

## 授業評価アンケートは何を評価しているのか

阿久津 洋巳

(2014年2月12日受理)

Hiromi AKUTSU

What are Evaluated by Lecture Evaluation by Students ?

<キーワード> 授業評価 尺度構成 測定 教育改善

### 要約

岩手大学教育学部は学期末に授業評価を実施するが、実施された結果を詳細に検討した報告書はまだ公表されていない。筆者は授業評価が測定している構成概念と測定される概念間の関連、および評価が高い授業の特性を調べるために、2013年度前期に集められた実際の評価データ（1680ケース）に多変量解析を適用して分析した。「教員の授業の技量と学生の満足度」「授業に対する学生の興味と意欲」「授業外の活動」「シラバスの評価」の4つの要因が見出された。技量・満足度と興味・関心の間に関連があり、また技量・満足度とシラバス評価の間にも関連があった。評価の高い授業は、話し方が明瞭で聞き取り易く、よく準備されていて、学生の理解に配慮して適度な分量と速さで進むという特性をもっていた。一方測定尺度としての欠陥も見つかった。評価アンケートは回答が特定の値に集中し、分布は偏っていた。尺度の信頼性は検討されていなかった。信頼できる評価尺度の開発が必要である。

### 背景と目的

#### 【背景】

1990年代以来日本では大学改革が進行している。その大学改革の一環として学生による授業評価の導入が提案され、推奨されてきた（1991年大学審議会答申；1998年大学審議会答申）。その結果、2002年には学生に授業評価を実施している大学は全国で574大学に上り、国公立大学全体の約84%に達した（文部科学省発表）。1998年の大学審議会答申以降、学生の授業評価の目的は、教員個人の教育活動を改善したり、学部と大学という組織の教育活動を改善する手段と考えられている（天野・南部，2004）。全国の国立大学86校の事務局学務課長宛てに授業評価に関する調査質問紙を送付した調査研究によると（回収率は93%）学生に対して実施される授業評価の結果の活用法は「ファカルティ・ディベロップメントを通じて、教育改善につながる資料として用いる」という解答が最も多く（複数回答可能）、約半数の40校がそのように回答した。検討中を含めると53校に上り、2番目の「各教員の評価結果を学生に公表する」の32校（検討中も含める）を大きく上回っていた（天野と南部，2004）。実際に行われている

授業評価の目的は、授業の質を高めて、大学教育の改善を図ることにあるといえるであろう。

実証的データに基づいて大学教育の改善を目指すという点で、授業評価は *institutional research* (IR) の側面も持っている。IR とは、大学等の計画立案、政策形成、意志決定に役立つ情報を提供する目的で、当該組織内で実施される調査研究である(中井・鳥居・藤井, 2013)。大学教育の質保証と管理運営の高度化に向けて、大学の活動に関する調査データを収集分析することは IR の中核を占める。授業評価は大学教育を改善し、そのコスト(投入される税金)にみあうよう教育の質を高める目的に使える情報を潜在的に含んでいる。しかし、現実にはその情報を取り出す方策がとられずに、教員に「あなたの授業に対する学生の評価は以下のとおりでした。」という程度の説明の後に質問項目ごとの平均点と学部の平均が掲載されるフィードバックに終わることが多い。大学の計画立案と意思決定に貢献することは少ない。授業評価に回答する学生のエネルギーは無駄に消費され、評価データに含まれる潜在的に有用な情報は、一度も日の目を見ないでコンピュータの記憶装置で眠り続ける。次の学期もまた同じことが繰り返される。この組織的無駄の一番大きな原因は、大学が十分な IR 要員を配置していないことにある。また、授業評価の実施主体が「実施したと報告するために実施する」側面も見逃せない。いずれにせよ、実際に行われている授業評価は、教員による授業の質を高めて、大学教育の改善を図るという目的に適合していない。

岩手大教育学部は、2002年以来専門科目の授業評価を続けてきた。筆者が知る限り2004年以来はほぼ同じ質問項目(20問)を使ってきた。授業評価アンケートは、教育学部内の委員会が評価に適すると思う項目(具体的質問)を選んで作成されてきた。質問項目を見直し、授業評価をより適正に実行するために、岩手大学教育学部は2013年度の学務委員会において、授業評価を再検討するワーキンググループ(WG)を作り、委員が2013年度の前期に実施された評価データをもとに、現行

の授業評価のデータを有効に使う方法を検討した。筆者はその一人であり、データ解析を担当したので論文の形で公表することにした。過去にもこのような試みがあったかも知れないが、著者が知る限りでは近年は公表されていない。

#### 【目的】

本研究の目的は、統計的手法を用いて授業評価のデータから岩手大学教育学部で実施されている教育の質を向上させるために使える情報を取り出すことである。具体的には、授業評価に使われた評価項目1~20(以下q1~q20)に対する学生の回答をもとに、(1) 授業評価アンケートは何を測定しているのか。(2) 測定された概念間にはどのような関係があるのか、(3) 評価が高い授業とはどのようなものか、という3つの疑問に答えようとした。

#### 方法

##### 【評価データ】

2013年度の前期期末に岩手大学教育学部において実施された授業評価のアンケート項目に対する評価回答を使用した。授業評価は1教員が担当する科目の1つについて実施されたが、担当科目が講義ではなく演習や実習であったり、受講生が少数の場合は、教員は授業評価を免除された。科目数は40、参加した教員数は39名であった。授業評価アンケートは20の質問項目から構成されていた。1683人分の回答が分析に使用された。自由記述は分析対象から除外された。

##### 【データ分析の方法】

通常心理尺度構成の手順に従い、(1) 項目に対する反応の共通性(相関)を手がかりに、類似の項目を1つにまとめる因子分析法の手法を採用した。(2) 抽出された因子別に、受講生の得点を計算した。(3) 受講生  $\times$  因子得点の行列から、因子間の相関を調べ、最後に回帰分析を適用して因子間の関連を検討した。

## 結果

データの分析には無料の統計ソフトRを使用した(R Core Team, 2013)。20の評価項目から少数の共通因子(評価アンケートが測定している心理学的構成概念,たとえば授業の満足度)を抽出し,通常の心理測定で用いられる方法で因子得点を求めた。以下にやや詳しくその手順を述べる。

全ての評価結果を縦に回答者(以下ケースとよぶ),横に質問に対する回答(質問1から質問20まで,以下項目とよぶ)の回答行列にまとめた。後の計算処理と解釈を容易にするため,「回答の値(1~4)を数値4~1に変更した(4:そう思う~1:そう思わない)。回答に欠損値がある場合は,ケース単位でデータを分析から除外した。最終的に1680ケースを分析対象とした。

(1) 因子の決定 主成分分析とスクリープロットの結果から,3因子,4因子,5因子が因子数の候補としてあがったので,3~5因子を仮定して探索的因子分析を繰り返した。因子の推定法は最尤法で,因子の回転には斜交promax回転を使った。1因子の項目数が3以上で,かつ多面的に評価回答を解釈できるように因子を選ぶことにした。この基準を使うと3因子と4因子が選ばれたが,項目を調べたところ4因子の方が解釈しやすかったので,4因子を採用した。

因子1は,「総合的にみて満足がいく」「学生の反応や理解度に配慮して進められた」「分量と進む速さは適切」「話し方は明瞭で聞き取りやすい」「新しい見方・考え方や知識を獲得できた」などの質問項目が含まれており,授業の技量と満足度を反映する因子と命名した。因子2は,「熱意をもってこの授業に臨んだ」「授業の目的,到達目標を達成できた」「新しい見方・考え方や知識を獲得できた」の項目であるから,学生の授業に対する興味と意欲を反映する因子と命名した。因子3は,「アイアシスタントを使う機会があった」「授業時間外に勉強する必要がある」「カリキュラムの中に位置づけて把握する機会があった」「ESDについて考える機会があった」の項目で

あり,授業外の活動と命名した。因子4は,「成績評価の方法や基準に関するシラバスの説明がわかりやすかった」「授業の目的や到達目標に関するシラバスの説明がわかりやすかった」「授業内容はシラバスに掲載されている予定と一致していた」の質問項目であり,シラバスの評価因子と命名した。因子分析結果を表1に示した。

### 尺度の信頼性

信頼性を調べるために,クローンバックの $\alpha$ 係数を算出したところ,4つの因子の尺度(4尺度)すべてで0.7以上の内部一貫性があった。

(2) 因子得点の計算 「授業の技量と満足度」の因子,「授業に対する興味と意欲」の因子,「授業外の活動」の因子,「シラバスの評価」の因子が見出されたが,これらの因子ごとに因子得点を次の手順で計算した。まず各因子の属する項目に対する反応(1~4)をケース別に合計した。幸い逆転項目はなかったので,合計値が因子得点の素点となる。次に,全てのケースを使って因子得点の平均と標準偏差を求めた。平均と標準偏差をもとに得点の標準化を行った。標準得点は平均50,標準偏差10とした。同様の手続きを4つの因子別に行った。表1には,抽出された因子に対応する質問項目の因子付加量のみを記載した。

### (3) 因子間の関連

因子間に関連があると予想できたので,ピアソンの積率相関係数を計算した(表2の相関行列)。表2から,第1因子(授業の技量と満足度)と第2因子(興味と意欲)および第4因子(シラバスの評価)の間には関連があり,第2因子(興味と意欲)と第4因子のあいだにも無視できない関連があったと読み取れる。第3因子(授業外の活動)は他の因子と関連が薄かった。

さらに,因子間の関連を重回帰分析を使って検討した。複数の変量間に相関がある場合は,実際には関連はないのだが真に関連がある変数を介して,見かけ上の相関が生じることがある(疑似相関)。標準偏回帰係数を計算することで,この疑似相関を回避して,ある変量が別の変量にどの程度影響を与えるかを推測できる。

ここでは、4つの変量（因子1～因子4）を順に目的変量として回帰分析を試した。

（1）因子1を目的変量、他の因子を説明変量とした重回帰分析結果

因子1の変動には、因子2と因子4がほぼ同

じ程度に影響したが、因子3は影響しなかった。

（2）第2因子を目的変量とし他の3つの変量を説明変量とした重回帰分析結果

因子2の変動には、実質的に因子1だけが影響した。

表1 因子分析結果

第1因子 授業の技量と学生の満足度	第1因子負荷量
問1 この授業は総合的にみて満足のいくものであった	0.578
問5 授業は学生の反応や理解度に配慮して進められていた	0.794
問6 授業の分量と進む速さは適切であった	0.790
問7 「周到に準備された授業」という印象であった	0.806
問8 先生の話し方は明瞭で、聞き取りやすかった	0.870
問9 私語がなく集中できる授業であった	0.513
問10 教科書、参考書、配布資料は学習の助けになった	0.368
問11 概ね定刻に始まり、定刻に終わる授業であった	0.546
第2因子 授業に対する興味と意欲	第2因子付加量
問15 授業の目的、到達目標を達成できた	0.531
問16 授業を通して新しい見方・考え方や知識を獲得できた	0.599
問17 熱意をもってこの授業に臨んだ	0.725
問18 この授業で学んだことを、さらに勉強したい	0.779
第3因子 授業外の活動	第3因子付加量
問12 この授業に関連してアイアシスタントを使う機会があった	0.533
問13 授業時間外に勉強する必要(予習・復習やレポートなど)があった	0.732
問14 予習や復習を十分にして授業に臨んだ	0.783
問19 この授業を卒業までのカリキュラムの中に位置づけて把握する機会があった	0.360
問20 この授業で「ESD」や「持続可能性」について考える機会があった	0.409
第4因子 シラバスの評価	第4因子付加量
問2 授業の目的や到達目標に関するシラバスの説明がわかりやすかった	0.849
問3 成績評価の方法や基準に関するシラバスの説明がわかりやすかった	0.903
問4 授業内容はシラバスに掲載されている予定と一致していた	0.530

表2 因子得点間の相関行列

	第1因子	第2因子	第3因子	第4因子
第1因子	1	0.67	0.3	0.68
第2因子		1	0.38	0.54
第3因子			1	0.32
第4因子				1

(3) 因子3を目的変数とし他の3つの変数を説明変数とした重回帰分析結果

因子2が少し影響するが、その効果は小さく、他の因子は影響しなかった。

(4) 因子4を目的変数とし他の3つの変数を説明変数とした重回帰分析結果

因子4の変動には、因子1だけが影響した。

以上の回帰分析の結果をまとめると、(1) 因子3 (授業外の活動) は、他の因子からほぼ独立であることが確かめられた。(2) 因子1 (授業の技量と満足度) の得点は、因子4 (シラバス評価) と因子2 (興味と意欲) の得点と関係がある。重回帰分析結果を表3に掲載した。

表3 重回帰分析結果

第1因子を目的変数とした			R2 = 0.59
説明変数	標準偏回帰係数	p	
第2因子	0.423	< 0.001	
第3因子	-0.009	0.568	
第4因子	0.469	< 0.001	
第2因子を目的変数とした			R2=0.48
説明変数	標準偏回帰係数	p	
第1因子	0.536	< 0.001	
第3因子	0.181	< 0.001	
第4因子	0.116	< 0.001	
第3因子を目的変数とした			R2=0.16
説明変数	標準偏回帰係数	p	
第1因子	-0.02	0.57	
第2因子	0.294	< 0.001	
第4因子	0.184	< 0.001	
第4因子を目的変数とした			R2=0.49
説明変数	p		
第1因子	0.561	< 0.001	
第2因子	0.109	< 0.001	
第3因子	0.107	< 0.001	

### 考察

結果に述べた分析にもとづいて、目的で提示した問題に順に答えていく。

(1) 授業評価アンケートが何を測定しているのか。

古典的な尺度構成の手順に従って授業評価アンケートのデータを分析した結果、授業評価アンケートは、①授業の技量と満足度、②学生の授業に対する興味と意欲、③アイアシスタントの使用、ESDについて考えるなどの授業外の活動、④シラバスの評価、の4つを測定していると考えられる。しかし、この解釈には以下に記す制限がつく。

本研究で使用したデータからみると、使用された授業評価のアンケートには測定道具の観点から望ましくない特性がある。それは、多くの質問項目に見られる反応の偏りである。図1に20の質問項目に対する回答をboxplotを使って表示した。回答の方向(1~4)を逆転して4を1番望ましく、1を1番低い評価とした。図から明らかのように、多くの回答は3か4に集中していた。言い換えれば、回答の分散が小さい。回答の分散が小さければそれだけ回答に含まれる情報は少なく、多変量解析などの統計技法を用いても、抽出できる情報がはじめから限られている。それ以前に、このように偏った分布をもつ質問項目の回答(1~4の選択はあくまでカテゴリカル反応)をそのまま数値(等間隔尺度)として用いることの妥当性は保証されない。より正確に言えば、質問項目の回答は順位をあらわすことは確か(順位尺度)であるが、順位を数学的演算を適用するための等間隔尺度上に数値化する手続きはとられていない。しかし、この問題は本論文の目的からそれるので、ここでは暫定的に質問項目に対する回答を等間隔尺度として扱う。

分散が小さいことの問題点をより具体的に実感するためには、次の図2を見るとよい。図2では平均(●)の上下に標準偏差の0.67倍の幅をTのマークとして付加した。標準偏差の0.67倍の大きさを選んだ理由は、平均値の上下に0.67標準偏差の区間を設定すると、その区間に正規分布では50%のデータが含まれるためである。平均と比較するために2つの授業評価例をプロットした。上向きの三角(△)は、総合点が一番評価が高かった1つの授業の得点であり、下向きの三角(▽)は、総合点が一番低かった1つの授業の得点である。

多くの場合△と▽が.67標準偏差の区間に入っていることがわかる。もちろん総合点が一番評価が高かった授業がすべての質問項目で1番良い平均を得ているわけではないし、総合点が一番低かった1つの授業がすべての質問項目で1番悪い平均を得ているわけではない。分散が小さいために(加えて、分布の歪みが大きいため)、授業間の得点の違いは小さい。その現実的な影響は、教員にフィードバックされる評価得点はどれも平均に近く理解しにくいことであり、授業の改善に役に立ちにくいことである。さらに、これまでは全体

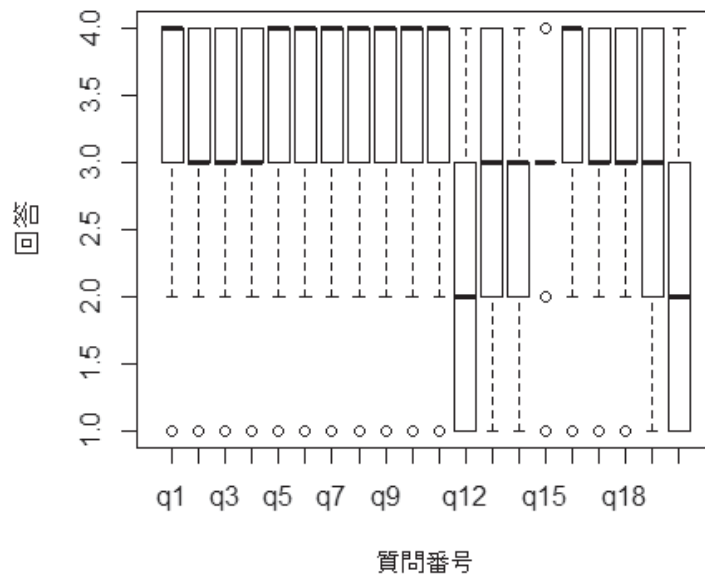


図1 質問1～20に関する boxplot

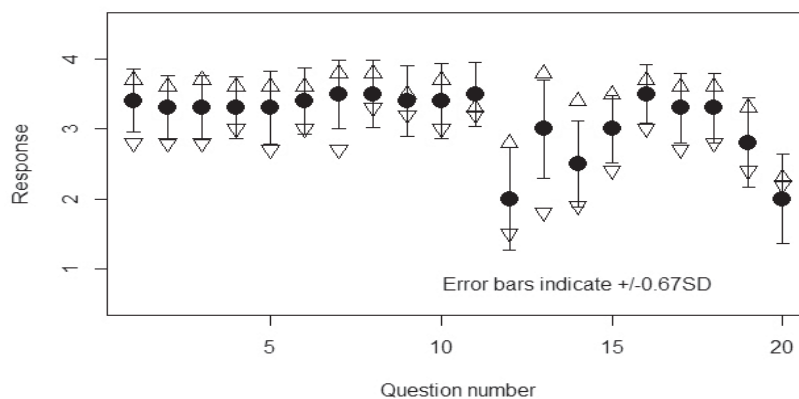


図2 質問1～20に対する回答の分布

●は平均、Tと逆Tは標準偏差の0.67倍を示す。△は総合点で一番高い評価の授業、▽は最も低い評価の授業の評価値を示す。どの質問も回答が特定の値に集まっていることが読み取れる。

の平均と担当授業の平均だけがフィードバックされた（平均値だけを見せられて自分の授業の評価が平均と同じかあるいは高いのか（低いのか）判断していた）。評価質問項目の改正の第一歩として各質問項目に反応のばらつきを大きくする工夫が必要であろう。

（2）測定された評価因子間にどのような関連があるか。

因子3は他の因子からほぼ独立であるが、因子1, 2, 4, は互いに関連していることがわかった。これら因子間の関連のうち強いと思われるものを図に示した（図3）。結果の項の表2と3に示された数字から、関連の強弱はわかるが因果の方向はわからない。図3も同様である。教員の授業の技量（または学生の満足度）とシラバスに対する評価には関連があり、また、学生の興味と意欲は教員の授業の技量（または学生の満足度）に関連していた。ここで観察された変数に限って解釈するならば、よく準備計画された授業は、学生の興味と意欲を引き出す、という解釈が可能であるが、逆に学生が興味を持っていると、その授業をよく評価するという可能性もある。これまでに報告された研究は、よい授業が学生の意欲を引き出すという因果の方向を構造方程式のモデルを使って示している（星野・牟田, 2003）。シラバスに対する評価が、授業の満足度に関連があったのは、やや意外であった。学生はシラバスを丁寧に読んでいるということであろうか。それとも、授業に満足した学生が好意的にシラバスを評定したのであろうか。

授業外活動（「授業時間外に勉強する必要があった」を含む）が授業の技量と満足度の因子に関連がなかったことも意外であり、後述する工学部の授業評価結果と異なる。

（3）評価が高い授業とはどのようなものか

第1因子の項目を再度詳細にみていくと（表1参照）、浮かび上がってくるのは、「話し方が聞き取り易く、適切な説明がなされ、分量と速さに配慮した講義」である。加えて、教員の授業準備も評価の高さに反映されている。評価が高い授業の特徴は、岩手大学工学部マテリアル工学科が2013年に報告した授業アンケート調査結果（未発表資料）におおむね一致している。一方、マテリアル工学科の調査では、授業に対する高い評価と学習時間の間に負の相関があり、成績評価比率と正の相関があった。簡単にまとめれば、マテリアル工学科の調査では、上記の教育学部で見つかった高い評価の要因に加えて、教室外学習時間が少ない授業と優と良の成績が比較的多く与えられる授業という要因が、高い評価に結びついていた。よい授業なので学習時間が少なくてすむと解釈したが、逆に学習時間が少なくてすむのでよい授業と評価されるのか、あるいは両者に直接の関連はないのかもしれない。与えられた結果からは、どの解釈が妥当か判定できない。教育学部の調査では、時間外学習の必要性の要因は、評価の高さと関連がなかった。今後さらに検討すべき要因である。教員の授業技量のほかに、単純な学習環境の要因が「技量と満足度」の因子に含まれていた。私語がないこと、開始と終了の時間を守る、教科書の選定と資料の配布、などの項目である。

学生の授業満足度に直接かかわる第1因子項目は、今日の多くの大学教員が心がけている事項と考えられる。その意味で本研究で見つかった評価が高い授業の特徴は、多くの教員が日ごろ考えている特性が正しいことを裏付けている。

（4）その他

研究目的では挙げなかったが、今回の授業評価の分析において第2因子として学生の「興味と意欲」が抽出された。学生の興味と意欲を高める授

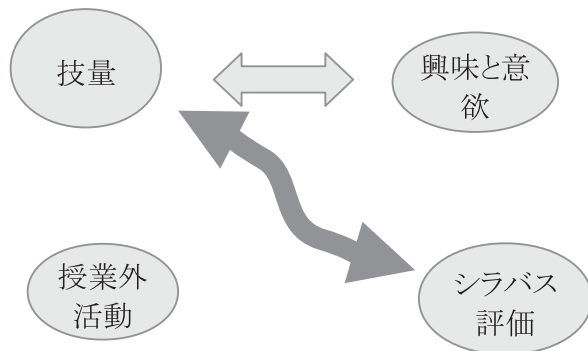


図3 評価要因間の関連

業がどのような授業であるかは、当然のことながら、きわめて重要である。調査データの分析から、第1因子（技量と満足度）と第2因子の関連は高く、ピアソンの積率相関係数は0.67であった。また回帰分析から、興味と関心を目的変数とし授業の満足度を説明変数とすると、満足度は0.536の偏回帰係数で興味と関心に関連していた。単純に満足度から興味関心という影響の方向を仮定するならば、学生の授業に満足する程度が1単位上昇すると興味関心は半分程度上昇することを意味する。教員の教授努力（「声が明瞭」「板書が読みやすい」「教員の熱意」「教材資料が適切」）が、学生の授業満足度と意欲に影響するという結果は、他の研究でも得られている（星野・牟田，2003）。もし、本研究の調査が信頼できる数値を生み出しているならば（その評価は別の研究に譲る）、教員の授業努力が学生の意欲に影響するのであるから、教員はよい授業を心がけるべきであるという提言ができる。先行研究の結果を合わせて考えるならば、この解釈が正しいかもしれないが、現在のデータをこのようにだけ解釈するのはデータが与える情報の範囲を逸脱している。

#### （5）解決すべき課題

考察のはじめに記したように、本研究で使用した授業評価アンケートの項目と評価法は、尺度構成法の観点から見ると看過できない欠陥がある。平均だけ関心があるとしても、得られたデータを数量化するためには、心理学的尺度構成の必要条件を満たさなければ、結果の価値は限られる。尺度構成の必要条件には、尺度の信頼性と妥当性を確保することも含まれる。授業評価アンケートにおいて、信頼性と妥当性を確保することは容易ではない（例えば大山，2003；藤田，2005）。

尺度としての必要条件を満たさない調査結果は、どんなに沢山集めても、あるいは長期間実施しても役立つ結果よりも誤った理解と誤った知識を生む可能性がある。少なくとも、信頼できる調査データと呼ばれる資格はない。そのような結果が調査資料としてなんらかの公式資料（もしくは報告書）に記載されて、一般に流布されることは

大学にとって好ましいことではない。IR機構を作り、心理学的測定の実験家と統計の実験家の知恵を集めることにより、「信頼できる調査データを収集する」授業評価アンケート作成の体制を整えることが、今大学に望まれる。

**【謝辞】** データ解析の段階で、教育学部の平田光彦先生と本田卓先生から有益な助言をいただきました。記して感謝の意を表します。

#### 引用文献

- 天野智水，南部広孝 2004 わが国の国立大学における学生による授業評価の展開 広島大学高等教育開発センター 大学論集 第35集 229-243
- 岩手大学工学部マテリアル工学科 2014 アンケート調査結果 岩手大学大学教育センター専門教育等連携部門会議資料（未発表）
- 大山泰宏 2003 「大学教育評価論」京都大学高等教育研究開発推進センター編「大学教育学」培風館 39-62項
- 中井俊樹，鳥居朋子，藤井都百 2013「大学のIR Q&A」玉川大学出版部
- 藤田哲也 2005 授業評価に対する心理学的アプローチ 名古屋高等教育研究 第5号
- 星野敦子，牟田博光 2003 大学生による授業評価にみる受講者の満足度に影響を及ぼす諸要因 日本教育工学会論文誌 日本教育工学雑誌 27 (Suppl.) 213-216.
- R Core Team 2013 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.