 プログラム・セオリー評価のイメージ



いは同一の「取組」が異なる「成果」につながることも当然あり得る。

Rogers (2000) の言葉を借りれば、「取組」の評価(図表の①)において問題が見つかった時には、実施の失敗(implementation failure)ということになり、必然的に期待された成果は得られないことになる。他方、「取組」が計画通りに実行されたにもかかわらず、「成果」の評価(図表の②)で十分な結果につながっていないことが明らかになるということは、セオリーの失敗(theory failure)であり、想定していた「取組」と「成果」の因果関係が不適切であったということ意味する<sup>7)</sup>。

セオリーの失敗は、プログラム・セオリーの論理構成の妥当性が、教育・学習に関する研究の成果などに照らしてあらかじめ検証されている場合には起こらない。そのためプログラム・セオリーを顕在化させるのと同時に、それを検証すること自体も重要な作業であり、「プログラム・セオリーの評価(Assessment of program theory)」(Rossi et al. 2004)という形で、プログラム・セオリー評価とは区別されている(図表の③)。しかしながら、「取組」と「成果」のすべての関係に、ある程度確立された理論的な根拠が伴っているとは限らず、経験や慣行に依拠したプログラム・セオリーの下で実践されている教育があることも事実である。

従って、評価を行うことにより事後的にプログラム・セオリーの不備が判明したときには、意図していた「成果」に対応する「取組」を変更することが必要となる。この作業が、まさしく本論が主題としている、教育成果による評価に他ならないことは言うまでもないであろう。つまり、プログラム・セオリーを明確にした状態で教育成果の評価を行うことにより、評価結果を根拠にどこを改善すべきかが明らかになるため、暗箱評価に陥るのを防ぐことが可能となるのである。

#### 4. 結論と今後の研究課題

本稿では、教育成果を根拠とした大学教育の形成的評価を行うために、どのような課題を克服する必要があるかという視点から、ふたつの側面について議論を展開した。

第一の側面である教育成果を把握し評価する段階に関しては、自己評価報告書の中に見られた教育成果の把握方法を、アウトプットと、情報の性格及び収集の方法によって分けたアウトカムに関する4類型とに区別し、後者の使われ方について批判的に考察した。いずれも、潜在的には教育成果を把握するための有効な手段であるにもかかわらず、客観的情報としてはプログラムの目標との対応がわかりづらい情報が集められていたり、主観的情報を利用する把握方法においては、教育の「取組」についての情報を収集することに力が置かれていたりするといった課題を抱えていることが明らかとなった。

第二に、評価結果を教育改善に活かす段階については、現状で多く見られる項目羅列型の評価に伴う問題点を指摘し、プログラム・セオリー評価を援用することにより、「成果」の評価を通してどの「取組」を改善すべきかが示され得ることを見てきた。大学における教育活動は非常に複雑であり、プログラム・セオリーを明示することは簡単ではないが、その作業は教育評価に際して有益であるだけでなく、プログラム構成員の共通認識を築くという意味においても重要なことである。

教育成果による評価は、いまだに多くの教育プログラムにおいて有効に機能しているとは言い難い。しかし、これまでの「取組」重視であった教育評価を改め、「成果」の把握に積極的な事例も見られようになってきている。今後の研究課題は、教育成果の評価結果を用いた教育改善の成功には、現実的にどのような条件が必要とされるのかについて、具体的な事例分析などを通して解明していくことである。大学教育の評価においてプログラム・セオリー評価が機能しうるのかを明らかにする点でも、そうした研究は有意義であると考えられる。

## 【注】

- 1) これまでの日本における大学教育評価論では、評価の単位として教育機関全体か個別授業を設定することが多いが、評価結果を教育改善に活かすことを勘案した場合、機関評価だけではあまり役に立ちそうもないことは自明である。また、修学の責任を全面的に学生に負わせることが難しい今日の状況を考慮すると、「個々の授業が効果的になれば、プログラム全体としての教育の質も向上する」という前提を適用できる大学は、極めて限られているというのが現実であろう。本論が教育プログラムに照準を合わせ、プログラム評価論の知見を援用しようという所以はここにある。
- 2) 本稿においては、「アウトプット」を「プログラムを修了した学生の数」、「アウトカム」を「プログラムを修了した学生が得た利益」と定義し、両者を含む表現として「教育成果」を用いている。
- 3) 専門分野は、理・工・農・医療系を「理系」とし、それ以外を「文系」とした（不明2学部）。設置年は、自己点検・評価が努力義務として規定された1991年を境に、それ以前と以後でわけた。入試難易度は、代々木ゼミナール（2005）によって各学部の偏差値を調べ、49.9以下を「易」、50.0～55.0を「中」、55.1以上を「難」とした。データが不明な学部もあったため、合計は420学部になっている。
- 4) 評価報告書の中で、JUAAの枠組みを利用したことに言及していた学部は、私立では約半数、公立は3割、国立は5%程度であった。ちなみにNIADの枠組みを利用していた事例は、国立で14件、公立9件であった。
- 5) このように教育目標ごとの達成度を学生に管理させる方法は、様々な能力を経験値により発達させる点で、コンピューター・ゲームの一種であるロール・プレイング・ゲーム（RPG）にも似ている。単に履修の目安とするだけでなく、複数の教育目標の達成度を卒業要件とするような、「RPGモデル」とでも呼ぶべき制度ができれば、卒業試験の合格と規定単位取得に続く第3の大学卒業条件となりうるかもしれない。
- 6) 通常、評価の「対象」とも表現されるが、「対象」という言葉は非常に多義的であるため、ここではFitzpatrick et al.（2004）の用語にならい、「焦点」（focus）とした。
- 7) Rogers（2000）では上記ふたつの「失敗」のほかに、プログラム・セオリーが成立するための周辺的な状況が整わないことを意味する「不適切な状況」（unsuitable context）にも言及してい

る。ここでは議論を単純化するために説明を省略した。

## 【参考文献】

- 大山泰宏 (2001) 「大学教育評価の課題と展望」『京都大学 高等教育研究』第7号, 37-55頁。
- 串本剛 (2005) 「教育目的との対応にみる教育評価の妥当性—授業評価項目の分析を具体例に—」『大学教育学会誌』第27巻第1号, 124-130頁。
- 関正夫 (1989) 「教育評価の原理と方法に関する一考察」『大学論集』第20集, 3-31頁。
- 大学基準協会 (2000) 「大学評価の新たな地平を切り拓く (提言)」。
- 大学評価・学位授与機構 (2004) 「大学評価・学位授与機構が平成12年度から平成15年度までに実施した試行的評価に関する検証について—試行的評価に関する検証結果報告書—」 (<http://www.niad.ac.jp/>) (2005年09月22日アクセス)。
- 滝島繁樹・都留稔了・奥山喜久夫・吉田英人 (2005) 「広島大学工学部化学工学講座における JABEE 認定取得に向けた教育改善」(日本工学教育協会第53回年次大会発表資料, 2005/09/11於広島大学)。
- 弘前大学人文学部 (2002) 「弘前大学人文学部の現状と課題—教育に関する自己点検評価報告書—」。
- 三重県立看護大学 (2001) 「軌跡を見つめ未来を探る 自己点検・評価報告書 2001」。
- 三重大学 (2001) 「三重大学 その現状と課題9」。
- 三好皓一・森田智・藍澤淑雄 (2003) 「わが国における適切なプログラム・セオリーの構築を目指して—国際協力評価と政策評価に焦点を当てて—」『日本評価研究』第3巻第2号, 40-56頁。
- 武蔵野女子大学・武蔵野女子大学短期大学部 (2003) 「現状と課題 (平成14年度)」。
- 代々木ゼミナール (2005) 『2006年度大学入試 代ゼミデータリサーチ 2005年度入試合否調査 DATA—入試難易ランキング VOL. 1』。
- Black, K. and Kline, K. (2002) 'Program Review: a spectrum of perspectives and practices', in Trudy Banta (ed.), *Building a Scholarship of Assessment*, San Francisco: Jossey-Bass, pp. 223-239.
- Conrad, C. and Wilson, R. (1985) *Academic Program Reviews*, Washington: ASHE.
- Eisner, E. (2003) 'Educational Connoisseurship and Educational Criticism: an arts-based approach to educational evaluation', in Thomas Kellaghan and Daniel Stufflebeam (eds.) *International Handbook of Educational Evaluation*, Dordrecht: Kluwer Academic Publishers, pp. 153-166.
- Fitzpatrick, J., Sanders, J., and Worthen, B. (2004) *Program Evaluation: alternative approaches and practical guidelines (3<sup>rd</sup> ed.)*, Boston: Allyn and Bacon.
- Rogers, P. (2000) 'Program Theory: not whether programs work but how they work', in Daniel Stufflebeam et al. (eds.), *Evaluation Models (2<sup>nd</sup> ed.)*, Boston: Kluwer Academic Publishers, pp. 209-232.
- Rossi, P., Lipsey, M., and Freeman, H. (2004) *Evaluation: a systematic approach (7<sup>th</sup> ed.)*, London: Sage Publication.

# The Present State and Challenges of Program Evaluation in University Education: Toward the Establishment of Formative Evaluation Based on Learning Outcomes

Takeshi KUSHIMOTO\*

This article identifies the present state and challenges of program evaluation in Japanese university education from the perspectives of binary phases.

The first phase is outcomes assessment. As a result of analyzing self-evaluation reports of 516 faculty members, a total of 17 methods for outcomes assessment are identified and categorized into 4 groups. The categorization is done according to the characteristics of the information gathered - objective or subjective, and the ways of collection - 'entire' and 'partial'. 'Entire' methods are those used at the end of a program in contrast with 'partial' methods that are used at the end of each subject. Typical examples of methods using objective information are graduates' courses and their jobs. Since it is irrelevant to show the attainment of education in non-professional fields, it is required to develop objective tests and a transparent system to assess graduation theses. Those methods utilizing subjective information, such as questionnaires for students and other stakeholders, often seek information about 'efforts'; it would be preferable if they included more questions about 'results' in order to be used for evaluation based on learning outcomes.

The other phase is an application of assessment results to educational improvement. Since the traditional evaluation form of university education has been 'connoisseurship', the relationship between evaluation focuses, such as input, process and output, have not attracted sufficient attention. Consequently, evaluation based on learning outcomes is often seen as problematic because of the risk of a 'black box evaluation'. Program theory evaluation, introduced here from the field of program evaluation, could prevent risks of this kind. In program theory evaluation, one has to articulate a program theory, i.e. the relation between 'efforts' and 'outcomes', in advance of implementing evaluation. This operation makes it possible for members of faculty to change program 'efforts' appropriately according to the 'results' of outcomes assessment.

The remaining research topic is to reveal how the results of outcomes assessment may be linked to actual improvement of education by means of case studies.

---

\*Doctoral Student, Graduate School of Education, Hiroshima University  
Research Fellow, Japan Society for the Promotion of Science