

教育における「評価」の考え方とその機能

斎藤 裕

問題と目的

大学において、現在、教員側の研究内容や業績は言うに及ばず、教育環境・教育内容まで含めて、点検を行う動きが大勢となっている。前者においては、各教員の研究総覧の作成や研究発表・公開が主となっており、後者においては、授業内容の公開としての「シラバス」作成や授業の充実度を点検すると称して「学生による授業アンケート」の実施を行う大学が目立ってきている。ここでは、あえて「評価」という言葉は使用しなかったが、世間的には、(大学)「評価」と言えば、上述したようなことを行うことを指している。このようなこと(教員の業績一覧やシラバスの作成、及び学生による授業アンケート)を行い、それを『大学における自己(若しくは第三者)評価』と名づけていることが、数多く見られる。

「評価」は、誰のいかなる行為を指すものであろうか。何のために行うものであろうか。多くの人々は、教員による評価と言えば、学習者(学生)に対して行うもので、具体的には、期末テストの採点や成績つけを意味し、逆に、学習者(学生)による評価と言えば、「授業がわかりやすかったか」等のアンケートで、教員に点数づけをすることだと思っている。しかし、これが「評価」なのであろうか。ここには、2つの問題がある。1つは、教員が学生に対して、あるいは学生が教員に対して行うものが「評価」なのかということであり、2つめは、そのどちらの場合も「数値」化されることが多いが、その場合の数値は、いかなる“ものさし(尺度)”で測られ、どのような性質を持っているものなのかということである。第1の問題は、「評価」とは何のために行うのか(評価の目的)という根本的な問題であり、第2の問題は、その目的を達成するための手段としての「測定」は、どのようなものなのか(正しく測られているのか)という問題である。

評価の目的を吟味する前に、まず、この「測定」について考えてみたい。

「測定」がどう「評価」に関わってくるかは、後に述べるとして、学習者の能力を数値として表現するためには、どうそれを正確に測っているかが大きな問題となる。例えば、A君が中学校1年・1学期・数学の期末テストで残念ながら“0”点を取ってしまったとしよう。この数値“0”は、何を意味するのだろうか。「全く数学的能力がない」ということを意味するのだろうか。重さを測った時、0gならば重さがないということである。長さを測って、0cmならば、長さがないということである。それと同様に、数学0点は、数学的能力がないということだろうか。そうではないことは、明白である。「ない」のであったら、A君は、1桁の足し算も引き算も、全くできないことになってしまう。そんなことを、この0点は意味してはいない。同じ“0”という数値は、教育におけるテスト結果の場合と重さや長さを測った場合とでは、違う意味を持っているのである。

「測定」とは、測りたい対象に「ものさし」(尺度)を当てて、その目盛(数値)を読み取ることである。そして、我々は、以下の4種類の尺度を持ち、対象に合わせて使い分けている。

- ① 名義尺度 (nominal scale)
- ② 順序尺度 (ordinal scale)
- ③ 間隔尺度 (interval scale)
- ④ 比例尺度 (ratio scale)

名義尺度の数値は“ラベル”である。分類をするために便宜上数値を使用しているに過ぎない。別に数値を使用しなくとも、かまわないのである。学校で同学年にクラスが5つあり、1組・2組・3組・4組・5組と命名したときの数字が、この尺度のよる数値である。別に数字でなくともよいこと（「きく・さくら・すみれ・ゆり・ひまわり」でも何ら差し支えない）は、誰でもわかろう。この尺度のよる数値は、分類（区別）という働きしかなく、それ以上のものではないのである。

順序尺度の数値は、順序を表すものである。しかし、順序を表すだけで、数値の差の大小は、必ずしも測定対象の性質の差の大小と対応するものではない。尺度上の等間隔性は、全く保障されないのである。100m競争での1位・2位・3位とつけられた数字が、代表的である。この場合、「1・2・3」は順序のみを示しており、その差は全く斟酌されていないことがわかろう。

間隔尺度は、順序に加えて、等間隔性が保障された尺度である。代表的なものとして、温度計につけられている目盛が挙げられる。温度計は間違いなく等間隔に目盛がふっており、その意味では、10℃と5℃との温度差と25℃と30℃との温度差は同じ5度である。しかし、この尺度の特性として、絶対“0”がないことが挙げられる。0℃は、水の融点であるにすぎず、便宜上定められたものである。0℃で、温度という性質が無くなるわけではないことは明白であろう。

比例尺度は、間隔尺度に絶対“0”が加わったものである。「度量衡」が、代表的なものである。この尺度の目盛は、等間隔で、かつ絶対“0”を有している。0gならば重さがないということであり、0cmならば、長さがないということである。

このように4種の尺度が存在しているが、教授－学習活動上行われている「テスト」に用いられている数値は、どの尺度上の数値なのであろうか。前述したように、どのテストで0点を取っても、その“0”は絶対“0”ではないのであるから、比例尺度のよるものではないことがわかる。では、間隔尺度上の数値であらうか。100点満点で100点と90点の差10点と60点と50点の差10点とは、同じであらうか。とてもそうとは思えない。また、問題毎に傾斜配点がなされていることが通常であり、どの問題ができたか（あるいはできなかったか）によって、正答数が同じでも、得点に変化することになる。Table 1を見てもらいたい。A・B・C・Dの4人のあるテスト結果を示してある。問い毎の傾斜配点によって、正答数が同じでも総得点は全く違ってしまうのであり、そのことを考慮すれば、点数差は、理解度の違いを同一に反映したものでないことは明らかであろう。また、順序さえ怪しいものである。傾斜配点が妥当なものであれば、総得点の順序性は保持されるが、そこが誤っていれば、全く違う順序になってしまう。配点ⅠとⅡで、4人の順序が変わっていることがわかる。配点の重みづけが誤っていれば（傾斜配点が理解度を正しく反映するものでなければ）、提出される数値は、順序性さえ保持されるものでは無くなってしまふのである。そこにあるのは、トータルとして「テストができたか・できなかったか」のラベルにすぎないもの、つまり、名義尺度によるものでしかないということになってしまうのである。問題毎の傾斜配点の十分な吟味を欠いたテストは、測りたいものを正しく測っていない可能性が高いと言えよう。

それは、通常の科目テストに対してだけのことではない。「授業評価」と称される学生への授業アンケート

Table 1 正答数と総得点・順位の関係性

配点Ⅰのケース							
問い	1	2	3	4	5	得点 (100満点)	順位
配点Ⅰ	10	10	20	30	30		
A	○	○	○	×	×	40	3
B	×	○	×	○	○	70	1
C	×	×	×	×	○	30	4
D	×	×	×	○	○	60	2

配点Ⅱのケース							
問い	1	2	3	4	5	得点 (100満点)	順位
配点Ⅱ	30	30	20	10	10		
A	○	○	○	×	×	80	1
B	×	○	×	○	○	50	2
C	×	×	×	×	○	10	4
D	×	×	×	○	○	20	3

に対しても、言えることである。荒井 (2001) は、この点に関して「全項目の平均評定値や評定の合計点などという指標は、学生による評価項目への要求がすべて等価であるという前提が確認されていない以上、必ずしも得たものであるとは言い難い」と述べ、国際基督教大学 (ICU) と東北大学で用いられている授業評価アンケート項目の配点調査を行っている。その結果、項目間における約 2 倍の傾斜配点が報告されている。彼の研究結果は、アンケート項目を全て等価と考えることは、少なくとも、学生が何をよい授業と考えているかを正しく把握できないということを示している。また、荒井は「(当該) 授業内容がそれなりに理解されたと判断できる学習者は、授業の内容的側面の配点が高く、そうではない学習者は授業の進行面についての配点が高いという結果となった」と述べているが、①必ずしも 1 回だけの授業内容・再認課題の正否だけでは、学生の特性把握としては不十分ではないか (現に、荒井の授業自体のアンケート結果では、そのような関係性は見られなかった)、②両大学のアンケート項目自体、そのような視点から組み立てられたものとは思えない、という点で、より詳しい研究が必要であろう。

授業は、「教師が、ある内容を、学習者に教え・学ばせるプロセス」である。その意味では、学生 (学習者) にアンケートを取るならば、少なくとも「教員 (教師)」・「内容」・「プロセス (進行)」という視点から項目選択及び型分けがなされなければならないと考える。そうした結果を踏まえてどう項目間に傾斜配点を行うかが、「授業」を測定していく上で、重要になるのではないだろうか。

今回、幾つかの大学で行われている授業アンケートを参考にしつつ、上記の視点で項目を選択・作成し、学生は、それらの項目にどう傾斜配点するかを調査することを 1 つの目的としたい。

次に、もう 1 つの「評価」の大きな問題、「誰が・何のために・何を測定するのか」という問題を考えてみたい。

アメリカの教育心理学者リー・クロンバックは、1960年代のアメリカにおける自然科学を中心とした教育内容刷新運動の中で、「授業の評価は、授業がどのような変化を生じさせたかを確かめ、改善を要する点を明らかにするものでなければならない」と述べ、「評価とは、1つの教育計画について決定を下すための情報の収集とその使用である」と定義づけている。この定義は、教育界において清冽で、画期的なものであり、日本で

も彼の影響を受け、多くの教育心理学者が、新たなる「教育評価」の考え方を打ち出している。代表的なものは、東京大学で教育心理学の教鞭をとっていた東洋によるものである。彼はクローンバックの定義をより発展させ、「教育評価とは、教育活動にかかわる意志決定の資料として、教育活動に参加する諸部分の状態、機能、所産などに関する情報を収集し、整理し、提供する過程である」と規程している。また、東北大学で教育心理学を担当し、現在も現場の教育実践研究の第一人者である細谷純は、クローンバックの定義における主語の明確化を意識し、「教育の評価とは、一定の教育計画の実現に責任を持つ主体が、自己の権利・権限において、制御要因について決定を下すためになす（主として）目下遂行中の自己の計画実現結果からの情報の収集と使用のことである」と定義している。今現在、これらの定義が教育学会や教育心理学会の定説となっていると、言ってよいであろう。

ここには、評価において、教師側が学習者側に対して成績をつける（ラベルを貼る）という意味合いは、全くない。また当然、その逆もない。「評価」は、他者に対して行うものではないからである。評価は常に自己評価なのである。評価は、一定の目的を追求しようとしている全ての活動に必ず付随している活動なのである。芸術家でも、科学者でも、自らの追及する目的を達成するために、自己の活動（芸術家なら、作品等）を評価していよう。そうでなければ、よりよい研究成果や作品は生まれようが無いはずである。

授業も、同じである。授業は、教師側から見れば教育（教授）活動であり、学習者側から見れば、学習活動ということになる。教員は教えたい内容（目的）があり、学生は学びたい内容（目的）があるはずである。したがって、教員は教育活動を評価し、学生は学習活動を評価するということになる。教員が学生の学習活動を、学生が教員の教授活動を評価するということは、ありえない。

しかし、残念ながら、このような考え方は、一般に理解されているとは言えないようである。佐藤康司（1989）は、教員を目指す私大（文系）学生（43名）を対象に、「教師は①何を（評価の対象）、②何のために（評価の目的）、評価するのか」について調査を行っている。その結果、学生らは、1）その対象として「子どもの理解度や学業成績」を半数以上が挙げるのに「教師自身の教え方」は誰も挙げていない、2）目的は子どもの理解や動機づけ・情報の提供で教師自らの次なる指導方法の模索などは殆ど挙がってこない、という事実が示されている。この考え方は、明らかに教育における「評価」とは異なるものである。このような考え方が、学生にも、教員にも、一般的だとすれば、授業アンケートをいくら取っても、お互いにレッテル張りをするにすぎず、何ら評価活動にはなり得ていないことになるろう。

また、先の「測定」との関係で言えば、確かに評価活動は測定活動と不可分な関係にある。しかし、不可分な関係にあるだけで、「測定」は「評価」ではない。教師は、自らの教授活動を、設定した（教授）目標との関連で評価しなければならない。その際に、教師は、「どの程度目標に迫れたか」知るために測定を行うのである。もちろん、「何をどのように測るのか」こそ、根本問題ではある。永野重史は、「昔は教育評価のうちには数え入れることができなかつたようなことで、実際には教師にとってたいへん大切な情報もある。例えば、授業中に生徒の顔が生き生きしていたとか、ざわついていてすっかり退屈している様子だったというような情報がそれである。試験の出来具合だけが評価だというように考えないで、その種の情報を大切にしたいほうが授業は良くなっていくであろう。」と、述べている。どんな情報を収集するかは別として、教師側は、自らの教授活動を評価するため、何らかの形で測定を行い、そしてそれを評価に生かしていかなければならないのである。

科目テストの点数は、測定の結果であり、評価ではない。言わば、健康診断における体重測定と同じである。

体重評価とは、誰も言わない。これは、体重に関する測定なのであって、問題は、その数値をどう読み取り、今後の健康維持にどう生かすかということであろう。「太りすぎたな、少し運動しよう」等の判断及び実際の行動があって初めて、健康増進という目的達成のための「評価」となるのである。

このような認識を、教員・学生とも共有しているであろうか。前述の佐藤の調査は、現役の小・中学校の教師にも行っているのであるが、評価の対象として、学生には全く見られなかった「教師自身の教育活動」を多くの者が挙げるといこともわかった。教員が「評価」を正しく理解していても、学生が誤った理解のままならば、その成果はあがらない。この点について、その認識程度を、教員・学生とも調べる必要があると考える。

したがって、授業評価を実施する前提として、教員・学生の「評価」イメージを調査し、両者がどのような「評価」観を所持しているのかを調べることを、本研究の第2の目的としたい。

I 授業評価項目に対する学生の意識調査

Table I - 1 が山梨大で、Table I - 2 が東北大で、Table I - 3 が国際基督教大 (ICU) で用いられている授業評価表 (項目) である。それぞれ、表現は微妙に異なっているが、提示されている項目内容は、ほぼ同様なものと言ってよいだろう。

Table I - 1 山梨大学・授業評価項目

評価項目	
1	授業内容の量はどうか
2	授業の速度についてはどうか
3	判り易い説明でしたか
4	この授業に対する教官の準備はどうか
5	この授業の内容について新たな興味を持ちましたか
6	授業は全体としてまとまっていたか
7	質問などを含めた学生に対する教官の対応はどうか
8	熱意が感じられる授業でしたか
9	この授業の休講回数はいかがでしたか
10	この授業内容に対するあなたの満足度はどうか

Table I - 2 東北大学・授業評価項目

評価項目	
1	内容が系統的に良く整理されている
2	授業がよく準備されている
3	説明が理解しやすい
4	講義の声がとく聞き取れる
5	授業を進める速度が適切である
6	板書の文字や数式が読みやすい
7	授業概要 (シラバス) にそって授業が行われる
8	視聴覚機器の利用が有効である
9	宿題 (レポートを含む) で授業内容の理解を助ける
10	教科書・プリントの分量や内容が適当である

Table I - 3 国際基督教大学 (ICU)・授業評価項目

評価項目	
1	講義が全体としてよくまとまっている
2	講義の内容が創造性に富むものである
3	講義から触発されることが多い
4	講義によって自分の期待していたものが満足される
5	教員が学習の目標をはっきりと示す
6	教員と学生との間にコミュニケーションが十分成り立つ
7	教員が講義の内容について十分な知識を持っている
8	教員が周到な準備をし熱意をもって授業を行う
9	授業の進め方の時間的配分が適切である
10	理解を助けるために各種の補助手段 (プリント・視聴覚教材等) を適切に用いる

Table I - 4 授業評価項目・配点表

評価項目	
1	教員が十分な準備をし、熱意をもって授業を行うこと
2	教員が授業内容の分野に十分な知識を持っていること
3	教員と学生との間に十分なコミュニケーションがとれていること
4	講義内容が十分魅力的なこと
5	講義内容の量が適切であること
6	講義内容がわかりやすいものであること
7	教員の声のトーンなど、授業のスピードが適切であること
8	板書がきれいであること
9	授業の理解を助けるため、プリント・テキスト・AV教材の使用がなされていること
10	授業が時間どおりに始まり、時間どおりに終ること

これらを参考に、「①教員の資質や授業意欲等に関わるもの・②授業内容自体に関わるもの・③授業方法や進行面に関わるもの」に型分け・整理し、プロトタイプとして以下のような「授業評価項目・配点表」を作成した（Table I - 4 参照）。項目 1 - 3 が教員の資質や授業意欲等に関わるものであり、4 - 6 が授業内容自体に関わるものであり、7 - 10 が授業方法や進行面に関わるものである。これらの項目に対して学生がどう配点するのか・どの項目を「よい授業」の指標として考えているのかを、調査したい。なお、荒井は当該授業の再認課題も併せて回答させており、その成績との相関を調べているが、今回は当該科目ではなく、調査対象者（学生）の履修全科目の成績（平均）との相関を、併せて調べることにする。

方 法

(1) 調査対象学生

県立新潟女子短期大学生活科学科生活福祉専攻1年生—43名

(2) 実施手続き

上記質問紙（「授業評価項目・配点表」）を学生に配布し、各自のペースで回答を求めた。所要時間は、約20分であった。回答内容は、提示されている10項目を100点満点で配点するものである。なお、質問紙には、記名が求められた後、以下のような指示が明記してある。

「授業評価は、今いろいろな項目について、みなさんのような学生を対象にアンケートが実施されています。みなさんは、『よい授業』について、いろいろな考えを持っているでしょう。どのような授業がよいと考えていますか。以下の表に、様々な大学で取り上げられている項目を整理してみました。この授業のような講義科目について、よい授業だったかどうか、あなたが採点するとしましょう。これらの項目について、100点満点になるように、あなたの考えで、配点して下さい（この授業に対する要望ではなく、あなたが受けている講義全体に通じるように配点して下さい。）」

結果と考察

Table I - 5 に、項目別の配点平均を示す。1人として均等配点（各10点）した者はおらず、各自の判断で傾斜配点を行っている。以下に、上位3項目と下位3項目を示す。

1位：講義内容がわかりやすいものであること（15.8）、2位：講義内容が十分魅力的なこと（13.3）、3位：教員が授業内容の分野に十分な知識を持っていること（12.4）

8位：教員の声のトーンなど、授業のスピードが適切であること（8.7）、9位：板書がきれいであること（5.3）、10位：授業が時間どおりに始まり、時間どおりに終わること（4.9）

1位と10位とでは、3倍以上の重みづけの開きがあることがわかる。これまで多くの大学で用いられてきた授業評価項目に対する学生の要求度合いは、全く等価でないことが判明した。教員の資質・授業内容・授業進行方策の型で見ると、授業内容関係項目に高い配点がなされており、進行方策系はさほど問題とされていない結果となっている。

全体で見るとそのような結果となっているが、成績別に見ると、やや様相は異なってくる。成績・履修全科目平均80点以上を「成績・高群」、70点以上80点未満を「成績・中群」、70点未満を「成績・低群」とし、各群の要因（型）別配点平均を見たものがTable I - 6 であり、成績と要因との相関を調べたものが、Table I - 7 である。これを見ると、①成績が低い者ほど授業進行要因に高い配点をしている（ $p < .05$ ）、②成績が高い者

Table I—5 項目別・配点平均

	評価項目	配点平均
1	教員が十分な準備をし、熱意をもって授業を行うこと	11.7
2	教員が授業内容の分野に十分な知識を持っていること	12.4
3	教員と学生との間に十分なコミュニケーションがとれていること	9.3
4	講義内容が十分魅力的なこと	13.3
5	講義内容の量が適切であること	8.5
6	講義内容がわかりやすいものであること	15.8
7	教員の声のトーンなど、授業のスピードが適切であること	8.7
8	板書がきれいであること	5.3
9	授業の理解を助けるため、プリント・テキスト・AV教材の使用がなされていること	10.1
10	授業が時間どおりに始まり、時間どおりに終わること	4.9
	計	100

ほど教師要因に注視している（統計的に有意なほどではない）、ことがわかる。荒井は、当該授業の再認課題で同様な指摘をしていたが、今回、学業成績全般でも同様な関係性が見られたことは、興味深いと言えよう。

学習者の内的条件と教授方略は、交互作用を持っている。R. E. スノウらの研究が有名である。彼らは、大学生を対象に物理学の授業を行う際、実際の実験を行わせるやり方と映像を見せるやり方の2種類を用意し、それらと学生の内的条件（対人積極性）との交互作用を調べたのである。その結果、対人積極性が高い者は実験的方法の方が、低い者は映像使用方法の方が、内容をよく理解したのである。指導方法だけを考えていたのでは、その効果を十分に果たせないし、また、学習者の特性を考慮しすぎても、効率的ではない。スノウの研究は、指導方法と学習者の特質との組み合わせを考慮して初めて、有効な学習支援ができるということを示している。この種の交互作用は、一般に「適性処遇交互作用」(ATI: Aptitude Treatment Interaction) と呼ばれている。

同じ意味で、「よい授業」に対する認識も、学生の内的条件に左右されるのであり、同時に、それを考慮して初めて、「よい授業」への建設的な方策が建てられるのである。今回明らかになったことをベースに考えれば、授業内容が高度な場合・あるいは学習の初期段階において、教員は授業の進行に配慮する（例えば、板書を丁寧にする・ゆっくり話す等）必要があると言えるだろう。また、学生が何らかの理由で学力が低いと予想される場合も、同様である。逆に、授業回数を重ねて内容が理解され始め、十分についてこれることがわかったら、教員は、授業進行に気をつかうよりも、下準備やさらなる授業内容の深化へ配慮をする方が、学生も満

Table I—6 項目別・配点平均

	教師要因	授業内容要因	授業進行要因	計
成績・高 (13名)	36.3	38	25.7	100
成績・中 (27名)	32.3	38.2	29.5	100
成績・低 (6名)	31.7	34.3	34	100

Table I—7 成績と各要因との相関

		教師要因	授業内容要因	授業進行要因
成績	Pearsonの相関係数	0.145	0.102	-0.322
	有意確率（両側）	0.337	0.502	0.029
N		46	46	46

足し、よりよい・充実した授業となっていくであろう。

少なくとも、我々は、ただ単にこれらの項目を等価と判断するのではなく、どのような学生がいて、どの程度内容を理解しているのかを考慮しながら、その授業内容・手法等を判断していく必要があるのである。「優れた指導法があって、どんな学習者でもおしなべて効果を生む」というようなことも、「学習者が悪ければ、どんな優れた指導法も効果はない」ということも、ありえない。よりよい授業の建設も、また同様なのである。

Ⅱ 「評価」に対するイメージ調査 — 学生・教員にとって「評価」とは —

前述したように、授業は、教員と学生の共同作業である。教員側は教授目標を持ち、学生側は学習目標を持ち、各々その達成に努めなければならない。その中で、両者とも、自己の目的的活動を評価するという事態が望まれるはずである。しかし、必ずしもそのような「評価」観を両者とも所持していない可能性が高いという指摘がある（前出 佐藤）。

したがって、今回、佐藤の研究（自由記述）の結果抽出された「評価イメージ（目的と対象）」を整理し、教員・学生が「評価」をどう考えているか—①教員は何を（対象）・何のために（目的）行うのか、②学生は何を（対象）・何のために（目的）行うのか—を調べることを、まずその目的とする。そして、その結果をもとに、今後授業評価どう行っていけばよいのか、その方向性を模索していきたい。

方 法

(1) 調査対象学生

県立新潟女子短期大学全学科・専攻1年生—321名

県立新潟女子短期大学教員—28名

(2) 実施手続き

i) 調査内容

佐藤の調査結果から浮かび上がってきたイメージと「評価」の定義とを照らし合わせて整理し、「評価の対象・項目群」と「評価の目的・項目群」を用意した。それぞれに、“①教員の評価としてどうか・②学生の評価としてどうか”について「そう思う・どちらと言えばそう思う・どちらかと言えばそう思わない・そう思わ

Table II—1 授業評価・調査項目

評 価 項 目		評 価 項 目	
評 価 の 対 象	学生の教授内容に関する理解度	評 価 の 目 的	学生の考え方・理解度を把握するため
	学生の学び方（勉強方法）		学生の授業内容に対する理解度を促進するため
	学生の学習意欲		学生の学習意欲を高めるため
	学生の性格		学生に授業内容理解度の実情・結果を示すため
	学生の授業態度		学生の成績をつけるため
	教員の教授意欲		教員が今後の指導方法の参考にするため
	教員の教え方		教員が授業進度の調整等の目安にするため
	教員の教える内容		教員が授業内容の適切具合を知るため
	教員の授業態度		教員の教授意欲を高めるため
	教員の性格		教員の成績をつけるため

ない」に分類してもらうことになる。Table II - 1 に調査項目を示しておく。なお、教員には、これらの項目に併せて教員状況項目：『・教員免許の有無・学生時代における教育学関係科目履修の有無・研究分野（理系／文系／その他）・教職歴（5年未満／5－10年／10年以上）』の回答も、お願いしている。

ii) 手続き

教員は、各メールボックスに調査用紙を配布し、回答を願った後、回収した。学生は、全学1年生科目担当教員の協力を得、授業開始直後に回答をお願いしている。所要時間は、15分程度であった。

結果と考察

「そう思う」に4点（以下、・どちらと言えばそう思う：3点－・どちらかと言えばそう思わない：2点－・そう思わない：1点）を与え、回答結果を得点化（4点満点）し、各々の項目の平均点を示したものが、Table II - 2 である。なお、Table II - 3 に、回答をいただいた教員の状況（「教員免許の有無」等）を示しておく。どの教員状況項目も、評価項目得点に関して統計的に有意な程の差異は見られなかった。

調査結果を見ると、(評価の対象・目的とも、「教員による評価」における“教員側要因（項目6-10）”に、

Table II - 2 授業評価・調査項目に対するイメージ

	教員による評価			学生による評価				
	学生 (321名)	教員 (28名)	検定 (t)	学生 (321名)	教員 (28名)	検定 (t)		
評価の対象	1	3.3	3.8	**	1	3.4	3.8	**
	2	2.5	2.6		2	2.8	2.6	
	3	3.2	3.5		3	3.3	3.2	
	4	1.7	1.5		4	1.9	1.6	
	5	3.2	3.2		5	3.2	3	
	6	2.9	3.4	*	6	3.1	3.6	**
	7	3	3.3		7	3.5	3.8	
	8	3.1	3.4		8	3.4	3.6	*
	9	2.9	3		9	3.3	3.3	
	10	2.3	1.7		10	2.4	1.9	**
評価の目的	1	3.5	3.7		1	3.3	3.7	*
	2	3.2	3.5		2	3.1	3.4	*
	3	2.8	3.4	**	3	2.9	3.4	**
	4	3	3.1		4	3	2.9	
	5	3	2.4	**	5	2.8	1.9	**
	6	3.1	3.5	**	6	3.1	3.6	**
	7	2.8	3.1	*	7	2.9	3	
	8	2.9	3.3	**	8	3	3.6	*
	9	2.3	2.7	*	9	2.6	2.9	
	10	2	1.7		10	2.3	2	

** : p<.01 * : p<.05

Table II - 3 回答教員の状況

教員免許		教育系	科目履修		研究分野			教職歴		
有	無	有	無	理系	文系	その他	5年未満	5-10年	10年以上	
19名	9名	20名	8名	9名	17名	2名	5名	10名	13名	

教員の方が学生よりも高得点を与えていることがわかる。教員は自己評価を意識していると言えよう。評価の対象として「教員の教える内容」や目的として「指導方法の参考にするため」が高得点になっていることは、明らかにそのことを支持している。また、「学生による評価」では、(評価の)対象・目的とも、「学生側要因(項目1-5)」に、教員の方が学生よりも高得点を与えていることもわかる。教員は、学生自らが、「自身の考え方や理解度を把握するため」(目的)に、「自身の教授(学習)内容の理解度」を、その対象とすべきだと考えているのである。ここにも、教員が「評価とは自己評価である」との意識を持っていることを、見ることができるのである。

もちろん、教員も、教員による評価の対象や目的に学生側要因に高得点を与え、学生による評価の対象や目的に教員側要因に高得点を与えてはいる。しかし、それと同程度に「自らの活動について評価するのだ」という認識が持たれているということは、注目に値すると言えよう。教員は、「評価」を他者(学生)へのレッテル張りとは決して思っていないのである。それは、教員による評価の目的の項目で、学生は「学生の成績をつけるため」に3点をつけているが、教員は2.4点しかつけていないことに顕著に現われている。教員は、学生の成績をつけることを評価の目的とは捉えていないのである。

大学の教員は、必ずしも教員免許を必要としていない。そのため、「教育学」的トレーニングを受けておらず、独善的だという批判がある。しかし、今回の調査結果からは、そのようなイメージは誤ったものであると言える。免許所有者と非所有者とで、あるいは、教育学関係科目履修者・非履修者とで、何ら統計的な差は見られていない。免許を所持しているか否かにかかわらず、教育学関係科目を履修しているか否かにかかわらず、アンケートに回答をよせた県立新潟女子短期大学・教員は、「評価」について、佐藤の調査で示された小・中学校の教員とほぼ同様な認識を持っているのである。「大学の教員は教育がわからない」という指摘・批判は、少なくとも県立新潟女子短期大学には当てはまらないと言ってよいであろう。

また、学生についても、思っていた程、他者評価一辺倒ではないということもわかった。佐藤は、自由記述でアンケート調査を行っているため、自発的には自己評価的内容が表出しなかったのであろう。しかし、項目を提示して問うた今回の調査では、彼らは一定程度評価について自己評価的理解を示していると、言ってよいのではないだろうか。もちろん、教員ほど自己評価的な意識はない。しかし、学生らは、教員の評価の対象として「学生の学び方」(2.5)よりも「教員の教え方」(3)の方に高い得点を与えているし、目的としても「教員の指導法への参考」について高得点(3.1)を与えている。同様に、学生による評価も、その目的の最高得点(3.3)は「学生自身の考え方や理解度の把握」となっているし、対象を見ると「学生自身の教授(学習)内容の理解度」に3.4の高得点をつけている(1位—「教員の教え方」:3.5)。

学生も、自己評価的な考え方を修得する素地は十分にあると言えるだろう。「評価は他者に対する値踏みではない」ということを、学生は十分に理解できると考えられるのである。もちろん、現状のままでは、このギャップは必ずしも埋められているとは言えない。学生は、教員の評価の目的として「学生の成績づけ」に3点という高得点を与えている(教員は2.4である)ことからわかるように、「評価とは点数づけである」という旧来的な評価イメージを払拭しきってはいない。教員よりも、一段と古い評価観に立っていると考えられる。しかし、今後「評価」について明確な説明を与えていけば、新しい評価観を獲得することは十分できるのではないだろうか。そのような期待は十分伺わせる調査結果である。

教員についても、同様である。教員も、「評価とは自己評価である」という認識が十分なものであったとは、言い難いからである。まず、教員が新しい評価観を獲得し、自己の活動を十分に評価していかなければならな

い。そして、その上で、学生が新しい評価観を獲得し、自己の活動を評価できるよう、支援していかなければならない。なぜならば、学生は、様々な分野で未だ学習途上であり、新しい評価概念を獲得しても、自らの評価を行うことが教員の支援無しには困難であろうからである。教員は、自己の活動を評価するだけでなく、学生が学生自身の活動を正しく評価できるよう支援しなければならないのである。

そのような活動の一環としての「授業評価」は、教員・学生両者にとって意義深いものとなっていくであろう。そうでなければ、旧来の誤った評価観に縛られままでは、ただお互いにレッテル張りをし合うにすぎないものになってしまうであろう。それは、建設的ではなく、全く不毛なものである。そうならないよう、我々は評価に意識的でなければならないし、また、十分にできるということを、今回の調査は示しているのである。

全体的考察

我々は、何のために「評価」を行うのであろうか。それは、言うまでもなく、あくまで自らの授業をよりよくするためである。学生にもっともらしい数値を当てはめることでは、決してない。

前出の永野は、以下のように述べている。

試みに、次のような問題を解決するのに、生徒1人ひとりに適当な数値をあてがう作業がどの程度重要な意味を持っているかを考えてみるとよい。

○平行四辺形の面積の指導はどのようにするのがよいか（指導法の決定）。

○分数の大小の比較を指導するのは何学年が適当か（学年配当の問題）。

○化学を嫌う生徒が多いのだが、指導の内容をどのように改善したらよいか（教育内容の改善の問題）。

○中学2年の終りになっても、英語の簡単な疑問文を作ることができない生徒がいるが、この生徒にはどのような指導をしたらよいだろうか（個人指導に関する決定）。

ここに列挙したような問題を解決するには、成績のつけ方（つまり、個人の学力に数値をあてがう作業）についていくら努力しても、それだけでは十分ではない。教育評価とは生徒に点数をつけることだと考えていたのでは現実の教育問題はかたづかないのである。

また、細谷も以下のように述べる。

教師が、教室で、教材を、子どもたちに教え学ばせる一連の過程と遊離したところで、いかに評価の技法を工夫・開発してみたところで、それは「授業」の評価に用いえないし、用いてはならないという原則こそが、先ず強調されるべきではなからうか。

「授業評価」という名目で現在各大学において実施されている「評価」は、永野や細谷の言う、「教育『評価』」になっているだろうか。このような考えに基づかない評価であれば、それは教育評価ではない。もっと言えば、「測定」にすらなっていないかもしれない。

今回の調査で、学生も教員も古い評価観にある程度しがらんではいるが、新しい（本来の意味での）評価を十分に行い得る素地があるということが、確認されている。その意味では、「評価」を教員も学生も共通に意識しあえる環境は十分に構築できるし、そうして初めて「授業評価」も機能すると言えよう。教員も学生も、各々の目標達成のために、どのような情報を収集し、活用したらよいかを常に考え、かつ実践していかなければならない（もちろん、未熟な学生のために、教員が学生自身の評価を支援するということも、忘れてはな

らない)。

そのような活動、即ち「評価」活動を通してのみ、よりよい授業が創り上げられていくことになるのである。

参考文献

- 荒井龍弥 学生の授業評価に及ぼす参加意識の違い 北海道・東北心理学会 2001
- 佐藤康司 よりよい教育評価を求めて 学校文化分析 第1号 盛岡大学教育方法学研究会 1989
- 細谷 純 「教育評価の使命と方法」 藤永保他編『能力・適性・選抜と評価』 第一法規 1975
- 細谷 純 「プログラミングのための諸条件」 斎賀久敬他編『講座現代の心理学3 学習と環境』 小学館 1983
- 東洋 『子どもの能力と教育評価』 東京大学出版会 1976
- 永野重史 『教育評価論』 第一法規 1984
- リー・クロンバック 「授業の改善のための評価」 ヒース編(東洋監訳)『新カリキュラム』 国土社 1965
- Snow,R.E. *et al.*, Individual Differences and Instructional Film Effects *J.Edu.Psychol.*, pp315-326 56 1965