

学生による授業評価は何に役立つのか¹

松尾 太加志・近藤 倫明

What student evaluation of teaching is useful for?

Takashi MATSUO and Michiaki KONDO

Abstract The purpose in this paper is to assess the utility of "Student Evaluations of Teaching" (SET). We compared the results of SET for the course "Psychological Experiments" during the academic years 2002 and 2003. There were significant differences found in several items among the SET results when comparing 2002 and 2003. These differences were due not to changes in instructor or course characteristics but to individual student characteristics. Prior student interest in psychological experimentation affected both the SET results and individual student course grades. SET results are simply products of student introspection and cannot be used to evaluate the effectiveness of teaching methods. The student-instructor relationship differs significantly from that of customer-purveyor. For this reason, SET results cannot be used for personnel evaluation. Closed-ended instruments are not useful for improving instructional practices.

1. はじめに

現在、国内の多くの大学では学生による授業評価を実施している。文部科学省の調査(文部科学省,2004)によると、2002年度の時点で約84%の大学で学生による授業評価が実施されている。一方、学生による授業評価に対する研究は、Algozzine, Beattie, Bray, Flowers, Gretes, Howley, Mohanty & Spooner (2004)によると、アメリカなどでは1920年代から始まっており、近年では伝統的な方法とは異なる手法も検討されてきている。現在多く行われているのは伝統的な学生による授業評価 (conventional student evaluation of teaching) であり (Sproule, 2000)、授業の最後の時間に選択式の回答で評定尺度で評価をさせるものである。学生による授業評価には主として2つの利用目的が考えられる (Algozzine, et al., 2004)。ひとつは授業改善であり、もうひとつは教員の人事査定である。国内においては、人事査定まで至らないものの、教員の個人評価として利用されている実践的取り組みがすでに報告されている (高等教育情報センター, 2003)。

しかし、学生による授業評価が有効であるかどうかについては、必ずしも、統一した見解があるわけではない。1980年代のGaski (1987) の論文では、学生による評価の妥当性を支持するものとして9本の論文を、支持しないものとして5本の論文を紹介している。1990年代のGreenwald & Gillmore (1997) の論文では、学生による授業評価を支持する立場とそうでない立場の文献をそれぞれ11本、7本紹介した上で、筆者自らは中間的な立場であると述べている。さらに、Algozzine et al. (2004) では、概括的に、多くの論文でどのような研究がなされ、今後の課題としてどのようなものがあるかをまとめている。このように、

¹ 本研究は、2004年度北九州市立大学教員特別研究費の助成を受けました。また、本研究の一部は、九州心理学会第65回大会 (2004年11月) で発表しました。

多くの研究がなされていても、その妥当性に関して、十分にコンセンサスがあるわけではない。学生による授業評価の論文の多くは、たんなる実践的な報告であったり、実証研究ではあっても、ただどのような要因が関与しているかを示したにすぎないものが多く、授業改善や教員評価に有効であるかを実証的に示した研究はあまりみられない。

筆者らは、2000年度から、いくつかの担当科目において試行的に学生による授業評価を実施しており、その結果をWeb上でも公開をしている²。しかし、このような取り組みが授業改善にどの程度役に立つのかは疑問である。伝統的な学生による授業評価結果だけではどのような授業改善が必要なかの情報を提供してくれない。ただし、授業評価の結果を複数年度で比較した場合、その結果に違いがあれば、それは教員の教え方の違いを反映していると考えられ、授業改善のヒントが示唆される可能性もある。

そこで、本論文では、2002年度と2003年度に実施した2度の学生による授業評価の結果を比較し、その比較によって何がわかり、授業改善に向けて何ができるのかを考え、さらに、学生による授業評価のあり方を検討した。

2. 方法

2002年度と2003年度の授業評価をどのようにして行ったのかについて述べる。また、授業評価と成績との関係をみるため、成績についても比較を行った。

2.1 授業の概要

対象とした授業科目「心理学実験」は、北九州市立大学文学部人間関係学科の心理学系の学生の必修科目で、2年生の1学期に開講される。2人の教員で担当しており、心理学実験の方法論について、実習を通して学び、さらに実験レポートの書き方を学ぶことを目的としている。シラバスでは授業のねらいを以下のように定めている。

実験科学としての心理学の基礎的な実験法を学習する。心理学の基礎実験を計画、実施し、そのデータ整理、レポート作成という一連の流れを実際に体験し、心理学的に人間を測定するための方法論を実践を通して学ぶ。3、4年次での演習における基礎的な研究能力を養うことを目的としている。

学生には、心理学の基礎的な実験を行わせレポートの提出を課し、それによって成績評価を行う。2コマ続き(90分×2)の授業が週に1回開講される。受講生は、30名程度である。

2002年度と2003年度で実施した授業の内容についての違いを表1に示した。主な違いは、2002年度に実施した「まばたき条件づけ」の授業を2003年度は行わなかったことである。これは、2003年度は天候不良により大学全体が休講になるなどの理由で、授業回数が少なかったためである。また、実験のまとめについて、1回分の授業をとれなかったことも異なっている。ただし、レポート課題を義務づけた4つの実験課題の内容については、授業時間および内容にほとんど違いはなかった。

2.2 被調査者

北九州市立大学文学部人間関係学科2年生で、2002年度28名(男性7名、女性21名)、2003年度30名(男性6名、女性24名)。受講登録者数は、表1に示したようにそれぞれ29名と31名であったが、調査に回答したものは、それよりも少なかった。

² <http://homepage1.nifty.com/tmatsuo/hyoka/>

表1 2002年度と2003年度の授業の比較

	2002年度	2003年度
受講登録学生数	29	31
シラバスの掲載 内容の違い	まばたき条件づけの記載なし	・まばたき条件づけの実施を記載 ・実験のまとめの回に、「人権」とい うキーワードを付加 ・実験の順序を変更する可能性がある ことを注記
レポート課題	長さの弁別閾 盲点の大きさの測定 幾何学的錯視 フィッツの法則	長さの弁別閾 盲点の大きさの測定 幾何学的錯視 フィッツの法則
授業内容の比較	<ol style="list-style-type: none"> 1. オリエンテーション 2. 触2点閾 3. 盲点の大きさの測定① 4. 盲点の大きさの測定② 5. 心理尺度の構成 6. 弁別閾① 7. 弁別閾② 正規確率紙の説明はレポート 返却時に 8. まばたき条件づけ 9. 幾何学的錯視① 10. 幾何学的錯視② 11. フィッツの法則① 3名でグループ 12. フィッツの法則② 13. 自習 14. 実験のまとめ 	<ol style="list-style-type: none"> 1. オリエンテーション 2. 触2点閾 3. 盲点の大きさの測定① 4. 盲点の大きさの測定② 5. 心理尺度の構成 パソコンのプレゼンテーションソ フトを利用 6. 弁別閾① 7. 弁別閾② 正規確率紙を使う意味を説明 8. 幾何学的錯視① 9. 幾何学的錯視② 10. 自習, ゼミ実験の被験者として参 加 11. フィッツの法則① 2名でグループ 12. フィッツの法則②, 実験のまとめ

2.3 学生による授業評価の質問内容

質問紙は、選択回答で、授業内容（Ⅰ）についての9項目、教え方（Ⅱ）についての14項目、資料・機器利用（Ⅲ）についての7項目、設備環境（Ⅳ）についての6項目、総合評価（Ⅵ）の4項目、受講者自身の評価（Ⅶ）の3項目で、合計43項目（図1参照）で、いずれも、「よく当てはまる」から「全く当てはまらない」までの5件法で回答を求めた。ただし、総合評価の中の1項目については0点から100点までの点数で評価を求めた。また、自由記述の欄を別に設けた。

2.4 レポート課題による成績評価

表1に示したように、いずれの年度でも、「長さの弁別閾」、「盲点の大きさの測定」、「幾何学的錯視」、「フィッツの法則」の4つのレポートを課した。レポートの評価は、A、B、C、Dの4段階で行い、学生に評価結果を添削したレポートとともに返却した。評価がDであったものは再提出を義務づけ、評価がCになるまで何度でも提出を求めた。再提出は、他の評価を受けた学生であっても可能であった。

評価は、「長さの弁別閾」、「盲点の大きさの測定」は2人の教員で評価し、「幾何学的錯視」、「フィッツの法則」に関しては1人ずつが担当した。

学生による授業評価は何に役立つのか

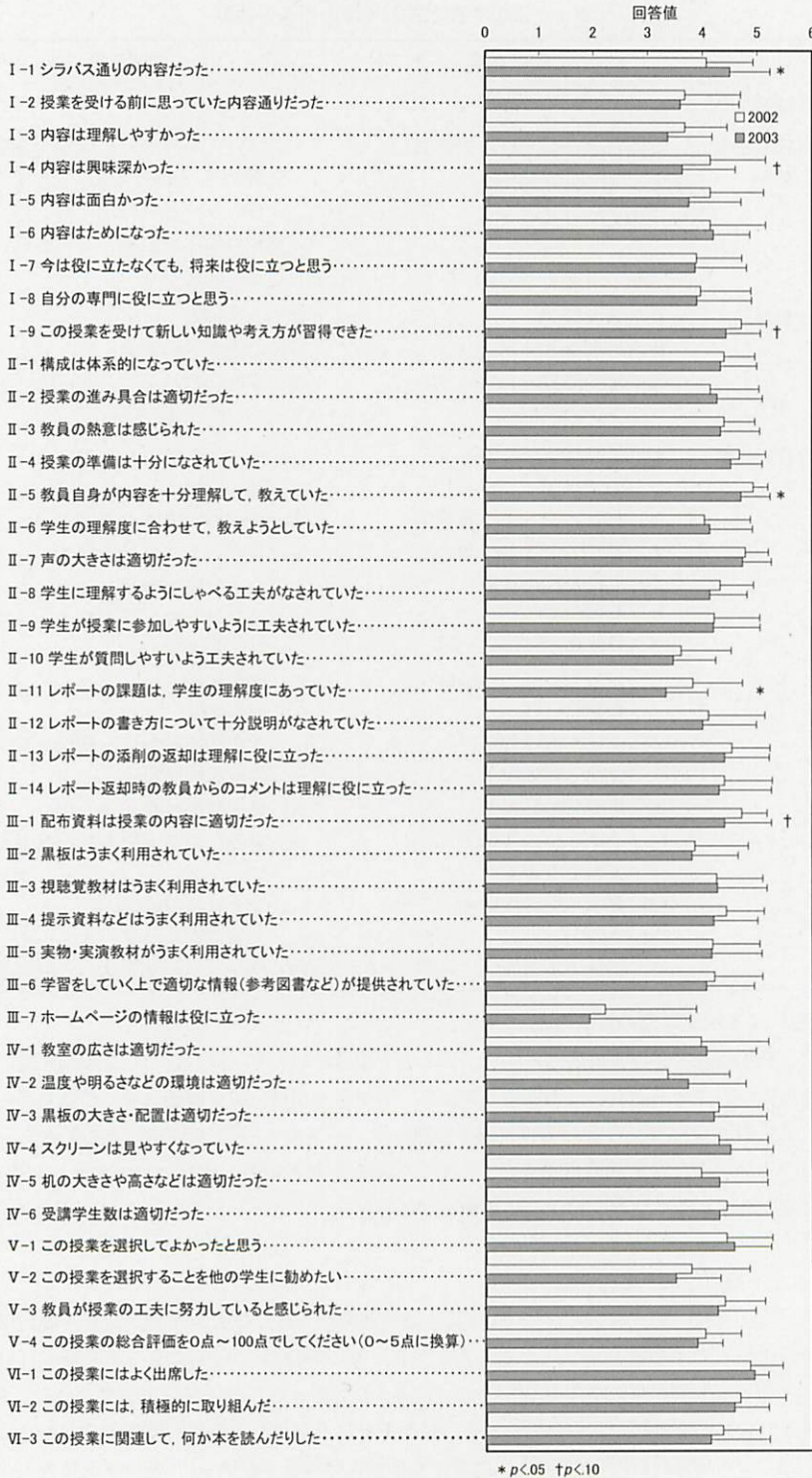


図1 2002年度と2003年度の学生による授業評価の平均回答値の比較

表2 年度間の差に傾向または有意差が見られた項目間での相関係数

	I-4 内容は 興味深かった	I-9 新しい 知識や考え方が 習得できた	II-5 教員自 身が内容を十分 理解	II-11 レポー トは理解度に あっていた	III-1 配布資 料は適切
I-1 シラバ ス通りだった	-0.02	-0.06	0.16	0.08	0.08
I-4 内容は 興味深かった		0.43**	0.30*	0.36**	0.27*
I-9 新しい 知識や考え方が 習得できた			0.16	0.40**	0.17
II-5 教員自 身が内容を十分 理解				0.15	0.46**
II-11 レポー トは理解度に あっていた					0.31*

* $p < .05$, ** $p < .01$

2.5 授業評価質問紙の実施手続き

各年度の授業の最後の時間に質問紙によって、学生に授業に対する評価の回答を求めた。授業担当者が質問紙を配布し、その場で回収した。質問紙は無記名での回答であった。

3. 結果

学生による授業評価とレポート課題の成績評価について、2002年度と2003年度を以下のように比較した。

3.1 学生による授業評価の比較

いずれも、選択回答の「よく当てはまる」を5点、「少し当てはまる」を4点、「どちらでもない」を3点、「少し当てはまらぬ」を2点、「全く当てはまらぬ」を1点として、各質問項目についての平均値を各年度で算出した(図1)。ただし、「V-4 この授業の総合評価を0点~100点でしてください」については、他の質問項目と比較しやすいように、20で除すことによって、0~100点までの点数を0~5点に変換した。

2002年度と2003年度を比較するために、回答の平均値についてt検定を行った。その結果、「I-4 内容は興味深かった」、「I-9 この授業を受けて新しい知識や考え方が習得できた」、「III-1 配布資料は授業の内容に適切だった」において差に傾向があることがみられ($p < .10$)、いずれも、2002年度のほうが当てはまるという回答であった。また、「I-1 シラバス通りの内容だった」では、2003年度が当てはまるという回答が高く、有意な差がみられた($p < .05$)。「II-5 教員自身が内容を十分理解して、教えていた」、「II-11 レポートの課題は、学生の理解度にあっていた」では、2002年度のほうが当てはまるという回答が高く、有意な差がみられた($p < .05$)。

また、年度間の差に傾向または有意な差がみられた項目について、相関をとったところ、表2のようになった。

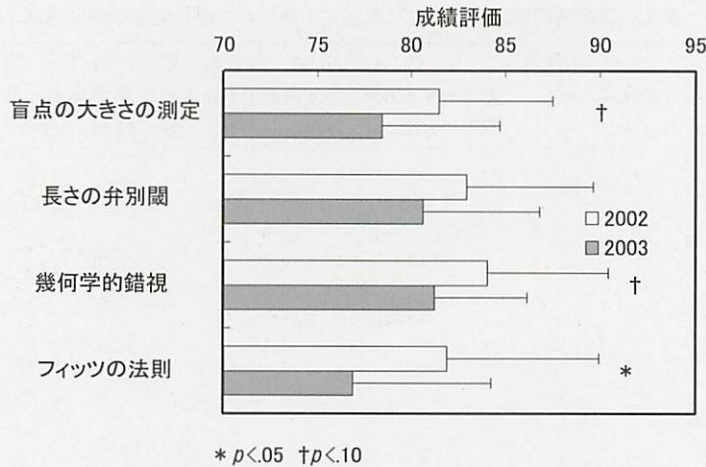


図2 2002年度と2003年度のレポート課題の成績評価の比較

3.2 レポート課題の成績評価の比較

担当教員が行ったA,B,C,Dの4段階評価を90,80,70,0と数値化したものを成績評価とした。D評価の場合0点だが、Dの評価を受けた場合、再提出を義務づけ、C以上の評価になるようにしているため、D評価の0点の学生は実質的にはいない。2002年度の学生と2003年度の学生の4つのレポートの成績評価の平均値を算出した。その結果を図2に示した。さらにt検定によって2002年度と2003年度の成績評価に差があるかどうかをみると、「盲点の大きさの測定」、「幾何学的錯視」においては、差に傾向があることがみられ ($p < .10$)、「フィッツの法則」では、有意な差がみられた ($p < .05$)。いずれも、2002年度の成績が高く評価されていた。

4. 考察

学生による授業評価では2003年度のほうが相対的によくない回答結果であった。2003年度は2002年度に比較して、内容が興味深くない、新しい知識や考え方は習得できなかった、配布資料が不適切であった、教員自身が内容を十分に理解していなかった、レポート課題は学生の理解度にあっていなかったという回答結果であった。全体的に2003年度のほうが授業内容を学生が理解できなかったことを示している。これは、レポート課題の成績評価にも表れてきている。2002年度と2003年度のレポート課題の成績を比較したところ、2003年度の学生のレポート成績は2002年度よりも相対的によくない結果を示した。

4.1 教員の教え方の問題

このような結果になった理由としてまず考えられるのは、教員の授業の教え方の問題である。教え方が十分ではなかったために、学生の理解度や興味関心が低下し、その結果、2003年度のほうがレポートの課題の成績も低下したと考えられる。

しかし、2002年度と2003年度では、ほぼ同じような授業を行っており、教え方に大きな違いがあったことは考えにくい。利用した教材や配布した資料についても、ほとんど同じである。2002年度ではまばたき条件づけの実験を行ったが、2003年度は行わなかっただけである。この授業内容の違いは、レポートの課題に関わらないところであり、レポート課題に関わる授業内容についてはほとんど違いはなかった。にもかかわらずレポート課題の成績が異なるのは、授業の内容や教え方の違いとは考えられない。したがって、

教員の教え方が原因で、学生の興味関心や理解を低め、レポート課題の成績が低下したとは考えにくい。

もっとも、教員は教え方を変えていないつもりであっても、気づかないうちに教え方が雑になっていくことも考えられる。しかし、客観的に同じであるはずの配布資料の適切さにおいて2003年度が低い評価であったのは、教員の教え方以外の別の要因を考えなければならない。

4.2 学生の回答のバイアス

学生の評価が何らかのバイアスを受けて、2つの年度で異なった可能性も考えられる。学生による授業評価は、本学ではここ数年実施されるようになってきた。そのため、学生の授業評価に対する意識も変化した可能性が考えられる。学生による授業評価が開始された当初は、比較的好意的な評価が多かったかもしれないが、年度を追うごとに、学生の意識のほうがかちんと教員の授業のあり方を評価しようということになり、学生の評価も厳しくなった可能性も考えられる。

統計上の差異は見出されなかった項目でも、回答結果を平均値で比較してみると、設備環境の項目以外は、ほとんど2003年度のほうで評価が低くなっている(図1)。このように2003年度のほうが低い評価になったのは、2003年度の学生のほうが厳しい評価をしたという可能性が考えられる。

4.3 学生の理解度の違い

次に考えられるのは、学生の理解度の違いである。学生のレポート課題成績が2003年度で低かったということは、学生の理解度が2003年度はもともと低かった可能性が考えられる。理解度が低いと、両年度でまったく同じレポート課題や配布資料であったにもかかわらず、2003年度の学生は適切でないと感じ、十分に理解できなかったため、興味がなかったという可能性である。そして、新しい知識などが身についたとは考えられなかったのだろう。ただし、Marsh & Roche (2000) は、GPA得点と授業評価では相関が高くないことを指摘しており、もともとの理解度が授業評価結果に反映するとは限らない。さらに、同じ大学の同じ学科を志望してきた学生であり、年度の違いによって学生の理解度に大きな差がみられることは考えにくい。

4.4 学生の興味・関心の違い

次に考えられるのは、学生の授業に対する興味・関心である。「内容は興味深かった」の回答が2003年度は低い傾向を示しており、この学年の学生は、心理学実験の授業そのものに対する興味・関心がもともと高くなかった可能性が考えられる。そのため、この授業に対する動機づけが低くなり、相対的に理解が低下したことが考えられる。つまり、理解度そのものは2002年の学生とは変わらないが、授業の取り組みへの意欲が相対的に低いことが結果として授業への理解の低下を招いたということである。

心理学を志望する学生の多くはカウンセリングなどの臨床心理学の分野を希望する学生が多く、このような学生は、今回の調査対象科目の「心理学実験」などの実験系の心理学に対する興味・関心が必ずしも高くない。臨床系を志望する学生が多ければ、全体としてのデータをとったとき、その成績が低くなる可能性がある。2003年度のレポートの課題成績が低かったのは、臨床系を志望する学生が多かったことがひとつの要因として考えられる。

授業の最初の時間に行ったオリエンテーションで学生に自己紹介をさせた際、どのゼミに行きたいのかを尋ねた。明確にゼミや分野を答えた学生のうち臨床心理学を希望している学生は、2002年度が23名中12名で約50%、2003年度が20名中16名で80%の学生であった。非公式に尋ねたものであるため、信頼性は高いものではないが、2003年度の学生のほうが臨床心理学に対する興味を持っていた学生が多かった。つまり、臨床系とは異なるイメージを持たれやすい心理学実験のような授業に対する興味が、2003年度の学生

は相対的に低かった可能性が考えられる。

興味・関心が理解や成績に及ぼす影響は、回答された項目間の相関からも裏付けられる。年度間に統計上、差の傾向が見られたり、有意差がみられた項目の相関(表2)を見ると、授業に興味を持ったという項目と他の理解に関する項目との正の相関が相対的に高い値を示している。これは、興味が高いと理解も高く、興味が低いと理解が低いという結果を示している。したがって、理解の低さは教員の教え方の問題よりも、学生がもともと持っていた興味・関心が低かったことが考えられる。

授業に対する興味・関心や動機づけが授業評価に影響を与えることは過去の研究でも立証されている。Marsh & Roche (2000) は、学生の成績や事前の興味などのいくつかの背景変数と授業評価の相関を見ている。その中で、事前の学生の興味が授業評価結果と相関が高かったことを報告している。また、浦上・林・石田(1999)は、受講動機の違いによって授業評価結果が異なることを報告している。浦上らは、受講動機として、卒業に必要な単位だったから、興味深そうだったから、時間が空いていたから、友達が受講するから、資格をとるために必要だったからの5つの中から最も大きな理由を選択させ、その動機理由と授業評価結果を比較している。その結果、興味深そうだったからという理由を挙げた学生は授業満足度が高く、卒業に必要な単位だったからという理由の者は授業満足度が低かった。また、興味深そうという理由をあげた学生の授業改善要求は授業満足度と関連がなかったが、他の理由を挙げた学生は、授業満足度の低さと改善要求の強さに関連がみられた。

つまり、もともと当該の授業に興味を持っている学生は、授業に対する取り組みへの動機づけが高くなっている。一方、動機づけが低い学生は、授業への満足度の低さを自分の動機づけの低さに帰属させるのではなく、授業の問題に帰属させてしまっている。

さらに、田中・藤田(2003)は、授業評価結果と学生の達成目標との関連を検討した結果、達成目標の中でも、学習や理解を通して能力を高めることを目指すマスタリー目標と授業評価結果に相関があることを見出している。さらに、授業評価結果が学業遂行結果に影響を及ぼすというパスモデルを提起している。つまり、マスタリー目標志向性が高い学生ほど、授業に興味・関心を持ち、教員の授業の仕方や内容を高く評価するという結果を示しており、さらに、そのような学生ほど試験で良い成績を取っていた。

本論文の結果でも、心理学実験という授業に対する興味・関心がどの程度あったかに授業評価の結果が左右されていると考えられ、その興味・関心をもった学生が2003年度は相対的に少なかったことがこのような結果を導いたのではないかと考えられる。

4.5 学生の評価の信頼性の低さ

以上のように、2つの年度での違いは、受講した学生の興味・関心の違いではないかと推測されるが、学生の評価自体の信頼性が低かったことも考慮すべきであろう。2つの年度では、「シラバス通りの内容だった」に有意な差がみられたが、これは信頼性の問題だと考えられる。

2003年度のシラバスには、2002年度にはなかったものが追加されているだけである(表1参照)。2003年度には、授業回数の都合で「まばたき条件づけ」の実験を行うことができなかったが、2002年度は、シラバスに掲載されていなかった「まばたき条件づけ」を行っている。むしろ2003年度のほうは予定されていた実験が実施されなかったわけであるため、2003年度の学生のほうが相対的にシラバスの内容とは異なると回答してもよいはずである。2003年度のほうがシラバスと一致しているのは、実験の順序を変更する可能性の掲載だけである。2002年度も2003年度も実験の順序はシラバス通りではなかった。2003年度は注を設けた分だけシラバスの内容に近いとも考えられるが、それは大きな要因とは考えにくい。

また、今回の調査では、43項目についてt検定を行ったが、ここで統計上有意差がみられたのも、第一種の過誤によるものであるかもしれない。つまり、実際には差異がないものの確率的に統計上は有意な差

が現れたにすぎない可能性もある。評価は、学期の最後に行っており、約4ヶ月の授業を振り返って評価をすることが求められる。そのため、必ずしも正確に判断できるとは言えない側面を持っている。

5. 総合考察

授業の評価においては、学生による授業評価が重要視されている。その形式の多くは、本調査で実施したような形式のもので、授業の学期の最後に質問紙調査で、授業の興味深さや理解などについて評定尺度の評価がなされている。Sproule (2000) は、こういった伝統的な学生による授業評価の問題点を指摘した上で、それにもかかわらず、多くの大学で実施されている理由を3つ挙げている。ひとつは、大学管理側のご都合主義的な能力測定という政策的な意図である。第二は、数量化することが客観科学だという神話に基づくもの、第三には、心理学や教育学といった科学的枠組みに基づいた科学的手法であると思われることである。つまり、伝統的な学生による授業評価が科学的で合理的な手法だと信じられているのである。数量化をしたりある学問領域の枠組みに則った手法をとったりすることは科学的であるかもしれない。しかし、すべてが科学的であるわけではなく、個々の手法については真に科学的であるかどうか検証しなければならない。

5.1 学生による授業評価の結果は何を表しているか

学生による授業評価は、一般に、学生の特性、教師の特性、授業の特性の3つの要因が影響を及ぼしていると言われる (Mason, Steagall, & Fabritius, 1995; Algozzine et al., 2004)。

本論文で2つの年度の学生による授業評価の結果を比較したところ、まったく同じ教員が同じ教材を使って行った授業であるにもかかわらず、異なる評価結果を示した。授業は同じ科目であるため、授業の特性の要因とは考えにくく、異なる結果を示したのは、評価をした学生の要因と考えることが妥当な解釈である。影響を及ぼした要因として、学生の回答バイアス、学生の理解度、学生の興味・関心といったことが検討され、教員の教え方の要因は相対的に低いことがわかった。学生による授業評価は、授業をした教員についての測定ではなく、むしろ、授業を受けた学生について測定したものと考えることができる。教員の教え方の評価ではなく、受講した学生についての測定になっていると考えるべきであろう。

宇佐美 (2004) は、学生による授業評価で学生に回答を求めているのは、評価や意見ではなく、学生の自分の「頭の中」の状態に関わる申告 (報告) にすぎないと指摘している。学生は、授業に関する専門家ではないため、授業に対して評価をしたり意見を述べたりすることはできないと指摘する。心理学的に考えると、学生による授業評価といわれているものは、学生の主観的判断である内省報告に過ぎず、学生の回答がそのまま授業評価に直結するものではない。たとえば、学生が「理解できましたか」という質問に対して、「そうではない」と回答したときに、その原因が学生側にあるのか教員の教え方なのか、あるいはそれ以外の要因なのかは、授業評価の回答結果だけではわからない。ただ、確実なのは、その回答が学生の内省であることだけである。

授業評価として学生の内省を求めることは意味あることかもしれないが、その内省を發することになった原因は、別に検討しなければならず、その内省の回答が直接教員の教え方に規定されていると考えるのは、一般的な調査手法の観点から考えると明らかな間違いだといわざるをえない。

それでは、学生による授業評価という内省は学生のどのような要因を反映したものであろうか。

5.2 学生の短期の利益を反映

授業評価は、本来、授業がその目的を達成されているかどうかで評価しなければならない。しかし、教育の目的が達成されているかどうかは、短期的にはわからない。教育は将来に対する投資である。学生の

立場に立つと、興味・関心が無く、受講したくないと思う科目であっても受講しなければならない。つまり、この時点ではコストであると捉えられる。しかし、どのようなコストをかけることが将来の投資となるのか、将来の役に立つのかを判断することが難しい。そのため、場合によっては、投資としてのコストではなく、無駄なコストだと捉えられてしまう。そうすると、授業に対する動機が低くなり、授業を理解しようとする努力を怠ったりしてしまふ。大学を卒業することは投資だと考えられても、個々の授業の内容について、今、その授業内容について勉強することが投資につながるという判断を下すことは難しい。

また、仮に、授業を受けることが投資だと考えられたとしても、その投資コストを少なくしたいと考えるのは当然であろう。実際には、コストをかけたほうが将来役に立つことがあるはずだが、その判断は難しいため、低いコストの授業を求めてしまふ。今現在興味を持っていて、今の自分の能力で努力しなくても理解しやすい授業のほうが好まれる可能性は高いと考えられる。

学生による授業評価は、実は、この短期のコストを尋ねているだけにすぎないのかもしれない。将来それが投資につながっているかどうかの判断はなされていない。というよりも、それは現役の学生にはできない。宇佐美(2004)は、学生による評価は卒業した後になされるべきだと述べている。現役の学生にとって、将来に当該の授業がどのように役立つかを見通すことは難しい。

Greenwald & Gillmore (1997) は、授業評価結果に影響を与える要因として、教育の質、学生の能力、学生の動機づけ、成績の取りやすさの4つの要因を用い、探索的に共分散構造分析モデルを構築して分析を行った。その結果、授業に対する学習負荷が高いか低いか、つまり、よい成績を取りやすいかどうかは授業評価に影響を与えるというモデルがもっとも適合したことを報告している。この結果はコストを低くしたいということが授業評価に対するバイアスとして働いているということである。一方で、Marsh & Roche (2000) は、学習負荷は授業評価にバイアスとして働くものではないと主張しており、むしろ、学習負荷と授業評価の間に正の相関があることを示している。

本調査における心理学実験の授業はかなりコストの高い授業である。4つのレポートを期日までに提出しなければならず、評価がD評価であると、評価がCになるまで、何度も再提出を求められる。他の授業に比べて学習負荷が高い授業である。このような授業において、臨床心理学を志望している学生にとってはかなりのコストだと考えられ、授業評価は低くなるであろう。一方、実験系の心理学を志望している学生は、十分な投資だと考え、授業評価は高くなるであろう。2002年度において授業評価が低かったのは、2002年度の学生は投資価値があると判断したためではないだろうか。

学習の負荷は投資であるはずだが、それをコストと捉えてしまふと学生の内省としての授業評価の結果は低くなると考えられるが、学習負荷が十分に投資価値のあるものであり、それによって自分の学力や知識が向上したと考えられるのであれば、学生の内省としての授業評価は高くなると考えられる。

5.3 学生による授業評価は顧客満足とは異なる

授業評価は、学生の内省にすぎないかもしれないが、顧客満足度を知るために学生による授業評価は重要だという考えもある。大学は授業というサービスを提供し、その顧客である学生の満足度を授業評価で行うのは当然であるという論理である(三田,1999)。

しかし、一般の顧客満足調査と授業に対する学生の授業評価とは根本的に異なる。一般の製品やサービスは、その製品やサービスをできる限り多く売ることに目標がある。したがって、顧客満足調査の結果として売れる数が見込めない場合は、数多く売れるように改善する必要がある。場合によっては、売れない製品やサービスは廃止したほうがよいという考えに至ることも必要である。そのため、一般の顧客満足調査の場合、全体を集計した結果として顧客の満足が高いかどうかは、有効な指標となる。少しでも顧客満足度の高い製品やサービスを提供して、相対的に満足度の低いものは提供しないようにすることが必要と

なる。そのため、全体のデータを集計して、平均的にどの製品やサービスで顧客満足度が高いかを知ることは重要である。

同じ論理を授業評価に当てはめると、平均的に人気のある内容だけに淘汰されてしまう。全体のデータとして集計された結果は、平均化されたものであるため、数が多いか少ないかが結果を左右する。数多くの学生の満足度が高くなると、授業評価結果は高くなる。たとえば、心理学関連の授業で考えると、一般に、臨床系の内容に興味を持つ学生が多いため、授業の内容に臨床系のものを多くすれば、授業評価結果は高くなることが想像される。極端な場合、科目自体をすべて臨床系にしまえばよい。そうすれば授業評価結果は高くすることはできる。学生という顧客の満足は高くなる。

これは企業の論理であり、書店などの場合、売れる本を置くことが必要で、心理学関連の書籍として臨床系ばかりを集めればよい。しかし、大学制度におけるカリキュラムの場合、そのような判断はできない。栗田・宇田川(2003)は、大学の授業を大学が提供するサービス商品だとみなし、学生が消費者であるという一般の企業と同じ論理を大学運営に利用することを問題視している。一般の企業では、顧客が満足する商品は無制限に増産することは可能であるが、大学の授業の場合、カリキュラムや時間割りの制約上、学生が望まない授業を担当せざるを得ないことはある。

このような制約は企業でも多かれ少なかれ存在する。売れるものが生き残る企業論理が優先されてしまうと、需要が少ないものは切り捨てられてしまう。たとえば2つの製品があって、一方は9割の顧客が満足を示し、もう一方は1割の顧客しか満足を示さなかった場合、コスト上、2つの製品を作ることが困難な場合、1割の顧客しか満足を示さなかった製品は切り捨てられてしまう。これが企業の論理である。

しかし、現実には、その製品がなくなってしまうと、困る人が出てくるため、一般の企業であっても、利幅は少なくとも、その少数者の要望に応えるため、製品やサービスを提供する。このような論理で製品やサービスが提供されるのは、利潤追求という企業の論理ではなく、公共の論理に基づいたものである。公共の論理の必要性は、一般の顧客満足調査結果では見えてこない。むしろ、無視されてしまうことになってしまう。

授業に関しても同様である。授業には、企業の論理ではなく、優先されるべき別の論理がある。それは教育の論理である。授業内容としてどのような内容が適切であるのかを、需要の高低の論理で判断してしまうと、適切な教育は行えない。ひとつの科目の中で、どのような内容を盛り込む必要があるのか、カリキュラムとしてどのような内容の科目が必要であるのかを考えなければ教育はできない。しかし、学生による授業評価での結果はマジョリティを示すだけになってしまう。教育の論理の必要性は見えてこない。科目内での内容構成やカリキュラム全体としてみたときの科目の必要性といった教育の論理は無視される可能性がある。

さらに、中原・遠藤・ワクター(2004)は、学生による授業評価が特定教員の独特の教育風格や個性的な魅力を抹殺してしまう恐れがあることや、学生の意識として、学生が教員の授業の仕方を自分の成績の不評のせいにして自己弁解する学生を生んでしまうことも指摘している。そして、中原らは、サービスマーケティングの立場に立って、授業評価の適切で公平な評価システムの構築が求められると考えている。

5.4 学生による授業評価を教員評価に利用することの問題点

学生による授業評価の利用目的として教員に対する評価が考えられている。現実には、授業評価の回答結果がそのまま教員評価に利用されている上に、授業評価の回答項目の中でも「全体として、この授業の教員はどう評価されますか」といった"single-most-important question" (SMIQ) といわれる特定の項目だけが活用されてしまっている (Sproule, 2000)。しかし、これまでに議論してきたように、学生による授業評価は、学生の自省にすぎず、学生の要因に起因しているものであり、教員評価に利用することは妥当

ではない。

今回の調査では、同じ授業について2つの年度の学生による授業評価の回答を比較したが、その回答値にはいくつかの点で差異が生じていた。この調査のように、同じ教員で同じ授業科目での回答の違いの場合、それが学生の要因であることは、容易に想像がつく。しかし、現実の教員評価の場合、教員や科目が異なった回答結果として比較される。そのため、その要因が教員の要因、科目の要因、学生の要因のいずれであるのかは判断がつかない。にもかかわらず、安易に教員の要因だと判断されて教員評価に利用されてしまっているが、これまでの議論のように、実際には学生の要因に起因するところが大きいのである。

学生がもともと興味・関心を示している授業内容や、もともと理解しやすい授業内容の授業評価は高くなる。たとえば、心理学の分野で考えると、一般には臨床系の心理学の科目への興味・関心が高い学生が多い。そのため、臨床系の科目の授業評価結果は高くなることが想像される。それに対して、実験系の科目の評価は低くなるであろう。また、心理統計などの数学的基礎が要求される科目は、数学的基礎を有していないと理解度が低くなる。しかし、一般に心理学を希望する学生は、心理学を文系の分野と意識していることが多いため、数学的基礎が理工系を志望する学生に比較して高くない。そのため、心理統計を受講する学生の中には、もともと理解度が低い学生が少なくない。そのため、心理統計のような科目は、他の科目に比較すると授業評価が低くなる恐れがある。つまり、学生の要因と科目要因が学生による授業評価に影響を与える。田中・藤田(2003)は、そのため、授業評価の解釈においては、全体的な平均値だけを見るよりも、背後にある学生の達成目標パターンを考慮することが重要だと指摘している。

このような科目のもつ要因は教員の努力で改善できるものではない。Algozzine et al. (2004)は、教師のコントロールが及ばないところでの評価が含まれているにもかかわらず、回答値の合計値でもって教師の教授能力が測定されてしまっているところに誤りがあると述べている。栗田・宇田川(2003)も、個人の教員の努力が及ばない要因に規定されることがあるため、授業満足度の結果を教員の処遇に反映することの危険性を指摘している。

5.5 学生による授業評価は授業改善につながるか

学生による授業評価は、実際に授業を受けた学生に回答を求め、それを授業改善に役立てるという考え方であり、もっともらしく聞こえるが、実際にその結果を授業改善に役立てることはできるだろうか。

今回の結果を見てもわかるように、学生の評価が低かった場合、それがどのような要因によるものかを知ることは容易ではない。学生の要因によるものなのか、教員の要因によるものか判断が難しい。仮に、今回の調査結果の解釈として教員の教え方の問題だと判断がなされたとしても、実際に教え方の問題がどこにあるのかは、選択回答の数量的な集計データからは何も見えてこない。

一般に学生による授業評価は、すべての授業が終わった段階でなされる。その評価で学生が理解できなかったとか興味をもてなかったという評価結果が出たとしても、授業の中のどの内容について学生がそう回答しているのかを特定することは困難である。つまり、授業のどの内容のどのようところが問題であるのかについて何か情報を提供してくれるわけではない。また、たとえよい評価結果が得られたとしても、どのような教え方がよかったのかもわからない。

授業改善のためには、授業評価結果以外の情報が有用である。本論文で調査対象とした「心理学実験」の場合、学生が課題として提出したレポートを見ることによって、どこかの理解が不足しているのか、そしてその理解不足がどのような原因であるのかがわかる。実際に、筆者らは、レポートの内容を見て、教え方のどこに問題があるのかを検討し、次の授業の時間に補足をしたり、次年度の改善につなげたりしている。さらに、「心理学実験」は、心理学の中心的な授業科目であり、他の科目との有機的な関係性が重要であるため(松尾・近藤・原口、1995)、その授業改善は単独に完結されるものではない。そのため、個々

の科目の授業評価結果を検討しても、授業改善に役に立つとはいえない。

5.6 学生からの情報を有効に活用するための提案

学生による授業評価は、学生からの情報として本来意味をもつものであるはずである。ところが、現在一般に行われている学生による授業評価は、授業改善に役立つようなものにはなっていない。学期の最後に選択回答の方式で回答を求め、それを数的に処理している。その結果だけからは、問題の所在が、学生の要因なのか、教員の要因なのかかわからず、授業の改善に役立つこともできない。また、個々の学生の多様な動機を汲み取ることもできず、平均化された結果として、全体として人気度の高い授業だけに淘汰されてしまう可能性を持っている。

必要な情報は、どのような学生がどのような内容に興味を持ち、どのような点についての理解が足りないのかといった情報である。その情報があるのはじめて、授業の進め方としてどこに問題があるのかの分析が可能である。これまで述べてきたように、学生は多様な動機を持っており、個々の学生の理解度や興味・関心は異なっている。そのため、まったく同じ事柄（たとえば、ある工夫したやり方）に対して、肯定的に評価されることも否定的に評価されることもある。現在行われているような選択回答では、ただ平均化された結果が示されるだけで、ほとんど何も情報を提供してくれない。

渡辺（2001）は、伝統的な授業評価では、学生を受動的な姿勢に引き込んでしまう可能性があり、積極的な参画者として学生の役割を変化させなければならないことを指摘しており、定められたアンケートに記入するという方法に対して再検討の必要性を指摘している。

実際に授業を行っている教員は、学生による授業評価を行わなくても、教え方の問題を把握できる情報を持っている。授業中の学生の反応、学生の発表や意見の内容、テストやレポート課題の成績などである。それらの情報に加え、さらに補うものとして、質問カードなどを使い、学生の質問や意見を受けることが必要であるだろう。

授業改善には、授業に対する学生からのメッセージが必要であることは誰も否定しない。その方法のひとつとして伝統的な学生による授業評価があるにすぎない。しかし、この授業評価は、これまで述べてきたように、授業改善のための有効な材料としてはあまり役に立つものではない。必要なのは、学生と教員とのコミュニケーションである。匿名の選択回答で学期末に行う授業評価がコミュニケーションであると到底考えられない。むしろ、教員と学生が顔を突き合わせて意見を交換できるコミュニケーションの場をうまく作っていくことが、授業改善につながるはずである。

6.まとめ

この論文では、学生による授業評価がどのように役立つのかを検討した。2002年度と2003年度の「心理学実験」の授業で実施した学生による授業評価の結果を比較した。その結果、いくつかの項目において、統計的に有意な差がみられた。ただし、この有意な差は、教員の教え方の違いによるものではなく、学生の要因によるところが大きかった。「心理学実験」という授業に事前に興味をもっていたかどうかに影響され、それがさらには授業の成績にも影響を与えたと考えられる。学生による授業評価は、「評価」ではなく、学生の「自省」にすぎない。したがって、その結果が直接、教員の教え方を反映しているとは考えるのは困難である。学生による授業評価を支持する考え方に、学生を顧客だと見る考え方があるが、教育という場面において、そのような考え方をするには無理がある。学生による授業評価を教員評価に利用することはむしろ危険である。さらに、多くの場合授業の各学期の最後に評定尺度による選択回答の質問紙を行っており、その結果だけでは、授業のどこに問題があるのか、何ら情報を提供してくれない。そのため、授業改善にはほとんど役に立たず、授業評価の指標として妥当性を有しているとは考えにくい。学生から

の情報を有効に活用するには、学生とのコミュニケーションの場を設けることが必要不可欠である。

参考文献

- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G. and Spooner, F. 2004 Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching*, 52, 134-141.
- Gaski, J.F. 1987 On "Construct Validity of Measures of College Teaching Effectiveness". *Journal of Educational Psychology*, 79, 326-330.
- Greenwald, A.G. and Gillmore, G.M. 1997 No pain, No Gain? The Importance of Measuring Course Workload in Student Ratings of Instruction. *Journal of Educational Psychology*, 89, 743-751.
- 高等教育情報センター 2003 教員評価制度の導入と大学の活性化 地域科学研究会
- 栗田真樹・宇田川拓雄 2003 学生による授業評価と満足度測定の問題点. 流通科学大学論集人間・社会・自然編, 15, 59-72.
- Marsh, H. and Roche, L. 2000 Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92, 202-28.
- Mason, P., Steagall, J. and Fabritius, M. 1995 Student evaluations of faculty: A new procedure for using aggregate measures of performance. *Economics of Education Review*, 12, 403-416.
- 松尾太加志・近藤倫明・原口雅浩 1995 心理学教育における統合的教育の実践報告. 北九州大学文学部紀要(人間関係学科), 2, 31-42.
- 三田誠弘 1999 大学はどこに行くのか 安岡高志・滝本喬・三田誠弘・香取草之助・生駒俊明 授業を変えれば大学は変わる プレジデント社, Pp.193-242.
- 文部科学省 2004 大学における教育内容等の改革状況について
(http://www.mext.go.jp/b_menu/houdou/16/03/04032301.htm)
- 中原龍輝・遠藤誠二・K, ワクター 2004 学生による授業評価に関する日米大学間異同点の一考察——富士常葉大学とThe University of Mississippiとの比較を中心に. 富士常葉大学研究紀要, 4, 157-186.
- Sproule, R. 2000 Student Evaluation of Teaching: A Methodological Critique of Conventional Practices. *Education Policy Analysis Archives*, 8 (50). (<http://epaa.asu.edu/epaa/v8n50.html>)
- 田中あゆみ・藤田哲也 2003 大学生の達成目標と授業評価, 学業遂行の関連. 日本教育工学雑誌, 27, 397-403.
- 浦上昌則・林雅代・石田裕久 1999 受講動機別にみた授業評価と満足度. アカデミア人文・社会科学編(南山大学), 70, 515-540.
- 宇佐美寛 2004 大学授業の病理 FD批判 東信堂
- 渡辺勇一 2001 学生による授業評価をどう見るか. 生物科学, 52, 209-216.

北九州市立大学文学部紀要

(人間関係学科)

第 12 卷

目 次

出口由美・富山浩三：マラソン大会参加者のランニング・コミットメントとランニング障害 — スポーツアディクションの視点から — ……………	1
黒田耕司：21世紀初頭におけるわが国の教育評価の特質……………	11
須藤 廣：ハワイにおける日本人観光発展史 — 海外旅行についての観光社会学のために — ……………	21
高西敏正：北九州市立大学生の肥満に関する意識と実態調査……………	33
徳田光広・近藤倫明：運動機能障害者における動作評価の基礎的研究 — 運動機能障害の症例を通して — ……………	43
松尾太加志・近藤倫明：学生による授業評価は何に役立つのか ……………	51
名古恵理・近藤倫明：失語症患者の音読訓練に関する症例研究 ……………	65
田島 司：外集団との対比が促進する集団内の協力行動について ……………	75
今田 文・竹川大介：現金贈与の社会装置 — キリバス共和国におけるビンゴゲームの役割 — ……………	83
脇園賀子・竹川大介：壺中の聖水 — インドにおける豊穰性と浄化性の自然観 — ……………	99

北九州市立大学文学部

2005年 3 月発行