

階層的分析法(AHP)による少人数授業における相互評価の試み

植 田 義 幸

(平成17年9月30日 提出)

大学における少人数の授業で、AHP(階層的分析法または階層的意思決定法)の考え方に基づく相互評価を試みた。AHPは「総合目的」―「評価基準」―「代替案」という階層構造による意思決定モデルを仮定し、総合目的から見た評価基準の重要性、各評価基準から判断した代替案の優先度を比較判断して、最終的に総合目的から見た複数の代替案の優先度を評価するものであり、主観的な判断から定量的な判断をすることを目的とするオペレーションズ・リサーチの一手法である。本研究では、大学の演習形式の授業で、受講者7名にAHPに基づくアンケート調査を実施し、受講者が重視する評価観点を明らかにし、受講者相互を評価することを求めた。また、相互評価の際はAHPの短所である比較回数の幾何級数的増大を抑えるために、限定した範囲での一対比較を求めた。その結果、受講者の評価観点の判断は首尾一貫していることが見いだされた。相互評価についても首尾一貫していることと教授者の観察とある程度一致することが明らかとなり、AHPを少人数授業の評価手法として利用できる可能性が示唆された。ただし、評価観点項目の選択は、受講者の意見を取り入れて取捨選択して改善する余地がある。また、AHPによる判断は評価の参考にはできるが、評価そのものとして用いる方法は見いだせていない。

キーワード：階層的分析法 AHP 相互評価 一対比較の改良

1. 目的

1.1 学習への参加とその評価

大学審議会の答申を引くまでもなく、高等教育において、厳格な指導と評価が求められている。その表れとして、FD(Faculty Development)が大学教員に求められている。FDは教員(団)の能力開発を意味するのではあるが、高等教育の改善を意図してのものであることは言うまでもない。シラバスの整備や資料の提示方法の工夫といった授業運営の改善だけでなく、学習成果の評価も重要な要素である。

いくつかのFDに関する文献を検討したが、評価についての記述は、主に大人数の講義形式や実験・実習を伴う授業において、基準の明示の必要

性や試験の実施上の留意点、レポート等の取り扱いや保管の注意等が中心であり、少人数授業において学習者の評価をどのような基準で行うべきかについては、ほとんど言及されていない(片岡・喜多村編1989、浅野1994、経済学教育学会編1998、池田・戸田山・近田・中井2001、高等教育情報センター2003など)。

もちろん、高等教育における多様な教授内容について、一律の評価基準や方法が定められないことは言うまでもない。外国語や技能の修得を目的とした科目のように、水準を決めればそれに伴ってある程度までは修得すべき知識量や技能が定められ、知識の定着度や課題の達成度を筆記試験や実技試験等で確認することが主眼となる科目もあ

れば、知識・技能獲得に収斂せず、一定の知識の習得よりもむしろ学習方法・研究方法の習得・慣熟や後の専門的な学習への動機付け、ものの考え方や学習態度の涵養が主目的となる科目もあり、一概に述べることができないことも確かである。

このような科目では、教授者から学習者への一方向の働きかけにとどまらず、学習者が積極的に授業過程に参加する学習が求められる。参加から更に一步踏み込んだ「参与」型・「参画」型の学習が唱えられてもいる（林1994）。教授者が学習者をあらかじめ定めた基準に従って測定し評価を下すという、従来の評価だけでは不十分な場合も多い。その一方、客観的な基準が定められない科目では、評価は主観的で曖昧になりがちである。評価者が異なれば、同一の評価を下すことはできない。

また、教授者が持つ単一の視点で評価できる場合ばかりではない。Atkinのいう羅生門的アプローチが評価の分野でも求められている（佐藤1996、東2001）。その方法として相互評価（同僚評価）が試みられることがある。従来から、いくつかの観点で評価するものや、文章表現（評言）を中心とする相互評価がみられる。しかし、このような評価には留意すべき点がある。

①基準のばらつき・ぶれ

相互評価は、学習者自身が行うため、人によって基準が相当異なり、「甘い」評価や「辛い」評価が生ずる。同一人においても、評価の時期や順序によって異なる評価があり得る。

②視点の多様性

何を評価するかを示さずに評者に評価をゆだねた場合、その視点は多様であり、統一的な評価は難しい。たとえ教師が示しても、ある者は説明の巧みを重視し、ある者は知識の量を問題にすることもある。もちろん、①の基準のばらつきも含めた多様性が学習のあり方を検討する手がかりに

なるというのが、羅生門的アプローチの主張するところであり、一概に否定できるものではないが、多様性を認めながら、全体として統一した基準による評価を行うことは困難である。

③評言の難しさ

学習者は、まさに学習の途上であり、適切な言葉を持ち合わせていないことも多い。全てを理解した教授者のみが評価を行えるというわけではないが、評価者が学習内容を十分に理解していない場合「がんばっていました」「声が大きくよく聞こえました」等の主観的でほめる一方のものが多数を占めることになりかねない。評言による評価により、型にはまらない自由な評価もできるし、評価能力の涵養が期待できる。特に、筆者の担当科目は教員免許取得希望者が受講するものが多い。教員の資質能力の一つとして安定した的確な評価能力を高めることも重要であり、簡略な方法を追求すべきだとも言いきれない。しかし、一般的に、評言による相互評価によつて的確な評価をすることは難しいと言ってよいだろう。

また、従来は、いくつかの観点にしたがって絶対評価による基準で評価し、総合的に単純に加算することが多かったと考えられる。たとえば、観点Aについて20点中12点、観点Bについて20点中16点などと評価し、合計100点中70点等とする評価方法は広く用いられているだろう。しかし、この方法には合算の根拠が乏しい。単純加算では、遅刻や欠席をせずに出席しているかどうかという点と、討論で他の受講者を知的に刺激するような発言をしたかどうかという点とを同等に扱うことになる。配点に軽重をつけても、その配分割合の根拠が欠けている。

高等教育における評価の厳格化とは、単に高い要求水準を求め、それを満たす者と満たさない者とを厳密に分けることだけを意味しない。どのような評価であれ、その方法が透明であること、つ

まり検証が可能であることも求められている。

1.2 本研究の目的

本研究では、少人数授業における相互評価を根拠のあるものとし、また、評価者による恣意的な差異を軽減する方法として、AHP（Analytic Hierarchy Process 階層的分析法または階層的意思決定法）の応用を試みる。

AHPは1970年代にSaatyが提唱した、不確定な状況や多様な評価基準のもとでの問題解決型意思決定法であり、評価基準や代替案の重要性についての主観的判断による比較をもとに、真の重要性を推測し、意思決定に導くオペレーションズ・リサーチの一手法である。あいまいで漠然とした状況下における意思決定や、客観的で一次元的な基準の設定が不可能な状況下で評価が求められる場合に、直感的な判断を客観的で定量的な情報に変換する手法と言い換えてもよいだろう（木下2000、木下編2000、Saaty 1980）。

AHPでは、複雑で多義的な問題を「総合目的」－「評価基準」－「代替案（選択肢・候補・評価対象）」の階層関係でとらえる。そして「総合目的」からみた「評価基準」のウエイト（重要度）を求め、各「評価基準」からみた「代替案」の評価をする。この後で、「代替案」のウエイトと「評価基準」のウエイトから「総合目的」に対する「代替案」の評価に換算するというプロセスを経る（具体的には行列の演算という形式になる）。各々のウエイトの評価は、「評価基準」や「代替案」といった要素同士を意思決定者が（多くの場合主観的に）比較して数値化した後、固有値・固有ベクトル問題を解き、行列の演算によって求める。また、モデルの整合性をC.I.（Consistency Index 整合度指数）という指標によって計ることができる。整合性の低いモデルは要素を取捨選択したり、グループ化したりして、より正確なモデルを探索することができる。

木下は、AHPの特徴を

- ①人間の持っている主観や勘が反映されるようなモデルが作られていること
- ②多くの目的を同時に考慮できるようなモデルであること
- ③あいまいな環境を明確に説明できるようなモデルであること
- ④意思決定者が、容易にこのモデルを使えることと整理している。（木下2000 pp.2-3）

AHPの適用領域として、経済問題と経営問題、エネルギー問題、医療と健康、紛争処理、プロジェクト選定、政策決定など、質的要素が重要な役割を占める意思決定問題がある。また、感覚情報による機械システムの信頼性・安定性の解析や、ソフトウェアの使いやすさという主観的判断による品質評価、人事評価など、あいまいな部分をもつ主観の評価を安定的・客観的な判断に変換する技法への応用が試みられている（木下2000、刀根・真鍋編1990）。教育分野では大学進学の際の選択基準（鈴木1996）、大学の授業における課題の評価（宮地1993）、大学の授業評価（宮地2001）への適用が見られ、応用可能性が探られているところである。

本研究では、少人数の演習形式で行った大学の授業について、AHPによるアンケート調査を実施して、評価を行う際の一助とすることを試みる。受講者に、階層構造モデルに基づいた評価観点の重要性と、その観点ごとに他の受講者と自分自身とを比較する調査への回答を求める。

これによって、受講者自身が、学習にあたってどのような点（評価の観点）を重視しているかを探る。また、受講者同士の相互評価の特徴を検討し、この手法の妥当性を考察する。

2. 方法

2.1 AHPの方法

AHPの手続きを簡単に説明し、本研究でAHPをどのように用いるかを説明する。

(1) 問題のモデル化

まず、取り組むべき問題（意思決定する問題）を階層構造に表現する。階層の最上位は「総合目的」である。その下のレベルは、問題の解決に、どのような視点があるか、どのような基準で総合目的が判断されるかを、いくつかの要素に分解したものであるとして「評価基準」（レベル2）を置く。評価基準は、さらに下のレベルの評価基準に分解することもある。たとえば物品の購入に際して「経済性」という評価基準をさらに「燃費」「初期費用」「税金」といったように分けて考えることもできる。階層の最下位には「代替案」を置く。代替案は、選択肢、候補者、評価対象など、実際に選択あるいは評価すべき対象を意味する。

(2) 要素の重みの評価

次にレベル2以下の各要素の重み付けを行うために、同じレベルの要素同士の対比較を行う。今、要素1のウエイト（真のウエイト）を w_1 、要素2のウエイトを w_2 、…要素 n のウエイトを w_n とする。また、要素 i と要素 j とを比べた比較値を $a_{ij} = w_i/w_j$ とする。逆方向の比較値は、その逆数とする。

同レベルに属する n 個の要素を総当たりで対比較し、得られた比較値を行列 A 、真のウエイトをベクトル W として次のように表す。

$$A = \begin{pmatrix} 1 & w_1/w_2 & w_1/w_3 & \cdots & w_1/w_n \\ w_2/w_1 & 1 & w_2/w_3 & \cdots & w_2/w_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_n/w_1 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

$$W^T = (w_1 \ w_2 \ \cdots \ w_n)$$

行列 A に、右側からベクトル W を掛けると

$$AW = nW$$

が成り立つ。

実際の意味決定問題では W が未知であり、 w_i を直接知る方法はないものとする。しかし、要素間の主観的な比較 a_{ij} を判断することは可能であるとし、これを a_{ij} の推定値として行列 A を考える。このとき、主観的な比較の値として、Saatyは、同じくらい重要なら1、少し重要なら3、かなり重要なら5、非常に重要なら7、極めて重要なら9とすることを提案する。要素1が要素2よりも少し重要ならば、 $a_{12}=3$ とし、 $a_{21}=1/3$ となる。本研究でもこの数値化の基準に従ったが、調査票の文言は変更した。

主観的な対比較で得られた A を用いて、固有値問題

$$A'W' = \lambda W'$$

を解いて A' の固有値 λ および固有ベクトル W' を求め、 W の推定値とする。ただし、固有値は λ のうち最大の $\lambda_{\max}$ のみを採用する。

ところで、問題の状況が複雑になると、判断の整合性がくずれ、主観的判断で真の比較値を推定できなくなり、信頼性が低下するおそれがある。AHPでは、 A' の整合性を考慮できる。つまり、推移律 $a_{ik} = a_{ij}a_{jk}$ がすべての i, j, k について成り立つとき、意思決定者の判断が首尾一貫していると考えられる。整合性を計る尺度として、整合度指数（Consistency Index）

$$C.I. = (\lambda_{\max} - n) / (n - 1)$$

を考える。整合性が完全に保たれているとき、 $C.I. = 0$ 、つまり $\lambda_{\max} = n$ である。整合性が失われると、 $C.I.$ の値は大きくなる。木下によれば、経験的に0.15以下であれば整合性のある評価だとしてよいとする（Saatyは0.1以下を求めるべきだとする（木下2000、Saaty 1980）。この指数を用いることによって、各レベルの判断の信頼性を検討し、モデルの再構成を行ったり、要素の取捨選択作業を通じて、よりよい意思決定に近づけることがで

きるのがAHPの特徴の一つである。

(3) ウェイトの総合

各レベルの要素のウェイトを推定した後、全体の重み付けを行えば、総合目的に対する各代替案の優先順位が決定できる。「総合目的」に対する「評価基準」のウェイトがベクトル v で表されるとする。また、各「評価基準」に対する「代替案」のウェイトがそれぞれベクトル $v_1, v_2, \dots, v_n$ で表されるとき、 $v_1 \dots v_n$ によって行列 V を考えると「総合目的」に対する代替案のウェイト W^T は、 $W^T = Vv$ で表すことができる。

2.2 本研究のモデル構成

本研究では、相互評価によって他の受講者を評価するとき、いくつかの観点から優劣を判断し、総合的な評価をしていると考えた。これをモデル化することを試みる。

対象となる科目の評価を「総合目的」(レベル1)とする。受講者は、評価者として、授業時における説明資料のわかりやすさ、発表内容の充実度、授業への参加などの評価観点にしたがって、他の受講者を評価するものとする。

この科目は、学校現場で最近取り入れられている新しい授業方法について述べられた文献を分担講読して発表する演習形式で行なわれた。毎回、1人が文献の担当部分を精読し、内容の要約と関連する情報を提供する資料を作成し配付する。担当教員である筆者と他の受講者を交えて内容について検討するという演習形式をとった。

授業の性格を考えて、評価観点項目(レベル2)として、①説明資料(発表レジュメ等)のわかりやすさ(以下、略記するときは「説明資料」)、②発表内容の充実度(同じく「内容」)、③討論の際の新知識の提供(同じく「新知識」)、④討論の整理や方向づけ(同じく「討論整理」)、⑤授業への参加度(同じく「参加度」)の5項目をあげた。これらのうち、①や②については、さらに細かい

評価項目に分割することも考えたが、限られた時間でアンケート形式による回答を求めるため、分割せず、①～⑤を同じレベルの独立した要素と考えることにした。

「代替案」(レベル3)は受講者である。当初8名であったが、この調査を実施した時点では7名になっていた。以下では受講者はA～Gの記号で示す。

このモデルを図1に示す。

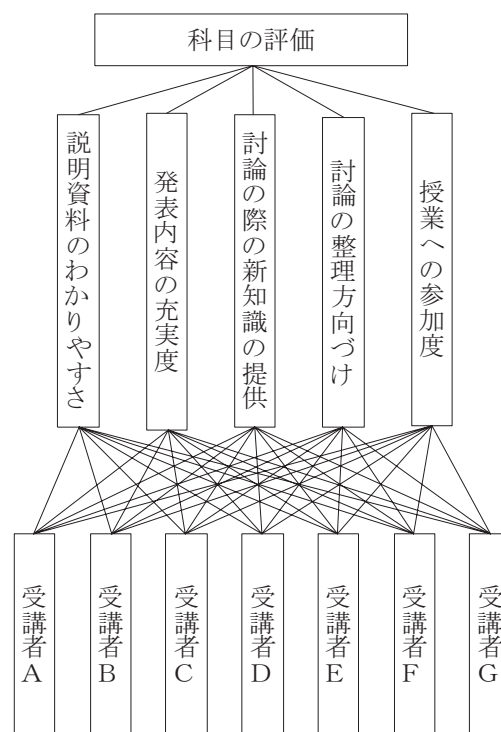


図1 相互評価による受講者の評価モデル

2.3 アンケート調査の概要

前述の科目の受講者7名に対して、AHPに基づくアンケートを実施し、相互評価を求めた。対象者は本学教育学科3、4回生の学生である。

レベル2の評価観点項目とレベル3の受講者のウェイトをたずねるために、図2のような調査票に回答を求めた。ここでは、各受講者の評価は

「説明資料」のみ示したが、「内容」「新知識」「討論整理」「参加度」のそれぞれについても同形式で行っている。

この科目を受講して得た知識や関心にもとづいて、あなたがこの科目の成績をつける立場にいて、

○説明資料(発表レジュメ等)のわかりやすさ
○発表内容の充実度
○討論の際の新知識の提供
○討論の際の整理や方向づけ
○授業への参加度(出席を含む)

の5つの基準について、この科目の評価をする際に、どれが最も大切だと思うか、2つずつ取り上げて比較してみてください。たとえば、資料のわかりやすさが発表内容に比べて少し重要だとすると、

資料 9—7—5—3—1—3—5—7—9 発表内容

というふうにしてください。

	非 常 に	か な り	少 し	わ ず か に	同 等	わ ず か に	少 し	か な り	非 常 に	
資料	9	7	5	3	1	3	5	7	9	発表内容
資料	9	7	5	3	1	3	5	7	9	新知識
資料	9	7	5	3	1	3	5	7	9	整理
資料	9	7	5	3	1	3	5	7	9	授業への参加
発表内容	9	7	5	3	1	3	5	7	9	新知識
発表内容	9	7	5	3	1	3	5	7	9	整理
発表内容	9	7	5	3	1	3	5	7	9	授業への参加
新知識	9	7	5	3	1	3	5	7	9	整理
新知識	9	7	5	3	1	3	5	7	9	授業への参加
整理	9	7	5	3	1	3	5	7	9	授業への参加

次に、それぞれの基準について、自分と他のメンバーを比較して、どの程度優れていたか、劣っていたか、貢献していたかを評価してください。自分の名前のところは書かなくてください。

○説明資料のわかりやすさについて

	非 常 に	か な り	少 し	わ ず か に	同 等	わ ず か に	少 し	か な り	非 常 に	
(受講者A)	9	7	5	3	1	3	5	7	9	自分
(受講者B)	9	7	5	3	1	3	5	7	9	自分
(受講者C)	9	7	5	3	1	3	5	7	9	自分
(受講者D)	9	7	5	3	1	3	5	7	9	自分
(受講者E)	9	7	5	3	1	3	5	7	9	自分
(受講者F)	9	7	5	3	1	3	5	7	9	自分
(受講者G)	9	7	5	3	1	3	5	7	9	自分

(※実際には、(受講者A)～(受講者G)の箇所には個人名が記入してある。)

図2 調査票(一部)

ただし、このアンケートで評価する範囲は学期の後半にあたる部分のみとした。成績は学期全体を通した平常の評価と最後に提出するレポートに対する評価を合わせて筆者の責任で行い、アンケートによる評価は参考とするが、主に学術目的による調査であることを明言し、成績に直結するものではないことを説明した。調査は、平成17年1月の最終回の授業時に実施した。受講者はA～D

が3回生、E～Gが4回生であった。また、2名が男性、5名が女性である。

まず、評価観点項目が他の項目に比べてどの程度重要であると考えられるかの判断を求める。図2(前半部)の用紙を用いて、①～⑤の観点項目の重要性について、9段階の評価による一対比較を求めた。たとえば、①「説明資料のわかりやすさ」と②「発表内容の充実度」とを比較し、どちらの要素がどの程度重要であると考えるか、該当する位置に○印をつけてもらった。①が②に対して少し重要だという回答であれば、①から②に対する数値として5を与える。逆に、②から①に対する数値は逆数の1/5とする。

次に、①～⑤の観点項目ごとに各受講者が受講者同士を比較し、どちらが優れているかを判断して、観点項目と同じく、9段階の評価による一対比較を求めた(図2 後半部)。ただし、本来のAHPの方法では、比較は総当たりで行うが、受講者については、自分自身と他の受講者との比較に限定した。

自分自身との比較に限定したのは、評価の簡易化と安定を狙ったものである。本来の一対比較によるAHP(相対法)では、対象が比較的少ない場合でも、幾何級数的に比較回数が増加する。

この調査の場合、レベル2の評価観点項目が5つであるから、レベル2の一対比較は(5×4)/2=10回となる。受講者7人をすべて一対比較すると(7×6)/2=21回の比較が必要となり、それを評価観点項目それぞれについて行うので、結局、10+21×5=115回の判断が必要となる。しかし、ここで用いる方法のように、自分自身を基準とすると、各評価項目における受講者の評価は、自分と他の受講者6人との比較であるから6回になり、合計で10+6×5=40回の判断ですむ。

さらに、自分よりも優れていたか劣っていたかは直感的に判断できるため、安定した評価が期待

できると考えた。他の受講者同士の比較をするためには、評価することを目的として、相当注意を払いながら授業に参加していなければ、自信をもって評価することはできない。担当教員が比較する場合はそうすることが必要だろうが、受講者は評価するためではなく、学ぶために授業に参加しているのであり、過大な負担を強いることはできない。また、わからないからといって、いい加減な回答をするおそれもある。

AHPで問題となる循環問題、いわゆる「三すくみ」の現象が起こりにくくなる点も利点である。循環問題とは、要素Aが要素Bよりも優れていると判断し、要素Bが要素Cよりも優れていると判断した場合に、要素Aが要素Cよりも優れていると判断すべきなのに、要素CがA要素よりも優れていると判断してしまい、三つ(ないしそれ以上)の要素に対する判断が矛盾を起こす問題である。この問題の解決のために、各要素に対する評価を、比較ではなく絶対評価にすることも提唱されている(たとえば、木下2000 pp.87-101を参照)。しかし、ここで用いる方法では、要素同士の比較範囲を限定するため、循環問題の発生を避けられる。

2.4 ウェイトの決定

次に、前節の方法によって実施したアンケートの結果をもとにした一対比較行列の固有値と固有ベクトルを求め、各要素のウェイトの算出(重み付け)を行った。

評価観点項目(レベル2)は、全員の回答の幾何平均によって一対比較行列を作成した。つまり、評価観点項目①と②を比較するとき、ある受講者が9と回答したときに、別の受講者が1/9と回答すると、両者の判断は相殺されて $9 \times 1/9 = 1$ となる。算術平均では $(9 + 1/9) \div 2 = 41/9$ となってしまう、意味をなさない。

受講者(レベル3)の重み付けをするための一対比較行列は次のように考えた。幾何平均を用い

ることは同様であるが、受講者Aと受講者Bとの比較判断を行うのは、当事者であるAとBのみである。Aが行ったAとBの比較を w_{ab} とし、Bが行ったBとAの評価が w_{ba} とする。Aを基準とすると、Bによる w_{ab} は、 $1/w_{ba}$ となる。この二つの値の幾何平均である $(w_{ab}/w_{ba})^{1/2}$ をAとBとの比較値とした。たとえば、Aが自分がBよりも非常に優れていると考えた場合、比較した値を9と評価する。一方Bは自分の方がわずかに優れていると判断した場合、1/3をAに対して与える。この回答の幾何平均を行い、 $(9 \times 1/3)^{1/2} = 1.732 \dots$ を w_{ab} として採用する。

このように、一人ひとりが行った他の受講者と自分との比較から全体の一対比較の値を推計し、行列の形で整理した。

3. 結果と考察

3.1 評価基準の重み付け

まず、レベル2の評価観点項目について、各々のウェイトを求める。表1は一対比較行列と固有ベクトルの成分を示したものである。固有ベクトルの成分が各評価観点項目のウェイトを示す。

先に述べたように、評価観点項目は受講者全員の意見を取り入れるために、各人の判断を幾何平均したものを一対比較行列の要素とした。全員の意見を元にして重要性を決めたのは、クラス全体における評価として考えたいからである。つまり、受講者ごとに重視する観点項目は違っており、自分が重視する項目に力を入れて学習するだろうから、個々の受講者ごとに算出すると、相互評価ではなく自己評価の側面が強くなってしまわないかと考えた。全体の意見を取り入れれば、観点項目の重要性をクラス全体の基準としてとらえることができる。

この行列の固有値のうち最大のものを $\lambda_{\max}$ とすると $\lambda_{\max} = 5.09$ である。また、整合性C.I.=0.02とな

り、評価の整合性は充分高く、矛盾が少ないことがわかる。ウエイトは成分の和が1になるように基準化した。

ウエイトの数値を見ると「討論の整理や方向づけ」が最も高く「授業への参加度」が続く。「発表内容の充実度」が中間にあり、「説明資料のわかりやすさ」や「討論の際の新知識の提供」はあまり重視されていないという結果になった。

指定した文献は専門書ではなかったが、小・中・高校での、あまり知られていない授業方法を

紹介するものであり、学生の関心は高かった。筆者はできるだけ口を挟まず、一方的に知識を与えることは慎んでいたが、各の受講者が担当する部分の内容を他の受講者に伝えようとする熱意は高かった。内容は高度ではないが、しばしば分かりにくい記述があったため、担当者の説明ではわからず「これはどういう意味だろう」「具体的に、どのように教えているのだろうか」といった議論の後、受講者が納得するといった場面がしばしばあった。

表1 評価観点の一対比較とウエイト

	説明資料	内容	新知識	討論整理	参加度	ウエイト
説明資料	1	0.855	0.865	0.367	0.679	0.136
内容	1.170	1	1.81	0.767	0.655	0.192
新知識	1.156	0.553	1	0.394	0.655	0.133
討論整理	2.728	1.305	2.536	1	0.855	0.287
参加度	1.472	1.526	1.526	1.170	1	0.252

$\lambda_{\max}=5.09$

C.I.=0.02

ここからうかがえるのは、この授業においては、討論への貢献や参加が重要な要素だと考えられていることである。一方、全員が一つの文献を所持してあらかじめ読んでいるため、必ずしも担当者の説明にのみ頼っているのではないことが推測できる。また、文献に書かれていない事柄まで紹介するということは、さほど求められなかったことを示すものと考えられる。

3.2 各人の観点別評価と総合評価

次のステップとして、受講者の評価を行う。レベル2の各評価観点項目について考えたとき、受講者Aと受講者Bとではどちらが優れているかという判断を求める。

オリジナルのAHPでは、すべての受講者の組み合わせについて比較を行うことを求めるが、2.で述べたように、自分自身と他の受講者との比較のみを求め、当事者の評価の幾何平均を採用した。

各人と他の受講者との比較のみから全体の一対比較を構成して一対比較行列をつくった。5つの評価基準ごとに、受講者7名のウエイトを求めた結果が表2である。たとえば、説明資料については、受講者G(0.302)が最も優れており、D(0.206)が続き、Cは他の受講者からの評価は低いことを意味する。一対比較行列は省略する。

どの観点項目に対する重み付けも、C.I.は0.15以下であり、整合性は高い。受講者の判断は信頼してよいだろう。

この結果を見ると、各受講者の特徴が表れている。受講者Aはどの項目も平均的な値を示す。BとDは参加に高い値を示すが、これは出席率が非常に高かったことを反映しているものと思われる。Dは文献で紹介されている授業について、教具を自作して実演し、相互評価でも優れていると判断されている。Fは文献で取り上げられている

教育研究活動に実際に参加しており、その活動内容を紹介するパンフレットを持参していた。これらの結果は、筆者自身の観察ともよく合致してい

る。ただし、筆者の判断とは異なる点もある。これについては、考察で述べる。

表2 各受講者の評価観点項目ごとのウェイトと総合評価

受講者	説明資料	内 容	新知識	討論整理	参加度	総合評価
A	0.081	0.132	0.122	0.150	0.091	0.119
B	0.071	0.093	0.101	0.108	0.266	0.138
C	0.042	0.038	0.046	0.046	0.024	0.039
D	0.206	0.195	0.171	0.202	0.268	0.213
E	0.121	0.104	0.135	0.114	0.126	0.119
F	0.177	0.158	0.193	0.151	0.112	0.152
G	0.302	0.280	0.232	0.228	0.112	0.220
	$\lambda_{\max}=7.618$ C.I.=0.103	$\lambda_{\max}=7.174$ C.I.=0.029	$\lambda_{\max}=7.285$ C.I.=0.048	$\lambda_{\max}=7.579$ C.I.=0.097	$\lambda_{\max}=7.803$ C.I.=0.134	

3.3 考察

本研究では、相互評価によるAHPに基づくアンケート調査により、次の点が確認された。

- (1) 評価観点項目については整理・方向づけや参加度という、討論の過程が重視されている。また、その比較判断には矛盾がほとんどないことが確認できた。
- (2) 各評価観点項目に対する受講者の相互評価も整合性が高く、矛盾はほとんどないことが確認できた。

また、受講者の一対比較の際に全ての受講者を総当たりで比較をしなくても、評価者自身と他の受講者との比較のみによって、一対比較行列全体の推定が許されることが示唆された。

3.3.1 調査の問題点

アンケートが終わった後、学生に感想を求めたところ、拒否的な反応は出なかった。回答は15分程度で終了しており、過剰な負担にはならないだろう。

評価観点項目（レベル2）については特に意見はなかったが、レベル3の他の受講者との比較で、

自分が優れていると書くことは遠慮があるようであった。また、受講者が7名と少ないのでお互いに顔と名前が一致していると考えており、この点は問題はなかったが、授業中の活動についての記憶が曖昧になっているようであった。各人が作成した資料等を取り出して記憶をたどって回答してくれていたが、難しかったかもしれない。アンケートを行ったのが学期の授業の最終日（1月上旬）で、評価すべき期間の始まり（11月上旬）から2ヶ月程度経過していたことにもよると考えられる。授業を3回程度行う都度実施するなどして、評価すべき期間の短縮を検討する必要があるだろう。

また、筆者の目からはある点で明らかに優れている学生が、他人との比較で遠慮したのか、自己評価が低くなっている場合があった。特に、同学年、同性の受講者に対しては「同等」という評価が目立った。総合評価では、受講者AやEの値があまり高くないが、筆者の観察では、他の受講者に比較して劣っていたとは判断していない。受講者の自己評価の高低を調節するプロセスとして、

具体的にどのような点が優れていたかを明記してもらったり、個別にアンケートに答えるのではなく、議論の機会を設けて全体の意見を調整したりすることを試みる必要があろう。

総合的には、AとBが同程度で、D、E、Fがそれより少し優秀、Gがもう少し優れているというのが、筆者の主観による判断であった。AHPの結果との違いは自己評価が低かったことのみではなく、評価観点項目のウェイトの影響も大きい。筆者は、資料や討論の際の発言内容が重視される知識獲得型の授業として考えていたが、実際には受講者は表1にあるように、討論の過程を重視している。担当教員と受講者との思惑は必ずしも一致していないことを示している。

3.3.2 比較の方法について

すべての対象を一対比較せずに、自分自身を基準とする方法は、他の受講者の対応する回答を、その受講者自身にとっての評価として考えるときには逆数にするという点で、調査実施上の難しさがあった。

たとえば、受講者Aの回答用紙に、

B 9—7—5—3—1—3—5—⑦—9自分

とあれば、 $w_{ab}=7$ と読むのに対し、受講者Bの回答用紙に、

A 9—7—5—3—1—3—5—⑦—9自分

とあれば、 $w_{ab}=1/7$ と読まなければいけない。つまり、受講者Aは自分の方が「かなり優れている」と考えているのに対し、受講者Bも「自分の方がかなり優れている」と考えているから、2人の回答の幾何平均は $(7 \times 1/7)^{1/2} = 1$ となり、相殺される。今回は筆者が調査用紙の整理を一人で行ったため、転記の誤り等はないはずであるが、回答用紙の形式を工夫する必要がある。

また、循環問題は必ずしも解消すべき問題ではないという意見もあり、可能であれば、全員の一対比較をおこない、比較する必要もある。今後の

課題としたい。

3.3.3 AHPの応用の可能性

AHPは、単純な仮定から出発し、数学的には単純な手法であるから実際の状況に適合するかどうかは、モデルの妥当性にかかっている。あまりに多くの要素を比較しようとしても比較そのものができない組み合わせも生ずる可能性があるため、各レベルの要素の数は多くとも9程度までにすることが推奨されている。AHPの専用ソフトである「Expert Choice」でも各レベルで扱える要素は9以下に制限されている。これは人間の認識能力の限界を反映しているものと思われる。したがって、考慮すべき評価基準が多い場合、レベルを増やし、何段階もの複雑なモデルが作られることがあるが、本研究のような問題を扱う場合、評価を行うのが受講者自身であり、また、考慮すべき対象となる活動は、授業での行動や態度に限られるため、単純なモデルに留めた。ただし、今回は筆者があらかじめ決定した評価観点項目を元に試みたが、この点は検討する必要がある。KJ法やブレーンストーミングを利用して、受講者自身の意見を求めることも考えたい。その作業を通して、受講者がどのような学習活動が求められるのかを自覚ようになることも期待でき、評価のみならず授業全体の改善に資する方法になる可能性もある。

当初、AHPによって得られた重要度や各受講者に対する評価値を科目の成績をつけるための参考にできるかもしれない、つまり、評価の点数としてそのまま用いることもできるかもしれないと考えていた。しかし、順位付けのために用いるか、受講者をグループ化する目安にはなるようであるが、成績として得られた数値をそのまま評価点に反映させることは難しいようである。得られたウェイトの数値の扱いについては、先行研究も妥当な考えを述べていない。たとえば、予算配分問題にAHPを応用したとき、ウェイトの数値に比例し

た金額を配分すべきかどうかまで示されるわけではない。代替案となる施策の性質や行政の役割に対する見方などにかかっているのである。AHPは意思決定を助けるツールではあるが、意思決定を任せるものではない。評価の参考にできることを示すにとどめる。

4. 結論

本研究では、大学における少人数の演習形式授業での受講者の相互評価について、AHPの適用可能性を探った。その結果、評価観点項目は同じ程度の重要性しかもたないのではなく、受講者が重視している側面としていない側面とがあり、その判断は首尾一貫していることが示された。

また、受講者の相互評価による重み付けで評価した結果は、ある程度、担当教員の主観的な評価を裏付けるものであった。

従来の相互評価に比べると、評言による曖昧さがなく、評価観点項目の重みを示すことができ、より合理的な評価が可能になることが示唆された。ただし、安定した相互評価のために、評価する過程の検討がさらに必要である。

なお、今回受講者の協力を得た授業は、順位付けや選抜が目的ではなく、通常の授業であり、科目の成績評価についても、従来通りの方法で、担当者である筆者が行ったことを確認したい。

今後、評価観点項目の取捨選択を行ったり、受講者個別に相互評価を行ったりするだけではなく、評価のための討論などを試み、AHPが教育評価に有用であることを示せるものと考えている。

参考文献

- 浅野誠1994『大学の授業を変える16章』大月書店
東洋2001『子どもの能力と教育評価』東京大学出版会
林義樹1994『学生参画授業論 人間らしい「学びの場づくり」の理論と方法』学文社

- 池田輝政・戸田山和久・近田政博・中井俊樹2001『成長するチップス先生』玉川大学出版部
片岡徳雄・喜多村和之編1989『大学授業の研究』玉川大学出版部
加藤幸次編著2001『総合評価に活かすポートフォリオ評価の実践』金子書房
経済学教育学会編1998『大学の授業をつくる 発想と技法』青木書店
木下栄蔵編著2000『AHPの理論と実際』日科技連
木下栄蔵2000『入門AHP』日科技連
高等教育情報センター編2003『成績評価の厳格化と学習支援システム』地域科学研究会
宮地功1993「AHPによるコンピュータ実習の報告書の定量的評価と分析」『教育情報研究』Vol.15 No.1 pp.15-22
宮地功2001「AHPによる大学の授業評価」『日本教育情報学会年会論文集』Vol.17 pp.162-165
Saaty, T. L. 1980 The Analytic Hierarchy Process, McGRAW-HILL
佐藤学1996『教育方法学』岩波書店
関谷和之2003「ANPを組み込んだAHPの適用：教員の評価を例題として」『オペレーションズ・リサーチ』2003年4月号 pp.259-264
鈴木規夫1996「AHPによる大学・学部選択における意思決定モデルの構成」『教育心理学研究』Vol.44 pp.287-295
刀根薫・眞鍋龍太郎編著1990『AHP事例集』日科技連

