

## 学生による授業評価の妥当性と有用性<sup>1</sup>

— 試験成績との関連 —

松尾 太加志

The validity and utility of student evaluation of teaching: Relation of grade of examination.

MATSUO Takashi

**Abstract** The purpose of this paper is to examine validity and utility in the student of evaluation teaching. The understanding level and the interesting level were asked per hour of the class. The relation between them and the final examination results was seen. As a result, the correlation was not seen in the entire student. But, a positive correlation was seen for the student with a lot of attendance frequencies in the result of an understanding level, an interesting level, and the final examination. As for the interesting, it was possible to prompt studying. However, the student with a little attendance frequency cannot correctly recognize the understanding level. In addition, the interesting did not promote study. Therefore, student evaluation of teaching was not necessarily valid. Moreover, there was little utility to the teacher evaluation and the class improvement, too.

### 1. はじめに

学生による授業評価として一般的に行われているものは、“cafeteria-style”の授業評価システムと言われる(Algozzine, Beattie, Bray, Flowers, Gretes, Howley, Mohanty & Spooner, 2004)。学期の最後に自由記述と評定尺度による質問紙の形式で行われ、主として評定尺度を点数化して、教員の教え方の指標として用いられている。理解できたかとか興味を持てたかといった内容が尋ねられる。このような方法の妥当性については、さまざまな研究がなされているが(Cohen, 1981; Marsh, 1984; Koon & Murray, 1995; McKeachie, 1997; Johnson, 2000; Sproule, 2000など)、本研究では、その評価の妥当性と有用性に関し、以下のような問題について焦点を当ててみたい。

#### 1.1 妥当性と有用性の問題

カフェテリアスタイルの授業評価では、評価者である学生は1学期の授業を振り返って総合的な評価をしなければならず、正しく評価することは容易ではない。また、かりに評価ができたとしても、その評価は学生自身の内省にすぎない(宇佐美, 2004)。授業に対して興味を持つことができたかとか面白かったかといった内容の場合は、学生の内省がある程度有用な情報になるかもしれないが、理解や将来役立つかといった内容の場合、正しく認識できているどうかはわからない。わかったつもりになっているだけであっ

<sup>1</sup> 本研究は、2005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66回大会（2005年11月）で発表しました。

たり、将来どのような形で役に立つのか学生にはわからないこともある。さらに、後述するが、学期末であると、ある程度自分の成績が予測でき、成績との関連によって評価が左右されることがある。このように、教員の教え方や授業内容そのものに対する評価として、カフェテリアスタイルの授業評価の妥当性については疑わしいところがある。

また、このような学生の評価が妥当であったとしても、その評価がどのような形で有用であるかという問題がある。ある授業科目に対する評価結果が評価値で示されたとき、その授業を担当した教員は、どのような基準で授業の良し悪しを判断するのであろうか。よくない評価で改善すべきと判断するのか、あるいは、すぐれた教え方で改善の必要がないと判断するのか、明確な基準が存在しない。回答にはばらつきがあるため、よい評価をした学生もいるだろうし、そうでない評価をする学生もいるであろう。評価結果として示されたものが平均的に極端な値を示さない限り、判断は難しい。他の授業との比較をすることは可能であるが、教え方の要因以外に左右されるため(松尾・近藤, 2005a)、たとえ評価値が低くても教え方の問題でないこともあり、評価値が高いからといって、教え方がよかったわけではないこともある。

さらに、ある評価結果に基づき、改善が必要だと判断された場合、どこをどのように改善すべきなのかの情報がもたらされないという問題がある(松尾・近藤, 2005a)。たとえば、理解できたかという設問に対して、低い評価がなされたときに、1学期間のどの日のどの内容に対して理解できないと学生が感じたのかは、1学期を通して総合的に評定された評定尺度値の回答では何ら有用な情報を得ることができない。

カフェテリアスタイルによる授業評価では、上記のように妥当性及び有用性に関して問題があり、授業評価を実施すること自体意味がないとも考えられる。そこで、本研究では、ここで指摘した問題が実際に存在しているのかを、学生による授業評価と試験の成績との関連を見るという方法によって検討することにした。

## 1.2 授業評価と成績との関係

学生自身による評価そのものの妥当性を確かめるために、本研究では、学期末に実施した試験の成績と授業評価との関連を見ることにした。たとえば、理解できたかという項目に対しての学生の評価結果と試験の成績の関連を見ることによって、学生自身の理解度が正しい評価であるのかどうかを知ることができる。授業の最大の目標は、授業で教えられた内容が理解されているかどうか、授業で与えられた知識が学生の知識となっているかどうかである。それは、試験の結果という形で表れる。授業の目標としては、学生の試験の成績が高いことが求められ、授業の成果の指標となりうるものである。

学生による授業評価と成績の関連については多くの研究がなされており、その多くでは正の相関があると言われている(Feldman, 1989; 安岡・吉川・高野・峯崎・成嶋・光澤・道下・香取, 1989; 松田・三宅・谷村・小嶋, 1999; Eisler, 2002など)。学生の授業評価と成績との間にどのような関連があるかに関してはいくつかのモデルが提起されている(Howard & Maxwell, 1980; Gaede & Lewis, 2005)。

まず、考えられるのは、教師の教授法が学生の成績を上げ、その結果、授業評価も学生の成績も向上するというモデル(teaching effectiveness model)である。授業評価と学生の成績に共通原因として教員の教え方を設定するモデルである(ex. Brown, 1976)。学生による授業評価を教員評価や人事査定に利用することが妥当であるという考え方の根底をなす考え方である。

しかし、実際には、学生による授業評価の結果は、教員の教え方だけに影響を受けるものではない。次に考えられるモデルは、学生の要因に起因するという考え方である。学生のパフォーマンスを上げる要因として学生の動機づけや興味が影響するとしたモデル(prior characteristics model)である(ex. Hoyt, 1973; 松尾・近藤, 2005a)。学生の動機づけが高いことが学生のパフォーマンスを上げ、同時に成績や授業評価を高めると考えるモデルである。

教え方や学生の動機づけが学生による授業評価や成績に影響を与えるという考え方自体には間違いはないであろう。しかし、実際に測定結果としてみることができるデータは、これらの要因だけではなく、社会的な関連性の要因に影響を受けることも考えられる。成績は教員が学生を評価するものであり、一方、学生による授業評価は学生が教員を評価するものである。そこには、潜在的に互恵性が働くことが考えられる。そのような考え方は、grading leniency modelにみることができる(ex. Bausell & Magoon, 1972; Greenwald & Gillmore, 1997)。いろいろな変形版はあるものの、基本的には学生が自分の成績が高いと期待したり実際に成績が高いと、その学生の授業評価が高くなるとするモデルである。このモデルは、よい成績を取りやすいと感じることが授業評価を高くするという考えである。単位が取りやすい授業ほどよい授業という評価を受けやすいことになる。このモデルにしたがうと、教員が学生から高い評価を得ることを望むと、成績のインフレーションが生起するという考え方にまで発展する(ex. Nelson & Lynch, 1984; Eiszler, 2002; Centra, 2003)。

さらに、原因帰属理論による説明もなされている(Marsh & Roche, 2000)。成績の高低がどのような原因によって生起すると帰属させるかによって授業評価が変わってくるというものである。帰属理論にしたがえば、成績がよい場合、自分の内的な要因に帰属させ、成績が悪い場合は、外的な要因に帰属させてしまいがちである。この考え方を適用すれば、成績がよい学生は自分の能力に帰属させ、教員の教え方がよいという評価をしない。一方、成績が悪い学生は、教員の教え方が悪いと帰属させてしまう。いずれにしても、教員の教え方に高い評価は与えられないことになってしまう。

grading leniency modelや原因帰属理論による説明では、ある程度自分の成績がどの程度であるのか期待できる段階あるいはすでに成績結果がわかっている段階での授業評価の場合に解釈として成り立つ考え方である。つまり、カフェテリアスタイルの学期末に行う授業評価の場合であれば、このような説明の可能性は十分に考えられる。しかし、本研究では後述するように、毎時間ごとの授業評価を行うため、自分の試験の成績がどの程度であるのかまで意識はしないと思われる。したがって、grading leniency modelや原因帰属理論による説明は、本研究の枠組みには適用しにくい。teaching effectiveness modelかprior characteristics modelのいずれかの枠組みで考えることが妥当である。

田中・藤田(2003)は、毎時間で授業評価を行い、学業遂行(試験の成績)との関連を検討している。それによると、直接的な相関係数でみると、授業評価と学業遂行の間には有意な相関は見出していないが、パス解析において、授業自体の評価と学業遂行には正の関連があるとしている。ただし、田中・藤田の分析では、単純に授業評価と学業遂行とに正の関連があるとしているわけではなく、学習や理解を通して能力を高めるというマスター目標を持つことが、授業自体の評価を介して高い学業遂行を導くことを指摘している。つまり、当該の授業に対して達成目標を持っているかどうかによって規定されることになる。これは上記のモデルの中ではprior characteristics modelを支持するような結果を示している。教員の教え方を反映しているというよりも、動機づけが高いことが授業評価を高くしてしまっている結果であると考えられる。つまり、学生による授業評価のひとつの目的であるteaching effectiveness modelには合致していない。

### 1.3 妥当性と有用性の検討

田中・藤田の研究では、学業遂行全体と授業評価全体との関連を見ている。授業全体としてみたときには、学生が持っている動機づけの要因が左右し、授業評価や学業成績に影響を与えることは十分に考えられる(松尾・近藤, 2005a, 2005b)。そこで、本研究では、試験問題を個別に分析し、試験問題にかかわる内容の当該授業日の授業評価との関連を検討することにした。

そのため、授業評価を毎時間実施することとした。毎時間の評価を行うため、試験問題の設問項目との

対応を見ることによって、どの内容に対して理解が不足していたのかどうか、どこに教え方の問題があったのかどうかを見ることができる。こうすることによって、1学期全体を振り返っての評価における妥当性の欠如を無くすこともできる。ただし、回答の負担を軽減するため、尋ねる内容は理解と興味だけとした。

授業評価が単なる学生の自省ではなく、教員の教え方の良し悪しを反映しているのであれば、授業評価が高い授業での試験は良い成績であり、評価が低い授業の試験の成績は低くなると考えられる。ただし、ここでは複数の授業の間での比較をするのではなく、学生間での比較を行う。同じ授業に対して、授業を高く評価した学生にとっては、その授業の教員の教え方はよかったものであると考えられる。したがって、試験の成績は高くなると思われる。一方、授業を低く評価した学生の場合、その学生に対しての教え方に問題があったと考えることができ、試験の成績は低いことが考えられる。このような結果になれば、学生による授業評価は有用であると考えられる。

以上のように、本研究では、毎時間ごとの授業評価と試験の成績の関係を見ることによって、学生による授業評価の妥当性と有用性を検討する。

## 2. 方法

### 2.1 被調査者

2001年度のコミュニケーション論の授業の試験を受けた北九州市立大学学生110名（男性55名、女性55名）。学年は2年生から4年生である。

### 2.2 授業及び試験

対象とした授業は、文学部人間関係学科及び経済学部経営情報学科の専門の選択科目として開講された科目である。2つの学科の学生が同じ授業を受講している。2001年度の1学期の授業で週1回の授業がなされた。講義は12回行った。各回の授業のテーマは表1に示した。

表1 授業内容

日付	内容	コミュニケーション カードの提出
4月11日	導入	×
4月18日	理解過程①	○
4月25日	理解過程②	○
5月2日	言語	○
5月9日	社会的認知①	○
5月16日	社会的認知②	○
5月23日	表情	○
5月30日	視線	○
6月6日	身体動作など	○
6月13日	対人距離①	○
6月20日	対人距離②	○
6月27日	電話	×
7月4日	試験	×

試験は、100点満点の6問からなる記述式の試験であり、各問の内容及び配点は表2に示した。授業は、教科書中心で進められ、試験時には教科書を含めすべて持ち込み可で行った。

表2 試験問題の内容

試験問題	内容	配点
設問1	理解過程①	16
設問2	社会的認知②	16
設問3	対人距離②	16
設問4	視線	16
設問5	電話	16
設問6	自由課題	20
合計点		100

### 2.3 授業に関する評定

授業に関する評定は、授業の毎時間に配布したコミュニケーションカードで、理解度と興味度について7件法で回答してもらった。1～7までの回答で、値が高いほうが理解度及び興味度が高いことになる。コミュニケーションカードは、学生の出席をとることを目的としたものではなく、そのときの授業に対する学生からの質問や意見を教員に伝達するコミュニケーションの役割を果たすものである。B6サイズで、学籍番号、氏名、所属などを記入する欄もある。12回の講義のうち、コミュニケーションカードを利用したのは最初と最後を除く、10回だけであった。最初の授業はイントロダクションであり、まだ授業を履修するかどうかを決定しない段階であり、最後は無記名式の授業評価を別に行ったため、コミュニケーションカードは配布しなかった。

## 3. 結果

### 3.1 学期を通しての分析

理解度及び興味度について、それぞれの評定値に関して、学期を通しての各個人の平均値を算出した。試験の成績は100点満点の素点を成績として表3に示した。理解度、興味度、試験の成績の間の相関も表3に示した。理解度と興味度の相関は高かったものの、理解度、興味度と試験の成績との相関係数は低く、有意な相関は見られなかった。理解度、興味度と試験の成績との相関図を図1及び図2に示した。

表3 理解度、興味度、試験成績の出席回数別の平均値及び相関係数

	全員	出席回数			
		0, 1	2~4	5~7	8~10
人数	110	28	20	31	31
理解度平均	5.06	5.45	4.89	5.21	4.88
興味度平均	5.26	5.09	5.13	5.42	5.24
試験成績	49.3	41.0	38.2	50.8	62.5
理解度－興味度相関	.553**	.084	.768**	.534*	.815**
理解度－試験相関	.031	.425	-.522*	.172	.349+
興味度－試験相関	.048	-.006	-.476*	.013	.480**

\*\*  $p < .01$  \*  $p < .05$  +  $p < .10$

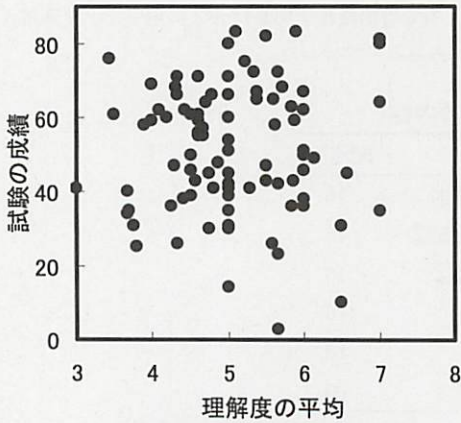


図1 理解度と試験成績の関係  
(全員の学生)

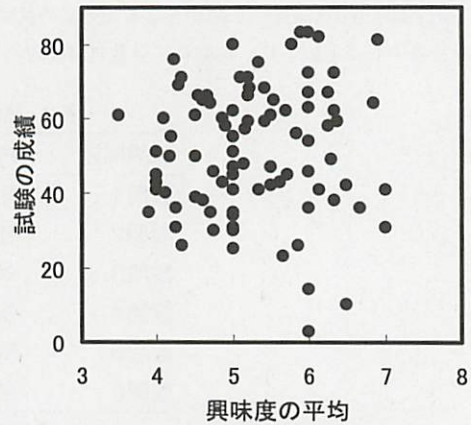


図2 興味度と試験成績の関係  
(全員の学生)

そこで、学生の出席回数に応じて、4段階に分けて分析を行った。出席回数は、コミュニケーションカードを提出した回数でカウントした。12回の授業のうちコミュニケーションカードの提出を求めたのは10回であるため、出席回数の最大は10回である。各段階がほぼ同数になるように、出席回数を0～1回、2～4回、5～7回、8～10回に分けた。出席回数段階別の試験の成績、理解度、興味度を表3に示した。出席回数段階間の違いをみると、興味度においては差はなかったが( $F=.574, df=3/103, n.s.$ )、理解度と試験成績にはそれぞれ有意な差がみられた( $F=3.23, df=3/103, p<.05$ ;  $F=16.25, df=3/106, p<.01$ )。Bonferroni法による多重比較の結果、5%水準で有意な差がみられたのは、理解度においては出席回数0～1回と8～10回の間であった。試験成績においては、出席回数0～1回と2～4回の間、0～1回と5～7回の間を除いたところですべて有意な差がみられた。

理解度と興味度の平均評定値と試験成績との関係については、図3～図10に示した。相関係数を算出したところ、出席回数が8～10回の学生は、理解度や興味度と正の相関(.349; .480)を示したものの、出席回数が2～4回の学生は負の相関(-.522; -.476)を示した。しかし、それ以外では有意な相関は見られなかった(表3)。

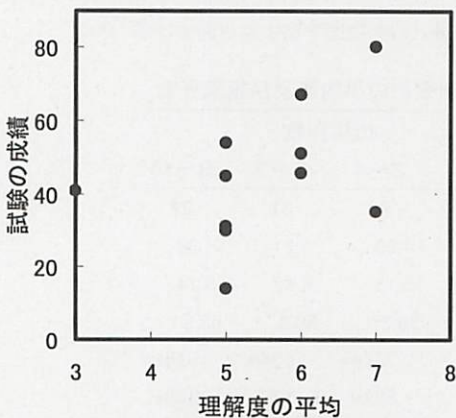


図3 理解度と試験成績の関係  
(出席回数1回の学生)

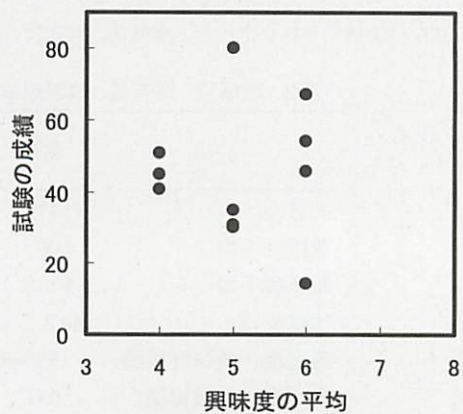


図4 興味度と試験成績の関係  
(出席回数1回の学生)

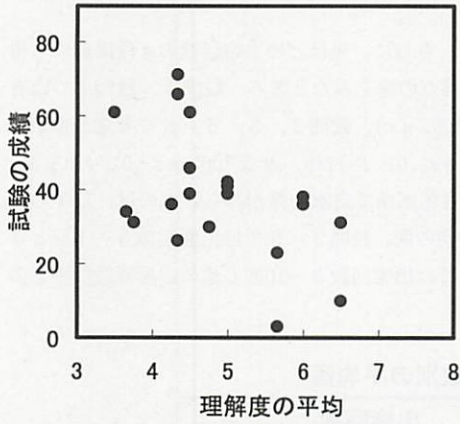


図5 理解度と試験成績の関係  
(出席回数2~4回の学生)

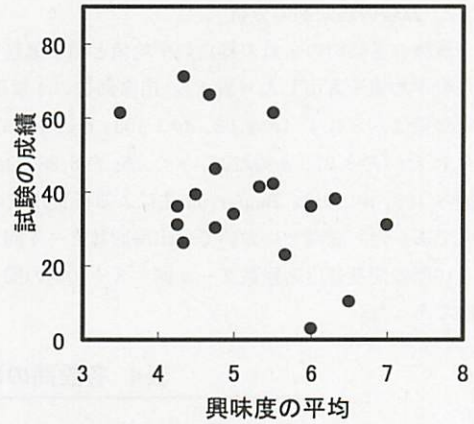


図6 興味度と試験成績の関係  
(出席回数2~4回の学生)

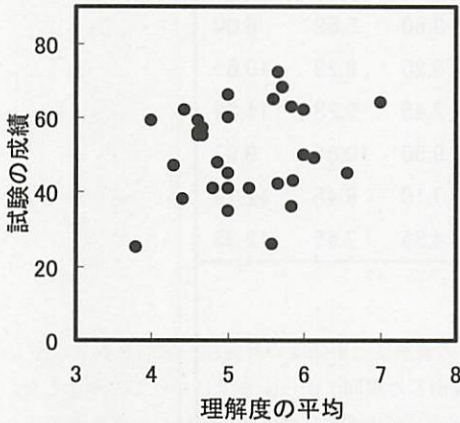


図7 理解度と試験成績の関係  
(出席回数5~7回の学生)

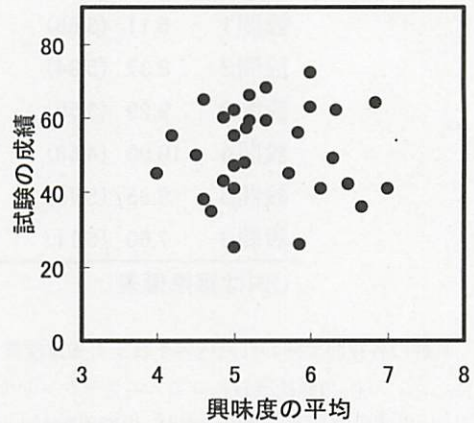


図8 興味度と試験成績の関係  
(出席回数5~7回の学生)

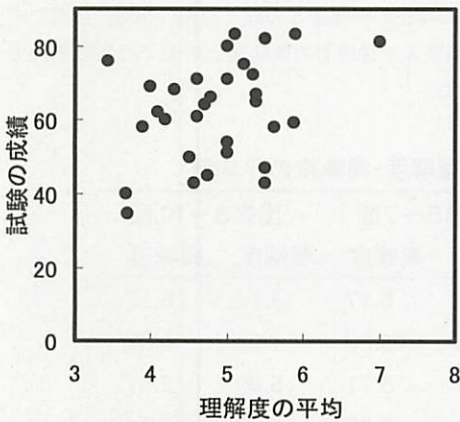


図9 理解度と試験成績の関係  
(出席回数8~10回の学生)

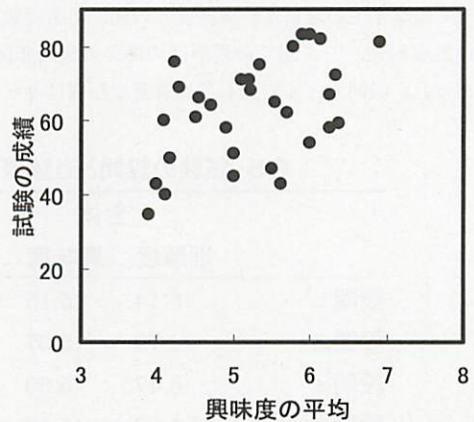


図10 興味度と試験成績の関係  
(出席回数8~10回の学生)

### 3.2 試験の設問別の分析

試験の各設問の全員の得点の平均値と標準偏差を算出し、さらに、先ほどの出席回数4段階別にも得点の平均値を算出した(表4)。出席回数4段階間での得点の差をみたところ、設問1、設問4では有意な差はみられず( $F=2.02$ ,  $df=3/106$ ,  $n.s.$ ;  $F=.57$ ,  $df=3/106$ ,  $n.s.$ )、設問2, 3, 5, 6で有意な差が見られた( $F=3.87$ ,  $df=3/106$ ,  $p<.05$ ;  $F=6.88$ ,  $df=3/106$ ,  $p<.01$ ;  $F=7.46$ ,  $df=3/106$ ,  $p<.01$ ;  $F=13.84$ ,  $df=3/106$ ,  $p<.01$ )。Bonferroni法による多重比較の結果、5%水準で有意な差がみられたのは、以下の通りであった。設問2においては出席回数2~4回と8~10回の間、設問3, 5では出席回数0~1回と8~10回の間及び出席回数2~4回と8~10回の間、設問6では出席回数8~10回と他の出席回数段階との間であった。

表4 各設問の出席回数別の平均値

	全員	出席回数			
		0, 1	2~4	5~7	8~10
人数	110	28	20	31	31
設問1	5.11 (3.66)	4.75	3.60	5.52	6.00
設問2	8.37 (5.04)	7.57	6.20	8.23	10.65
設問3	9.29 (3.59)	8.36	7.45	9.23	11.39
設問4	10.00 (4.58)	9.61	9.30	10.84	9.97
設問5	8.95 (5.75)	6.11	7.10	9.45	12.19
設問6	7.60 (6.11)	4.57	4.55	7.55	12.35

( )内は標準偏差

試験の各設問で問われた内容を教えた当該授業日の学生の理解度と興味度の評定値の平均を表5に示した。ここでは、当該授業日にコミュニケーションカードを提出した設問1から4までについてのみ示した。また、出席段階ごとに算出したが、出席回数が4回以下の場合、当該授業日に出席していない学生が多く、人数が少なかったため、出席回数が5~7回及び出席回数8~10回の学生についてのみを示した。さらに、理解度と興味度と各設問の相関を見た(表6)。その結果、5%水準で有意な相関がみられたのは、設問1では全学生の理解度、興味度との間、出席回数8~10回の学生の興味度との間での正の相関、設問2では出席回数5~7回での理解度の間での負の相関、出席回数8~10回での興味度との間での正の相関であった。設問3, 4に関しては有意な相関はみられなかった。

表5 試験の設問と当該授業日の理解度・興味度の平均値

	全体		出席5~7回		出席8~10回	
	理解度	興味度	理解度	興味度	理解度	興味度
設問1	5.14	5.15	5.08	5.17	5.14	5.17
設問2	4.79	5.07	4.95	5.00	4.79	5.31
設問3	5.47	5.60	5.71	5.71	5.39	5.57
設問4	5.09	5.13	5.71	5.35	4.89	5.14

表6 試験の設問と当該授業日の理解度・興味度との相関係数

	全体		出席5～7回		出席8～10回	
	理解度	興味度	理解度	興味度	理解度	興味度
設問1	.299*	.327*	.399	.198	.212	.439*
設問2	-.037	.182	-.503*	-.279	.270	.553*
設問3	.016	.151	.188	.436	-.028	.014
設問4	.139	.164	-.088	-.072	.288	.306

\*  $p < .05$ 

#### 4. 考察

##### 4.1 理解と興味の関係

一般的には理解と興味の間には高い相関があることが予測される。理解できたから興味深く感じたということもあるし、逆に、興味関心がもともとあったため、理解もできたということも考えられる。つまり、理解と興味の間には相互因果性があると考えられる。さらに、本調査では、毎時間、コミュニケーションカードで理解度と興味度を同時に尋ねることとしたため、2つの間にはいっそう高い相関がみられることが予測された。実際に得られたデータを表3に示したが、理解度と興味度の間には高い相関が見られた。

##### 4.2 理解に対する認識

一般には理解が高ければ試験の成績はよいと考えられる。しかし、学生の回答した理解度と真の理解が必ずしも一致しているわけではない。学生が理解を高く回答しても、単にわかったつもり（西林, 2005）になっていてわかっていないこともありえる。また、理解していないと回答したとしても、低レベルでの理解不足ではなく、かなりの部分はわかっていても、ある部分がわからなかったために理解できなかったと回答したかもしれない。さらに、回答したときと試験実施時期の時間的隔たりも大きな影響を与えらると思われる。授業のときは理解していなくても、その後の授業の受講あるいは自分での勉強の結果、理解ができた可能性は十分にある。以上の点を考慮して本調査の結果を考察してみると、以下のようなことが考えられる。

学生全体においては、理解度と試験成績の間に相関は何も見られなかった（表3）。出席回数の段階別に分析を行うと、出席回数が8～10回の学生の場合、正の相関がみられたが、出席回数が2～4回の学生の場合は負の相関がみられた。出席回数が8～10回の学生は自己の理解を正しく認識していたと考えられる。一方、出席回数が2～4回の学生で理解度を高く評定した学生は、わかったつもりになっただけで、自分の理解度について正しく認識できていない可能性が考えられる。理解度を低く評定した学生は、理解が低いと自覚したため、自分で学習することによって試験の成績が高くなったのではないかと推測される。出席回数が0～1回や5～7回の場合は、上記の要因が複合的に関係しあって、有意な相関が見出されなかったものと考えられる。

個別の設問ごとに理解との関係をみると、設問1の場合のみ、全体との理解度との関係において低いながらも有意な正の相関がみられたが、それ以外の設問2～4の場合、相関関係は見られなかった（表6）。これは、問題自体の持つ要因が影響しているものと考えられる。設問1は、同じ配点である設問2～5に比較して、平均点が低くなっている（表4）。設問1は正しく理解しておかなければ解けない問題であり、処理能力依存型(Norman & Bobrow, 1975)の問題になっていたことが考えられる。そのため、授業を行ったときにある程度理解しておかないと、試験でよい成績をとることができなかったと考えられる。設問2

に関しては、出席回数5～7回の学生で有意に負の相関を示している。これは先に述べたように理解に対する誤認識が作用していると思われる。設問2の全体の相関が低かったのは、理解度を正しく認識できていない学生がいたためだと考えられる。設問4は、他の設問に比して平均点が高くなっている。設問4は得点の分布を見ると、高い得点者が多く、理解度にあまり依存しない問題で、データ依存型(Norman & Bobrow, 1975)の問題であった可能性がある。比較的やさしい問題であり、試験時に教科書を持ち込むことができたこともあり、あまり理解していない学生であっても、それなりに解答できたと考えられる。設問3に関しては明確な判断材料はないが、得点の分布が設問2に類似していたため、理解度の誤認識である可能性がある。

さらに、本調査の方法論的な問題も考えなければならない。授業評価の回答が直接的に当該の設問に対する評価では必ずしもなかったという問題がある。毎時間で授業評価を行ったが、それも90分の授業全体に対する評価であるため、90分の中で話された授業の個々の内容に対するものではない。そのため、当該の設問の内容についての評価にはなっていない。そのため、各設問の試験成績と授業評価との関係が一貫した形ではなかったと考えられる。

#### 4.3 興味による学習の動機づけ

興味に関しては、上記の理解と興味の相互因果性を考えれば、理解度と試験の成績との関係に見られたような関係を示すことが考えられる。それは、出席回数が2～4回や出席回数8～10回の学生において、理解度と成績の関係に同様の関係性(表3)が見られたことから明らかである。

しかし、興味が無くても、内容として理解できることもあり、逆に興味があっても、内容として理解できないこともあり、必ずしも、理解と成績との関係と同じような関係が、興味と成績の間に見られるわけではない。ただし、授業を受けた時点で内容理解が十分でなかったとしても、興味があれば、後で復習をしたり、試験に向けた勉強を行ったりする可能性が、興味がない学生に比べて高いことは考えられる。

設問ごとに興味度と試験の成績の関係を分析したところ(表6)、設問1においては全体の学生で有意な正の相関( $r=.327$ )がみられ、出席回数が8～10回の学生では、さらに高い相関( $r=.439$ )が示されている。これらの相関は、理解度と試験の成績で示されている正の相関よりも高い値を示している。授業を受けた時点では理解できなかったものの、後で勉強をすることによって理解でき、試験の成績が高くなったと解釈される。そのため、理解度と試験の成績には相関が見られなかったものの、興味度と試験の成績に正の相関が見られたと考えられる。同様の傾向が設問2の出席回数が8～10回の学生において見られている。

いずれの場合においても、出席回数が8～10回の学生において、興味度と試験成績の間に高い相関が見られた。これは、単に授業評価において興味があったと回答した人すべてが勉強をしたのではないことを示している。表1に示したように、出席回数が8～10回の学生だけが興味を高く回答しているのではなく、出席回数が少ない学生の中にも興味度を高いと回答した人もいる。しかし、出席回数が低い学生においては、興味度と試験の成績の間に相関はみられなかった。つまり、興味度が高くかつ出席回数が多い人のみが試験勉強をしたということが伺える。出席回数が多い学生とそうでない学生の違いは、田中・藤田の述べたマスタリー目標における違いがあったと考えられる。出席回数が多い人は、マスタリー目標が高く、そのため、興味があるといった授業評価は高く、それが試験成績を高くしたと考えられる。

## 5. 総合考察

### 5.1 授業評価と成績との関係

本研究では、学生による授業評価の妥当性を検討するために、試験の成績と授業評価の関係を見てきた。

しかし、実際には授業評価と試験の成績の間にはいくつかの統制できなかった背景要因が関与しており、単純な関係性をみることはできなかった。ただし、そこで明らかになったことは、以下のような点である。理解に対する認識が必ずしも正しくはないこと、興味が高いことが試験の成績を向上させるという点である。さらに、これらの関係性は出席回数という要因によって異なるということである。

背景要因としてまず問題になるのは、評価を行った時点と試験を行った時点での時間的なずれである。評価後（授業受講後）に勉強をすることによって試験の成績は異なったものとなる。授業を受けて理解できていなくても勉強することによって試験の成績は向上する。逆に授業時に理解していても復習などの試験勉強をしなかったために、授業で聴いた内容を忘れてしまい、試験の成績は悪かった可能性もある。全学生での理解と試験の成績でほとんど相関が見られなかったのはこのような事情によるものであろう。

さらに問題なのは、理解に対する認識の正しさの問題である。学生の回答が必ずしも額面通りではないということである。理解していると回答していたとしても、わかつたつもりになっているだけで、実際にはわかっていない可能性もある。また、ある程度理解していても、逆にわからないところが顕在化してしまい、理解の評定として相対的に低く評価されてしまうこともある。本調査の理解度については、出席回数をもっとも多い学生のほうが相対的に低い評価であった（表3）。にもかかわらず、試験の成績はもっとも高い点数を示していた。内省として理解できなかったという回答をしたのは、まったくわからないということではなく、疑問点が多く出てきたということであり、問題意識を持つことができるようになったとも考えられる。

一方、興味に関しては、興味があることが学習意欲を動機づけると解釈することができた。試験の成績との関係は、興味度の関係のほうが理解度よりも高い結果を示した。授業を受講後、勉強して理解できるようになり、試験の成績が高くなった可能性が考えられる。ただし、興味との関係において試験の成績と正の相関を示したのは、とくに出席回数が8～10回という学生であった。出席回数が多い学生は、もともと授業に対して積極的であり、マスタリー目標が高かったのではないかと推測される。田中・藤田が報告しているように、マスタリー目標が高いことが授業評価を高くしたことにつながり、試験の成績も高くなったのではないかと考えられる。

試験の成績と授業評価の関係について最初に4つのモデルを挙げ、そのうち2つのモデルのteaching effectiveness modelとprior characteristics modelのいずれかで本調査の結果が説明できるのではないかと考えられた。本調査の結果からは、教員の教え方ではなく、学生がもともと持っている特性に左右されていると考えることができる。

## 5.2 学生による授業評価の妥当性

学生による授業評価に関しては多くの研究がなされてきている。そして、必ずしも一貫した結論が示されているわけではない(Gaski, 1987; Greenwald & Gillmore, 1997; Algozzine et al., 2004)。学生による授業評価の結果には、様々な要因が複合的に関係しているからである。個々の論文では、個々の結論を最終的に提起してはいるものの、多くの論文では、それぞれの適用限界を最後に論じている。ある条件が整ったときには、ある結果を導くが、現実の授業評価においては、個々の論文で統制できていない要因に左右されてしまう。多くの要因が関与していることを論じて、それらの要因を統合的にまとめあげたグローバルなモデルの提案も試みられているが(Abrami, d'Apollonia & Cohen, 1990; Marsh & Roche, 1997)、それらですべてが説明されるわけではない。

学生による授業評価では、評定尺度による選択回答と自由記述がなされるが、その妥当性に関しては、評定尺度の場合と自由記述の場合では分けて考える必要がある。本研究では、評定尺度による評価の問題を取り扱った。評定尺度の場合、学期末に実施され、その評定値の平均値などの代表値が教員にフィード

バックされ、それを授業改善の資料として役立てたり、その評定平均値で教員の教え方を査定することになる。本研究では、1学期間を振り返っての回答が容易でないという問題点を解消するために、毎時間、授業評価を行った。しかし、90分の授業の中のどの内容についての理解・興味を問うのかまでは特定できなかった。そのことが各設問と当該授業での相関において一貫した結果が得られなかった理由でもあった。

さらに、試験成績との関係を見たところ、理解に対する学生の認識はまちまちであり、学生の自省にすぎない回答は、正しく学生の理解度を反映しているわけではないことがわかった。理解度が高いという評定結果が出たとしても、それは表面的に理解しているだけであったり、わかったつもりになっているだけかもしれない。そのような評価結果によって、授業そのものが高く評価されたり、教え方が高い評価を受けてしまうと、誤った判断をしてしまいかねない。そして、授業評価と試験の成績の関係については、teaching effectiveness modelではなくprior characteristics modelで説明されるのではないかということが示唆された。つまり、授業評価に影響を与えているのは、教師の教え方ではなく、学生の特性要因であるという結果である。

学生による授業評価の回答は、実際には理解していなかったり、あるいは、精緻化が十分になされていない形でのわかったつもりの理解であったりすることもあり、表面的な学生の評定値で授業を評価することは危険である。必ずしも妥当ではない学生の授業評価の評定値を高くすることが目標となってしまうと、授業本来の目標を失いかねない。

### 5.3 教員評価における有用性

学生による授業評価によって教員の教え方の評価を行う場合、評定尺度による学生の回答を平均化した値で判断されるのが一般的である。その数値によって、教え方が適切であるのか評価されたり、あるいは、他の教員との比較がなされる。

しかし、学生による授業評価の数量化されたデータは、複合的要因によるものであるため、評価値で教員の教え方を評価することは困難である。授業評価研究の流れを見ても、学生による授業評価の教員評価への利用においては慎重でなくてはならない。研究としてなされる場合は、いくつかの要因を統制した結果であり、何が統制できていて何が統制できていないのかは、ある程度捉えることはできる。しかし、実際の現場で利用された場合、そこで出された授業評価の結果にどのような要因が関連しているのか知ることとは極めて困難である。したがって、その結果を教員の人事査定などに利用すること自体は、科学的であるとは言えない。

さらに、平均化した値で評価することも問題である。授業の良し悪しは、平均値で表されるわけではない。少数の学生でも、興味を示した学生がいれば、授業としては優れていると考えられる。本研究で明らかになったように、興味を示すことが勉強への動機づけを促進することがわかったが、その授業に対してすべての学生がマスター目標を持っているわけではないと考えられる。カリキュラム上、学生はやむを得ず受講しなければならない授業科目は少なからず存在している。もともと興味関心を持たずに受講する学生は決して少なくない。平均化された値で評価されてしまうと、興味を持った学生が多くいる科目が高く評価されるだけであって、担当教員の教え方の評価ではなくなってしまう。

### 5.4 授業改善における有用性

学生による授業評価の評定尺度による結果を直接授業改善に役立てることは困難である。最初に述べたように、評価値の代表値の値に対する判断基準があるわけではなく、授業改善が必要なかどうかの判断を行うことは難しい。

ただし、学生によって評価されるということ自体が授業改善をもたらすことは考えられる。評価結果そ

のものはほとんど有用な情報にはならないが、評価されることが授業改善の動機づけとなる。授業は、常に何らかの反省材料を抱えている。授業評価を受けなくても、実際に授業を行っていれば、どこに問題があったのかわかることが多い。ただし、実際に次年度以降の改善につなげるのかどうかは別問題である。改善しなければならないという気持ちを持っていながら、前年度と同じ授業を行ってしまうことがある。しかし、学生から評価されるとなると、改善しようという動因が高まることが考えられる。授業評価を実施することによって授業評価の評価値が上昇したといったことが報告されているが(石浦, 2005)、それは、このような動機づけの効果によるものと考えられる。

したがって、学生による授業評価を導入し始めた当初は、評価値が上昇することは考えられる。しかし、授業評価をある程度継続していくと、上昇はなくなるであろう。もちろん、評価値が高い値で安定することは好ましいことであり、高い値を維持していくために、学生による授業評価を継続的に行うことは効果的であるかもしれない。ただし、学生による授業評価が妥当なものであるという前提が必要である。しかし、これまで述べてきたように学生による授業評価は必ずしも妥当なものではない。評価値を高くすることが目的となってしまうと、授業評価で評価される項目だけを改善することが目標になってしまう。テクニカルに評価値を上げることが目標となってしまう、目先の目標にとらわれた授業になってしまう。学生に迎合するような授業になってしまう可能性もあり、授業が本来持っている理念や目標とは異なるところで授業改善を行ってしまいかねない危険性を持っている(サックス, 2000)。

授業改善にとって必要なのは、具体的にどこの内容がどのように理解できなかったのか、どのように興味を持たなかったのかであり、それらがわからなければ、どのように改善してよいかかわからない。学生の評価による評定値を見ただけでは何も情報は得られない。評定値の高低も学生側の要因が大きい(松尾・近藤, 2005a) ことが考えられ、直接的には役に立たない。

本調査で用いたコミュニケーションカードは、本来学生が理解ができなかった点を教員に伝えるためのカードである。具体的に学生がどのような点がどのようにわからなかったのかわかるため、次回の授業で補足説明をしたり、内容や教え方の改善に役立てることができる。今回の理解や興味の評定は、付加的に尋ねたにすぎず、評定を目的としたものではない。授業中に実際に学生に何がわからなかったのか、興味を持たないのかを尋ねればよいかもしれないが、口頭で伝えることに対して躊躇があったり、そのための時間がとられるため、書いてもらうやり方が効率的であるにすぎない。

本質的には、学生とのコミュニケーションが必要であるにすぎない(松尾・近藤, 2005a)。学生とのコミュニケーションができれば、「授業評価」といったスタイルをとる必要はない。

#### 参考文献

- Abrami, P.C., d'Apollonia, S., and Cohen, P.A. 1990 Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.
- 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.
- Bausell, R.B. and Magoon, J. 1972 Expected grade in a course, grade point average, and student ratings of the course and instructor. *Educational and Psychological Measurements*, 32, 1013-1023.
- Brown, D.L. 1976 Faculty ratings and student grades: A university-wide multiple regression analysis. *Journal of Educational Psychology*, 68, 573-578.
- Centra, J. A. 2003 Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44, 495-518.

- Cohen, P.A. 1981 Student ratings of instruction and student achievement: A meta analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Eiszler, C. F. 2002 College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43, 483-501.
- Feldman, K.A. 1989 The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 583-645.
- Gaede, C.S. and Lewis, K.G. 2005 Considering the validity of the interpretation of student ratings of teaching. *MEC Online Reports*.  
Available at: <http://www.utexas.edu/academic/mec/publication/reports.html>
- 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.
- Howard, G.S. and Maxwell, S.E. 1980 The correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72, 810-820.
- Hoyt, D.P. 1973 Measurement of instructional effectiveness. *Research in Higher Education*, 1, 367-378.
- 石浦章一 2005 東大教授の通信簿—「授業評価」から見えてきた東京大学 平凡社
- Johnson, R. 2000 The authority of the student evaluation questionnaire. *Teaching in Higher Education*, 5, 419-434.
- Koon, J., and Murray, H.G. 1995 Using multiple outcome to validate student ratings of overall teacher effectiveness. *Journal of Higher Education*, 66, 61-81.
- Marsh, H.W. 1984 Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H.W. and Roche, L.A. 1997 Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52, 1187-1197.
- Marsh, H.W. and Roche, L.A. 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, 202-228.
- 松田文子・三宅幹子・谷村亮・小嶋佳子 1999 学生による授業評価と自己評価, 授業選択態度, 及び成績の関係—教職必修科目「生徒指導論」の場合—. 広島大学教育学紀要第一部 (心理学), 48, 121-130.
- 松尾太加志・近藤倫明 2005a 学生による授業評価は何に役立つのか. 北九州市立大学文学部紀要 (人間関係学科), 12, 51-64.
- 松尾太加志・近藤倫明 2005b 学生による授業評価における学生側要因. 日本心理学会第69回大会発表論文集, 1281.
- McKeachie, W.J. 1997 Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
- Nelson, J.P. and Lynch, K.A. 1984 Grade inflation, real income, simultaneity, and teaching evaluations. *The Journal of Economic Education*, 16, 21-37.
- 西林克彦 2005 わかったつもり 光文社
- Norman, D.A. and Bobrow, D.G. 1975 On data-limited and resource-limited processes. *Cognitive Psychol-*

ogy, 7, 44-64.

サックス, P. 後藤将之(訳) 2000 恐るべきお子さま大学生たち—崩壊するアメリカの大学 草思社  
(Sacks, P. 1996 *Generation X Goes to College: An Eye-Opening Account of Teaching in Postmodern America*. Open Court Publishing Company.)

Sproule, R. 2000 Student evaluation of teaching: A methodological critique of conventional practices.  
*Education Policy Analysis Archives*, 8 .

Available at: <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.

宇佐美寛 2004 大学授業の病理 東信堂

安岡高志・吉川政夫・高野二郎・峯崎俊哉・成嶋弘・光澤舜明・道下忠行・香取草之助 1989 学生による  
講義評価—成績と講義評価の関係について—. 一般教育学会誌, 11, 99-102.

# 北九州市立大学文学部紀要

(人間関係学科)

第 13 卷

## 目 次

原 賢治・近藤倫明：ヴィジュアルフォーマット教材の視認性評価 .....	1
亀山徳彦・津枝美・加藤由美子：子どもの遊びの場として魅力的な公園 —小倉南区内の15の公園を例として—.....	11
小賀 久：デンマークにおける障害者の地域生活支援.....	23
黒田耕司：「道德の時間」における評価の特質と課題.....	35
楠 凡之：生活指導実践における「枠づけ」の意義 —「解離」の問題を抱える中学生への指導に視点を当てて— .....	47
松尾太加志：学生による授業評価の妥当性と有用性 —試験成績との関連—.....	63
野井英明・太田泰弘・梅崎恵司：福岡県北九州市黒崎城跡における完新世の環境 変遷 —特に縄文人の環境適応についての—考察—.....	79
須藤 廣：観光文化とアウラの喪失 —ポスト・モダニズム論から観光文化へのアプローチ—.....	89
田島 司：公認される社会的カテゴリーへの所属と自己の社会的定位について .....	99
田中信利・黒木美沙・税田慶昭：共同注意行動の出現時期と発達の順序性 .....	107

---

北九州市立大学文学部

2006年3月発行

