

〔論 説〕

授業評価アンケートデータから見えてくるもの — 講義科目に焦点をあてた分析 —

聞 間 理・菅 野 禎 盛

要 旨

本報告は、平成16年度に全学で実施された授業評価アンケートデータを分析したものである。専門講義科目および総合講義科目に焦点を当てて分析を行った結果、現行のアンケート項目では、授業内容の評価に関わる設問12～設問22の評定値の相関が非常に高いため、授業の総合評価を良くするためにはどのような点を改善すべきかといった詳細な分析をすることは難しいことが分かった。また、設問12～設問22の評定値を平均した「総合スコア」（授業の総合評価に相当する指標）に対して受講人数と教室の規模が及ぼす影響について分析した結果、受講人数が100人までは受講人数の増加にしたがって「総合スコア」は単調下降し、100人を超えると下がりきってフラットになった。また、授業間の「総合スコア」のばらつきを調べると、受講人数が100人を超える場合は、授業ごとに「総合スコア」の差がはっきりと出てしまうことが示唆された。さらに、授業が実施される教室の規模も「総合スコア」に影響を及ぼしており、教室の収容人数が100人を超えると「総合スコア」が「良い（評定値4）」を割り込むことが示された。

1 はじめに

授業評価アンケートは大学教育を充実させるための有力な方法と考えられており、多くの大学で行われるようになってきた。九州産業大学も例外ではなく、授業評価アンケートを行っている。授業評価アンケートは授業内で行われ、回収された授業評価アンケートは集計されて、各授業の担当教員に集計結果表が送られている。その集計表に含まれているのは、アンケートに設けられた「各項目（22項目）の全学平均値」と「各項目（22項目）の当該授業科目の平均値」、および「自由記述に書かれたコメントの一覧」である。授業評価アンケートによって、各教員が集計表のデータを参考にして、担当授業の充実を図ることが期待される。

しかしながら、現在採用されている集計方法では、授業の形態や内容による違いによって生じる影響などが考慮されておらず、各教員が担当授業の充実を図ろうとしたときに、参考にし難いものとなってしまっている。また、授業評価アンケートの結果を、学部レベルでの授業計画に反映させていくことも可能であると考えられる。しかし、現在採用されている集計方法で

はどのように授業計画に生かしていけばよいかが見えてこない。

本報告は、学部長の許可のもと、平成16年度の授業評価アンケートのデータを分析したものである。この分析が、各授業の運営および経営学部の授業計画に対して、何らかのインプリケーションを示すことができれば幸いである。

2 調査概要

対象授業 平成16年度における九州産業大学経営学部の前期・後期の科目

調査期間 平成16年6月18日～7月2日（前期）、平成16年11月30日～12月14日（後期）

調査科目数 552科目（平成13年度にカリキュラムの変更が行われているため、一部の科目においては旧カリキュラムと新カリキュラムとの合同クラスになっている。その場合には、同じ科目として扱っている。）

回収調査票数 19021部

2.1 教示および設問項目

授業の担当教員は、学生に対して以下のような教示を行うことになっている。

このアンケートは、授業担当教員が今後の授業の方法や内容の改善を考えていく上で参考資料とし、教育内容の充実を図ることを目的に実施するものです。個人のアンケート内容が、この目的以外に使用されることはなく、あなたが不利益を被ることはありません。あなたの素直な気持でお答え下さい。ただし、記入にあたっては誠意をもって行ってください。

調査票の設問項目を表1に示す。学生は、自由記述を除いた全ての設問に対して、以下のような5段階で評定を行っている。

- 評定値5 とてもそう思う（大変良い）
- 評定値4 ややそう思う（良い）
- 評定値3 どちらともいえない（普通）
- 評定値2 あまりそう思わない（やや悪い）
- 評定値1 全くそう思わない（悪い）

表1. 調査票の設問項目

1. この授業の履修について	
設問1	履修した動機は、学習目標のために役立つ科目だったから
設問2	履修した動機は、講義概要（シラバス）を読んで興味を持ったから
設問3	履修した動機は、就職のために役立つ科目だったから
設問4	履修した動機は、友人あるいは先輩からの勧めがあったから
設問5	履修ガイドの講義概要（シラバス）はわかりやすくよい内容でしたか
設問6	この授業の受講生数は授業形態（内容等）からみて妥当と思いますか

2. あなたの受講態度について	
設問7	予習や復習をして授業に臨みましたか
設問8	授業に遅刻することなく受講しましたか
設問9	授業中に私語や居眠りをすることなく受講しましたか
設問10	質問等を積極的に行いましたか
設問11	出席率はどうでしたか

3. 授業の内容と進め方について （講義科目用）	
設問12	テキスト・配布資料または情報機器は授業の理解に役立ちましたか
設問13	板書・提示された教材は見やすかったですか
設問14	声や言葉は明瞭で聞き取りやすかったですか
設問15	授業は履修ガイドの講義概要に沿って行われましたか
設問16	専門知識や用語の説明は理解できましたか
（実験実習演習科目用）	
設問12A	（実験実習科目用）実験実習機器等の操作について、丁寧に指導されていましたか
設問12B	（演習科目用）報告、レジュメ、レポートの準備・作成・内容について、丁寧に指導されていましたか
設問13	出される課題（作品）またはレポートの量は適度でしたか
設問14	学生からの質問・意見に対して真剣に応答してくれましたか
設問15	授業は受講生の反応をみながら進められましたか
設問16	専門知識や用語の説明は理解できましたか

4. 担当教員について	
設問17	教員は授業開始・終了の時間を守りましたか
設問18	教員は熱意をもって授業を行いましたか
設問19	教員は学習目標をはっきり示しましたか
設問20	教員は私語・携帯電話の使用等によって授業が妨げられない環境づくりに配慮しましたか
設問21	教員は一方的な授業ではなく、学生の参加（質問、発言等）を促しましたか

5. 総合評価について	
設問22	この授業の総合評価をして下さい

6. 授業で興味深かった点、役に立った点、改善すべき点について自由に記述してください。（建設的な意見を書いてください。）	
--	--

2.2 授業区分の設定

授業の形態および内容により授業評価の結果に質的に大きな違いがあると考えられるため、経営学部の授業科目を以下の6つのカテゴリに分類し、この6つのカテゴリのうち、講義科目（専門講義と総合講義）に焦点を当てて分析を行った。各授業区分の科目数および回収した調査票数を、表2に示す。分析の対象となる講義科目（専門講義と総合講義）の科目数は187科目、調査票数は13146部である。

専門講義 専門科目のうち、講義形式の授業が含まれる。

専門演習 専門科目のうち、ゼミナールを除く少人数演習形式またはパソコンを使う実習形式の授業が含まれる。

ゼミナール プレゼミナール、ゼミナールⅠ、Ⅱ、Ⅲ、Ⅳ、が含まれる。ゼミナールは、演習科目の中でも特に教員と学生のコミュニケーションが密であることが多く、その他の演習科目とは異質であると考えたので「専門演習」から分けて別カテゴリとした。

総合講義 総合科学科目のうち、体育実技を除く講義科目が含まれる。

スポーツ 体育実技科目が含まれる。

外国語 外国語科目が含まれる。

表2. 各授業区分の科目数と調査票数

授業区分	科目数	調査票数
専門講義	124	8992
総合講義	63	4154
専門演習	55	1355
ゼミナール	150	1712
スポーツ	8	269
外国語	152	2539
合計	552	19021

3 調査データの分析

3.1 予備的分析

授業ごとにアンケート回収数が異なることから、評定値をそのまま分析した場合は授業規模（履修人数）の影響が出てしまう。そこで、授業ごとに評定値を平均し、この平均評定値に対して分析を行った。分析は、「専門講義」と「総合講義」に絞った。

設問12～設問22の評定値の相関分析 設問1～設問11までは、学生の履修動機や授業態度に関するものであるのに対して、設問12～設問22は授業内容の評価に関するものである。本報告の目的は、授業内容をいかにして改善できるかを提案することにあるため、学生の動機づけや授業態度に関する設問1～設問11については、分析を保留し、設問12～設問22に焦点を当てて分析を行った。

設問12～設問22の間の相関分析を行ったところ、これらの設問項目の間の2変量間の相関はいずれも非常に高かった(表3)。これらの設問のうち、設問22は「総合評価」であることから、設問12～設問22はすべて「総合評価」として統合可能である可能性が考えられる。統合可能かどうかについては後述する信頼性分析で改めて検討する。

一方で、設問間の相関が高いことは、例えば「総合評価(設問22)」に対して「板書の見易さ(設問13)」と「教員の熱意(設問18)」のどちらの要因が大きく影響しているか、などといった詳細な分析が困難であることも示している。具体的には、「総合評価(設問22)」の評定値(従属変数)を、「板書の見易さ(設問13)」と「教員の熱意(設問18)」の評定値(独立変数)から重回帰式により予測するような場合、2つの独立変数間の相関があまりにも高い場合には、独立変数にかかる偏回帰係数の統計的安定性が低くなるという問題(多重共線性)が生じる(芝、南風原, 1990, p.161)。多重共線性がある場合は、従属変数に対する独立変数の寄与の推定値が信頼のおけないものになってしまうのである。

このように相関が高くなった原因としては、学生がアンケートに対して「いい加減」にしか回答していないことが大きな要因として挙げられよう。図1は、講義科目(専門講義と総合講義)を対象として、調査票ごとの設問12から設問22までの評定値の標準偏差の値の分布をヒストグラムで表したものである。ヒストグラムの階級の幅は0.01でとっている。

図1のヒストグラムでは、階級値0.005に最頻値1431が存在している。階級値0.005には標準偏差0.00のデータしか含まれていない。すなわち、驚くべきことに調査票の約11%(有効回答数12977のうち1431)で設問12から設問22の全てに対して同じ評定値で回答されている(標準偏差が0)のである。このデータこそ、学生が「いい加減」にしか回答していない証拠である。

また、図1のヒストグラムで、最頻値0.005の次に度数が多い階級値は0.525である。標準偏差が0.525となるのは、設問12～設問22までの11個の設問のおよそ半分に対してある評定値をつけ、残りの設問に対してはその評定値と1段階だけ異なる評定値をつけたような場合である(例えば、6つの設問に対して「5」を、残りの5つに対して「4」をつけるような場合)。このことから、多くの学生が設問12～設問22の各設問項目に対して同じような評定値しか与えていないことが分かる。

表3. 授業ごとの平均評定値に関する設問12から設問22の間の相関係数

設問	12	13	14	15	16	17	18	19	20	21
13	.722									
14	.731	.878								
15	.739	.746	.784							
16	.733	.812	.865	.822						
17	.498	.565	.550	.571	.582					
18	.690	.692	.795	.773	.822	.664				
19	.754	.790	.839	.838	.886	.667	.930			
20	.653	.661	.751	.679	.743	.551	.814	.828		
21	.658	.724	.753	.731	.779	.492	.808	.852	.732	
22	.767	.832	.868	.838	.910	.659	.892	.944	.769	.848

専門講義と総合講義のみを対象とした (N=187)。
 全ての相関係数は $p < .001$ で有意。

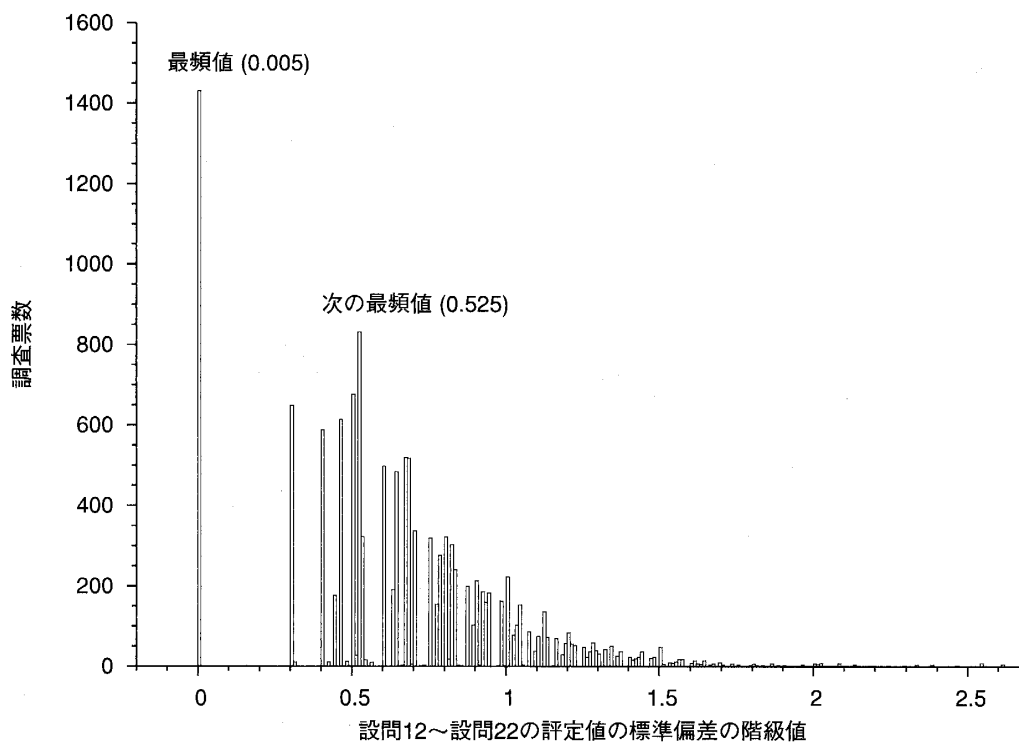


図1. 調査票ごとの設問12から設問22の評定値の標準偏差のヒストグラム。専門講義と総合講義のみを対象とした。ヒストグラムの階級幅は0.01である。

学生がアンケートに対して「おざなり」な評価しかしていないような結果はなぜ生じるのだろうか。まず、同じアンケートを授業ごとに何回もさせられるため、面倒であると感じているのではないかということが考えられる。また、学生にとって授業評価をすることで自らにどんな利点があるのかわからない、ということから積極的に回答する気になれないということもあるだろう。

しかしながら、設問の内容に問題がある可能性も否定できない。設問12～設問22は、一見すると授業の様々な側面を別々に尋ねているようにみえるが、結果として授業の総合評価を繰り返し評定させることになりやすい。例えば、学生がある授業を面白い（良い）と感じていたとして、それは「総合評価（設問22）」だけを高くつけるわけではない。たとえ、その授業の中で教員が少しの板書と少量のプリントしか使っていなくても、結果としてその授業を面白いと感じているのであれば、「配布資料やスライドの活用（設問12）」は上手くて「板書の見易さ（設問13）」は高く、「専門知識や用語の説明の理解（設問16）」もしやすいと答えてもおかしくない。さらに、面白いと感じる授業であれば、学生は集中して聞いていることになって「声や言葉の明瞭さ（設問14）」や「学習目標の明示（設問19）」をより感じると答えることになり、それは学生にとって「一方的な授業ではない（設問21）」ということになりやすいだろう。結果として、そうした教員は学生にとって授業に前向きなように映り、「熱意をもっている（設問19）」と評価されるだろう。

つまり、設問12～設問22は、その設問の内容が同じような評定値を導きやすいものになっているのである。設問12～設問22は、「授業の様々な側面を別個に調査したい」とのアンケートの作成者の意図とは裏腹に、単に「授業の良し悪し」を繰り返し測定しているだけなのかもしれない。もしそうであるなら、各設問に対して同じ評定値をつける学生に「いい加減に回答するな」と要求するのは酷であるといわざるを得ない。

いずれにせよ、現状の授業評価アンケートデータは、授業のおおまかな良し悪しは測定可能であるが、授業の総合評価を良くするためにはどのような点を改善すべきか、といった詳細な分析に耐えられるものではない。

設問12～設問22の評定値の信頼性分析：総合スコア指標の作成 信頼性分析により、設問12～設問22を統合可能であるかを、項目間の内的整合性の指標であるクロンバックの α 係数を用いて検討した。クロンバックの α 係数は、複数の項目の評定値の合計（もしくは平均値）と各項目の評定値との相関が高いほど値が1.0に近づく。経験的にはこの値が0.8以上の場合に項目間の内的整合性が高く、各項目を統合可能であると判断される（林，山岡，2002，p.55）。

「専門講義」と「総合講義」について、設問12～設問22についてのクロンバックの α 係数を算出したところ、0.968と非常に高い値を得た。さらに、これらの設問項目のいずれか1つを除いた場合の信頼性係数を比較したところ、設問17（「教員は授業開始・終了の時間を守りましたか」）を除いた場合、信頼性係数が0.968から0.970へと若干増大することが示された。しかしながら、その他の項目を1つだけ除去した場合には信頼性係数の増大は認められなかった。このことは、設問12～設問22までの項目を合計ないしは平均するなどして1つの指標へと統合することが十分に妥当であることを示している（ただし強いていえば、設問17を除けばさらに頑健な指標が得られる）。

信頼性分析の結果より、設問12～設問22が統合可能であることが示されたので、これらの設問項目の評定値を全て平均し、授業の良し悪しを反映した統計的に頑健な指標を作成した。以下、この指標を「総合スコア」と呼ぶ。

3.2 授業の「総合スコア」の分析：受講者数と教室規模の影響

授業担当者および授業計画の作成者ができることは、その内容を充実させることだけではない。使用教室および受講者人数をある程度コントロールすることも、選択肢の一つとして考えることができる。そこで、教室収容人数および受講者数（回答者数）と「総合スコア」との関係プロットした。

さらに、各教室をその収容人数に応じて表4のように5つにレベル分けし、教室規模と「総合スコア」の関係性を分析した。5つの教室規模レベルのそれぞれに該当する科目数を授業区分ごとに表4に示す。

まず、実際の受講人数と総合スコアとの関係性をみてみたい（図2）。図2からは、全ての授業区分でみると、受講人数が100人までは受講人数が増えるにしたがって総合スコアが下降する傾向にある。また、受講人数が100人をこえたあたりから、総合スコアは下がりきってフラットとなる。さらに100人以上の授業では、授業ごとに総合スコアの差がよりはっきりとできるようになる。この人数は、「良い」と評価される授業を実施するための一つの臨界値と考えられる。

図2で、専門講義と総合講義に着目すると、総合スコアが下がりきってフラットになる値がすべての授業でみたばあいは100人ほどであったのに比べて、50人のあたりまで前倒しになっている。さらに、専門講義と総合講義は、授業ごとの総合スコアのバラツキが外国語やゼミナールといった他の授業カテゴリに比べて大きい。確かに受講人数は授業評価と関連しているようであるが、その影響を見て取れるのは多めにみても50人以下までというかなり限定された範囲

のことである。また、このようなバラツキがなぜ生じるかについてであるが、すでに3.1節の予備的分析で示したようにデータの質が悪いということもあり、これ以上の分析は今回は断念した。

つぎに、授業で使用する教室の規模と総合スコアとの関係をもてみたい(図3, 図4)。これらの図からは、「良い」と評価されるためにはどの程度の規模の教室を使用すべきかという一つの目安を得ることができるであろう。

図3は、授業で使用された教室の収容人数に対して、総合スコアがどのように推移しているかを授業カテゴリ別に示したグラフである。教室の収容人数が増加するにつれて総合スコアも下降傾向にあることが分かる。総合スコアが「良い」を割り込む(4.0以下)のは、教室収容人数がおよそ100人を超えたあたりであることが示されている。

さらに明確な傾向を把握するため、教室を収容人数を基準にして5つのレベルに分け(表4)、この5つのレベルの教室規模に対して総合スコアがどのように推移しているかを図示してみた(図4)。図4は、教室規模が大きくなるにしたがって総合スコアが下降する傾向にあることを示している。総合スコアの全体の平均点は、教室規模のレベル2まで総合スコア4.0以上を保っているが、レベル3以降では総合スコア4.0を割り込んでいる。とりわけ、専門講義と総合講義は他の授業カテゴリに比べて、教室規模の増大に対する総合スコアの落ち込みが顕著である。やはり、教室収容人数が100人を超えたあたり(レベル2からレベル3)で総合スコアの急激な落ち込みが生じている。

4 結論と提案

これまでの分析から、以下のことが明らかとなった。

1. 現行の授業評価アンケート項目では、授業の総合評価を良くするためにはどのような点を改善すべきか、といった詳細な分析をすることは難しい。
2. 専門講義と総合講義において、より大きな教室を使用すれば、授業の評価が低くなる傾向にある。
3. 専門講義と総合講義において、受講人数の増減は、多めにみても50人あたりを境に授業の評価にほとんど関係がなくなる。
4. 専門講義と総合講義は他の授業カテゴリに比べて、授業ごとの総合スコアが大きくばらついている。

これらのことから、各授業の運営および経営学部の授業計画、そして授業評価アンケートに

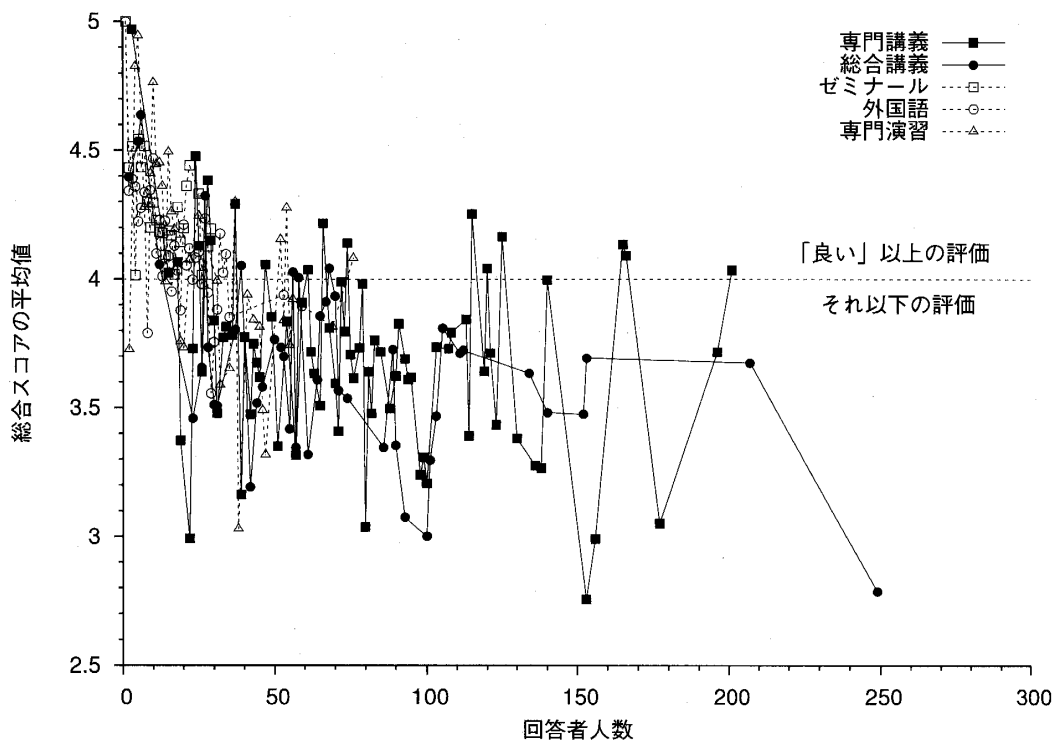


図2. 総合スコア（設問12～設問22を平均した指標）を授業カテゴリごとに、授業の受講人数に対してプロットしたグラフ。総合スコアの値が4以上の場合には平均で「良い」以上の評価であることを示している。

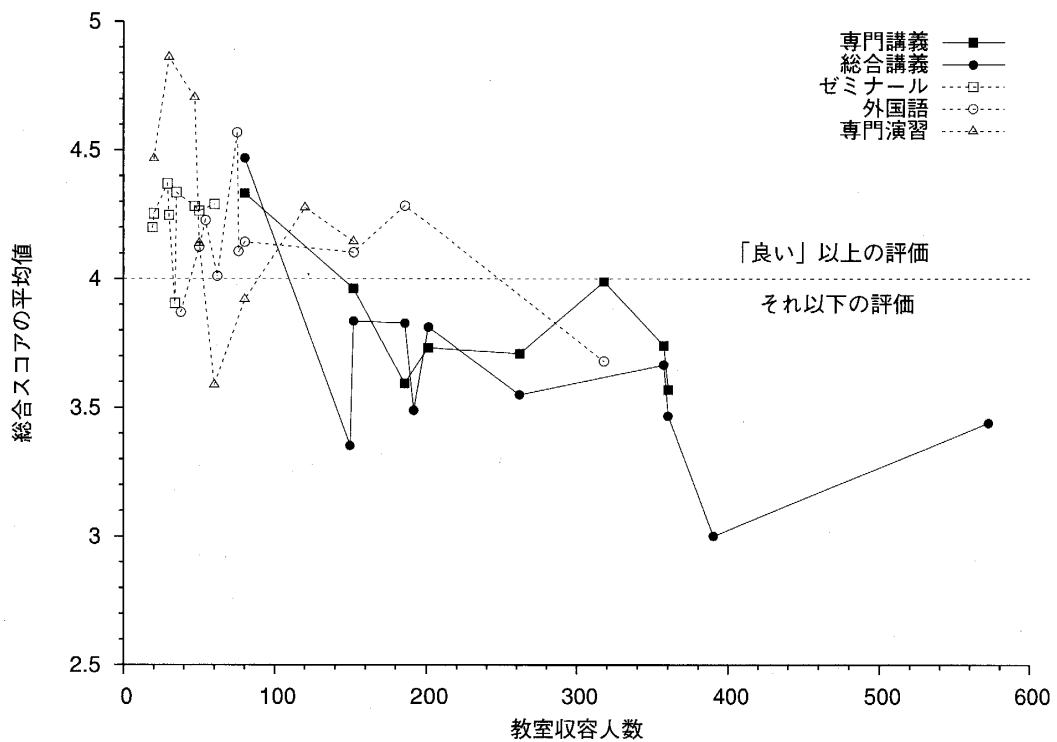


図3. 総合スコア（設問12～設問22を平均した指標）を授業カテゴリごとに、授業が実施された教室の収容人数に対してプロットしたグラフ。総合スコアの値が4以上の場合には平均で「良い」以上の評価であることを示している。

表 4. 教室規模レベルと授業区分からみた科目数 (スポーツを除く)

教室規模	レベル分けの基準	ゼミ	外国語	専門演習	専門講義	総合講義
レベル 1	収容数50人以下	139	31	20	0	0
レベル 2	収容数51人~100人※	11	101	30	4	6
レベル 3	収容数101人~200人	0	20	5	7	23
レベル 4	収容数201人~300人	0	0	0	64	19
レベル 5	収容数301人以上	0	0	0	48	15

※5501教室は54人だが、レベル 2 に含めている。

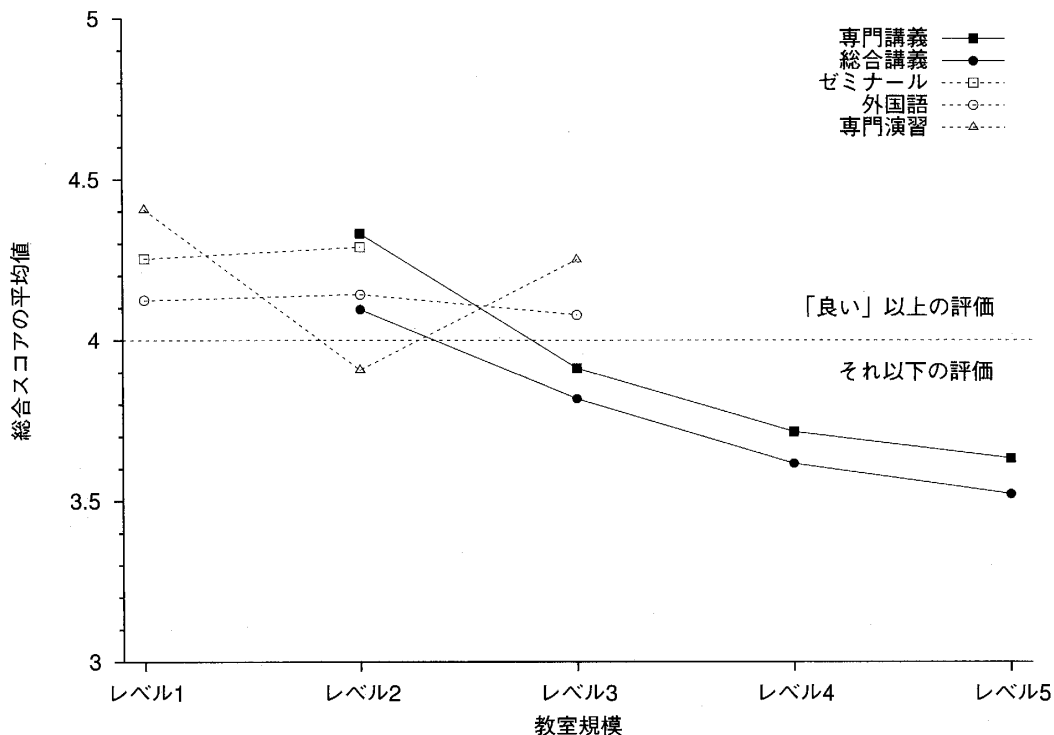


図 4. 総合スコア (設問12~設問22を平均した指標) を授業カテゴリごとに、授業が実施された教室規模のレベルに対してプロットしたグラフ。総合スコアの値が 4 以上の場合は平均で「良い」以上の評価であることを示している。教室規模レベルは、「1」が50人以下、「2」が51人~100人、「3」が101人~200人、「4」が201人~300人、「5」が301人以上、の教室であることを示している (表 4 参照)。

対していくつかのことを提言することができる。

4.1 アンケートの項目の整理・統合

現行のアンケート項目の設問12～設問22は、一つの指標にまとめてもほぼ同じ結果がでる可能性が高いこと、および効率性の観点からみて、一つの指標にまとめたほうがよい。今回の分析では動機付けに関する（設問1～設問11）項目を分析の対象外としているが、設問12～設問22と同様に項目を整理・統合できるか検討してみる価値はある。より具体的には、授業の総合評価を先に聞き（評価グレードも「よい・わるい」の2段階ぐらいで十分である）、そののちにその評価をした理由を選択式で選んでもらうようにすればよい。

より詳細にアンケート項目の整理・統合を検討するときには、以下の点に注意を払うべきであろう。

1. 似たような設問が他に存在しているかどうか
2. 複数のことを同時に聞くような質問になっていないかどうか
3. 学生に聞かなければわからない項目であるかどうか
4. 期末という時期に調査すべき項目であるかどうか
5. 集計結果がでたのちに、それを授業の改善に活かせる項目であるかどうか

4.2 自由記述の活用

各項目の点数をみても「担当している授業のどの点を改善すべきか」についてはほとんどわからない。また、学生の声をしゅくい上げる際に現行の項目ではしゅくいきれない側面も多く存在しているであろう（例えば、授業の合間にする教員の雑談が積極的な受講態度を保つのに役に立っているなど）。したがって、今のところ自由記述が最も授業改善の参考になりやすい情報である。このようなことから、回答者に対しては、今まで以上に自由記述欄に感想・要望・改善点等を積極的に挙げてもらえるような工夫をするべきである。また、授業担当者は各項目の点数よりも、自由記述に書かれた感想を参考にするほうがよいであろう。

4.3 小教室・少人数クラス編成の戦略的導入

小教室もしくは少人数クラスであるほど授業の評価は高くなりやすい。しかし、すべての授業においてこれを実施するにはさまざまな制約があることもまた確かである。そこで、学部で「総合スコア」の最低到達ラインを決め、それを下回る授業を対象にして、翌年度の授業で受講制限をかけたり、クラス分割をさせるといった方法が考えられる。目安としては第一段階と

して、クラス2レベルの教室で実出席が50人以下（回答率からみて1クラス100～120編成）となるようにすることを目指したほうがよい。それでも「総合スコア」の最低到達ラインを下回る場合には、第二段階として、25人以下（回答率からみて1クラス50～60人編成）を目指すべきである。この小教室・少人数クラスの戦略的導入は、授業ごとの評価のバラツキが大きい専門講義と総合講義において有効であると考えられる。

5 おわりに

われわれがこの授業評価アンケートの分析を進めていく中で、何度も立ち戻らなければならなかった観点がある。それは、この授業評価アンケートは何のためにされるのかという点である。例えば4.1節の「アンケート項目の整理・統合」における提案の前提には、普段の授業の中では気づかない改善点を見付け出すために授業評価アンケートをしているとの考え方があ

る。もし、そうではなくて教員が行動規範通りに授業を行っているかどうかを監視するものとしてアンケートをすることを前提とするならば、その前提に沿った項目を立てるべきであろう。さらに、大学運営の費用対効果の観点で現行のアンケートをみれば、廃止も十分に検討する価値がある。要するに、今後の授業評価アンケートの姿を考えると、まず問われるべきことはわれわれの大学観や教育観をどう定めるかということなのである。

謝辞

最後に、平成16年度授業評価アンケートデータの利用許可を与えていただいた樺山忠雄経営学部長、そして分析のためのアイデアやアドバイスをくださった経営学部の同僚の皆様にご

の場を借りて御礼を申し上げます。

引用文献

- [1] 芝祐順, 南風原朝和 (1990) 行動科学における統計解析法, 東京大学出版会.
- [2] 林文, 山岡和枝 (2002) 調査の実際—不完全なデータから何を読みとるか—, 朝倉書店.