

学生による授業評価の信頼性と妥当性に関する検討

南 学

問 題

近年、日本の大学をめぐる状況の大きな変化にともなって、各大学で学生による授業評価が注目・実施されるようになってきた。大学も教育に力を入れな
いといけないと言う考えが広まってきたのだと思われる。そのためには、まず
学生自身の声を聞くところから始めようという考えだろう。この背景には、少
子化やそれにとまなう経営環境の悪化、いわゆる「学力低下」論、学生気質の
変化（井上，1993）などがある。今後もしばらくは学生による授業評価が広ま
る傾向が続くと予想される。

しかし、学生による授業評価は、必ずしもすべての大学教員に歓迎されてい
るものではない。その最大の理由の1つとして、「学生による授業評価は信頼
できない」というものが挙げられる（大槻，1993；住田，1996；安岡・及川・
吉川・山本・高野・光澤・香取，1994 a）。しかし、これとは正反対の意見も
あり（守・野口・筒井・川島・小松，1996；安岡ら，1994 a；安岡，1999），
議論は錯綜している。議論が整理されない理由の1つには、「学生による授業
評価が信頼できない」というときの「信頼」の概念が混乱していることがあり、
もう1つに、実証をふまえず観念論に終始している点が考えられる。そこで、
まず授業評価に対する信頼性・妥当性の概念の整理をおこなう。

心理学的な評価理論の観点からは、信頼性とは、同一の対象に対して、同一
の条件下で同一の評価テストを実施した場合に、同一の結果がでる程度のこと
を指す。それに対して、妥当性とは、測定対象を適切に測っている程度のこと

を指す。一般的に「学生による授業評価が信頼できない」という場合、信頼性ではなく、妥当性を指していると考えられる。また、その「信頼」が評価者である学生自身に向けられる場合は、上の信頼性とは異なるものを指している。そこで、以下ではこの場合を信用性と呼ぶことにする。信用性と妥当性は必ずしも対応しているとは限らない。たとえば妥当性が高くても信用性は低いということは十分にあり得る。

教員の側からの学生による授業評価の妥当性への疑問については、数多く目にするところである（井上，1993；安岡ら，1994 a，1994 b；住田，1996；安岡，1999）。ただし、教員への意識調査の場合、これらの指摘は主観的なものになることが多く、必ずしも根拠を持ち、実体を反映しているとは限らない。実際安岡ら（1994 a）では、予想よりも評価が高かった場合とそうでない場合を比較し、低い場合に妥当性への疑問の度合いが強まることを示している。このように、学生による授業評価に対するこうした批判は、真に授業評価に対する疑義を反映しているのではなく、自らの授業の評価が悪いことへの言い訳、あるいは授業評価の実施を避ける口実として用いられている可能性がある。

他方、実証的検討をおこない、授業評価が別の要因によって変化するという意味で、その妥当性に疑問を呈する知見もある。たとえば、教師に対する好感度が高いほうが授業に対する好感度を高く評価する（松本，1996）、男子学生に比べて女子学生のほうが概して授業評価が高い（榊原，1993；松田・三宅・谷村・小嶋，1999）、成績の良い者ほど概して授業評価が高い（榊原，1993；松田ら，1999；ただし安岡・吉川・高野・峯崎・成嶋・光澤・道下・香取，1990では相関なし）などの傾向があることが確認されている。しかし、これらが安定した傾向であるならば、技術的に修正可能であるはずで、妥当性を否定するものとはいえない。

評価者としての学生に対する信用性に関する疑問も根深いものがある。たとえば安岡ら（1994 a）によると、学生による授業評価を信頼できない理由として、「学生が不真面目である」という意見が18.1%選ばれている。これは具体

的には「すべての項目で1をつけるような学生がいる」というような意見に代表されると考えられるのだが、こうした意見には合理性がない。なぜなら、大講義などでは確かに一部にこうした学生が現れることは認めざるをえないかもしれないとしても、全体の評価を下げるほどに多くはないという事実を無視しているからである（安岡ら，1994 a）。

安岡ら（1994 a）によると、つぎに「学生に評価能力がない」という意見が17.3% 選ばれている。これに対して安岡らは、学生に評価能力がないのは学生の側の基礎学力不足などのほかに、教員側にも問題があるのではないかと反論している。しかし、これは十分な反論とは言えないだろう。本当に学生に評価能力がない、すなわち学生による授業評価は信頼性がないのか否かを問わずに、評価能力がないことの責任がどちらにあるかということだけを問うているだけであるからである。

このように、学生による授業評価をめぐる議論は、実証的検討がないままに進められていることが多く、それゆえ議論が錯綜したり、水掛け論に終始したりすることになる。とくに妥当性に関する議論は、少なくとも信頼性があることを検証していないのであれば意味がない。ある程度の信頼性を満たすことが妥当性を論じる上での前提条件となるからである。

学生による授業評価の信頼性を検証した研究は、多くない。安岡・高野・成嶋・光澤（1986）が総合評価について2年間の比較を、臼井（1994）が学部全体の2年間の比較をおこなっているにすぎない。そこで、本研究では、学生による授業評価の信頼性・妥当性に関する実証的な検討をおこなう。妥当性の議論をする前に信頼性について確認しておく必要があるからである。そのために、1つには、同じ科目について2年にわたって同一の授業評価をおこない、全質問項目の評価のパターンを比較することを通して、その結果の安定性について検討する。もう1つは、いくつかの指標との関係を比較することを通して、妥当性について検討する。

方 法

被験者 私立文系大学（経済，経営，人文，法の4学部がある）で心理学と教育心理学を受講する大学生で1999年度は243名，2000年度は224名（それぞれの年度内で重複する受講生はいなかったが，年度間では多少いる）。受講生の男女比は約半数ずつであるが，質問紙では尋ねていなかったため，正確な割合は不明である。なお，両科目は同一教員が担当しており，心理学は主として全学部の1年生を対象としているのに対し，教育心理学は全学部の2年生以上の教職課程履修者を対象としている科目である。

手続き 両科目の講義最終回に，質問紙を配布し，回収した。回答時間は約15分であった。まず，授業の方法や内容に関する質問項目34個（うち，1項目は総合評価）に対し，それぞれ5段階評定で回答するよう求めた。その際，まず単位が認定されるという想定での評価をおこない，続いて単位が認定されないという想定での評価を求めた（ただし，本研究では分析の対象としない。この点についての分析は南（執筆中）を参照のこと）。34個の質問項目は，従前から実施されている複数の授業評価の項目を参考にして選んでいる。

次に，1999年度生に対して，授業に対する学生の姿勢として，(1)出席回数，(2)授業以外の時間における指定教科書の読書頻度，(3)授業中に紹介した文献の読書量，(4)授業に対する態度11項目について尋ねた。(2)，(4)は5段階評定によって回答を求めた。2000年度生には別の質問をしており，今回は分析の対象としていない。なお，調査にあたっては，無記名とし，回答が成績評価に影響を与えることはない点を強調した。

結 果

Figure 1に，2年分の認定／不認定条件ごとの授業の方法や内容に関する34の質問項目の各平均値を示した。Figure 1から，翌年のものとの比較においてきわめて類似していることが見いだせる。例外的に，「授業を静粛に保つ配慮

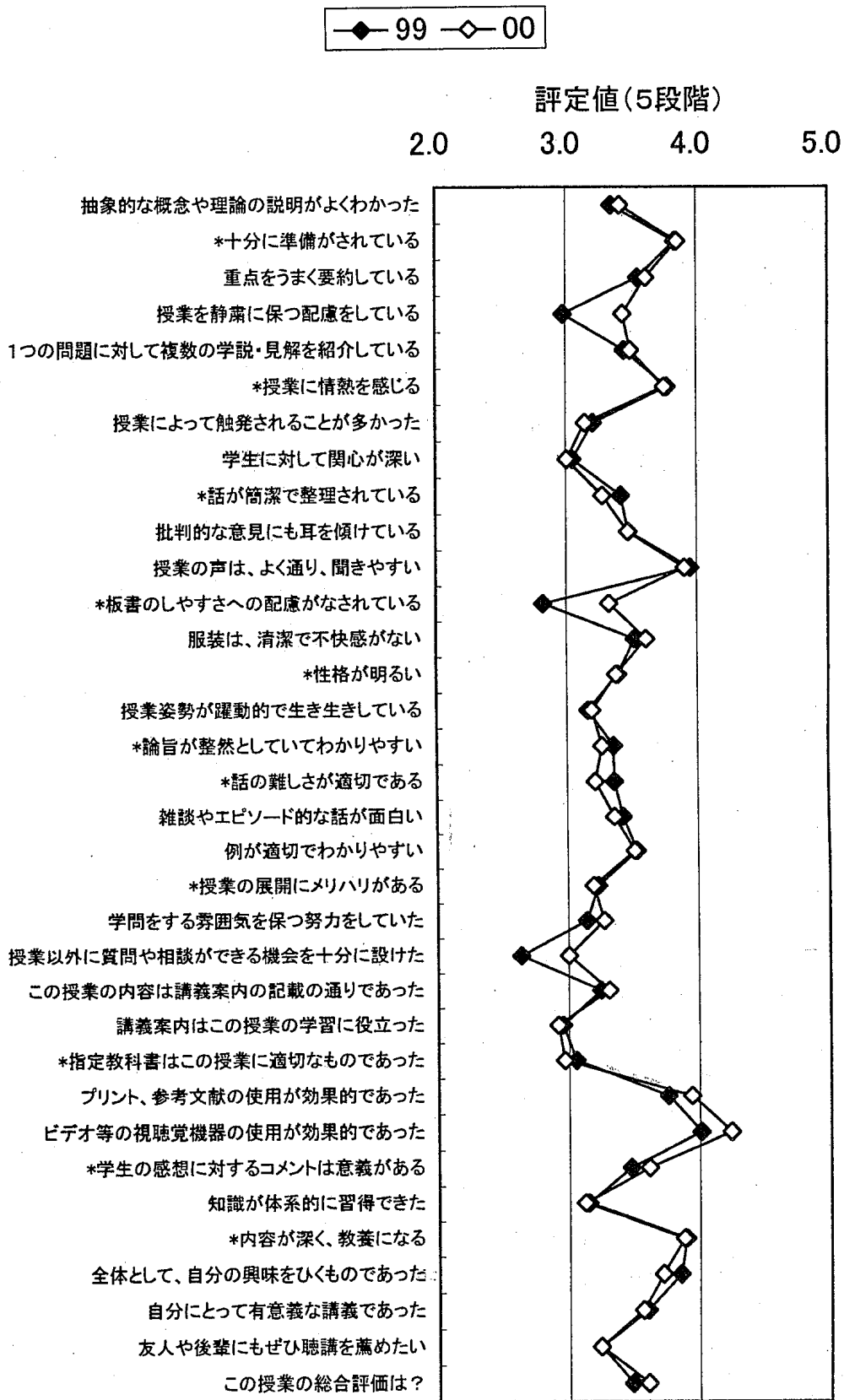


Figure 11 授業評価に関する項目の評価

をしている」「板書のしやすさへの配慮がなされている」「授業以外に質問や相談ができる機会を十分に設けた」などにおいて、1999年度よりも2000年度の評価平均が高くなっている。Figure 2に、学生自身の授業に対する態度11項目の各平均値を示した。Figure 2から、平均値が翌年のものとの比較においてきわめて類似していることが見いだせる。

次に、1999年度のデータを元に、総合評価の項目を除いた授業の方法や内容に関する33項目の回答に関して、単位の認定条件と不認定条件それぞれについて、因子分析(主成分法・バリマックス回転)により7因子を抽出した(累

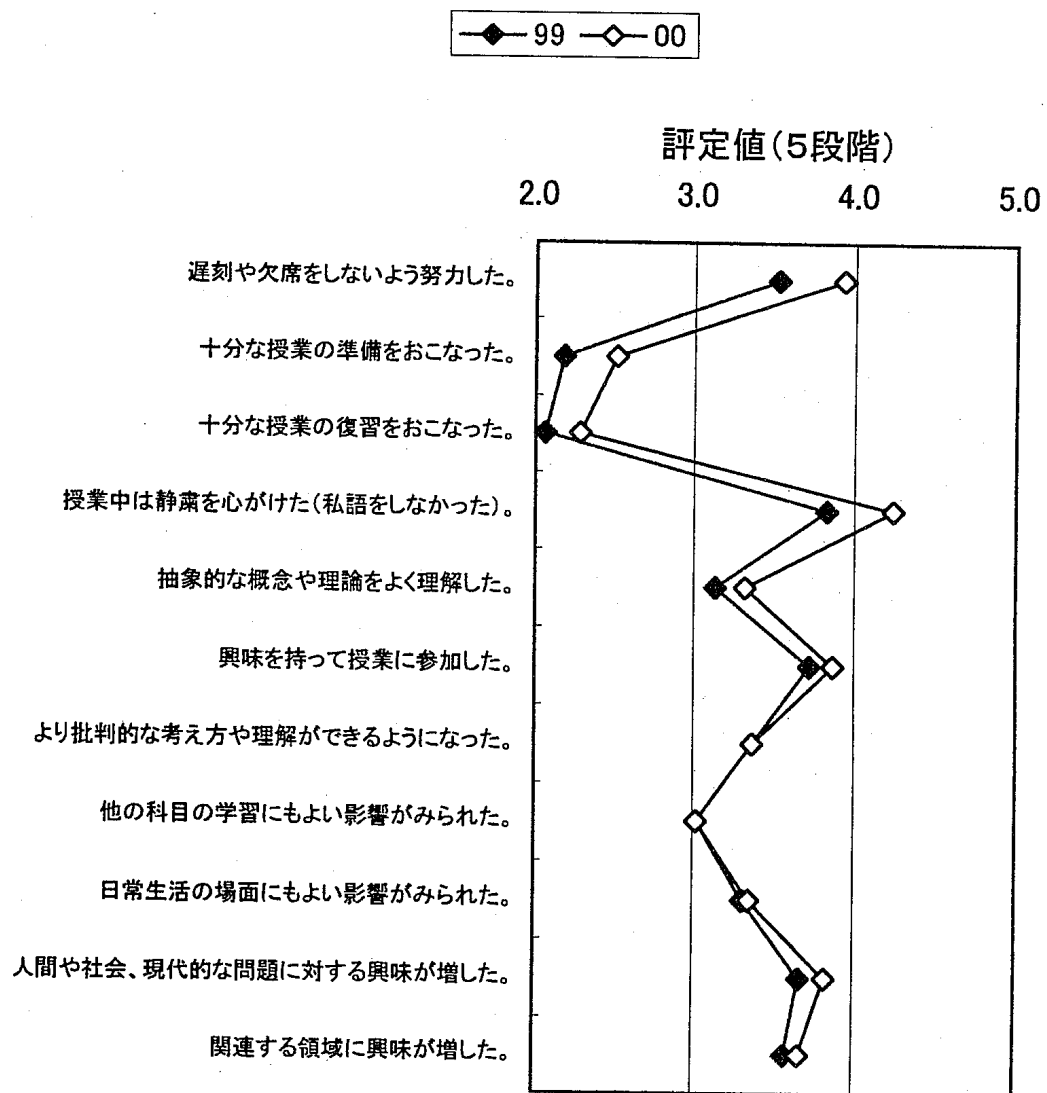


Figure 2 自己評価に関する項目の評価

積寄与率それぞれ 49.2%, 52.4%)。このとき両条件において因子構造が同一であったため、以下では、各因子に共通する項目 (計 24 項目) を用いて分析をおこなった。各因子は、以下のように命名された (Table 1)。因子 1: 習得感, 因子 2: 授業姿勢の前向きさ, 因子 3: 授業の受けやすさ, 因子 4: 授業内容の充実度, 因子 5: 学問的雰囲気醸成, 因子 6: サービス性, 因子 7: 教材の有効性。

Table 1 授業評価の各因子に所属した質問項目

<p>因子 1: 習得感 抽象的な概念や理論の説明がよくわかった。 授業によって触発されることが多かった。 例が適切でわかりやすい。 知識が体系的に習得できた。 全体として、自分の興味をひくものであった。 自分にとって有意義な講義であった。 友人や後輩にもぜひ聴講を薦めたい。</p>
<p>因子 2: 授業に対する姿勢 学生に対して関心が深い。 批判的な意見にも耳を傾けている。 授業の声は、よく通り、聞きやすい。 服装は、清潔で不快感がない。 授業姿勢が躍動的で生き生きしている。</p>
<p>因子 3: 授業の受けやすさ *板書のしやすさへの配慮が足りない。 *性格が暗い。 *指定教科書はこの授業に不適切なものであった。</p>
<p>因子 4: 充実/冗長性 *準備が不十分である。 *授業に情熱を感じられない。 *話がくどくて無駄が多い。</p>
<p>因子 5: 学問に対する姿勢 (アカデミズム) 授業を静粛に保つ配慮をしている。 1つの問題に対して複数の学説・見解を紹介している。</p>
<p>因子 6: サービス性 授業以外に質問や相談ができる機会を十分に設けた。 講義案内はこの授業の学習に役立った。</p>
<p>因子 7: 資料提示の有効性 プリント、参考文献の使用が効果的であった。 ビデオ等の視聴覚機器の使用が効果的であった。</p>

注) *は逆転項目

同様に、学生自身の授業に対する態度11項目についても因子分析（主成分法・バリマックス回転）をおこなったところ、2因子を抽出した（累積寄与率52.3%）。各因子は、以下のように命名された（Table 2）。因子1：授業のとりくみのまじめさ、因子2：知的収穫。

最後に、授業評価の7因子と自己評価の2因子の尺度得点を、学部別、出席回数別、教科書の読書頻度別、文献の読書量別に集計したものをそれぞれ Figure 3-6 に示した。それぞれについて分散分析をおこなった。学部別では因子の主効果 [$F(8, 1880) = 115.412, p < .001$] のみが有意であった。出席回数別では、出席回数、因子の主効果 [それぞれ $F(2, 239) = 14.898, p < .001$; $F(8, 1912) = 92.464, p < .001$]、および交互作用 [$F(16, 1912) = 2.005, p < .05$] が有意であり、授業因子6, 7, 自己因子1, 2で出席回数の単純主効果がみられた。教科書の読書頻度別では、読書頻度、因子の主効果 [それぞれ $F(2, 239) = 9.039, p < .001$; $F(8, 1912) = 110.516, p < .001$] が有意であった。文献の読書量別では、読書量、因子の主効果 [それぞれ $F(1, 240) = 21.550, p < .001$; $F(8, 1920) = 101.763, p < .001$]、および交互作用 [$F(8, 1920) = 4.167, p < .001$] が有意であり、授業因子1, 自己因子1, 2でとくに読書量の大きい単純主効果がみられた。

Table 2 自己評価の各因子に所属した質問項目

因子1：授業のとりくみのまじめさ
遅刻や欠席をしないよう努力した。
十分な授業の準備をおこなった。
十分な授業の復習をおこなった。
因子2：知的収穫
授業中は静粛を心がけた（私語をしなかった）。
抽象的な概念や理論をよく理解した。
興味を持って授業に参加した。
より批判的な考え方や理解ができるようになった。
他の科目の学習にもよい影響がみられた。
日常生活の場面にもよい影響がみられた。
人間や社会、現代的な問題に対する興味が増した。
関連する領域に興味が増した。

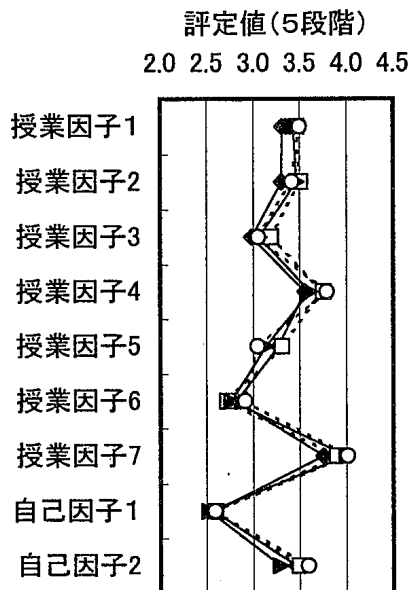
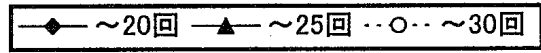
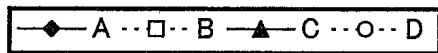


Figure 3 学部別の因子得点

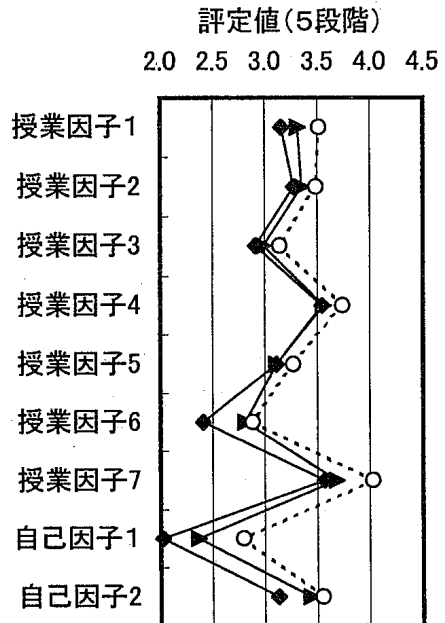


Figure 4 出席回数別の因子得点

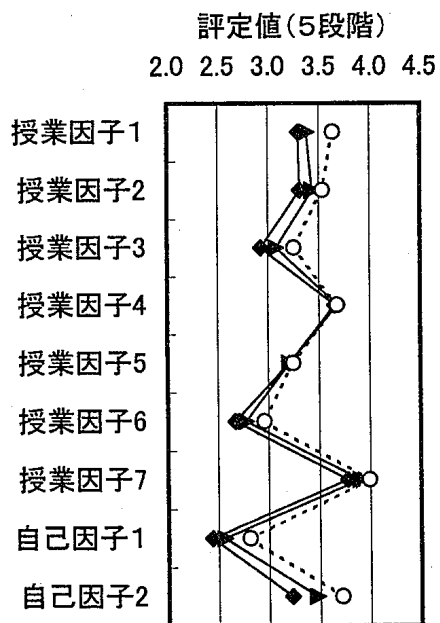
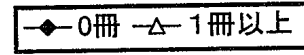


Figure 5 教科書の読書頻度別の因子得点

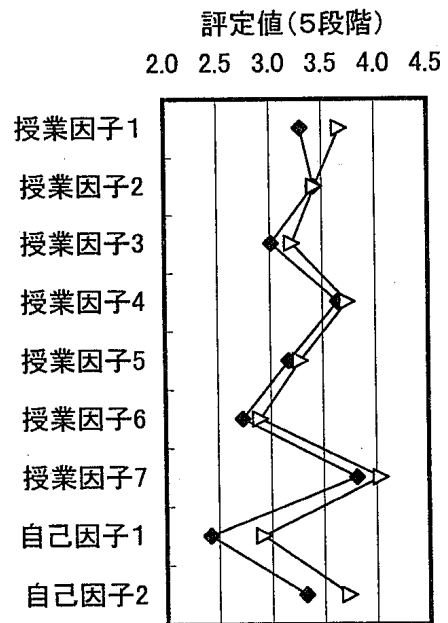


Figure 6 文献の読書量別の因子得点

考 察

授業評価の信頼性

本研究は、学生による授業評価の信頼性・妥当性に関して、実証的に検討することを目的とした。学生による授業評価をめぐる議論が水掛け論や理念の問題になりやすい理由の1つに、とくに信頼性に関する検証がほとんどおこなわれていないということがあるからである。

Figure 1, 2 から、同一教員が同一科目を担当した2年分の授業評価を比較すると、きわめて類似したパターンを示していることがわかる。ただし、Figure 1 では、「授業を静粛に保つ配慮をしている」「板書のしやすさへの配慮がなされている」「授業以外に質問や相談ができる機会を十分に設けた」などの項目において、1999年度よりも2000年度の評価平均が高くなっている。

2000年度の授業では、一部でプレゼンテーションソフトを導入しており、その結果として板書が読みやすくなったこと、ページの切り替えが多くなることから、ノートテイキングの時間が短くなり、私語をする時間的余裕がなくなったのではないかと考えられる。メール機能付きの携帯電話の普及も要因であるかもしれない。また、併せて、Web上で講義情報を提供できるような環境を整えたことが質問や相談の機会が増加したと評価されたと思われる。このように、一部の項目での変化は講義全体の改善によるものと考えられることから、全体のパターンとは別の解釈をすべきだと思われる。

また、補助的指標で被験者を分類した上で、各因子得点の比較をおこなった Figure 3-6 から、いずれの指標の分類であっても授業評価のパターンは酷似していることが示された。この結果からも、授業評価のパターン自体は特定の集団によって形成されたものではなく、授業そのものの評価を反映していると考えられることができる。

以上の結果から、年度間で重複する被験者も多少いるのでその点を割り引く必要はあるが、学生による授業評価の全般的信頼性はかなり高いといえる。す

なわち、学生による授業評価は、教員が授業改善のための情報収集活動の一環として用いたいのであれば十分信頼できる情報源の1つであるといえる。

授業評価の妥当性

授業評価の妥当性を検討するために、いくつかの補助的指標との関連性を検討した。Figure 3 から、学部別の因子得点はほとんど差異がないことが示された。本研究で対象とした講義科目はいずれも全学部を対象とした科目であり、また、いずれの学部の専門領域からも遠い科目であることを考えると、学部間で授業評価に差異が見られないことは、その妥当性を示唆するものであるといえよう。

Figure 4 - 6 から、出席回数、教科書の読書頻度、文献の読書量が増えるにつれて、概して授業評価の評定値が高くなることが示された。南（執筆中）や守ら（1996）、安岡・峯崎・高野・香取（2002）によると、授業の総合評価には、習得感・説明のわかりやすさがもっとも大きな影響を与えていることから考えると、出席回数、教科書の読書頻度、文献の読書量が増えるということは、間接的に習得感を高める方向につながると思われるので、それが結果として授業評価を高く評定することにつながっているのではないかと説明できる。このように考えるならば、これらの指標との関係は、授業評価の妥当性を示唆するものと解釈できる。

以上の結果を総合すると、学生による授業評価はかなり高い信頼性をもつと同時に、妥当性という点においてもそれを支持する結果がえられたと結論づけることができる。よって評価者としての学生の信用性も高いと考えるべきであると思われる。ただし、今回の知見は、1名の大学教員の担当講義科目において得られたものであるので、妥当性の議論のうち、収束的妥当性についてしか検討されていない点は言及しておく必要がある。妥当性の検討には、同一対象を測定する場合には同一の結果が出るという収束的妥当性と、異なる対象を測

定する場合には異なる結果が出るという弁別的妥当性が必要である。今後、弁別的妥当性を検討するために、他の教員が担当している科目に関しても検証が必要であるだろう。

また、今回は専門領域とは離れた科目での検討であったが、学生の専門性を考慮した授業評価の妥当性についても検討される必要がある。授業を適切に評価するためには、学生の側にも一定の力量が求められると考えられる(田中, 1998)。この考えに従うと、学生が専攻する専門科目とそうでない科目では授業評価において一定の傾向が見られる可能性があるからである。授業改善のために学生による授業評価を用いる場合には、こうした点についても検討しておく必要があるだろう。

引用文献

- 井上正明 1993 学生による授業評価の方法論的考察—大学の授業評価に関する実証的研究(8)—福岡教育大学紀要第4分冊, 42, 277-291.
- 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—広島大学教育学部紀要第一部(心理学) 48, 121-130.
- 松本恒之 1996 大学生による授業評価の信頼性に関する一考察 東洋大学文学部紀要 教育学科・教職課程編, 22, 175-182.
- 南 学 執筆中 単位の認定/不認定が授業評価に及ぼす影響
- 守 一雄・野口宗雄・筒井健雄・川島一夫・小松伸一 1996 学生の授業評価による望ましい大学授業の特質の解明(2)—総合評価を目的変数とする重回帰分析—信州大学教育学部紀要, 89, 65-73.
- 大槻 博 1993 多摩大学の学生による授業評価「ボイス」をめぐる考察 一般教育学会誌, 15, 47-49.
- 榊原禎宏 1993 学生による授業評価の特質と授業改善の課題—大学での授業評価の構造に関する一考察— 山梨大学教育学部研究報告, 44, 204-211.
- 住田幸次郎 1996 学生による「授業評価」に関する数量的分析 ノートルダム女子大学研究紀要, 26, 23-40.
- 田中幸代 1998 大学教員に求められる教育力向上のために—教育心理学が検討できる問題の展望—教育心理学研究, 46, 473-483.

- 臼井義春 1994 授業アンケートの実施とその結果 一般教育学会誌, 16, 55-60.
- 安岡高志 1999 それはたった四人から始まった 安岡高志・滝本 喬・三田誠広・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, 7-64.
- 安岡高志・及川義道・吉川政夫・山本銀次・高野二郎・光澤舜明・香取草之助 1994 a 学生による授業評価の信頼性に対する教員意識の調査 東海大学紀要 教育研究所, 2, 87-270.
- 安岡高志・及川義道・吉川政夫・山本銀次・高野二郎・光澤舜明・香取草之助 1994 b 東海大学の自己評価の基本方針および学生による授業評価の信頼性に関するアンケート調査 一般教育学会誌, 16, 51-61.
- 安岡高志・高野二郎・成嶋 弘・光澤舜明 1986 学生による講義評価 一般教育学会誌, 8, 50-60.
- 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 1990 学生による講義評価—成績と講義評価の関係—東海大学紀要 教育研究所教育工学部門, 3, 49-52.
- 安岡高志・峯崎俊哉・高野二郎・香取草之助 2002 授業アンケートにおける学生の達成感と総合評価の関係 大学教育学会誌, 24, 123-126.

※1 この研究の一部は、松尾浩一郎（聖カタリナ女子大学社会福祉学部助教授）との連名で、日本教育心理学会第42回総会（2000年）において発表されている。

※2 この研究は、平成13年度松山大学特別研究助成を受けている。