

補 論

1 学生による授業評価——現状と課題——

松下 佳代

大学の授業評価といえば、今日では、次のようなタイプの〈学生による授業評価〉が一般的である——学期末に実施される質問紙形式のもので、だいたい10～30くらいの質問項目について3～5段階で評定させる。自由記述欄がついているものも少なくない。評価結果は、個々の教員にフィードバックされるほか、報告書やWeb上で公表されることもある——。

ほとんどの大学教員は、一度ならず、こういう形の授業評価を受けたことがあるはずだ。しかし、こうした授業評価を積極的に受けとめている教員はそう多くはないだろう。組織的に取り組むことになったので、仕方なくやっている。実施・集計・報告書作成にかなりの時間をとられるが、何の役に立っているかはっきりしない。授業評価には、こんなふうに、強制感、徒労感がつきまとっている。

授業評価に意味が感じられないとすれば、それはなぜなのだろう。授業評価をもっと意味あるものに変えることは可能だろうか。そのためには何が必要か。以下では、こうした問題について考えてみたい。

(1) 評価の時代

大学の授業評価とは、大学の授業の価値判断をおこなう行為のことである。大学教員のなかには、以前から、毎回の授業後や最終回に、授業に対する感想や意見などを書かせることで評価情報を得ようとする人はいた。そこまでやらなくても、授業中の学生の表情や言動、あるいは試験やレポートの出来具合などの情報にもとづいて、授業がうまくいっているかどうかを判断することはなされてきた（もちろん、学生が授業をどうみているかなど我関せず、という教員もいなくはなかったが）。しかし、大学の授業評価、とりわけ〈学生による授業評価〉が現在のように盛んになったのは、それが組織的・制度的な大学評価の一環として取り組まれるようになってからである。

1991年の大学設置基準改訂、いわゆる「大綱化」以降、この15年足らずの間に、自己点検・評価の実施義務化（1999年）、大学評価・学位授与機構の設立（2000年）、第三者評価の義務化（2002年）、国立大学法人化（2004年）、認証評価制度の実施（2004年）などが立て続けに進められてきた。まさに、「評価の時代」である。その背後には、「事前規制から事後チェックへ」という文部行政の方針変更がある。

文科省の調査によれば、〈学生による授業評価〉を実施している大学は、2002年度現在、約84%にのぼっている。「評価の時代」の中で、この実施率は上がりこそすれ、下がることはないだろう。

(2) 授業評価を考える枠組み

このようなく学生による授業評価〉の広がりの中で、いつのまにか、

大学の授業評価＝＜学生による授業評価＞という図式ができあがってしまっているようにみえる。

しかし、現在の＜学生による授業評価＞は、実は、大学の授業評価の一つのかたちにすぎない。以下では、授業評価を考える枠組みを示しながら、＜学生による授業評価＞を相対化し、その問題点を指摘していこう。

〈1〉 評価の目的

授業評価の第一の目的は、授業についての情報を収集しそれをもとに授業の成否やその原因をさぐることによって、授業の質を「改善」することにある。しかし、＜学生による授業評価＞では、改善よりむしろ、「説明責任」の遂行が重視され、報告書作成のためのデータ収集で終わっている場合も少なくない。

「改善」と「説明責任」は、授業評価に限らず、大学評価全体を貫く二つの異質な理念である。オランダの大学評価の専門家フローインスティンは、この二つを両立させることの困難さを、「オデュッセイア」に出てくる海の怪物スキュラと大渦巻きカリブデイスの間を航行することにたとえている⁽¹⁾。そのくらい両者の両立はむずかしいのだ。たとえば、改善のための評価では、現在のシステムの問題点を見つけだすことが望まれるが、説明責任のための評価では、できるだけ問題点は目立たぬようにし評価結果をよく見せようとする力がはたらくだろう。両者の調停をはかることは果たして可能か、どうすればそれができるのかは、大学評価の大きな課題となっている。

〈2〉 評価の主体

＜学生による授業評価＞では学生が授業評価を行う主体として考えられているが、授業評価の主体になりうるのは学生だけではない。たとえば、同僚教員同士で授業を参観して、授業についての意見や感想を述べることも、授業評価の一つのやり方である⁽²⁾。

学生を大学という教育サービス（産業）における「顧客」とし、＜学生による授業評価＞を一種の「顧客満足度（CS=customer satisfaction）調査」とみなす考え方がある。大学を経営面から考えれば、学生が顧客・消費者であることは紛れもない事実である。だが、教育面から考えたとき、学生が顧客・消費者にとどまることを認めるべきではない。学生は、自ら学習を創り出していく生産者であるべきだからである。

大学が生き残りのために、学生のニーズをさぐり、それにあったカリキュラムや施設・設備を準備し、希望にそった学位・資格を取得させ、学生や社会から評価を受け、それを大学経営にフィードバックする——そういうふうにして、教育サービスを購入する消費者としての学生が、大学教育の主導権を握るようになることを、アメリカの社会学者リースマンは「学生消費者主義」と呼んだ。1980年代初めのことだ⁽³⁾。が、同時に、リースマンは、学生を「受動的な消費者」から「能動的な生産者」へと変えることが、大学教育の重要な課題であることも指摘していた。

日本の大学生たちは、子どもの頃から学校と塾・予備校というダブルスクールに通い、受動的消費者としての態度をしっかりと身につけている。そうした彼らに対して、＜学生による授業評価＞は、受動的な消費者主義的態度をいっそう強めるおそれがある。＜学生による授業評価＞が、単なる顧客満足度調査に終わらないためには、それが学生にとっても、自らの学習評価の機会を与えるようなものでなければならないだろう。

一方、教員の側は、〈学生による授業評価〉が評価情報の一部にすぎないという認識をもつ必要がある。学生や同僚から与えられるさまざまな評価情報を統合して授業評価を行い、それをどう改善に生かすか、というところで主体性を発揮すべきは、あくまで教員自身なのである。

〈3〉 評価の内容

〈学生による授業評価〉では、学生が教員の提供する授業の内容や方法をどう受け止めたかが評価の主な内容になっている。しかし、授業評価の内容は、本来、このような内容に限定されるべきではない。もっと多様な質や成果が評価に含まれるべきだろう。

小学校から高校までの授業評価と大学での授業評価を比較してかねがね不思議に思っていたのは、大学以外の授業評価では、(a)〈学習者が授業でどんな力をつけたか〉が重要な評価内容とされているのに、大学の授業評価では、(b)〈学習者が授業をどうみたか〉に評価内容が偏っていることである。

たとえば、初等・中等教育段階の物理教育から例をひくと、「仮説実験授業」では、授業の目標と成否の評価基準を3点に置いて、授業研究を進めてきた⁽⁴⁾。

①すべての子どもたちがそこで教育目的とされている科学上の概念・原理・法則を理解し使いこなせるようにする。

②クラスのすべての子どもたちが科学とこの授業とが好きになるように、授業を組織する。

[③は、本稿の内容に関係ないので省略]。

①は(a)〈学習者が授業でどんな力をつけたか〉、②は(b)〈学習者が授業をどうみたか〉にあたる。(a)(b)が両方そろって、授業の成否が評価できると考えられている。これは、仮説実験授業に特殊な考え方ではなく、初等・中等教育では広くみられる考え方である。

ところが、大学の授業評価では、〈学生による授業評価〉によって、ほとんど(b)のみが問題とされる。これはとても不思議なことだ。授業のよしあしを評価するには、学生が授業で何を身につけたのかも含めた授業全体の成果が評価されるべきだろう。

フローインスティンのいうように、質は、「単一のもの(the quality)」としてではなく「複数の質(qualities)」として語られなければならない。もちろん、〈学習者が授業でどんな力をつけたか〉と〈学習者が授業をどうみたか〉だけではなお、授業の質を十分掬いとっているとはいえないだろう。しかし、後者のみに偏るよりははるかにましである。

〈4〉 評価の方法

〈学生による授業評価〉では、質問紙を用いて、複数の評価項目に対し3～5段階で答えさせ、それらの平均値などで授業のよさを評価しているところが少なくない。しかし、この方法には、授業形態・内容、教育観の違いに関係なく一律の評価項目が課される、という問題がある。大学によっては、授業形態ごとに異なる評価項目を準備しているところもある。が、その場合も、1回だけの数値による評価では、抽象的すぎて、授業の改善につなげられるような具体的な学生の姿がとらえにくいという不十分さが残る。〈学生による授業評価〉では自由記述欄がいちばん役に立つという大学教員の意見は、そのことを反

補論1

対側から裏付けている。

質問紙で段階評価させるのは、説明責任の効率的な遂行には都合がよい。しかし、授業改善のための情報を得るには、むしろミニッツ・ペーパーのような、毎回の授業でおこなわれる記述式の授業評価の方が有効である。

〈5〉 評価の時期

学期末という評価の時期にもまた問題がある。まず、最後まで残った学生は授業に適応した学生であるため、評価が甘くなるなど、固有のバイアスがかかりやすくなる。それに何よりも、評価をした当の学生たちが評価にもとづく改善を享受できない。こうした問題を防ぐため、ハーバード大学デレック・ボク・センターでは、“early evaluation”と称して、授業開始後4回目あたりに授業評価を実施することを勧めている。これならば、バイアスもかかりにくく、また改善のための時間も取れるからである。

教育評価論では、授業の途中で実施され、授業の改善に役立つ情報を得るための評価を「形成的評価」、授業の最後に実施され、授業の最終的な成果をみるための評価を「総括的評価」と呼んでいる。説明責任の遂行には総括的評価の方が適しているが、改善のためには形成的評価でなければならない。〈学生による授業評価〉がなかなか授業改善につながらないという声をよく聞くが、〈学生による授業評価〉が授業の最後に実施される総括的評価である限り、改善の機能を果たしにくいのはむしろ当然なのである。

(3) 授業評価のオルターナティブ

以上述べてきた現在の〈学生による授業評価〉の問題点をふまえた上で、授業評価において改善と説明責任をどう両立・調停するか、その具体的な提案をおこなってみたい。

〈1〉 二つの評価の併用

まず、興味深い例として福島大学小野原雅夫の試み⁶⁾を紹介しよう。福島大学では2001年度後期から〈学生による授業評価〉を導入した。小野原の授業（「倫理学概論」）の評定値（総合満足度）は、開始当初は5段階評定で3.89だったが、2003年度後期には4.34に上昇している。これだけみると、〈学生による授業評価〉が授業改善につながったようにみえる。

しかし、改善を引き起こした仕組みは、実は、別のところにあった。小野原は、2003年度から、授業の始めに「シンキング・ワークシート」（A4判用紙の表）、終わりに「リフレクション・ワークシート」（A4判用紙の裏）を使うようになった。シンキング・ワークシートは、その時間の授業内容に関連した問い（例：「戦争とは何ですか。戦争を定義してください」）を考えさせ、リフレクション・ワークシートでは、授業を振り返って「今日の自己評価」「疑問や関心」「要望や感想」を書かせる。一般に、〈学生による授業評価〉には必ずといってよいほど理解度を問う項目が含まれているが、その評定値が低かったとしても、どこがわからなかったのかという改善に必要な情報は得られないし、学生に対して補足説明をする機会もない。小野原の場合は、リフレクション・ワークシートを使うようになって初めて、学生がどこを理解していないかがわかるようになり、補足説明もでき

るようになった。それに伴って、〈学生による授業評価〉での理解度の評定値も、期末試験の答案における理解の質も上がったという。

ここからいえるのは、〈学生による授業評価〉は、改善の原因というよりむしろ結果だったということである。〈学生による授業評価〉は、改善のインセンティブにはなるかもしれないが、原因にはなりにくい。改善の原因になったのは、毎回の授業で使われる記述式のワークシートの方であった。このように、この例では、改善を目的とする授業評価と説明責任を目的とする授業評価が併用されることによって、二つの目的の両立が果たされている。

〈2〉〈学生による授業評価〉そのものの改善

(a) early evaluation の機能と教員の主体性

だが、〈学生による授業評価〉自体にももっと検討の余地がある。私が前任校の群馬大学で2001年度後期に経験した「授業評価アンケート」は、改善と説明責任の調停をはかる一つの方法を示したものといえる。それは次のようなやり方でおこなわれた。①第1回は11月下旬、第2回は1月中旬に実施する、②各教員は、自分の担当する授業科目の中から一つを選んで、第1回のアンケートをおこなう、③事前に各教員に配布されたA4判1枚の「アンケート結果の報告書」に、第1回の結果、および改善すべき項目とその方策を記入する、④第2回のアンケートをおこない、「アンケート結果の報告書」に結果がどう改善されたか（されなかったか）を記入する、⑤両方のアンケート結果と「アンケート結果の報告書」を自己評価委員会に提出する。

この「授業評価アンケート」で注目されるのは、以下の点である。

- ・ early evaluation の機能をもっている。
- ・ 特定の回の授業に対する評価なので、評価と授業内容が対応させやすい。
- ・ アンケート結果を教員自身が分析し改善に生かすような道具立てがなされている。
- ・ アンケートを実施するのに適した授業科目を教員自身が選べる

この授業評価は、一般の〈学生による授業評価〉と同じく、自己点検・評価報告書作成の資料として使われる、つまり説明責任の遂行を目的としたものであった。が、ここでは、評価結果のみでなく、評価結果にもとづく改善も、説明責任の対象になっている。また、単に組織的な評価の押しつけでなく、評価・改善における個々の教員の主体性も一定程度確保されている。このようなかたちで、〈学生による授業評価〉において改善と説明責任の調停をはかることも可能なのである。

(b) 教員による授業評価と学生による学習評価

私の所属するセンターでも現在、工学部で実施する「授業アンケート」の原案を作成中である。その特徴をあげよう。

①「授業評価」と呼ばない——「授業評価」という表現を使わなかったのは、学生からの評価情報を受けて授業評価を行うのはあくまで教員自身であるという考え方に立っているからである。一方、学生に対しては彼らの学習評価の機会となることを意図している。質問項目は「自分自身の学習状況」（7項目）、「授業の内容・方法」（15項目）、「授業全体を通して得られた成果」（8項目）に分かれており、これによって、学生自身が、自分の

補論1

学習の中に授業をどう位置づけ、そこから何を得たかを振り返る機会となるよう設計してある。また、今回のアンケートでは、4段階で答える質問項目と自由記述の他に、「この授業を通して、重要であると思った概念・理論・キーワード等を5つあげて下さい」という質問も設けている。これも学習の振り返りを促すだろう。

②記名式にする——アンケートは記名式である。記名式には、回答に責任をもたせる、成績などとの関連がみられるといったメリットがある反面、成績に影響することをおそれて率直に書けなくなるという懸念もある。今回のアンケートでは、「回答用紙を授業クラスごとの回収封筒に直接入れ、最後に提出した学生が封をし、それを授業担当教員が事務に提出する」「教員には学生の特定につながる情報を除いた結果をフィードバックする」という手続きをとることで、その懸念が最小限になるようにした。

③成績と関連させる——記名式で成績との関連づけが可能になったことによって、<学習者が授業をどうみたか>だけでなく、<学習者が授業でどんな力をつけたか>という点からも、授業の質がとらえられるようになる。たとえば、アンケートで「授業は理解できた」と答えているのに、成績がよくなかったり、適切な「キーワード」をあげることができていないという場合、その原因として、学生の側に自分の理解度を把握するメタ認知が欠けている、学生の理解度についての目標が低い、レポートや試験が難しすぎる、といったさまざまな解釈が可能である。どの解釈がより妥当であるかは吟味が必要だが、いずれにしても、<学生による授業評価>だけを頼りにして授業評価をおこなうよりははるかに授業の見かたが広がるだろう。

この授業アンケートの結果は冊子にして、それを素材に、学生もまじえたシンポジウムを開催することを計画している。大学教育の改善は教員だけの責任ではない。教員・学生の両方が教育改善に取り組むためのツールになることを、この授業アンケートはめざしている。

*

*

改善と説明責任をどう両立・調停するか、教員と学生をともに教育の主体にしていくには何が必要か。これは、授業評価に限らず、大学教育全体にかかわるチャレンジングな課題なのである。

注

- 1) フローインスティン, A. I. (2003) 『大学評価ハンドブック』(米澤・福留訳) 玉川大学出版部.
- 2) キーグ, L. ・ワガナー, M. D. (2003) 『大学教員「教育評価」ハンドブック』(高橋靖直訳) 玉川大学出版部.
- 3) リースマン, D. (1986) 『高等教育論』(喜多村・江原他訳) 玉川大学出版部.
- 4) 板倉聖宣 (1974) 『仮説実験授業』仮説社.
- 5) 「大学授業ネットワーク」(<http://www.highedu.kyoto-u.ac.jp/jkp/index.html>) .

付記

本稿は、松下佳代「学生による授業評価——改善と説明責任——」『物理教育学会誌』(印刷中) を一部修正したものである。

2 学習コミュニティ形成に向けての 授業評価の課題*

大塚 雄作

(1) 評価の時代と授業評価の普及

大学評価時代の到来

加速度的に変化しつつある社会の動きに呼応して、大学は今、大きな転換点に立たされている。2004年度から、国立大学が「国立大学法人」として再出発したことは、まさにその一つの象徴的な出来事であろう。そうした大きな大学の変革は、高度情報社会、グローバル化、18歳人口の漸減、学習指導要領の改訂、行政改革等々の社会的な動きの中で、国立大学のみならず、すべての大学にとって避けて通れない重大な課題となってきた。

それらの変革の代表的な制度改革の一つに、大学評価の導入がある。1991年の大学設置基準の「大綱化」と共に、大学は「自己点検・自己評価」が義務づけられ、我が国でも大学自身に「評価」が課せられるようになった。その後、社会へのアカウンタビリティ（説明責任；accountability）が強調され、第三者による本格的な「評価」の実施が要請されるようになり、1998年10月の大学審議会による答申、いわゆる「21世紀答申」において、大学評価に関わる第三者評価機関の設置が提案され、それを受けて、2000年度に大学評価・学位授与機構（略して機構）が誕生した。既に、機構による「大学評価」は、足かけ4年にわたって、試行的段階の評価として、主として国立大学を中心に実施されてきたところである。

さらに、2002年8月の中央教育審議会の答申において、大学設置時の規制緩和を進める一方で、「大学の質の保証」に関して、事後的に、文部科学大臣に認証された第三者評価機関による「認証評価」を、7年以内に1度ずつ、国公私すべての大学に義務付けるという新しい評価制度が提案された。それによれば、認証評価機関はそれぞれ大学評価基準を定め、要請のあった大学に対して、その基準を満たしているかどうかに関する評価をおこなう。一方、基本的には認証評価とは独立に、文部科学大臣は大学に対して法令違反等がある場合には段階的是正措置（改善勧告・変更命令・閉鎖命令の3段階）をとることができるようになった。その際に、認証評価の結果が参考とされることは当然予想されることであり、基準を満たしていないと評価された大学には改善勧告等がなされる可能性もあるだろう。この提案は、その後すぐに法制化され、2004年度からその端緒についている。

その他にも、研究の拠点形成するための21世紀COEプログラム（COE；Center of Excellence。研究の先端的拠点を指す）、特色ある大学教育支援プログラム、現代的教育ニーズ取組支援プログラム等々の、大学の教育・研究を対象とした新しい評価の枠組みも

* 本稿は、『溝上慎一・藤田哲也（編）2005 心理学者、大学教育への挑戦 ナカニシヤ出版』の「第1章（pp.2-37）」として筆者が書き下ろした草稿（校正・修正前）を、了解を得て転載させていただいたものである。一部の記述や図の体裁など、その著書のものとは異なっていることをお断りしておきたい。

導入され、また、国立大学法人に対しては、それとは別に、中期目標・中期計画の達成度についての国立大学法人評価が6年に1度実施されることになり、今まさに、大学評価のブームが到来しているのである（大塚，2002）。

学生による授業評価の普及

こうした大学評価の始まりに軌を一にして、大学授業の改善とアカウンタビリティのための一つ的手段として、「学生による授業評価（以下、単に授業評価）」が、多くの大学で組織的に取り入れられ始めている。日本では、従来、授業評価は、基本的に大学の風土に馴染まないなどと言われ、個人の教員が自主的に実施している例は多く見られても、大学として授業評価を実施するケースは決して多くはなかった。1996年12月に文部省（当時）から出された「進む大学改革 ― 大学の新しいすがた」によれば、1992年においては、学生による授業評価をおこなっている大学は、国立大学9大学、私立大学29大学の計38大学となっている（表1参照）。それも、全学的に実施しているというよりも、

表1 授業評価の大学における実施状況

	国 立	公 立	私 立	計
1992年	9	0	29	38大学
1994	33	6	93	132
1996	64	8	150	222
1998	84	23	227	334
2000	93	36	322	451
2002	97	61	416	574

注：上表は、文部科学省「大学における教育内容等の改革状況について（2004）」などにより、筆者が表にまとめたものである。

学部単位でおこなわれているものを集計した統計であった。それが、10年後の2002年の時点では、国立大学では98%に上る97大学で実施され、公立大学61大学、私立大学416大学、計574大学となっており、全大学の約84%を占める大学で授業評価がおこなわれるようになってきている（文部科学省2004年3月23日発表「大学における教育内容などの改革状況について」より）。このように、授業評価は、今やほとんどの大学で実施されており、日本の高等教育における普遍的かつ基本的な教育活動の一つとして位置づけられる情勢になっている。

しかし、授業評価のここ10年来の急激な普及は、日本の大学の状況に適合した授業評価調査が開発されたからでも、また、大学の授業を改善していくための内発的な動機が高まったからでもない部分がある。多くの大学において、審議会答申や大学評価などで授業評価が取り上げられるに至り、いわばそれらの外圧のかかる中で、半ば受け身的に授業評価が取り入れられてきているのではないかと思われる。現に、大学の独自性や授業の多様性などに十分に配慮することなく、欧米や日本で先行しておこなわれている授業評価調査をほぼコピーしただけのような質問紙調査によっておこなわれていることも決して少なくない。そうなると、授業を改善したり、社会にわかりやすく教育の状況や成果を示すといった本来の授業評価の目的の達成は覚束なくなり、結局、大学評価のための授業評価に陥っていくことにもなりかねない。

授業評価は、言うまでもなく、大学の授業がどうあるべきかという「授業観」、**「授業**

モデル」に依存することになるが、そもそも、大学の实情に応じた授業のあり方について、それぞれの大学で十分な議論がされてきた形跡は多くを見出すことはできない。また、大学教育を改善する目的で、FD (Faculty Development : 大学教員などの教育を中心とした資質開発) などの大学教員のための研修が、授業評価同様、多くの大学で企画・実施されるようになってはいるものの、授業評価の結果を、実際の授業改善にどう活用するかということについては、まだ十分にコンセンサスが得られる知見として体系化されているとは言い難い段階にある。

そこで、本論では、筆者の授業評価の体験に基づいて、授業評価に関するいくつかの課題を浮き彫りにすると共に、その課題解決のための理論的背景として、実践コミュニティ (Wenger, et al., 2002) に関わる心理学的知見を取り上げ、大学の授業、および、その授業改善を一つの学習コミュニティ形成のプロセスと捉え、授業評価を学習コミュニティ形成のためのツールとして捉えることの有効性について論じていくことにしたい。

(2) 授業評価の測定論的課題

評価と測定

「評価」は、点数や評定値を付けることであると思われることが少なくないが、「評価」と「測定」ということは区別して捉えることがまずは肝要である。「測定」は、ある対象に対して、一定の基準や定められた方法によって数値や記号を割り当てることである。一方、「評価」は、「測定」を含むさまざまな情報に基づいて、その対象を価値づけることである。点数や評定値を付けることそのものは、その意味で「評価」というよりもむしろ、一種の「測定」として位置づけることができる。測定値は、ある目的の下では、大きい方が価値が高いかもしれないが、別の目的の下では、低い値の方が価値が高かったり、中程度の値が価値が高かったりもするであろう。活動の改善や方向付けは、まさに、測定値そのものではなく、その価値づけ、すなわち「評価」に基づいておこなわれることになる。

授業評価は、通常、授業の理解度、興味度、満足度などのいくつかの評定項目 (10～25項目程度が多い) を含む質問紙調査の形で実施され、その評定値のクラス平均値をもって、それぞれの評定項目が反映する授業の特徴を表す指標として利用されている。評定の段階数としては、4～5段階のものが多くと思われるが、その個々の評定値は、回答者である学生のそれぞれの項目に関する個人的価値判断に基づくものであって、その意味ではまさにある種の「評価」がなされているとみなすこともできる。心理学などの領域では、このように「評価」と「測定」は必ずしも明確に分離できるわけではない。しかし、教員の側が、クラスの評定平均値をもって、たとえば、学生の理解度の程度とみなすような場合には、それは、授業の理解度に関する測定値とみなすべきであり、その測定値を「よし」とするのか、「問題あり」とするかは、授業の目的などに沿って改めて価値判断すべきことである。

測定としての授業評価

授業評価を導入しようとする際に、必ずといっていいくらいに出てくる反論に、学生の授業評価能力に関するものがある。すなわち、学生の中には、いい加減な姿勢で授業に臨んでいる者も少なくないし、また、専門的な知識も乏しい学生に、授業の良し悪しなど評

価できるわけがないというのである。しかし、その授業に実際に不熱心な学生がどのくらい含まれているのか、専門的な知識をもたないような学生がどう授業を感じているのかといったことを的確に把握できなければ、授業をどのようにおこなうかということは、本来は決められることではない。そのような情報を収集する手段としても、授業評価は利用が可能なのである。

また、そのようなクラスの状態を把握することなしには、測定値としての授業評価結果を適切に価値づけることもできない。たとえば、不熱心な学生がクラスにある割合で含まれていることが授業評価から読み取れるとしたならば、必要に応じてそのような学生に照準を合わせた授業作りが求められる場合もあるだろう。その際に、授業評価の「理解度」に関する評定平均値は、そのような学生が含まれる割合によってかなり影響を受けることが考えられ、むしろ、一定のレベルの値が確保さえされていけばよいということにもなり得る。高い評定平均値になることは、必要以上に授業を易しくしてしまったり、内容を薄めてしまったりということから導かれているかもしれないからである。一方、もしその授業が最先端の研究の知見を共有し、先端的研究者を育てるというところに目的があるとすれば、まさに、少数の熱心で優秀な学生がその授業をどう捉えているのかという部分が大切になるという見方も可能である。その場合には、クラス全体の平均値ではなく、その一部の学生の授業に対する評定値を、授業の方向性を定める「評価」のための「測定値」として利用すべきということも起こり得るのである。

このように、授業に関するさまざまな測定値や観察結果に基づいて、それらの情報に価値づけがなされ、すなわち「評価」がされてはじめて、授業をどのように改善していくかという方向性が導き出せるのである。この考え方からすれば、近年、授業評価と呼ぶ代わりに、「評価」という言葉へのアレルギーということもあるのかとは思われるが、「測定」、ないしは、授業に関する情報収集という部分を強調する意味もあって、「授業調査」、「授業アンケート」などと呼ばれることも増えているようである。このように、授業評価結果を「測定」されたものと捉えることは、大学審議会や大学評価などで取り上げられる授業評価の趣旨にも沿っていることである。そこでは、授業の改善が導かれることが重視されているのであって、授業評価をやることそのものが強要されているわけではない。授業を改善するための手段として有効な方法が他に採用されているのであれば、通常の質問紙方式による授業評価をやらなければならないということはないのである。しかし、授業評価はそのような情報を比較的簡単に得ることができる手段でもあり、また、社会的に授業評価が避けられない情勢にある以上、学生の授業評価能力といったことを云々する前に、その情報のもつ意味を的確に把握しつつ、自らの授業実践にその情報を利用していく姿勢をもたねば損ということにもなるであろう。

評定平均値の信頼性

授業評価によって得られる統計量を「測定値」と捉えるならば、それが適切な測定値と言えるのかということが問題となる。測定値は、常に誤差を含むものであり、まずはそれがどの程度の大きさであるか、言い換えれば、測定値の「信頼性 (reliability)」がどの程度であるかということが問われることになる。特に、授業評価では、ご機嫌取りの反応をする学生もあれば、かなり厳しい評定を常に返してくる学生もいたり、また、授業中居眠りが恒例の学生もまことしやかに回答してくることもあるわけで、個々の評定値

にどれほど信頼が置けるのかということに疑義が出されることもしばしばである。

授業評価の個々の評定値自体は、このように個人差も大きく、授業の特徴を一貫して反映し再現し得るかという意味において、信頼に足るとは言えない部分を確かにもっている。しかし、同一授業において、受講生の構成もほぼ一定であるような場合には、授業評価結果として示されるクラスの評定平均値が比較的安定しているということは、授業評価を何回か経験してみると自ずと感じることのできることでもある。

そこで、評定平均値の誤差をどの程度に見積もればいいのかについて、クラスで得られる個々の評定値を、その項目の評定値に関する母集団から無作為抽出された標本とみなす、ごく単純な統計的モデルによって考えてみることにしよう。このモデルによれば、項目のクラス評定平均値は「標本平均」に相当することになり、その安定性は、「標準誤差 (standard error)」によって表現することができることになる。標準誤差は、標本サイズが大きくなるほど、小さくなるという特徴があるので、クラスの人数が大きくなるほど、評定平均値は安定した結果が得られるということが容易に推測される。50人前後のクラスであれば、4～5段階評定値は標準偏差が1前後の値になることが通常であり、その場合は、評定平均値を中心に、おおよそ±0.2～0.4程度の信頼区間(95%)を想定しておけばよいことになる。したがって、確かに詳細な差を云々できるほどの精度は持ち得ないが、とりあえず、評定平均値にその程度の範囲での安定性が見られるということは、十分に納得できることである。

評定値の統計的モデル化

しかし、授業評価の評定値にはさまざまな要因が関係しており、その評定平均値が何を反映するものであるかは、もう少し精細な評定値のモデル化を必要とすることになる。評定値に影響を及ぼす要因としては、まず、授業評価の主たる対象である教授者・授業の要因を挙げることができる。この要因には、教授者の教授技術や教師としての資質・性格、また、授業については、その内容や方法、利用される教材、カリキュラムにおける位置づけなど、さらに細かい要素を数多く含んでいる。また、学習者側の要因も、学習者のレディネスや学力、動機づけなど、やはり多くの下位要素が含まれる。それ以外にも、授業評価の実施時期や実施方法の要因、回答者のこの種の評定項目への反応傾向(中央の評定に反応しやすいとか極端に反応しやすいなど)、その他の偶然誤差要因などが評定値に影響をもつものとして挙げられよう。これら多くの要因が、ときには、それらが複雑に相互作用しつつ、授業評価の各項目の評定値が決められていくことになる。

その点で、統計的、測定論的なモデルに基づいて、体系的に評定値の検討をおこなった豊田・中村(2004)の研究は注目に値する。それによると、7科目を受講した75名の学生に6項目の授業評価調査を30分おきに3回おこない、教授者(授業)、学生、観点(項目)、順序の4つの要因とそれらの2次までの交互作用項によって評定値のモデル化を試み、分散分析によりそれぞれの要因による評定値の変動の大きさをについて検討している。その結果、取り上げた要因によって、評定値の変動の6割程度の変動を説明することが可能であり、その変動への寄与の大きい要因として、まず、教授者と学生の交互作用項が検出された。続いて、学生の要因、学生と項目の交互作用項、項目の要因などが評定値への影響が大きく、さらに、教授者の要因、教授者と項目の交互作用項がそれらに次いだ。なお、順序の効果はほとんど検出されていない。また、教授者に関わる要因が比較的小さ

補論2

な影響力としてしか検出されなかったのは、扱った科目が7科目という少数であり、授業そのものに大きな違いを確保しきれなかったと考察されている。

いずれにしても、ここで最も大きな影響力として検出された要因が、教授者と学生の相互作用項である点は興味深い。すなわち、ある学生はAという授業との相性がよく、別の学生はBの授業との相性がいいといったことが、評定値の変動を左右しているということである。この結果に基づけば、授業は教師が一人で作って一方向的に学生に与えるものではなく、教師と学生が共に相互作用しつつ作り上げられるものであるという授業観が、むしろ自然であろうことが窺える。その見方に従えば、先に述べた「授業評価能力」ということも、個々の学生のみならず、教師と学生の総体に帰属させるべきものであることが示唆されよう。つまり、授業は、講義形式の授業にせよ、演習方式の授業にせよ、教師と学生の間のある種の相互作用のあり方を無視することができないということであり、その点で、そのような人的つながりを概念化し（その概念を「構成概念」と呼ぶ）、大学の授業にアプローチしてみることは意義なしとはしないであろう。後節において、その構成概念を「学習コミュニティ」という語で表現し、それに対する心理学的な課題を論じてみることにしたい。

(3) 毎回授業評価の実践事例

毎授業ごとの授業評価の試み

授業評価結果を左右する要因は、以上のような統計的な分析や、さまざまな実験的研究によって明らかにしていくことがある程度は可能である。しかし、それを実際に授業改善に結びつけていくためには、評定平均値に影響を及ぼす要因において、個々の授業の文脈や背景の中で、どのような下位要素の影響力が大きいのかを具体的に抽出していくことが必要とされることになる。すなわち、学生の構成の違い、教授者の教育目的の違い、その授業を含むカリキュラムにおける位置づけなど、それぞれの授業には独自の文脈や背景があり、単に評定平均値だけから改善すべき点を具体的に見出すことは困難である。具体的な改善点は、授業のノウハウに関する一般的な知見も参考になるとはいえ、むしろ、自らの授業実践を通して、自らが生み出していく以外ない部分が少なくないのである。

授業評価が、個々の授業の文脈や背景に依存するという点を考慮すると、授業の内容や受講生の層などがほぼ一定に保つことのできる自らの授業の中で、授業の方法の違いがどのように評定平均値に関連していくのかを検討していくことが考えられる。通常、授業評価は、学期末に総括的におこなわれることが多いので、その場合には、学期を通した授業方法の違いが反映されることになるが、より具体的な授業の手法の効果を調べるために、毎授業ごとに簡単な授業評価を実施してみることが考えられるだろう。毎回の授業評価などは、大学全体での実施としては当然実行不可能なことであろうし、必ずしも毎学期継続する必要もないと思うが、一度は試してみる価値のあることと思う。実際に、筆者は自らが担当するいくつかの授業で、毎回の授業ごとに簡単な7項目の評定項目（「あてはまる」「ややあてはまる」「あまりあてはまらない」「あてはまらない」の4段階で4～1点と得点化）を学生に課して、評定値に影響を及ぼす要素を探ってみた。その内、3つの科目を選び、授業ごとの「理解度（わかりやすかった）」、「発見（新たな発見があった）」、「興

味度（興味深かった）、「満足度（総合的に満足できた）」の4項目に関する毎授業ごとのクラス評定平均値の推移（平均値は1～4点の範囲となり、中央の値は2.5点となる）を図1～3に示した。

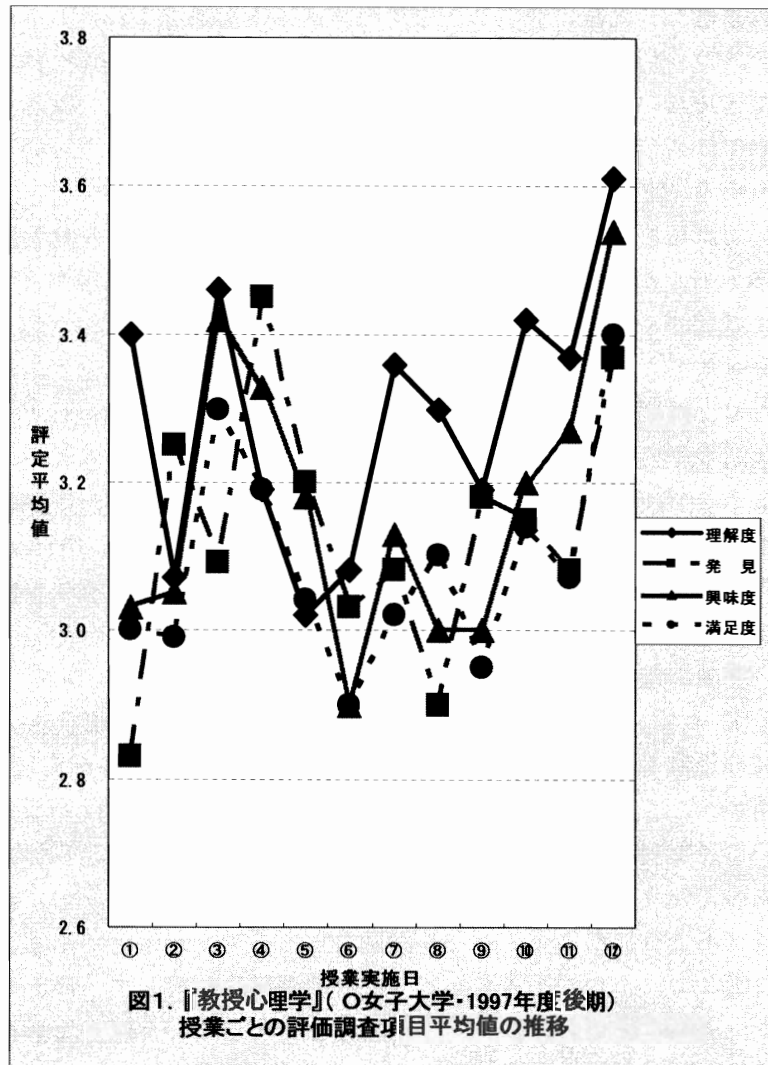
ビデオ視聴の「理解度」への効果

図1は、筆者が、〇女子大で実施した「教授心理学」の評定平均値の授業ごとの推移

である。この図を見ると、③、⑦、⑩、⑫回目の授業の「理解度（—◆—）」の評定平均値が高くなっていることが見て取れる。このように一つの授業科目でも、個々の授業における理解度はかなり大きく変化するのが通常である。この授業では、筆者一人が講義しており、受講生の出入りも大きな変動はなかったことから、評定平均値の変動は、主として授業の方法やその内容に関わる要素によるものであると考えられる。実は、その4回の授業は、授業方法の点で、いずれも、授業中にビデオを10～20分程度、学生に視聴させているという共通点をもっている。すなわち、視覚的に訴える教材を利用することによって、

少なくとも学生が理解しやすかったと感じてくれるということが、この評定平均値の変動から示唆されるであろう。なお、①回目の授業も高い理解度となっているが、これは、授業の概要を提示するなどガイダンス的な授業であって、ある意味でベースライン的な評定平均値と位置づけることができる。2回目以降の授業では、学生にとっては新しい内容が扱われることにもなり、一般的にはベースラインの「理解度」のレベルを維持することはそう簡単なことではないが、視聴覚教材を取り入れることによって、ベースラインに近いレベルにその値が上昇していることが見て取れよう。

「理解度」はあくまで授業の一側面に過ぎない。そこで、別の側面、たとえば、「発見」に関してはどうなのかというように、授業評価を多面的に見ていくことも肝要である。図1



において、③回目の授業でビデオを見せた際に、「理解度」は高くなっているが、「発見」はそれ程高くなっていないのに対して、⑩回目の最後の授業では、「理解度」も「発見」も高い評定平均値となっているという違いが見出される。これは、ビデオの内容が、③回目は比較的基礎的な「記憶」に関する内容のもので、ビデオはわかりやすく質の高いものであったにもかかわらず、受講生の多くにとっては決して新鮮な内容ではなかったということを示していると捉えることができる。その証拠に、自由記述などで「ビデオは耐えられません」と眠気を感じた学生のコメントも散見された。それに対して、最後に見せたビデオは、「オープン・スクール」に関するもので、これはビデオのでき自体はいいとまでは言えないものであったが、壁のない教室という新しい試みへのチャレンジが、学生にとっては新鮮であり、学生のニーズに合った教材であったということが、自由記述などからも窺うことができた。

これらを総合することによって、ただ単にビデオを見ればよいというものではなく、学生のニーズに応じた教材が、教材の質そのものを凌駕する場合もあるということを改めて確認することができたのである。このように、自らの実践と授業評価の結果を照らし合わせることの積み重ねを通して、授業評価の意味するところがより実際に即して明らかになり、授業の改善などにも結びつきやすくなる。研究というディシプリンからすれば、個人的な特殊性や体験はできるだけ排除することが求められ、一般的に共有できる部分を、あるモデルや統計量として表現していくことになるが、実際の授業実践では、むしろ、個々の授業の特殊性が大切になる。むしろ、そのようなモデルからのズレが授業の実践には重要であり、そのようなズレは、実践を通して、その意味を体験的に明らかにする以外にないということである。

学生の授業参加を促す工夫

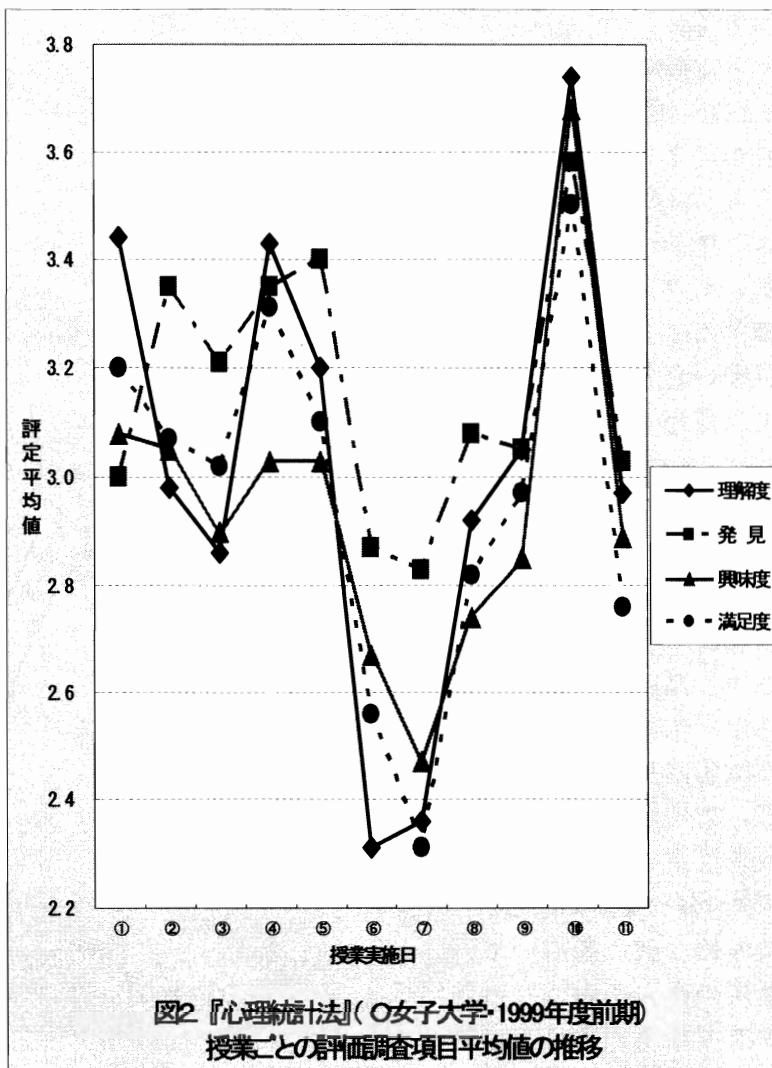
図2は、同じく〇女子大で筆者が担当した「心理統計法」に関する毎授業の評定平均値の推移を示している。これによれば、⑩回目（最後から2回目）の理解度の評定平均値がかなり高くなっているのに対して、⑥、⑦回目にはかなり理解度が下がってしまっている。文科系における統計の授業は、「理解度」の評定平均値が低くなる傾向を宿命的にもっており、分散の概念、標準化の概念、相関係数などと、新しい統計的概念が次々に出てくるようになると、かなり混乱してくる学生が増えてくる。

そこで、一つの工夫として、統計の講義に並行して、班を構成して、質問紙調査を作成し、実際にデータを集めて、それを分析するという調査実習を課題としてみたのである。この⑩回目の評定平均値が高くなっている授業では、班ごとに、作成した調査についての合評会をおこなっている。学生自身が作成した質問紙調査を説明し、それに対して、筆者もいろいろコメントはするが、学生同士の意見交換や質疑も奨励するといった、いわゆる「学生参加型」の授業形態となっている。筆者自身は、学生の発表や質疑応答は必ずしも体系的かつ効率的におこなわれたという感覚は得られずに、この種の形態の授業の進め方の難しさを感じたのであるが、少なくとも、学生にとっては、わかりやすく感じられた授業になったということが窺える。

これは、その授業で、数式などがほとんど出てこなかったということもあるのかとは思いますが、学生の自由記述から、自分自身の友人が、自分と同じような課題を抱えていて、それに対しての教師のコメントが、自分の課題にとっても有益なものとなっていると、ある

種の親近感をもって感じられたということも大きかったのではないと思われる。教師が一方的に説明するだけでは、「先生の言うことは私にはわからないだろう」といったような距離感が、「理解度」をますます低めてしまうということがあり得るのであるが、学生の説明や、それへの教師のコメントは、「友人がわかるんなら自分もわかる」といった「自己効力感 (self-efficacy)」の源泉にもなっている可能性がある。そして、それを裏付けるのが、授業時間外に班ごとにおこなう作業であって、それが個々の学生自らの学習を引き出し、その学習があつてこそ、自己効力感がそれなりの実感となつて感じられるということになる部分もあるだろう。

大学の授業は、「単位制」の制度の下に実施されているが、そこでは、講義時間に等しい予習・復習をおこなうことが前提となっている。しかし、一方的な講義に終始すると、「予習や復習の大切さを痛感した」などという授業評価の自由記述が散見されるなど、授業外の学習が結局十分におこなわれないうまま終わってしまうことも少なくない。班単位の実習などでは、友人同士、お互いに役割を分担することなどを通して、授業外に学習をおこなわざるを得ない状況を作り出し、また、相互に学習結果を評価し合うことで、学習の定着にも結びつくといった好循環を引き出す効果があるということが、図2に



示した授業評価結果から示唆された一つのポイントとなっている。

教授側の共同体制の効果

しかし、この種の実習を、一つの講義形式の授業のなかだけで十分に組み込めるかという、決してそうではない。効率的に、内容を学生に伝えるという意味では、やはり、一方的でも、講義形式の授業が有用であつて、実習はその意味では非効率的である。授業の形態は、それぞれ長所・短所があるわけで、一つの授業のなかで、さまざまな授業形態の特長をすべて盛り込むということは簡単にできることではない。

そこで、効果的な授業改善を試みるのであれば、学科なり、講座なりのレベルで、カリキュラムの構成からの工夫が望まれることになる。たとえば、統計の授業に並行して実習や演習の時間を設けるとか、統計の授業を2学期連続であったものを、もう1学期延ばして、3学期連続の講義にするとか、あるいは、ティーチング・アシスタントを付けて、課外の課題のチューターを担当してもらうようにするとか、一つの授業の枠組みを超えた試みを講じていくことがより生産的であろうということである。

図3は、やはり、〇女子大学の別の学期の統計の授業の評定平均値の推移を示したものである。この授業では、前年度までの授業構成を、2学期連続でおこなっていた統計の授業を、3学期連続の授業に変更したときのものである。その際に、他の統計の授業を担当している友人と、授業を分担することとし、彼自身が書いたやや高度な教科書を採用して、筆者が最初の2/3の部分を担当し、彼が後半の1/3を担当することにした。

図3に示した学期は推測統計に関する講義が中心で、高い「理解度」の評定平均値を示しているのは④回目の授業であるが、この回では、統計的仮説検定に関して、裁判アナロジーを用いるなど、具体的な例に基づいて、検定の導入をおこなった授業であり、文科系の学生が嫌う数式の展開がなかったのと、その具体例が比較的わかりやすかったことが、評定

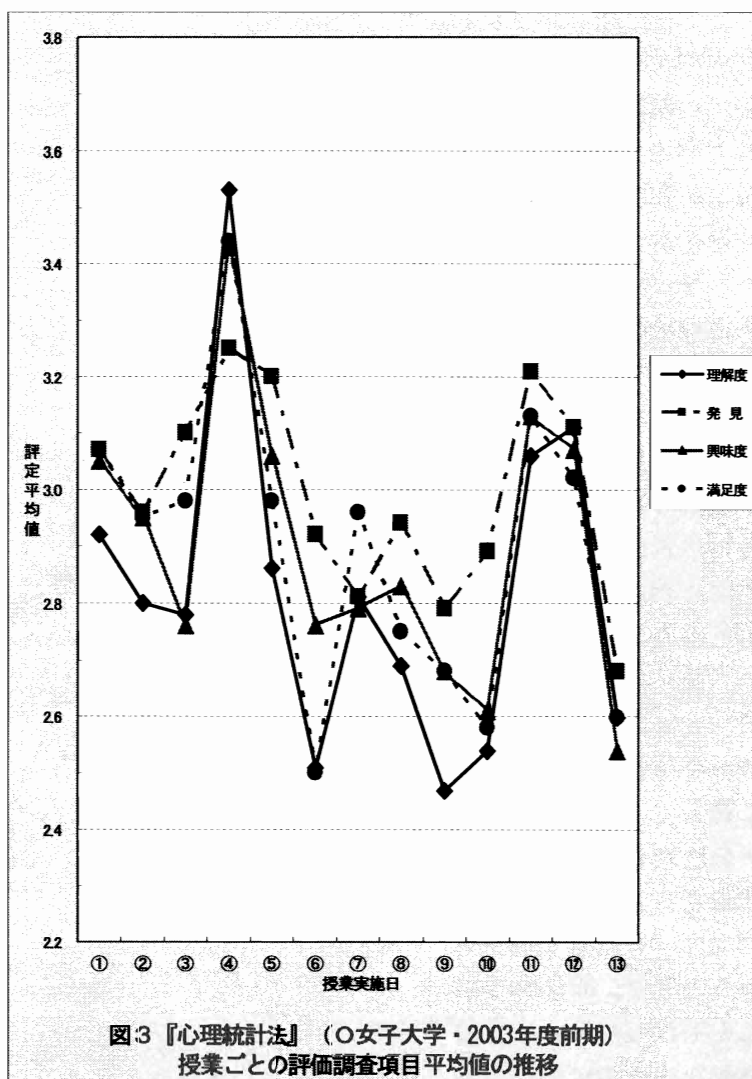


図3 『心理統計法』 (〇女子大学・2003年度前期) 授業ごとの評価調査項目平均値の推移

平均値の上昇に一定の寄与を及ぼしたのではないかと思う。

また、授業の後半の⑩、⑫回目の2回にわたり、その前後の回に比べて、比較的わかりやすく感じられている授業があったことも窺われる。実は、この評定平均値の上昇が、筆者にとっては、一つのインパクトとなった。このときには、統計量のベクトル表現についての解説をおこなっている。統計量のベクトル表現については、混乱を招く可能性もあると思ひ、それまでは文科系の学部 of 学生を対処とする場合にはまったく導入した経験がなかったが、採用した教科書に含まれていたこともあって、この学期にはじめて試みてみ

たものである。文科系の学生に伝わるかどうか、半信半疑で解説していたが、図 3 のように評定平均値は一定のレベルを維持することができた上に、「今まで学習してきた統計量が、思いもよらないベクトルによって表現できるということに感激した」といったコメントを授業後の感想に記載していた学生が何人か含まれていた。このような知的感動を得ることは、勉学への意欲を大きく増進させるのに有効であることは言うまでもないことである。ただし、③回目のレベルまでに到達していないのは、ベクトルに馴れていない学生も含まれており、ベクトルというだけで敬遠してしまうといった感想も自由記述欄に吐露する学生も存在したからである。いずれにしても、このような授業に関する「発見」を筆者自ら体験できたのは、与えられた一つの授業を開いて、友人との分担、言ってみれば、ティーム・ティーチングへと授業を発展させることができたからに他ならない。

(4) 授業評価から授業改善へ

評定平均値の妥当性

このように、授業後との授業評価を積み重ねていると、たとえば、「理解度」の評定平均値を比較的高くするために、自らの授業において有効であろういくつかの具体的な方法が浮き彫りにされてくるということがある。また、同じレベルの平均値が得られたとしても、「理解度」の意味するところは、科目によってはもちろん、個々の授業の内容や文脈によっても、かなり異なっているということも体験できるであろう。ただし、そこから授業改善にはこうすればいいというような一般的に通用するような授業の「要素」を単純に取り上げるということは容易にはできることではない。むしろ、それぞれの授業の文脈のなかで、どんな方法を授業に取り入れてみると有効であるのか、それが授業評価の評定平均値にどんな形で表われているのかということ、実践のなかで少しずつ体得していくことができるということである。すなわち、「理解度」の指標の一つとしても、こうした実践のなかで、その指標にどんな要素が寄与しているのかは、試行錯誤的にまた体験的に明らかにされていくものである。また、そのように自らの実践を通して自ら生み出された授業に関する知見が得られてこそ、授業評価が授業改善につながっていくということにもなるであろう。

評定平均値がどのような意味をもっているかということは、いわゆる「妥当性(Validity)」の問題と捉えることができる。妥当性は、往々にして測定値に固有の属性と捉えられる向きがあるが、同じ測定値を利用するにしても、目的の所在がどこにあるかによって「測定」したいものが異なることになり、妥当性自体も違ってくるということに留意しておく必要がある。

授業評価に関しても、近年は、個々の授業の改善のためだけではなく、大学全体で組織的に授業評価をおこなうことを通して、大学教育の質保証やアカウンタビリティのためにも利用されつつあり、その目的が広がってきている。たとえば、授業が「全体的」に一定レベルの「質」を維持しているかどうかを明らかにするために、「理解度」の評定平均値を利用する場合、毎回の授業評価の評定平均値は、個々の授業内容の要因の寄与が大きく、先に見たように各回ごとに大きく変動してしまうので、その目的に適した指標とは言い難い。その場合は、総括的におこなわれる授業評価の「理解度」に関する結果の方が、個々

の授業内容の変動が相殺されて、「全体的」な質の「一側面」をより反映してくれることになるだろう。

また、「質」ということを考えた場合、「理解度」はあくまでその「一側面」であって、どのような指標が「質」をよりよく反映するのかということも問題になるだろう。わかりやすい授業であっても、授業にまったく興味がわかないようであれば、すなわち、「理解度」は高くても「興味度」が低いような場合には、次の学習には結び付きにくいということもあるだろう。一方、授業の内容に関心が高ければ、難しさが感じられたとしても、「質」が低いとは言えないことになる。そういう意味で、「満足度」はさまざまな要因に関係しており、総合的な「質」を反映させるためにはよりよい指標と言えるかもしれない。授業評価項目に関する因子分析をおこなってみると、「満足度」はしばしば単純構造を満たさずに、いくつかの因子に高い因子負荷量を示すことになる。たとえば、「理解度」に関する因子と、「興味度」に関する因子が抽出されたとする、「満足度」はその両方の因子に関係が深いという場合が多いということである。「満足度」の項目は、その他のさまざまな授業評価項目がそれぞれ「満足度」にどのような寄与を及ぼしているかを検討するための従属変数的な位置づけができることもあって、多くの授業評価に定番的に含まれる項目の一つになっている。

授業実践に基づく実践的妥当化

もっとも、「理解度」や「興味度」をはじめ、授業評価の項目の多くは、それぞれかなり相関が高く、因子分析でも一因子性が強い結果が得られることが少なくない。図1～図3でも、取り上げた各項目の推移がほぼ同じような形になっているが、これもそれらの項目の関連性が高いことの現われと捉えることができる。

その意味では、「理解度」の指標で「質」の指標に代えることもまったく否定されることではない。その際には、「理解度」に関して、一定の水準が保たれていることを示すことが求められる。特に、文科系の統計関係の授業では、授業評価結果に待つことなく、「難しい」という定評があり、「理解度」に関する評定平均値は低い値となりがちである。それが高じると、筆者自身が、授業の最後におこなった総括的な授業評価の自由記述に書かれたことであるが、「統計の授業でわかったことは統計が難しいことだけだった」などということになりかねない。そのようなことがクラス全体に及ぶことになれば、授業の質を保証するとはとても言えないことにもなる。そこで、質を保証するための「理解度」評定平均値はどの程度であるかという目安、すなわち、測定値として得られる数値にどういう意味があるかを、ある程度は把握しておく必要がある。しかし、この水準も、一般的に決められることではない。学生の構成にも依るし、授業の内容にも依るわけで、これも実践の中で体得していくべきことであろう。

統計の授業の場合は、筆者の経験では、個々の授業において、4段階評定のクラス平均値が「3」を割り込むと黄信号であり、「2.5」を割り込むと何を質問していいのかもわからないほど重篤な不理解にクラスの大半が陥っているといった印象がある。質の保証という意味では、理解できるときもあれば、理解不能に陥ることもある中で、総合的には「2.8」程度の評定平均値を保持する必要があるのではと考えている。そして、このような基準は、心理学に関する授業では別であって、「3.0」レベルの「理解度」が望ましいと考えている。なお、このような基準は、あくまで、筆者の個人的な実践とその文脈や背景の中から導か

れたものであって、決して一般的なものではないことに注意されたい。

測定値の妥当性について検討すること、あるいは、測定値が何を反映しているかを明確にすることを「妥当化 (validation)」と呼んでいる。授業評価に関する妥当化は、以上に見てきたように、評価の目的や個々の授業の特徴・文脈などによって異なってくるわけであり、そういう意味で、授業評価研究者などの他者に任せられるものではない。まずは、自らの授業の中でそれぞれの評定平均値を意味づけていくことが重要ということになる。そこで、ここでは、自らの実践のなかで評定平均値を意味づけていくプロセスを「実践的妥当化」と呼んでおくことにしたい。

このように、自らの実践と授業評価の結果を照らし合わせることの積み重ねを通して、授業評価の意味するところがより実際に即して明らかになり、授業の改善などにも結びつきやすくなる。研究というディシプリンからすれば、個人的な特殊性や体験はできるだけ排除することが求められ、一般的に共有できる部分を、あるモデルや統計量として表現していくことになるが、実際の授業実践では、むしろ、個々の授業の特殊性が大切になるということである。すなわち、モデルからのズレが事業実践には重要であり、そのようなズレに関しては、実践を通して、その意味を体験的に明らかにする実践的妥当化のプロセスが必要になるということである。そして、その実践的妥当化の結果を、個々の授業の文脈の中で、明示的に表現していくこと自体、アカウンタビリティといった点から、今後ますます重要となっていくであろう。

授業評価の活用と FD 共同体

授業評価結果を的確に把握したとしても、実は、授業評価の結果に基づいて授業を変えていくことは、存外にたやすいことではない。

たとえば、文科系の統計の授業が、評価結果から難しいと示された場合、それに基づいて、どのように授業改善を試みていくかを考えてみると、まず思いつくことは、授業で扱う内容を減してみるということであろう。実際に、筆者自身の担当してきた統計の授業においても、当初の内容に比べて、年々、かなり内容を絞り込んでいった。しかし、その結果、「理解度」に関する評定平均値が高くなったかと言えば、少なくとも顕著にわかりやすくなったという評価結果を得ることはできておらず、せいぜい横這い程度の印象のものであった。「理解度」の評定値に影響を及ぼす要因は、先に見たようにいろいろあるものであって、少なくとも、内容をある程度減すだけではそれ程大きな効果をもたらさないというのが、筆者自身の経験から導かれた一つの結論でもある。

そこで次に考えられることが、図 2 で見たように、調査実習を学生に課すなど、学生自身がデータに触れる機会を導入するということであるが、この種の実習や演習は、一つの授業のなかに組み込めるかということ、決してそうではない。効率的に、内容を学生に伝えるという意味では、やはり、一方向的でも、講義形式の授業が有用であって、実習はその意味では非効率的である。授業の形態は、それぞれ長所・短所があるわけで、一つの授業のなかだけで、さまざまな授業形態の特長をすべて盛り込むということは簡単にできることではない。

そこで、授業改善を志すときには、より大きな視点、すなわち、カリキュラム全体の枠組みの中で工夫していくことが肝要ということになる。授業改善は一つの授業のなかだけでおこなえばいいということではなく、一つの授業で狙っている教育目的をより十全に達

補論2

成するためには、他の授業や教師との連携をいかに図るかという視点が必要不可欠になるのである。そのささやかな取組として、図 3 に示した実践事例を位置づけることができるであろう。

その場合、学部なり、学科なり、コースなりの組織における授業改善に向けての取組が問われることになる。大学では、往々にして、教員は自らの研究に関わる内容をそれぞれ勝手に講義していればよいというような風潮もある。しかし、そういった個別性、閉鎖性を超えて、教員同士が如何にお互いを補い合って、ある組織ごとに追求している教育目標の達成に向けて、授業をどう組み合わせればよいかという検討が必要になるということである。そのためには、そういう組織単位で、共通の目標を共有し、その達成に向けて組織の成員が協力していく体制が築かれる必要がある。

ただ、カリキュラムという視点からの授業改善に必要とされる協力体制は、既存の組織で上から与えられるよりも、むしろ、「FD 共同体」とでも呼べる「コミュニティ」が、その成員によって自発的に形成されることが望まれるであろう。もちろん、現実的には、自発性に待つだけでは、何も起こらないということがあるので、たとえば、授業評価に基づいた授業検討会を企画するとか、あるいは、Web などを利用した学習材を共同開発するプロジェクトに引き込むといった仕掛けを組み込むことも考えられる。コミュニティをはじめから意識することなく、それぞれ役割を分担しつつ共同作業を進めていく中で、自然にコミュニティが形成されるような仕掛けをトップダウン的に組み込むことも、少なくとも初期には必要とされることであろう。

いずれにしても、そのようなコミュニティが形成されていく際に、授業評価は、その FD 共同体における共通の言語として機能し得るのである。たとえば、「理解度」に関する項目の評定平均値が「4 段階評定で 3 点を下回る」という結果を共有するところから、「この授業は何か手を入れる必要があるだろう」とか、その一つの方策として、「実習を並行して取り入れよう」とか、そうした議論やカリキュラムの改訂に結びつけていくことができるというわけである。

授業評価結果の公表とコミュニティ

授業評価結果は公表すべきかどうかということが時々議論されることがある。公表しないと、結果がどうであれ、教員自身の自覚がなければ、そのまま埋もれてしまうということも当然起こるのである。しかし、公表すると、一種のアカウンタビリティは達成できるということはあるが、評定平均値の数値が一人歩きして、不当な管理に利用されたりといったことになりかねないという心配もある。そこで、その中庸の立場として、評価結果を共有する範囲を、FD 共同体といったコミュニティにとりあえずは制限するということが考えられる。そのようなコミュニティが形成されれば、個々の授業の特殊性のなかで意味をもつ統計量が、その文脈や背景をも含めた生きた情報として、コミュニティの範囲に共有され得るのである。そして、「統計」などの場合には、実習を組入れたことによって、「理解度」の評定平均値が「3」点を超えるかどうかということが、たとえば、その共同体における重要な「数値目標」として機能する場合もあるだろう。

ただし、この種の数値目標は、そのコミュニティのなかでのみ有意義に機能するのであって、コミュニティが異なればまったく意味をもたないということが通常である。それを、科目の特殊性も配慮することなく、教員の管理に利用することなどは論外であるし、また、

語学などの他の種類の授業や、同じ統計の授業でも異なる目的があるならば、別の基準が必要とされることにもなるであろう。そのように背景や文脈が異なる領域間では、すなわち、異なるコミュニティの間では、指標そのものの情報交換よりも、どういう状況において、どういう授業方法が効果的であったか、どんな状況でどういう工夫が役に立ったのかといった、文脈や背景をふまえた質的な情報の交換や、その状況を表わすストーリーの提示が、とりわけ授業の改善に向けて、より示唆に富むことになるであろう。

このように、授業改善には大学内にある種のコミュニティ形成が必要とされ、そのコミュニティのあり様によって授業評価の機能のあり方も違ってくる。そこに、「コミュニティ」と「評価」との関連性に関わる心理学的課題が浮き彫りにされてくることになるのである。

(5) 新しい学習観と評価観の共有に向けて

学習コミュニティ形成のための授業作りのチャレンジ

「コミュニティ」という言葉は、レイヴとウエンガー (Lave & Wenger, 1991) のいうところの「実践コミュニティ (community of practice)」に相当する語と考えるとよい。実践コミュニティについては、その定義及び形成過程などについて、ウエンガーら (Wenger, et al., 2002) に詳しくまとめられているが、ここでは、実践コミュニティは、「あるテーマに関する関心や問題、熱意などを共有し、その分野の知識や技能を、持続的な相互交流を通じて深めていく人々の集団」と定義されている。実践コミュニティは、その大きさや形成過程において、さまざまな形態のものがあり、また、一人がいくつかのコミュニティに重複して属したり、また、コミュニティとは別の組織やグループに同時に属しているということもあり得る。そして、コミュニティに属する人々は、コアになる人、アクティブに実践する人、また、周辺メンバーといったさまざまな立場の人を含んでいる。いずれも、コミュニティの価値を共有し、その価値の実現に向けて活動が積み重ねられていくことになる。

その例として、インターネットの電子会議室上に形成されたある種の実践コミュニティの形成過程を分析した山地 (2001) の研究は興味深い。ここでは、メディア教育開発センターによって提供された SCS (Space Collaboration System: 衛星通信を利用したテレビ会議システム) 研修講座「医学・看護学における学生参加型授業の工夫」の 3 人の研修講師を主要参加者として、その講座に並行して、また、その講座に引き続き、インターネットの電子会議室上で自由なディスカッションがおこなわれた。その電子会議室では、当初は、3 人の講師とコーディネーター、SCS 研修での討論に積極的に参加していた数人のメンバーが主要参加者となって議論が展開されていくことになる。その主要参加者はそれぞれの大学で講座なり同志のグループなりの「ローカルコミュニティ」をもっており、議論のプロセスの中で、徐々にその「ローカルコミュニティ」に含まれる人々が電子会議室で発言し始める。このような参加者が発言をし始めると、自ずと電子会議室は活発化し、その参加者によるある種の「パブリックコミュニティ」が芽生えていくことになる。そうすると、「ROM (read only member の略で電子会議室を読むだけの存在)」と呼ばれる周辺参加者も、次第に、会議室への発言を試みるようになり、「パブリックコミュニティ」

全体の知識の一部を構成する役割を果たしていくといったことが起こり始める。こうして、一つの実践コミュニティが形成され、そして、そのコミュニティの周辺参加者が、アクティヴになり、コアメンバーというように成長していくということも見られるようになる。

実践コミュニティが対象とする領域を学習の場とするとき、それは学習コミュニティと呼ぶことができる。学びということに焦点を当てる際に、その背景となる心理学的理論として、レイヴとウエンガー（1991）の正統的周辺参加論（*legitimate peripheral participation*）を取り上げることができる。正統的周辺参加とは、あるコミュニティに周辺的に参加しているその成員が、コミュニティの中心的人物を目標に、他の成員との相互作用を通して、コミュニティにおける確固たる役割を確保していくという参加の様態を指すが、その全人的な成長の過程そのものを「学び」と捉える学習観を提供するのが正統的周辺参加論の特徴となっている。すなわち、ある学びが成立するためには、そこに一定の学習コミュニティが形成されることが前提となり、そのコミュニティにおける自らの役割を明確にしていくプロセスを学びと捉えるのである。

学習コミュニティについては、学校教育において、授業を学びの共同体と捉えて、教育の再構築を試みる提案がされてきたが（たとえば、佐伯・汐見・佐藤，1992）、この考え方は、学ぶ意義が多様化する高等教育において、むしろ有効に適用できる余地があると思われる。受験という目標もなくなり、大学の教員は自分の専門に関わるテーマをマイペースで講義するということが少なくなく、学生にとっては、何でこの授業をとらねばならないのかという、その原点から揺らいでいる者も少なくない。そういう学生にとっては、教員側のもつコミュニティの境界線を如何に踏み越えて、少なくとも、その周辺参加者として加わっていくことができるのかということが、授業の最初の課題ということになるだろう。その授業が、自分の興味に合っていることが認識されたり、また、将来通らねばならない卒業論文などの比較的大きな課題への関連性が認識されたりすると、とりあえず、その学生は、授業に、言い換えれば、授業の学習コミュニティの周辺参加者として加わることができたと言えよう。

そんな学生が、授業と関わり合い、勉強を積み重ねる中で、少しずつ、学習コミュニティに自らがしっかりと位置付いている感覚が得られるようになる。たとえば、その授業に関連する領域の大学院に進みたいと考えたり、あるいは、その領域の課題を取り上げて卒論をまとめようなどと思ったり、また、そのクラスの他の学生とそのテーマに関して話合ったり、教え合い、学び合うということが起こり始めたりということが生じてくる。そのような学習コミュニティへの帰属がある形で起こり、それへの関与が高まること自体が「学び」というわけである。

そのような感覚が得られやすい一つの授業形態として、最近では、「参加型授業」と呼ばれる学生たち自身が班などを構成することによって、お互いに学び合い教え合う機会を導入するような形式の授業が多く大学の取り入れられ始めている。ただ、講義形式の授業であっても、学習コミュニティは形成し得るのであって、この学習観に基づけば、教師の役割は、まさに、学習コミュニティを授業の場に作り上げ、そのコミュニティの場に多くの学生を引き込んでくることということになる。

授業評価による学習コミュニティの表現の可能性

このように考えてくると、授業評価は、学習コミュニティ形成のツールとして、教師と

学生、あるいは、学生同士のコミュニケーションを確保するために利用すべきであるという見方が生まれてくる。そこで、筆者は、授業評価を総括的に最後の授業だけでおこなうのではなく、図 1 ~ 3 などに示したように、授業のプロセスのなかで実施し、その評定平均値や自由記述に寄せられた学生の意見を、クラスで共有するという試みにチャレンジした。一部の学生のコメントには、筆者のコメントも付して、ワープロ入力し、次週にプリントとして配布したのである。毎週、その入力作業に数時間はとられることにもなり、大変でないといえは嘘になるが、この試みは最後の総括的な授業評価においても比較的高い評価（たとえば、2004 年度の心理統計の授業における約 50 名による最終授業評価では、「前回の授業の他の人のコメントは刺激になった」という項目は 4 段階評定平均値 $m = 3.24$ 、標準偏差 $s = 0.73$ 、「自分の授業のコメントは授業をふり返るのに役に立つ」は平均値 $m = 3.24$ 、 $s = 0.76$ ）を学生から受けており、その労力を投じた甲斐はあったと思っている。特に、日本の学生は質問やディスカッションが苦手というか、積極的にはしないという傾向も見られるので、毎回、自由記述のコメントを次回に学生にフィードバックする試みは、日本の大学のそういう風土にも合致しているのではないと思われる。また、このような授業評価の機会が、学生自身の学習にとっても、その授業をある程度ふり返る場にもなっており、その意味では、学習の機会としても捉えられるのである。また、フィードバックされるプリント自体は、前回の授業の学生の疑問への回答をコメントで付したりもしているので、単なる授業評価ではなくて、学生にとっての学習材としての機能ももつことになるのである。

また、統計の授業では、筆者が一方向的に講義する形式で 1 学期間を通したが、その間、レポートを 2 回課し、その際、友人と一緒に取り組むこと、書いたレポートは友人と相互評価してみることなどを奨励した。このような学習コミュニティを意識した試みをいくつか授業に取り込むことで、「授業を通じて友人とのつながりが広がったり深まったりした」という項目は評定平均値 $m = 3.02$ ($s = 0.92$) というまずまずのレベルの値が得られた。ただし、標準偏差が若干大きいことから、友人とのつながりが深まった学生とそうでない学生の差も大きかったことが窺われる。さらに、「教師と学生が共同して授業を作り上げているという感覚がもてた」という項目は、筆者自身、ほとんど達成感はなかった授業であったが、評定平均値 $m = 2.71$ ($s = 0.75$) という結果が得られた。この平均値は、決して高い値ではないが、統計という内容と、一方向的な講義に終始した授業としては、2.5 を割り込むくらいの平均値を予想していたが、毎回学生がコメントを通じて授業づくりに参加した感覚も得られた部分もあったのか、それなりの値が維持できたのではないと思う。授業評価では、このように、自分自身の授業の特徴を表現し得る項目を含めて、それをしばらくの間定点観測的に利用してみるという姿勢も大切になる。筆者自身の授業観と、それに基づく授業の特徴をどのように表現していけるかということを追究していくために、今後、この種の学習コミュニティ形成を意識した項目の授業評価における有効性について、さらに検討していきたいと思っている。

実践コミュニティの心理学的表現

大学授業や FD などの大学教育の諸課題に対して、学習コミュニティ、FD 共同体といった、実践コミュニティの視点からの接近を試みることは、大学教育の領域のみならず、心理学の領域においても、非常にチャレンジングな問題提起をいくつか提供してくれてい

る。

まず、「実践コミュニティ」自体、目に見えるものではなく、そのコミュニティをどのように記述することができるのかという点について、まず、心理学的に明らかにしておく必要があるだろう。授業という枠組み、学科という枠組みなどが与えられてはいるが、通常は、それがそのまま「コミュニティ」となっているかどうかは保証の限りでなく、それはある場合には自然発生的に形成されたり、時にはいつの間にか衰退したりもするものである。考えてみれば、クラスとか学科といった枠組みであっても、それは実際に目に見える形のものではなく、ある時間にある教室に同じ人が集まって教授学習活動をおこなっているというところから「クラス」なるものが定義され、あたかも目に見えるかのように認識されるということもあるだろう。これは、輪郭をもった人間も同様であって、確かに、人間の体は目に見えるけれども、心理学で対象にしている性格や知能や学力といった構成概念は、決して目に見えるものではなく、ある心理学的理論に基づいて定義されたり、あるいは、それを測定するテストなどによって操作的に決められていたりするものである。実践コミュニティについては、その種の構成概念としての概略の定義は先に見たようにウェンガーら (Wenger et al., 2002) によって与えられてきているのであるが、その形成過程や促進要因などを分析するためには、もう少し、その「輪郭」を明らかにできるような心理学的表現方法を共有していく必要があるだろう。今の時点で、その回答を明らかに示すことはできないが、「人間」は、人之間と書かれるように、一人で生きているわけではなく、人同士のつながりの中で常に生きていることを考えると、今後の心理学の対象は、一個の人間の心理過程から、まさに、コミュニティの力動的プロセスの記述とモデル化にますます傾倒していくであろうことを、心理学を志す者は十分射程に入れておく必要があると思われる。

コミュニティ内における評価の真正性

次に、コミュニティの形成や成長、衰退に関して、どのような要因が影響を及ぼしているかについて検討する課題が浮上してくることになるだろう。その際の一つの視点として考えられるのは、コミュニティの内外における情報の流れとその役割を明らかにしていくことであろう。コミュニティには複数のメンバーが含まれることから、そのコミュニティ内のメンバー間の情報の流れにどのような特徴があるのか、また、コミュニティの枠内とその枠外との間に起こる情報の流れのあり方はどのようなものとなるのか、といった課題である。この課題、まさに、「評価」のあり方が関与することになる。

まず、コミュニティ内においては、ある「場」において、メンバーは大なり小なり共通の価値の追求に関わる実践活動をおこなっていくことになる。言い換えれば、コミュニティは、ある実践活動の背景や文脈を定める役割を果たしているとみなすことができる。したがって、あるコミュニティ内においてそのメンバー同志が発信したり受け取ったりする情報は、その文脈の中に自ずと位置づけられて解釈することができる。このように、ある文脈に即した評価情報がフィードバックされるような評価のことを、真正性 (authenticity) のある評価と呼ぶことができるだろう。真正性が高い評価では、被評価者は、ある実践の活動主体として、評価されているということさをさほど意識することなく、自らの活動を向上させるために必要な情報として、活動の中で自然に受け止められることになる。その実践が学習の場合には、「評価が学習に埋め込まれている」といった表現を

用いることがあるが、それがまさにこの真正な評価に対応する状況と言えるのである。

コミュニティ内の評価情報としては、たとえば、スポーツなどがそうであるように、価値基準が明確な定量的指標として表わされていることが、その活動を促進するために有効に機能することが少なくない。評価情報が定量的に表現される場合は量的評価と呼ばれる。量的評価で利用される一つの数値は、複雑な側面をもつ実践活動のある一面しか表わし得ないことになるが、目指す方向性が明確になり、その価値を表わす指標の有効性が確認され、それが共有されている範囲であれば、その価値を目指して数値を追い求めることが、複雑な実践活動の総体を促進していくことにもなる。しかし、そのような数値は、その価値を必ずしも共有し得ないコミュニティの外にある人々には、ほとんど意味がないということも少なくない。そればかりか、コミュニティ外の人に、その数値を高めるように強要することは、その一次元的な指標を選んだあるコミュニティからの押しつけと受け取られることもあるだろう。数値であれば客観的ということが常識的な見方としてあるが、そのような場合、数値ほど、主観的なものはないとも捉えられるのである。何故なら、複雑・多面的な活動のその一側面を切り出す意味が納得できず、そこに、その指標を選んだ恣意性を感じるようになるだろうからである。その点、コミュニティ内では、一次元的に選択された指標の意味を共有できるわけで、その意味で、数値の客観性は、あるコミュニティ内でのみ通用するという言い方もできるだろう。

異なるコミュニティ間の評価

この点からすると、たとえば、大学のあるクラスに属しているからと言って、その学習コミュニティのメンバーといえるかどうか微妙な学習者の場合には、テストの得点などの数値による評価を繰り返しても、そのことによって自らには意味もわからぬ序列化がされたりもすることもあって、評価に対してネガティブなイメージがもたれてしまうことにもなるし、学習そのものを必ずしも促進するとは限らないということは、教授・学習の場面ではしばしば経験されることであろう。そればかりか、その数値が一人歩きして、しばしば社会的にも見られる評価の弊害を生み出すことにもなりかねない。

そのような学習コミュニティの周辺に参加しようとしているような学習者に対しては、たとえば、学習の成果物をポートフォリオとして蓄積し、それを教師と学習者がふり返る（リフレクション reflection）などして、学習者が学習コミュニティの価値を共有するきっかけを作り出すなどの工夫が有効となる。この試みは、学習者を、学習コミュニティがもつ背景や文脈に取り込む働きかけをしているとみなすこともできよう。ウェンガーら（Wenger et al, 2002）では、コミュニティにおける評価のあり方として、因果の連鎖がわかりやすく示されるストーリーとして表現することが奨められているが、それもまさに、コミュニティの文脈を共有することに力点が置かれていると考えられる。このことは、コミュニティ外の人にとっては、自らにとっては異なる価値観や世界観がぶつけられることになり、その揺さぶりを通じて、そのコミュニティの境界が取り払われ、周辺参加者としてコミュニティに組み込まれていくきっかけにもなり得る。この種の言葉で表現される評価は、質的評価と呼ばれる。

このように、それぞれの評価のタイプの長短を把握して、それぞれの状況に応じた適切な評価を利用することが望まれることになるが、その適材適所を判別する視点として、「コミュニティ」がどのような役割を果たすかということが、今後の重要な研究課題の一つと

なると思われるのである。

理論と実践の橋渡しとしての授業評価

以上は、実践コミュニティに関わる心理学的な研究課題の一部に過ぎないが、そのフィールドとして、大学教育は非常に適したものである。教員、学生、事務職員などの大学に関わる人々や、それらを取り巻く、社会や地域など、多様な人的相互作用が常に起こっていて、そのあるまとまりをコミュニティとみなすことが有効である状況が少なくないからである。そして、心理学者は、田中（2002）の言うところの、授業実践者と研究者を兼ねているという大学教育のフィールドならではの特殊性を前向きに活かして、理論と実践を橋渡しする心理学的知見の蓄積と共有を図ることができる立場にある。言い換えれば、大学教育のフィールドに心理学コミュニティを形成し成長させていく、そういうチャレンジングな余地が残されているということである。

それを通して、本論のテーマである授業評価に関しても、コミュニティの視点からの評価観がより鮮明に浮き彫りにされ、それに関わる心理学的知見が実証的に積み重ねられていくことによって、より授業実践に密接に結び付いたツールとして進化していくことにもなるだろう。それはすなわち、授業実践という枠組みでは学習コミュニティの形成とその成長に結びつくツールとして、また、大学や学部、学科などの単位では、FD 共同体の形成と成長のためのツールとして、授業評価が健全に位置づけられるようになるということでもある。それは、大学の教育実践と心理学理論がまさに橋渡しされたということであり、授業評価はそのためのキーとなる役割を担うということである。少なくとも、そのような発展を期して、大学の教育、そして、授業評価に、心理学者としてチャレンジしていきたいものである。

文 献

- Lave, J. & Wenger, E. 1991 *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. (佐伯 胖 (訳) 1993 状況に埋め込まれた学習 — 正統的周辺参加 — 産業図書)
- 大塚雄作 2002 高等教育における評価の諸要素とその機能 — 改善志向の評価文化の形成に向けて — 大学評価 (大学評価・学位授与機構), 1, 25-66.
- 佐伯 胖・汐見稔幸・佐藤 学 1992 学校の再生をめざして2 — 教室の改革 — 東京大学出版会
- 田中每実 2002 大学授業研究から大学教育学へ 京都大学高等教育教授システム開発センター (編) 大学授業研究の構想 — 過去から未来へ — 東信堂、Pp.185-206.
- 豊田秀樹・中村健太郎 2004 大学における授業評価の信頼性 — 一般化可能性モデルと構造方程式モデリングによる4相データの解析 — 心理学研究, 75(2), 109-117.
- Wenger, E., McDermott, R., & Snyder, W. M. 2002 *Cultivating Communities of Practice*. Boston, MA: Harvard Business School Press. (野村恭彦 (監修)・桜井祐子 (訳) 2002 コミュニティ・オブ・プラクティス — ナレッジ社会の新たな知識形態の実践 — 翔泳社)
- 山地弘起 2001 学習コミュニケーションとメディア活用 坂元 昂 (監修) 教育メディア科学 — メディア教育を科学する — オーム社、Pp.70-82.