

お茶の水女子大学における学生による授業アンケート分析報告 2015-16

半田智久

お茶の水女子大学 教学 IR・教育開発・学修支援センター

Report on analysis of the results of course evaluation questionnaire by students in Ochanomizu university, 2015-16

Motohisa HANDA

Ochanomizu University; Center for Institutional Research, Educational Development, and Learning Support

はじめに

お茶の水女子大学では2015年度から自主開発したスマート・ウェブ・サーベイを用いて学生に対して授業アンケート nigala を実施している。この手法は名称が示すとおり学内 LAN 上でウェブブラウザを用いておこなう調査だが、それをあえて「スマート」と称しているゆえんは半田(2015)が詳述している。その要点はつぎのつおりである。

第一に、nigala ではウェブ調査の特性を活かし、とりわけその POR (Point Of Responding) 特性を最大限に活用して学生が回答した時点が事実上の集計となる仕組みをつくり、同時にその回答の時間管理を可能にすることで学生一人ひとりの行動特性も知りうるデータを捕捉している。この種のアンケート調査が質問紙でおこなわれていた時代にあつては匿名調査にすることで、気兼ねない率直な回答を担保するという見方が通用していた。だが、現今ではむしろその「率直な意見なり反応」という見方が宿す価値はともかく、調査の現実においてはその匿名性が無責任な回答や遊び半分の回答の発生を誘発しがちになり*1、また「(匿名の) 誰かさんたちが回答しているだろうから自分はよいだろう」といったモラル・ハザードを高めることにもなってきた。この調査の性質上、その目的の一部とは裏腹に、政治参画同様、回答した結果が目に見えて変化につながることを、にわかには実感できにくい調査の場合は、なおさら匿名性の長所は現実場面であらわれがちに短所の背後に縮退する。よってむしろこの種の調査では記名ないし誰が回答したかが「わかりうる」仕組み(本人認証を介したウェブ調査)のもとで実施し、実際に結果分析でその識別をおこなう

かどうかは別として回答者を特定できる環境下で実施することを明示し、責任を伴う回答、顔の見える積極的なコミュニケーション・ツールとしてのアンケートという特性を実施し、その利を活かそうというわけである。たとえそれによって生じうる配慮による下駄履き評価が生じがちになったとしても、評価の絶対値から離れ、授業間や科目群間の相対的な差異に観察や分析の焦点をおくことでその点は十分に乗り越えられる。

そうしたことからこの調査では回答者個人の反応特性を知りうるデータもあわせて抽出している。その匿名調査に比した情報量の豊かさと分析力の広がりにおいてこの調査はスマートなサーベイになっている。

第二に、反応時点管理の特性とは換言すれば無媒介の回答集計のことである。ここに無駄を廃した端的な効能を認めることができる。従来の質問紙調査では質問紙なりそれと切り離された回答用紙は回答者の直接の反応を得るための生のデータであり、調査結果の原点になる。だから、それは調査にとって最も大切に扱われる存在で、とりわけ回収されたそれらはたとえば入学試験であれば、まさに現金扱いの輸送体制がとられる。しかるにそれらはその後、読み取られて原データ化されたのちは一気にその存在価値が失われ、読み取り不備等のリスクヘッジとして倉庫にはしまわれるものの、時が経てば出すのも厄介な廃棄対象となる。そうした膨大な紙の束を目前にした経験が一度でもあれば、たとえば無機物とはいえ、一瞬にして天から地へと急落する用済みということの切なさを感じると同時に、この大いなる無駄は何とかならないだろうか、と思わざるをえなくなる。ここには回答抽出のための紙媒体という具体的媒介物があるだけでなく、これを印

刷する人、束ねる人、封入する人、運ぶ人、受け取る人、開封する人、仕分ける人、配る人、回答する人、回収にかかわる人、集まったものをチェックする人、また束ねる人、回収用紙を入力する人といった具合に実に多くの人手も動員、媒介される。その過程で何人も人が紙で手を切り、痛い思いもする。この労務コストもさることながら、これもまた一度でもこうした作業に携わればすぐに実感できるように、これらの労務のほとんどすべては単純反復の肉体労働で、これを事務ならぬジムのトレーニングとでも思い込んでおこなわないかぎり費やしている時間の無駄を惜しむことしきりの作業になる。要は回答を得ることだけなのだが、この多重媒介過程が超絶的に肥大する組織の体脂肪をもつくりだしている。

POR、反応時点管理とは回答者が反応したときの生のデータがそのまま原データとして機械に記録（管理）されるということである。上記の幾段もの媒介プロセスの始端と終端がワープする。だから、そのあいだにあった退屈きわまるすべての労務は消失する。そればかりか紙幣的価値から紙屑へと一気に急落する切ない紙の山もなくなる。ウェブ・サーベイは従来の観点でいえば、回答者が同時に回答読み取り者でもあるかのように回答するようなものなのである。だから、実査から集計・分析への動きが迅速化する。集計処理などは実査が完了した時点でほぼ終了する。まさにその点でこれはメタボ症候群からおさらばしたスマートなサーベイなのである。

スマートさの第三のポイントはこの第二点から派生し、これが授業アンケートに適用されることで生じるスマートさである。この調査実査が無媒介で実施されるということはアンケート用紙を印刷し、仕分ける等の事務作業も、授業ごとにそれを学生に配布し、あるいは回収する教員ないしTAの作業も不要になるということである。それはその労務や配慮が消えてなくなるという組織にとっての好都合に留まらない。そもそも授業の内容や方法について受講者に振り返り評価を求めることを、その授業の最終段階で時間を割いて実施すること、しかもできうればすべての授業でそれをするということには基本的な問題が孕まれている。「最後にこのアンケートに時間がとられることなく授業を受けることができたなら満点の授業でしたのに」などといった自由記述回答が単なる皮肉ではなく、教員としても相応の授業を実施した最後にアンケートを自らの手でするということは授業が満足のいくものとなり得た度合いに応じて興ざめするものだからである。

だから、nigalaのスマート・ウェブ・サーベイは授業に非侵襲で実施する。質問紙を配布する必要がないから、学生は授業中に回答する必要がない。授業期間が終了して以降、授業アンケート期間に回答する。授業中に回答せずに済むだけではない。本人認証を経て回答するから、本人が当期に履修した授業が設問毎に一覧表示され、受けた授業を相対的に比較しながら回答できる。これは授業ごとに質問紙が配布されて回答する形式では不可能なことだったが、この方法をとることで授業間、科目群間の比較分析も原データの水準で基本的にそれをするに足る比較評価の条件が満たされることになる。しかも、nigalaでは主たる回答反応スケールが一般にありがちな5～9段階の離散的ないわゆるリッカート尺度ではなく、-50～50までの101段階スケール上でなされる。事実上の連続量尺度である。

これは医療における痛みの測定でしばしば用いられているVAS (Visual Analog Scale) である (e.g., Melzack, 1975)。VASでは患者の痛みの程度を把握するのに、その個人がこれまで経験した最大の痛みと無痛のあいだの10cm程度の尺度上においてどのあたりであるか、指示を求める。単純な診断法だが、そもそも個人差が大きい痛みの主観評価を、同様に個人差が大きい言語による表現を用いずにできるだけ適切に把握するためには、まさに的を射た測定法として現場で利用されている。その利点は授業へのアンケートでも活かすことができる。臨床では個々の対面診断だから物差しを当てて読み取る。授業アンケートではそれが不可能なのでウェブ・サーベイが有効性を発揮する。Figure 1のごとくスライダーが表示され、そのつまみを動かすことで回答する。

スケールが101段階で構成されているというのは

残り7問：この授業は学生の理解度を把握しながら進み、全体の内容は質、量ともに適切でよく理解できた。

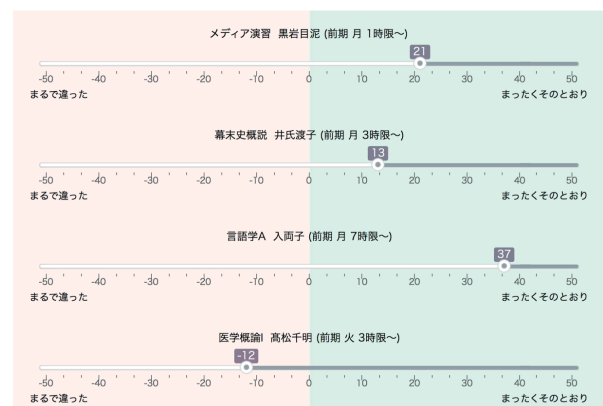


Figure 1 nigala 授業アンケートにおける回答スケール

それほど微細な反応差異を読み取れるという点に意味があるわけではない。この事実上の連続量スケールは、この種の主観評価に免れえない評価の仕方における多様な個人差を一定程度標準化するうえで有効なのである。端的な例でいえば、5～9段階の段階評価ですべての回答が「どちらともいえない」という結果になったとしても、その同じ「どちらともいえない」なかでも微妙な違いではあるが弁別的な評価ができる場合がある。そこに表現される違いが他者にとっては明らかな差異として相似的に回答されていることや、そうでないこともあってとくに前者の場合は統計的には確たる差異に近づく結果として浮き出すことになる。このことは成績評価において、たとえば全員Aではあるけれども、そのなかでまったく差異がないというわけではない、ということと同じである。だから、お茶の水女子大学のGPAは素点からグレード・ポイントを算定することでGPAにおいても同様の意義ある差異を測っている。同じ考え方が授業アンケートでも採られている。この繊細さという点でこれはスマートなサーベイなのである。また、このデータ・ドリブンの反応抽出は「言語表現に依拠しないループリック」を定量的に実現しているということでもある。そのことから、この手法をリニア・ループリック法と称してもいる。

当報告の目的

2015、および2016年度にお茶の水女子大学の学士課程で実施した各年度の前・後期、計4回の授業アンケートの結果を分析した結果を報告する。分析の焦点はつぎの5つである。

- (1) 結果の全学的な鳥瞰、設問間・学期間・年度間の結果差異
- (2) 科目群間の結果差異
- (3) クラスサイズ間の結果差異
- (4) 開講学期間の結果差異
- (5) 成績評価の仕方の相違と授業アンケート結果の関係

方法

調査対象

2015、および2016年度にお茶の水女子大学の学士課程で開講された全授業が調査対象であった。ただし、集中授業は開講やその時期の決定が一律に定まら

ずアンケート時期を事前に決めがたい場合があることから対象外とした。

調査対象数は2015年前期1200科目、後期1321科目、計2521科目、2016年前期1218、後期1348、計2566科目であった。通年開講授業と不定期開講授業は後期に含めてアンケートを実施した。なお、一般的なこととして毎年度多少のカリキュラム改訂がなされる一方で、学年進行で改定前の開講授業も実施される。そのようなことから、異なる授業科目名で同時に実施する授業科目も生じる。これは同一教員が同一日時に実施しているわけであるから、上記の数値ではそうした例では1授業として算定している。そのため、名目上の開講授業科目数と上記の数値には相違がある。

調査回答者

調査回答者は2015、および2016年度にお茶の水女子大学の学士課程（文教育学部、理学部、生活科学部）に在学していた正規学生全員であった。その他、大学院生で学士課程開講科目を履修した場合もアンケート回答対象者に含まれた*2。

回答抽出法

本稿冒頭に述べたスマート・ウェブ・サーベインigalaを用いて回答を得た。

設問

2015年度からここで説明しているnigalaスマート・ウェブ・サーベイを実施し、従前の質問紙配布回収による方法を改めた。伴って設問の内容も数も見直した。その経緯、変更内容や方法は半田（2015）が詳述している。

設問はリニア・ループリック法による8設問と自由記述による1設問で構成した。回答者は大学の本人認証システムを介してアンケートに着手した。そのため氏名、所属等の本人属性に関わる設問はない。リニア・ループリック法8設問と各スケール両端の表現はつぎのとおりであった。ただし、設問2の時間数については2015年度はスライダー上で6段階（0 1 2 3 4 5時間以上）の選択をするようにしていた。その後、結果の読み取りにおいて時間評定はより細かな値での採取が望ましいことがわかり、2016年度からは「() 時間 () 分」の形式で数値記述にした。

(1) この授業のシラバスは過不足なく、わかりやすく書かれており、目的や到達目標は実際の授業に合致していた。

【まるでその反対であった —— まったくそのとおりであった】

(2) この授業について授業時間外に学習した時間数は授業1回あたりおおよそ、

(3) この授業は学生の理解度を把握しながら進み、全体の内容は質、量ともに適切でよく理解できた。

【まったくそうではなかった —— まったくそのとおりであった】

(4) 授業に対する教員の熱意を感じ、今後の自分の学修や研究にとってたいへん意義があった。

【まったくそうではなかった —— まったくそのとおりであった】

(5) 教員の説明はわかりやすく、ていねいで、授業は創意工夫に満ちていた。

【まるでそういうことはなかった —— まったくそのとおりであった】

(6) 授業に学生の意欲が湧くような配慮があり、実際にとても意欲的に取り組めた。

【まるでそういうことはなかった —— まったくそのとおりであった】

(7) 授業の主題や内容に対する興味や関心がたいへん高まった。

【まるでそういうことはなかった —— まったくそのとおりであった】

(8) 成績評価の方法や基準はあらかじめ明確で適切であった。

【まったくそうではなかった —— まったくそのとおりであった】

(9) 各授業について教員に伝えたいことがあれば200文字以内で自由に記述してください【教員には記述者情報は【求めがあっても】知らせず記述内容だけが知らされます】。

これらは2015年度前期の仕様であった。調査方法をまったく改めたことと設問も全面的に改訂したことにより、結果の検討などをつうじて当初数年は継承性を保ちつつ、探索的かつ積極的に改善を進めた。その結果、上述した時間評定の他に、2015年度後期からはつぎの2点を改訂した。

(3) の設問文「全体の内容は量も」を「全体の内容は質、量ともに」に変更

(9) の設問文「200文字以内」を「200字程度」に変更

2016年度前期からは全体に選択肢の表現を簡潔に

し、(2) も含め、設問表記をつぎのように変更した。

(1) この授業のシラバスは過不足なく、わかりやすく書かれており、目的や到達目標は実際の授業に合致していた。

【まるでその反対であった —— まったくそのとおり】

(2) この授業について授業時間外に学修した時間は、平均すると毎週の授業1回あたり約()時間()分くらいだった。

(3) この授業は学生の理解度を把握しながら進み、全体の内容は質、量ともに適切でよく理解できた。

【まるで違った —— まったくそのとおり】

(4) 授業に対する教員の熱意を感じ、今後の自分の学修や研究にとってたいへん意義があった。

【まるでなかった —— まったくそのとおり】

(5) 教員の説明はわかりやすく、ていねいで、授業は創意工夫に満ちていた。

【まるで違った —— まったくそのとおり】

(6) 授業に学生の意欲が湧くような配慮があり、実際にとても意欲的に取り組めた。

【まるでその反対であった —— まったくそのとおり】

(7) 授業の主題や内容に対する興味や関心がたいへん高まった。

【まるで反対であった —— まったくそのとおり】

(8) 成績評価の方法や基準はあらかじめ明確で適切であった。

【まったく違った —— まったくそのとおり】

2016年度後期は改訂しなかった。

なお、設問9の自由記述は常に最後に呈示したが、それ以前の8設問は呈示順序効果の発生を回避するため、学生ごとにランダムな順番で呈示した。この措置がとれたのもウェブ・サーベイならではのことである。

設問9の自由記述設問は回答者の個人属性を示さずに、記述文そのままを各授業担当教員にフィードバックすることを目的にした設問である。そのため本稿では扱わない。

手続き

実施期間はつぎのとおりであった。2015年度前期アンケート7月14日から11月27日まで。同後期1月26日から5月13日まで、2016年度前期7月28日から12月1日まで、同後期1月26日から5月8

日までであった。

お茶の水女子大学での学修成果状況、すなわち履修単位数、取得単位数、授業成績、そのGPAの学内での位置づけや期間毎の推移などは、個々の学生が本人認証のもとで学内LANを介してwebによりいつでもチェックできるようになっている。これをalaginシステムと称している。alaginは授業から学生への成果情報提示のコミュニケーション・システムである。このコミュニケーションを双方向にするのが学生から授業への授業アンケートになる。そのことを象徴するようにシステム名称nigalaはalaginの逆綴りになっている。学生は自分の学修成果をみるにあたり、履修した授業に対するアンケートに応じておく仕組みである。上記のごとくアンケート実施期間が4ヶ月ほどの長期にわたっているのは、各授業期間が終わってからその成績が開示されるまで現状では2ヶ月ほど要していることによる。そのため、アンケート自体は授業期間の終了時には回答できるようになっているが、十分な回答数を得るには成績開示を待つて実際の双方向コミュニケーションができるようになる時期を待つことになる。

この実際のアンケート回答期間の長さは、当然のことながら授業終了時に質問紙を配布・回収していた方法と大きく相違する点である。常識的には授業が終了してから時が経つにつれて記憶が薄れ、適正な回答ができにくくなるという見方もある。だが、心理学の記憶の実験ではレミニセンス効果という現象が知られている。とりわけ有意義なことがらの記憶想起は直後再生よりもある程度、日数を経た後の方が正確さにおいて高まるという現象である。むしろ、その日数も数ヶ月先となれば別の話だが、授業ごとにまちまちに回答するのではなく、履修した授業全体を見比べながら相対的に回答していくnigalaの方法では、授業が終了してから一定期間経た後に、全体を落ち着いて振り返りつつ、記憶に残った授業とそうでもなかった授業という観点も加味しつつ、むしろ差異のコントラストを強めるかたちで回答することもできることになる。むしろこの点の実際の効能は今後の検証に委ねたいところである。

また、実務上の別の観点からいえば、従前の質問紙配布が全授業に対して実施されていたわけではなかったなかで回収率は学期平均57.1% (2011～13年度実績でレンジ49.1～69.9%)であった。これに対して、nigalaでは集中授業のみを除く全授業実施で、回収率は80～90%に達している。これを可能にして

るのがこの双方向のシステム・アーキテクチャである。しかもこの実施、回収に教職員はまったく関与せずに済み(この事態を称してステルス・サーベイと呼んでいる)、加えて毎回無経費で実施、完了している。

アンケート回収はシステムのデータベースに直接なされ、そのまま集計と分析を進める。教員はwebを介して必要に応じていつでもnigala結果フィードバックにアクセスし過去に遡ってその結果を閲覧できる。個人的に結果をファイリングして管理するような手間は解消された。フィードバックされる結果は複数のインフォグラフィクスで示される。その内容は半田(2015)が紹介している。個々の教員にではなく、全学や各種の授業分類ごとの結果はnigalaダッシュボードにより、年度毎にインターネット上で多種のインフォグラフィクスにより公開している(Figure 2)



Figure 2 インターネットにおいてダッシュボード形式で公開している毎年度のnigala授業アンケートの総括的な結果

(<https://crdeg4.cf.ocha.ac.jp/nigag>)。

回答データの集計と分析にあたっては、まず、アンケートにアクセスしたものの、すべての設問に対して無回答のままアンケートを終了したケースを無効回収として除外した。それ以外の回収が有効回収になった。

分析には基本的に2つの指標を用いた。ひとつは回答スケールに直接回答された値。これを本稿では「スケール値」と表現する。スケール値は0を含むレンジ-50～50の整数値である。もうひとつは各学生が一設問に対して履修した全科目に回答したスケール値を、設問毎に標準化した値である。これは一般的な偏差値に置き換えている^{*3}ので以降「偏差値」と表現する。この相対値を用いるのは上述したスマート・ウェブ・サーベイに関する説明のなかで述べたように、回答者ごとの多様な反応傾向を標準化して回答結果を読み取るためである。ただし、あまりにも少ない度数の値を標準化することは適切な処理とはいえない。そのため各設問における履修数が9件以上であることを条件に偏差値を算定した。よって、スケール値は当然、全有効回収について存在するが、偏差値はその部分になる。

主たる結果表出は2種類である。ひとつは有効回収された「個々の学生の」スケール値および偏差値を用いた結果で、対象とする集団（たとえば全学）における平均値である。本稿ではこの指標を「個別回答ベース」と表現する。もうひとつは「各授業科目ごとの」スケール値と偏差値の平均値である。たとえば、「 Semester科目」のように科目群別の平均像をみる際にはその群に該当する科目を履修した個々の学生の回答値を単位にしてその平均値を捉えると、そこに含まれる科目の履修人数の幅の大きさが平均像に影響してしまう。そのため、あらかじめ科目群に属する授業ごとの平均値をおさえ、それを用いて群の代表値を捉えた。本稿ではこの指標を「授業ベース」と表現する。

当然のことだが、どの授業でも履修生全員が回答するとは限らないし、回答しても上述のごとく無効回収になる場合もある。また、授業ごとのスケール値の平均値は当該授業の有効回収数が2以上であった場合に算定した。偏差値の場合は有効回収のなかで上述のように一学生につき履修数が9件以上であった場合に算定し、個別回答ベースの偏差値算定数が各授業において2件以上であった場合に授業ベースの偏差値の代表値を算定した。したがって、アンケートそのものは集中開講以外の全授業に対して実施したが、集計

と分析の対象はその部分になる。アンケート対象となった授業全体に占めたそれらの割合は次項に示す。なお、本稿では小数点下の丸めは四捨五入で処理した。

また、結果表出には基本的にグラフを用いたが、どのような結果であれグラフには高低差が表現される。そこには常にいわばサイコロ遊びの結果があらわれる。つまり、見た目の差異には全体に種々の偶然による変動が加味される。だから、差異があるということや、それが何ポイント上回った、下回っているといった評価をただグラフ形状の相違から語ることは、調査結果で賭け事をするようなことになる。本稿ではそのサンプル性を低減させるために、形状における見た目の差異は傾向的な差異や差異を読み取るためのひとつの手かがりにしか用いない。差異の存在を語る場合は、基本的には統計検定でCohenのdによる効果量をみて、その相対的な大きさが他に比して非連続的に大きくなる部分をあきらかな差異、それに続く連続的な差異をもつ一定の効果量（その都度示すが概ね $d > 0.3$ ）を示した場合を「差異あり」と判定し、そのうえで年度をあらためてもその差異に一貫性が認められたとき、その差異性を特性とみなしうると判定し、結果を述べた。

比較する差異は、ある対象群内（たとえば、「外国語科目群」）での評価項目間の回答差異である。対象群間（たとえば、「外国語科目群」と「教職科目群」）の比較は比較対の組合せが甚大になる一方で、ことからの性質上、多角的な比較手続きを踏む必要があることから、その作業と必要性の釣り合いが現実的にとれないため、特定の理由をもつ一部の例外を除いては本稿では扱わなかった。

結果と考察

回答回収結果

上項の調査対象に記した各年度各学期の授業数に、それぞれの履修者数を乗じた値の総計に対する回収数とその割合は、2015年度前期19826、回収率95.8%、同後期17563、回収率89.5%、2016年度前期20283、回収率96.0%、同後期17393、回収率89.9%であった。この4期をつうじた平均回収率は92.8%。集中授業を除く全開講科目対象のアンケートの回収率としてはかなり高い値であったといえよう。むろんこれは学修成果閲覧と対になった教学における授業と学生間の双方向コミュニケーションというアー

キテクチュアゆえにもたらされた結果である。

双方向のコミュニケーション・ルートを開いていても実際のコミュニケーションが片方になることは当然あり得る。この場合でいえば、アンケートにはアクセスしたものの、回答はせずにページ送りにしてそのままアンケートを終了したというケースである。それを除いた回収数に占めた有効回収率をみると、2015年度前期91.1%、同後期82.4%、16年度前期91.7%、同後期85.0%であった。毎回収したなかの5%程度はアンケートを素通りして学修成果にアクセスしていることがわかる。ただこの件でわかることはそれだけではない。この仕組みは本人認証を介しているから、実際に誰がそうしているかもわかる。だから、その事実が学修行動の一端として履歴に残りうること、そしてその種の情報価値は、人物の行動特性評価にとっては学業成績以上に高いものであることも、これは学生自身が推察的に了解しておく必要があるところである。だが、そのイメージーションに一定数及んでいない状況がつかめる。

授業履修数は必然的に1年生が最も多く、学年が進むにつれて減少する。したがって、アンケート回答内の学年比率もそれに対応した構成になる。有効回収に占めた各学年の比率は2015年度前期1年生48.9%、2年生30.8%、3年生16.5%、4年生3.8%、同順で後期46.8%、32.9%、18.6%、1.7%、16年度前期46.3%、29.3%、19.6%、4.8%、後期44.2%、32.0%、20.1%、3.6%であった。念を押せばこれは各学年の回収率ではない。授業履修数に律動した有効回収に占めた各学年の授業数の割合である。したがって、学期にかかわらず授業アンケートの全体をみる場合はおよそ半数に1年生の反応が映し出され、7～8割方は1、2年生の回答が反映していることになる。ちなみに2015年度の各期学年の年間履修登録数（教職関連科目を除く）の平均値は1年生50.0単位、2年生42.6単位、3年生29.7単位、4年生15.1単位であった。4年生は卒業研究を履修しているため履修科目数に比して単位数は多くなっている（学科によっては単位数12という卒業研究もある）。

ところで、授業と学生間の双方向コミュニケーションのアーキテクチュアとはいえ、回収率からすると、およそ1割の学生は自分の成績を確認していないのか、という疑問が生じるだろう。すでに述べたように、成績確認はalaginというweb上での学修成果チェックシステムを介しておこなわれている。その一方で、それとは別に成績証明書は独立の証明書発行機をつ

じてお茶の水女子大学の現況では無料で取得できるようになっている。その発行で成績や取得単位数の確認自体は事足りる。また、4年生の場合、後期成績は他の学生よりも早い時期に卒業認定とともに書面で配布している。そのためalaginにアクセスしない場合があり、おそらくその影響もあって後期の授業アンケートは前期に比較してやや回収率が低下している。

すでに十分な回収率を得ているのだが、さらにこのシステム・アーキテクチュアを活かし、一層の改善を図るとすれば、上記の点を改めることだろう。教員の成績評価入力の手続き自体は授業期間終了後の試験期間を経て間もなくである。だから、学務システム上は現況より1ヶ月ほど早く成績開示ができる仕組みになっている。それができていないのは一部の教員の成績評価入力が期日どおりになされない傾向があることと、書面をつうじて成績評価事務のやり取りを長く続けた時代の習慣が尾を引いていて業務フローの見直しがなされていないことであろう。このあたりは比較的最近に設立された大学であると学生への成績開示が試験期間終了後、当然のごとく時を隔てずなされているようである*4。だから実際は十分に改善余地があるといえよう。

手続きの項の末尾で記したように、集計・分析の指標は2種類であった。そのうち個別回答ベースの指標数は履修生数が1の授業についての結果も含むかたちで有効回収数に等しく、2015年度前期18046件、同後期14451件、16年度前期18620件、同後期14785件であった。授業ベースの指標はスケール値については、履修生数が2名以上の授業について抽出した。それが全体に占めた割合は15年度2190科目中の79.5%（1740科目）、16年度2233科目中79.0%（1763科目）、偏差値については2015年度66.3%（1451科目）、2016年度64.8%（1447科目）であった。年度によらず授業ベースのスケール値は集中開講科目を除く全科目のおよそ8割、同偏差値は2/3が対象になっている*5。以下の記述ではこれらのことわり書きを逐一記すことは煩瑣になるので、便宜的に各々これらをもって「全科目」と表現する。

結果の全学的な鳥瞰 — スケール値での全科目平均値と標準偏差による評価、年度毎、年度間差異

はじめに全設問の回答尺度（リニア・ループリックの設問のことを指す。授業外学修時間数を尋ねた設問以外の7設問）への直接の反応値であるスケール値の結果を、個別回答ベースの平均値と標準偏差の幅を

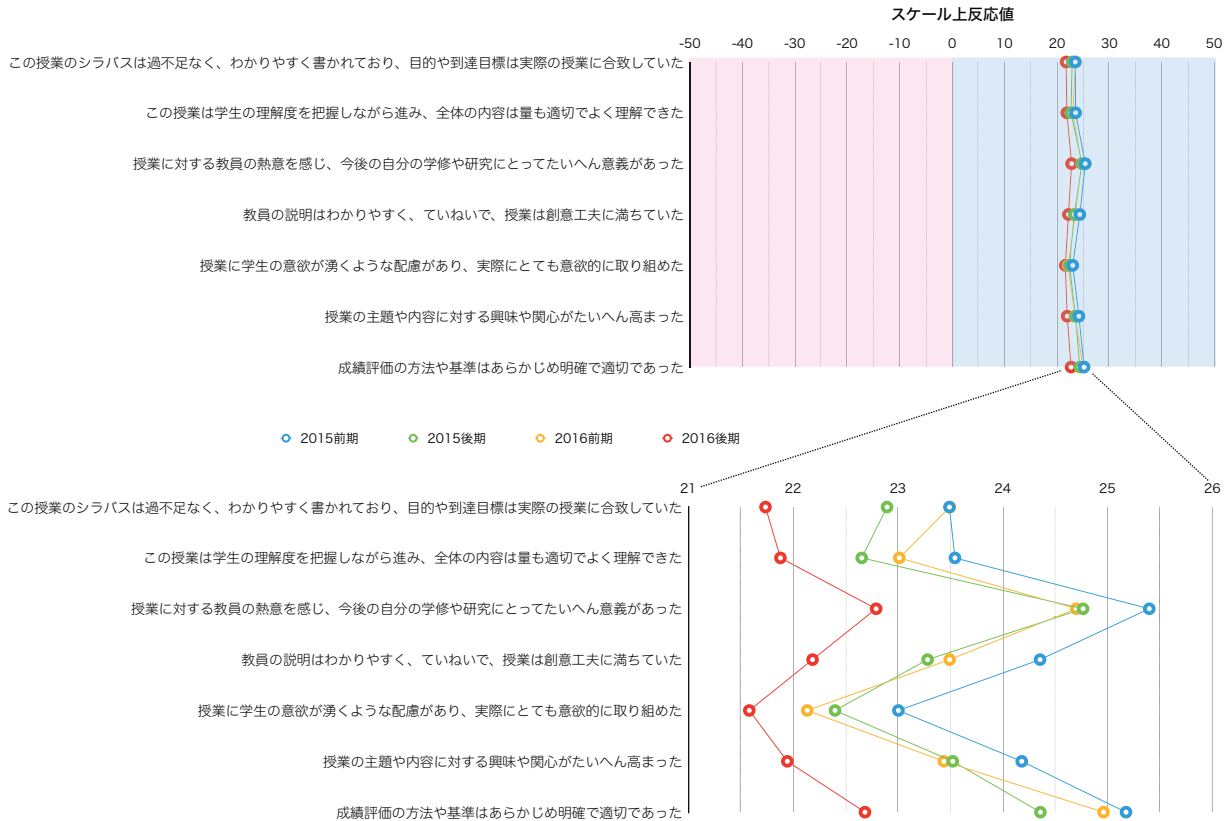


Figure 3 スケール値でのアンケート対象全科目の2015、16年度各学期の7設問に対する回答平均値、上図は実スケール上での位置、下図はその部分拡大

表現した Figure 3 で確認する。これにより履修生数が1名であった授業も含めた全体的な回答に関する正負の傾性を捉えることができる。

グラフからすぐにわかるように、学期・年度にわたる変動は少なく設問間の回答差異も4学期通算で平均値の最大と最小の設問間の開きは101のスケール幅において3.8。わずかであった。よって授業に対する評価の傾性を7設問それぞれの平均値をさらに平均した値をもって概括して語れよう。その値は4期全体平均で23.3、レンジ21.6～25.4であった。グラフ上の背景領域の色分けとグラフの布置からもあきらかなように、これは肯定域の中位にあたる。

Figure 1 に示したように、尺度上、背景色で負の領域を明示している0以下を、成績でいうところの不合格にあたるとみなし、0.1を合格最低点、50を事実上の満点として算定すると、スケール値23.3は78.6点になる。つまり、これは100点満点で表現した場合の値である。レターグレードでいえばB、f-GPAの算定でグレードポイントは2.36になる。ちなみに2016年度後期における全学士課程の平均GPAは2.58であった。学生に評定している成績と学生からの授業に対する評価はほぼ同様で、やや学生か

らの評価の方が辛いというところである。

標準偏差も設問間、学期間で大きな差異はなかった。平均20.2、レンジ20.6～19.7であった。よってこの4期間、7設問各々についての標準偏差幅を平均値にプラス・マイナスして得られる3.1～43.5(100点満点では53.1～93.5点)という範囲が大方の学生の授業に対する評価回答がなされた域とみることができる。このように、その全域が肯定域にあったことから、まずは全学的な観点での授業に対するアンケート評価は、2015および16年度について概ね良好であったと判定できる。

設問間・学期間・年度間差異

Figure 3 上図にあらわされた各年度/学期の7設問に対する回答結果を一瞥すれば、これは一般にこの種のアンケート調査の回答方法にある5～7段階の回答結果を例にすれば、「どちらでもない」と「まったくそのとおり」のあいだにある「そのとおり」に相当し、設問間、年度/学期間に表出しているわずかな差異は回答誤差の範囲と裁定しかねないところである。だが、この微細な差異で表出した各結果のグラフを Figure 3 下図のように拡大してみると、ここに誤

差とはいいがたい一貫した回答差異と体系的な変動を看取できる。その差異は101段階スケール上で0.5～2.5の範囲の違いではあるが、観点を異にした7設問のほとんどにおいて年度間、学期間において同様の相違をとって全体に変動している。よって、このパターン変動には統計的な有意差もうかがえる。実際、分析を進めると少なくとも以下のことがあきらかになった。

1) 7設問のうち、2年間4学期をつうじて一貫して最も高く評価されている項目があった。「授業に対する教員の熱意を感じ、今後の自分の学修や研究にとってたいへん意義があった」または「成績評価の方法や基準はあらかじめ明確で適切であった」の2項である。前者は認知的不協和の理論でいうところの履修者自身の履修選択に対する正当性を認めたいという傾性が反応に寄与したことも考えられる。それがあつたとすれば、この設問の前半にある「教員の熱意を感じ」が後者に引きずられた割り増し評価になっている可能性がある。だから、熱意の高さについて他の設問に比して相対的に常に高く評価されているとみるには留保がある。同時に、この曖昧性を除去することは今後の授業アンケートの設問設定で対応すべきところだろう。

後者の「成績評価の方法や基準はあらかじめ明確で適切であった」が相対的に高く評価されている点は主としてシラバスにおけるその記載と授業をつうじての伝達の両者にもとづくと思われる。この点が安定的に高く評価されていることの確認はこのアンケートにとっても好都合であった。授業アンケートは学生への回答負担をできるだけ抑えるために必要最低限の設問数に絞っている。このように一定の期間をつうじて安定して高評価を得てきた設問については、これを今後も続けて尋ねることは控え、別に必要性の高い設問に代えることができる設問とみることができる。

反対に2年、4学期間をつうじて相対的に常に低い評価であった設問も明白となった。「授業に学生の意欲が湧くような配慮があり、実際にとても意欲的に取り組めた」である。むろん念を押せば、この設問に対する評価値自体は4学期をつうじて100点満点でいえば71.5～73点であり他の設問と同様、良好であった。ただ、あえて他の設問と比較してみるならば、相対的に低かったということである。また、この結果から意欲喚起や実際に意欲的に取り組めるようにすることは授業運営上も、相対的にややむずかしい課題になっていることがわかる。しかし、このように課題もみえてくれば、その解決に向けた方策も明確になる。

たとえば、この全般的に認められた傾向に反し、この「意欲喚起」の設問においてとりわけ高く評価されている授業を探し、その授業運営を参考にするようなFD企画を組むことなどはそのひとつだろう。

2) この微細な差異のなかでもうひとつ両年度共に認められた差異は学期間差異であった。どちらの年度でも7設問すべてについて前期のほうが後期よりも評価が高かった。これは2年間にわたり比較14ケース(7設問×2)でいずれも同様であった。したがって、いずれの差異もわずかとはいえ、検定をまつまでもなく確率的には偶然にもたらされた結果からはほど遠い差異が認められたといえる。この結果の解釈はとりあえずつぎのように推測できる。新年度の前期は常に新入生を迎え、教員、学生ともにリフレッシュされた面持ちで授業に臨むことになる。これに対して後期は長期休み後とはいえ、前期を経ていることとアンケート時には少なくとも一年間の経験が累積されている。そのため評価は辛めになるのだろう。もっともこれはいくことではなく、年間の授業実施の配置はほとんど固定的であるから、前期よりも後期を構成している授業のほうが、まとめれば相対的に負の方向になったというコンテンツ構成の差異が反映されただけのことかもしれない。

3) この全学鳥瞰の結果にもうひとつ見いだされた体系的な差異があつた。年度間差異である。すなわち2016年度の各学期の評価は15年度の同学期に比して7項目すべてについて、つまり14比較対すべてについて評価が下回ったのである。この変動パターンからするとこれを反応誤差、偶然変動に帰することには無理が大きい。原因として回答者である学生のアンケートに対する慣れやある種の疲労感が影響したと考えることもできる。だが、このアンケートの回答者構成は既述のように自然な授業履修数の相違からおおよそ半数が1年生であった。つまりどちらの年度においても半数は調査について同条件のナイーブな対象者であった。だから前年度に引き続き調査に回答することに嫌気がさして、それが評価低下を導いたとはいえない。もっとも残り半数は2015年度ではじめてこの方法でのアンケートをおこない、次年度で引き続き経験したため、その種の影響が生じた可能性はある。

この点を検証するため、両年度の結果を常にアンケートに初経験となる1年生とそうでない2年生以上に分けて、あらためてこの結果を捉えた。その結果がFigure 4である。この両グラフから以上にあらわれた学期間、学年間差異の原因がほぼはっきりした。

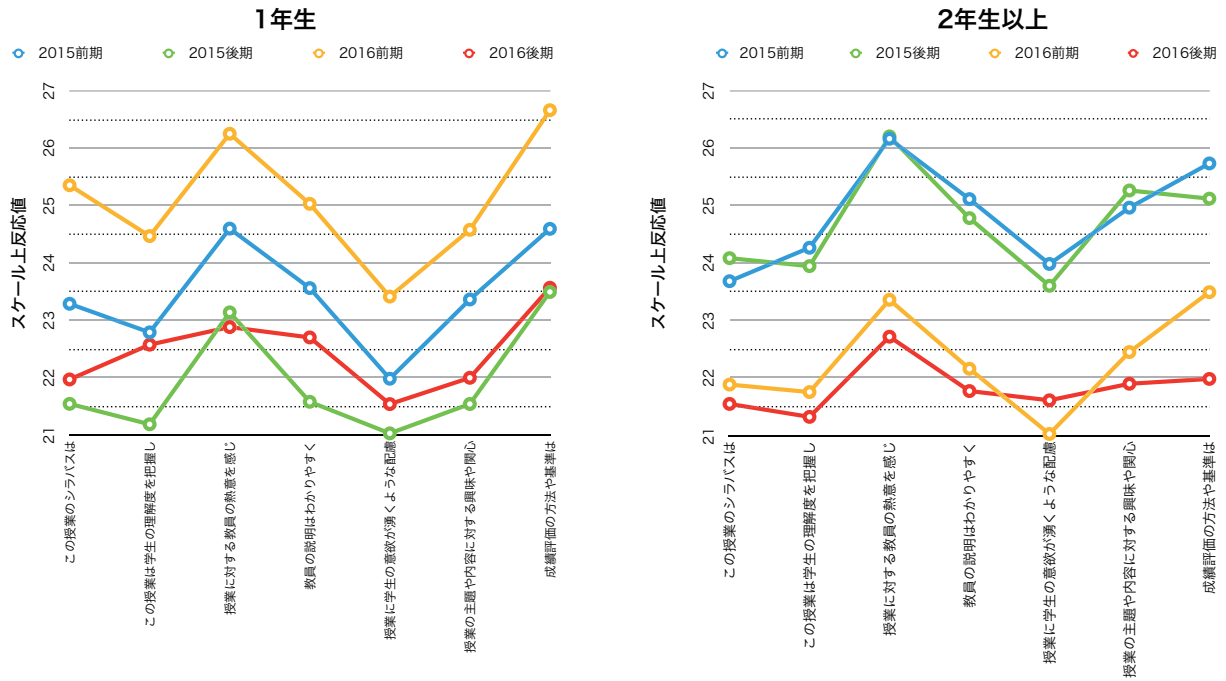


Figure 4 スケール値でのアンケート対象全科目の2015、16年度各学期の7設問に対する1年生の履修授業と2年生以上のそれを分けてみた回答平均値〈設問は表記上の便宜から短縮〉

まず、あらためて2)でみた学期間差異だが、これが明確に再現されたのは1年生であった。2年生以上では学期間差異は明白に縮小し、設問項目によってはほとんど差異が認められないことがわかった。このように1年生は両年度とも前期に比して後期で評価が低下した。2年生以上にはそれがなかった。理由はほぼあきらかである。アンケートにナイーブなのは1年前期のみ、後期からは「またあのアンケート」ということである。回答への抵抗感、モチベーション低下がナイーブな前期のときに比して評価を全体的に低下させたとみることができる。これらから全体をみたときに両年度ともに認められた学期間評価格差は主として1年生のアンケートに対する疲労感に起因したもので、授業に対する評価そのものとは切り離してみることができそうである。

また、Figure 4から、年度間の評価差異についても同質の原因を推測できる。まず1年生をみると15年度より16年度の方が評価が低下しているということはなく、反対に高まっている(14比較対(7設問×2年間)のうち12項で16年度の評価の方が数値そのものは高い)。他方、2年生以上は14比較対すべてについて、スケール値でおおよそ2ポイント程度ではあるが、16年度の評価が低下している。これらからこれもあきらかに同じアンケートの反復実施による回答モチベーションの低下が評価劣化を引導したとみる

ことができる。

アンケートをとる以上、対象とするすべての授業、すべての履修者からとりたい。なにもしなくても能動的に回答してくる履修者はアンケートをする側からすればありがたい。だが、それだけでとどめてしまえばかなりバイアスのかかった結果をみてしまう。マジョリティはどのようにみているのかを知りたい。半数ではもたたりない。残りの半数は果たしてどうみているのか。残りの3割は、。こうして8~9割方を回収できる仕組みをつくる。それは事実上の強制回答法になる。だから、回答する側に抵抗感や疲労感が生じるのは自然なことである。よって、このアンケートは授業から学生への一方的な成果評価だけでなく、学修における双方向コミュニケーションとして学生から授業に対して反応する(事実上、評価する)機会を開いているものと位置づける。それはおそらく正当なありようだから、「強制回答」はそうではなくて「当然の権利としての回答」と解釈、表現し直せる。そうすれば忌避感もだいたい消えるはずである。だが、それでも抵抗感が残るだろう。何であれ正直なところ回答作業は億劫なことである。よってこの種のをしようとするかぎり、この点は根本的には解決不能だろう。その不可能な残滓のところがFigure 4には微妙なカタチで、しかしはっきりとあらわれたということである。

このように比較的安穩に結論できるのは、アンケート

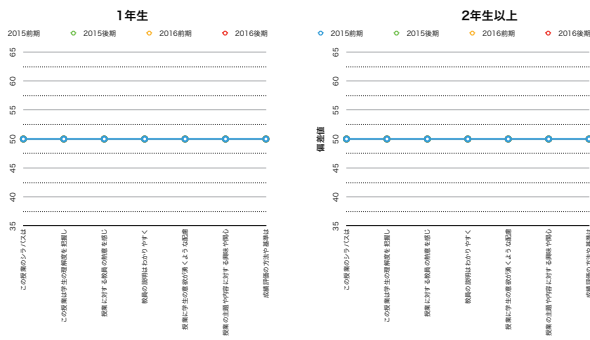


Figure 5 偏差値でのアンケート対象全科目の2015、16年度各学期の7設問に対する1年生の履修授業と2年生以上のそれを分けてみた回答平均値（設問は表記上の便宜から短縮）。2つの年度4学期分の値はどれも50になるためグラフは重なり合う。

ト反復による回答モチベーションの低下が評価劣化を導いているとはいえ、そのスケール上の劣化の程度は101段階スケールにおいて2ポイント前後、つまりこれが5～9の段階評定であったなら差異としてみいだせない程度の微細な差異であり、ただ高感度に反応を採取したゆえにみえたことであったからである。また、そのわずかな差異のなかでもすでにみたような設問間評価差異はどちらの学年のどのグラフでも同様に一貫性をもって表出した。このことから、回答モチベーションの低下とはいえ、反応感度を鈍らせるような質的变化をもたらしたのではなく、ほとんど回答者当人も意識化しないような潜在水準において回答にあらわれた結果とみることもできるからである。

とはいえ、それは解釈上の話であり、実際、事実データはこのような結果であったから、回答尺度に直接反応されたスケール値そのものは、このような全体的な傾向を読み取るうえでは有効だが、個々の授業における結果をみるときはもちろん、以下におこなうような授業科目の群間比較をおこなう際には、この値を用いてゆくと、回答に対するモチベーションの違いやそもそもの回答者自身の反応傾性の特性が値に影響を及ぼし、質的な相違を見誤る危険が高まることになる。そこですでに述べたようにnigalaでは個々の回答者の回答について設問ごとに回答を標準化（ということは回答者ごとに履修科目を相対比較した結果になる）し、平均50、標準偏差10の一般的な偏差値に置き換えて結果をとらえている。以下に示す結果はこの偏差値に依拠したものである*6。

念のためFigure 5はFigure 4に示したスケール値を偏差値に置き換えて同じグラフに表現し直した結果である。グラフには2015年度前期のグラフがどの設問についても偏差値50となって表出しているが、こ

こには他の3学期分も表出している。4つのグラフが重なり合っているため見えなくなっているだけである。個々の学生が履修した授業の平均値を個々の設問についてとれば偏差値は常に50になる。それを何人の履修者について集積してもむろん平均は50になるというわけである。これによりスケール値上にあらわれた個々の学生やその集団の回答モチベーションの相違などによる反応傾性の相違は消え、授業科目間に生じた相対比較をして質的な差異をみていくことができる。

科目群間の差異

お茶の水女子大学における学士課程カリキュラムの科目群は課程前半に全学必修の配当がある「外国語」「情報（関連科目）」「スポーツ健康」、課程全期にわたり履修可能な「文理融合リベラルアーツ（以下LAと略記）」「全学共通（科目）」、主として課程後半に履修配当のある3学部の「専門（科目）」、それに「教職（関連科目）」である。この9分類ごとにそれぞれに属する授業科目の設問毎の平均偏差値を、その分類内で平均した値を2015、16年度についてFigure 6に表示した。

評価項目間で評定の振れ幅が比較的大きな群とそうでない群がある。これは一般にその群を構成している科目数に依存する。その構成数が多いほど相互相殺して群内での項目間の分散は小さくなる。例外は「情報関連科目」で科目数は50以下だが、項目間の差異は2015年度では認められなかった。ただ、その全体の評価が科目群中最も低く、すべての項目にわたり平均偏差値が50を下回った。2016年度でも後者については変化がなかった。なかでも「教員の熱意」が他3項に比して統計的な差異をもって最も低く評価された。前年度この項は他項に比して有意な差異が認められなかったが、同傾向にはあったので課題の所在は比較的あきらかといえそうである。情報関連科目が必修指定であることにも低評価の原因の一端があるかもしれない。だが、同じ条件にある「スポーツ健康」はほぼ「情報」と正反対の高評価を得ている。だから、単に必修であることに低評価の原因を帰すことはできない。授業の内容、体制を含む総合的な対応が要請されている。この結果のみに依拠したわけではないが、「情報」科目群は2018年度に体制の見直しをおこなった。その効果のあらわれが期待される場所である。

これと対照的な結果を示した「スポーツ健康」は科目構成数がこの科目群のなかでは最も少なかった。お

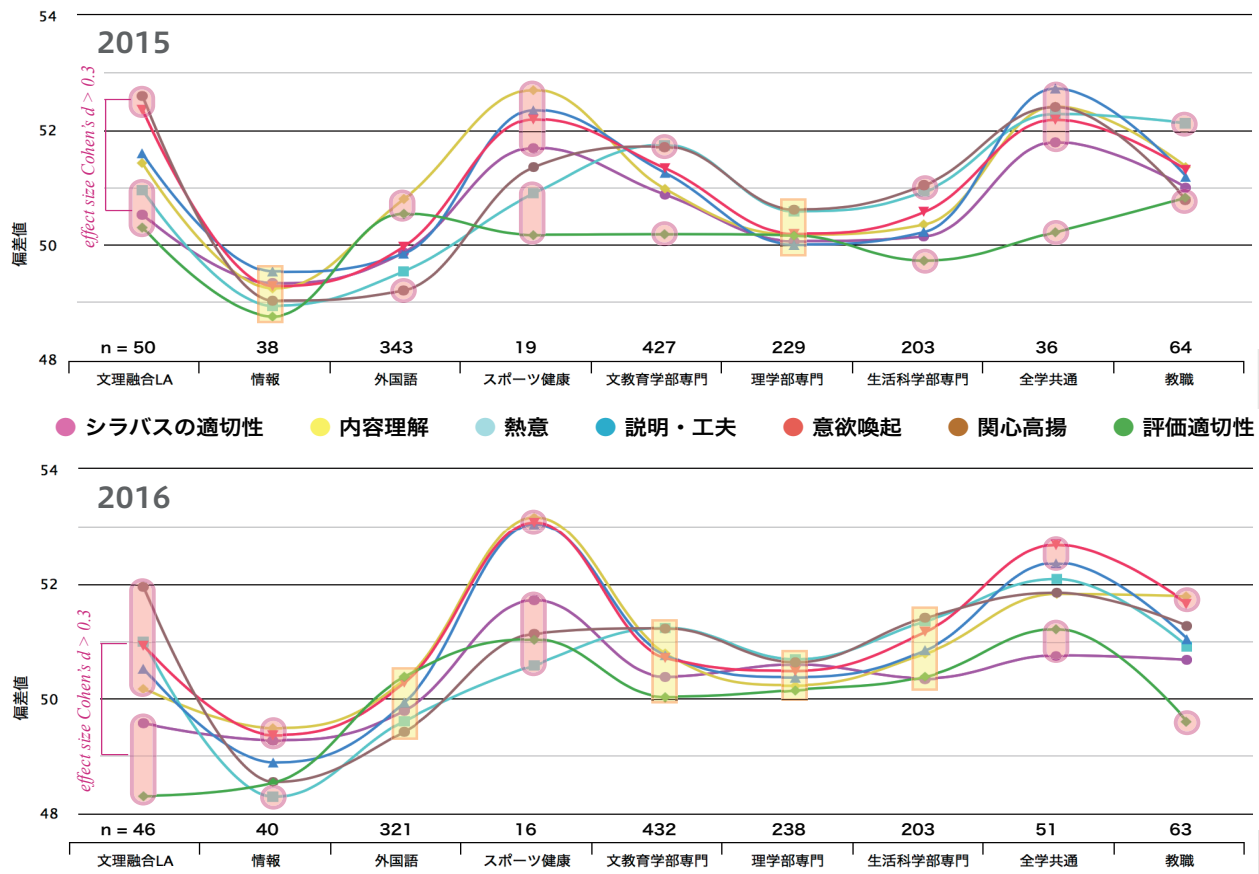


Figure 6 年度別、科目分類別の授業アンケート7項目の偏差値平均値の比較

そらくそのこともあって、内容一貫性が比較的とれて高い評価を得たといえそうである。授業内容そのものが「わかりやすい」という点では他の科目群とは性質そのものの違いもある。そのことが直截に評価されているともいえる。一時期、お茶の水女子大学に限らず「スポーツ健康」が大学の授業に存在することの意義が問われたこともあった。だが、それはエビデンスに基づかないある種の思い込みに発した論議であったとも振り返ることができそうである。この結果をみれば少なくともこの大学では学生からあきらかに高く評価されている授業に対して云々する余地はない。であるとすれば、もうひとつの思い込みは情報時代にふさわしく情報科目は必修科目、という考え方にもあてはまるかもしれない。

同じ必修のしほりがある「外国語」は2年間をつうじて情報科目群に次いで評価がすぐれなかった。しかし、この群は構成科目数が300を超える大所帯である。だから、そのすべてがこの平均像と似た評価になっているとは考えがたい。ここでは立ち入らないが個々の科目に降りて当該の科目群内での詳しい検討が望まれる。底上げ対応で全体の評価が高まる可能性が

あるためである。個別の評価項目では2015年度に評価の方法とその明示、および内容理解が関心よりもあきらかに高い評価を得ている。翌年度はその差異の効果量が落ちたものの同様の評価傾向の持続は認められた。「内容がよく理解できた」という回答は一般的には望ましい回答と受け止められるが、全体的な回答傾向が負の方向にある場合は、「内容が簡単すぎる」という回答を多分に含んでいると解釈をする必要もでてくる。

別に実施している同大学の学修行動比較調査の結果からは学修成果に関する自己評価で英語を含む外国語に関する読む、書く、聞く、話すの諸能力は入学時をピークに、大学での修学期間をつうじて劣化しているという評価が実施年度に抛らず表明されている。しかも、それらは他の能力の成長感と比してあきらかに否定的な評価としてあらわれている。これらの結果を総合すると、とくに英語については学生の要求に十分応え切れていない様子がうかがえる。もしそうであるとすればそれは学生の問題ではなく、授業実施そのものの課題になる。一般に大学における外国語あるいは英語の能力に関する学生の自己評価は入学時に比して劣

化していくことと、個人差が大きくなる傾向がある(半田,2016)。だが、なかには例外も認められている*7のでそれらの事例を参照にするなど改善の余地は十分にある。

「評価方法の適切性とその明示」は他の科目群では7設問のなかでは概ね最も低く評価されるが、外国語科目ではそれが反転しており、はっきりと高く評価されている。この点は大きな特徴である。プレースメントテストの結果で能力別クラス編成がなされ、そのもとで授業が実施されていることと、学修成果が明快なスキル評価になっている側面が強いので、そのことが反映された結果かもしれない。このあたりは養いの能力について同様の特性をもつ情報科目分野で参考できるところだろう。

学士課程前半での履修が同様に多いLA(リベラルアーツ)科目は2015年度はスポーツ健康と同様、すべての設問項において相対的に高い評価を得、とくに関心昂揚と意欲喚起についてシラバス記述の適切性や評価の適切性と明示に比してあきらかに高い評価を得ていた。この評価項に対する反応特性は翌年度も維持された。ただ、後2項については2016年度は偏差値50を下回り、他項に比して明確に低い評価を得た。

2015年度の評価結果にもとづき、お茶の水女子大学でのリベラルアーツ科目は全学のカリキュラムのなかでも学生から相対的に高い支持を得ているという自負を(同科目部会等で)担当教員間で共有し「なんらかの改善というよりも現状の推進を」という方向性を確認した。だが、そのことがすぐに災いしたとは思えぬものの、その矢先の同年度評価では学生から「いやいや、シラバス記述と成績評価法の適切性についてはNOです」という駄目出しを喰らったことになる。そもそも両年度とも7項目間の分散幅は大きい、項目間の評価構成の性質は安定している。したがって、この評価パターンに偶然による変動の影響は少ないとみてよい。いずれにせよ評価法の明示も含めて捉えれば、シラバス記述にもとづく授業実施がLA科目では強く求められている、ということと同科目部会で共有し、改善対象として認識する必要があるといえよう。今後の経過観察を要するところである。

LAという分類には入らないが同様の履修形態をとっている全学共通科目群もやはりLAと同様の2項が「意欲喚起」や「説明工夫」とのあいだであきらかな差異ないし差異の兆候を示した。ただし、この科目群は両年とも全学のカリキュラムのなかで相対的に最も高い評価を得ている。

成績評価方法の適切性や明示が全般的に課題になっていることはすでに述べたが、他項のいずれかとのあいだであきらかな差異をもって低く評価されている科目群には教職およびその関連科目もある。この科目群も全体に相対的には高く評価されているが、同項については両年度とも内容理解、意欲喚起、教員の熱意などに比してあきらかに低く評価されている。果たして教職科目の授業のなかでは「成績評価の適切性」についてどのように教授されているのか興味深いところである。

3学部それぞれの専門科目群に対する評価は構成科目数において他の2学部の約2倍を抱える文教育学部の専門科目群での設問項目分散が必然的にやや大きくなっている。ただ、全体に項目間の評価の振幅は狭く、概ね良好な評価を保っている。3学部共通した評価の質的特性として「教員の熱意」と「授業内容への関心昂揚」が相対的に高く評価された。2015年度では文教育学部と生活科学部においてこれらは評価方法の適切性や明示に比してあきらかに高く評価された。専門科目を主として履修する高学年次にそれぞれの専攻において教員との距離が一層近くなることを反映した評価と推察できるが、それが良好に評価されていることが確認できる。

クラスサイズ間の評価差異

つぎに授業の履修人数を50名以上、30～49名、20～29名、10～19名、4～9名、1～3名に分類し、それぞれにおける評価を比較した(Figure 7)。お茶の水女子大学では50名を超えればそれが「多人数授業」とカテゴライズできる。それが大学のひとつの特徴でもある。実際、以上のクラス群それぞれの構成授業数もFigure 7に示したとおりである。2015年度の場合、最多のクラスサイズは20～29名で448授業、つぎが10～19名の329授業、つまり10～29名という統廃合対象の小中学校クラス並みの授業が全授業の半数を占めている。翌年度も最大数と次点のクラスサイズはこの2群が入れ替わっただけであった。したがって、必然的に学生の授業に対する評価の観点は、この主たる授業の履修人数で運営される授業の雰囲気や感覚を基準になされることになる。

そのことを念頭において各群を7評価項の平均と比較すると、クラスサイズについてはサイズが小さくなるほど評価がほぼ高くなっていることがわかる。最小の1～3名の場合、それまでの評価上昇率は維持されていないが、これはここでのクラスサイズの分類

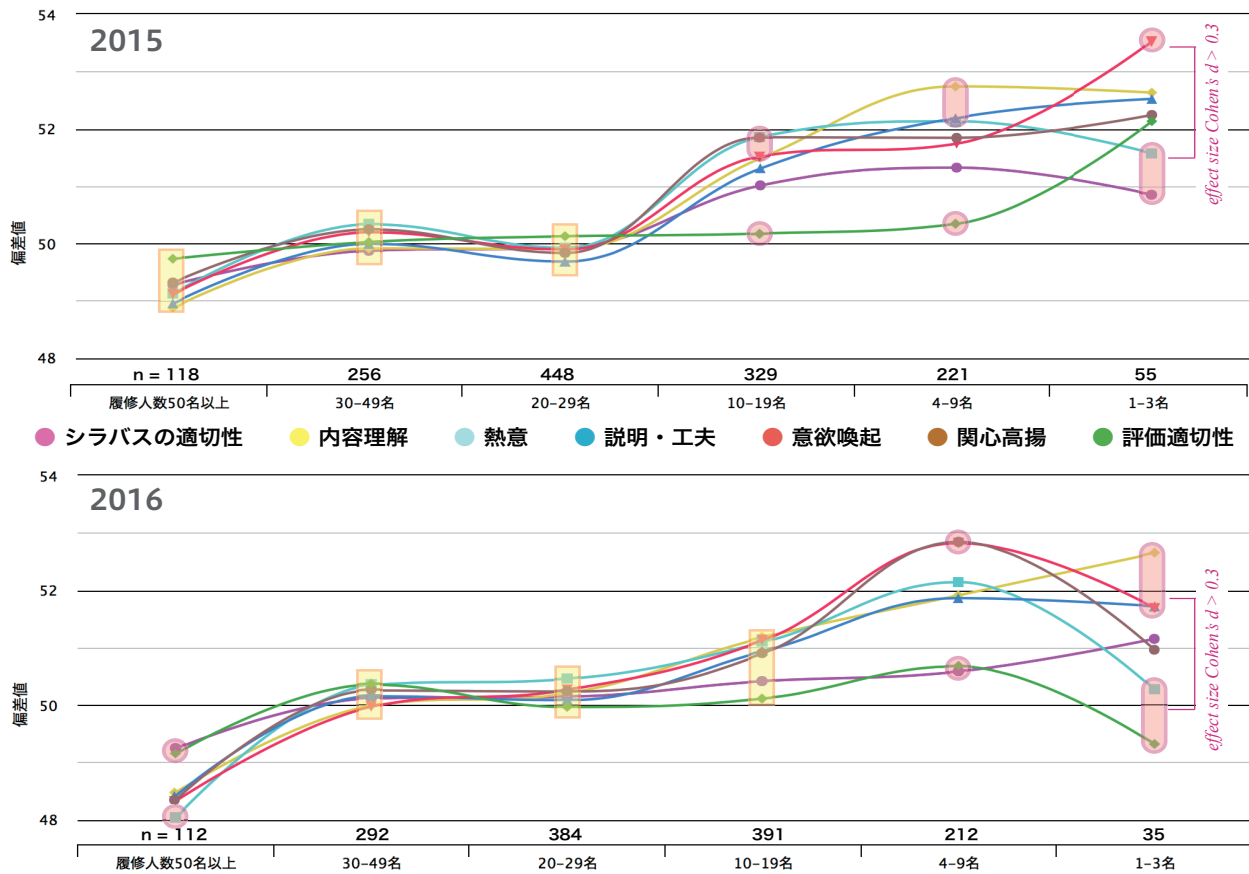


Figure 7 年度別、履修者人数分類別の授業アンケート 7 項目の偏差値平均値の比較

自体が一様になっていなかったことにもよる。また、評価項間のばらつきはクラスサイズが小さいほど大きくなり、とくに 19 名以下になるとその度合いが増す。これはクラスサイズが小さくなるほど群の構成授業数が少なくなることに加え、回答する学生数も少なくなるから、個々の学生の回答特性が色濃く反映してくることによる自然な結果とみることができる。

そうしたなかでも相対的に評価の高い小クラスサイズの授業において一貫して高く評価された項目は内容理解と意欲喚起、あるいは関心高揚であった。大半のケースでこれらは「評価方法の適切性や明示」とのあいだであきらかな差異を示した。また、年度一貫してあきらかな差異を示した項目として興味深いのはクラスサイズ 1～3 名の場合、「教員の熱意」が「意欲喚起」に比してあきらかに低く評価されている点である。ほとんど個人教授的になるこのサイズでは、たとえば発話の仕方他他の授業とは変化してくる。また、授業運営の自由裁量幅が増すため、なかには開講時間が教員と学生の都合によって柔軟に変更されたり、同様のことが授業内容にも起こる。そのため一般の授業とは異なる性質が強まり、結果的にはこの評価項の全文「授

業に対する教員の熱意を感じ、今後の自分の学修や研究にとってたいへん意義があった」か、と問われれば相対的には必ずしも高い評価として振り返ることはむずかしくなるという面があったのかもしれない。

もうひとつクラスサイズの違いによる評価差で特徴的な結果は最大のクラスサイズ群での評価が最も低く、しかも両年度ともにすべての評価項で偏差値 50 を下回ったことである。2016 年度評価ではとりわけ「教員の熱意」という現場特性が「シラバス」や「評価方法の適切性」といったいわゆる授業の外形的な事前評価に比してあきらかに低く評価されている。常識的には履修人数が多い授業というのは人気のある授業で、大学にとっては欠かせない代表的な授業のように受け止められる。だが、学生から振り返られた結果的な授業評価をみると、少なくともお茶の水女子大学においてはどの授業をなくせば、端的に授業全体の学生の満足度を上昇させることができるか、といった過激な外科的処置をすれば、それは履修生が 50 名以上の授業を切除することという結果になっている。むろん、そのような馬鹿げた荒唐治が必要な状況にこの大学の授業運営が陥っていないことは全体の結果を

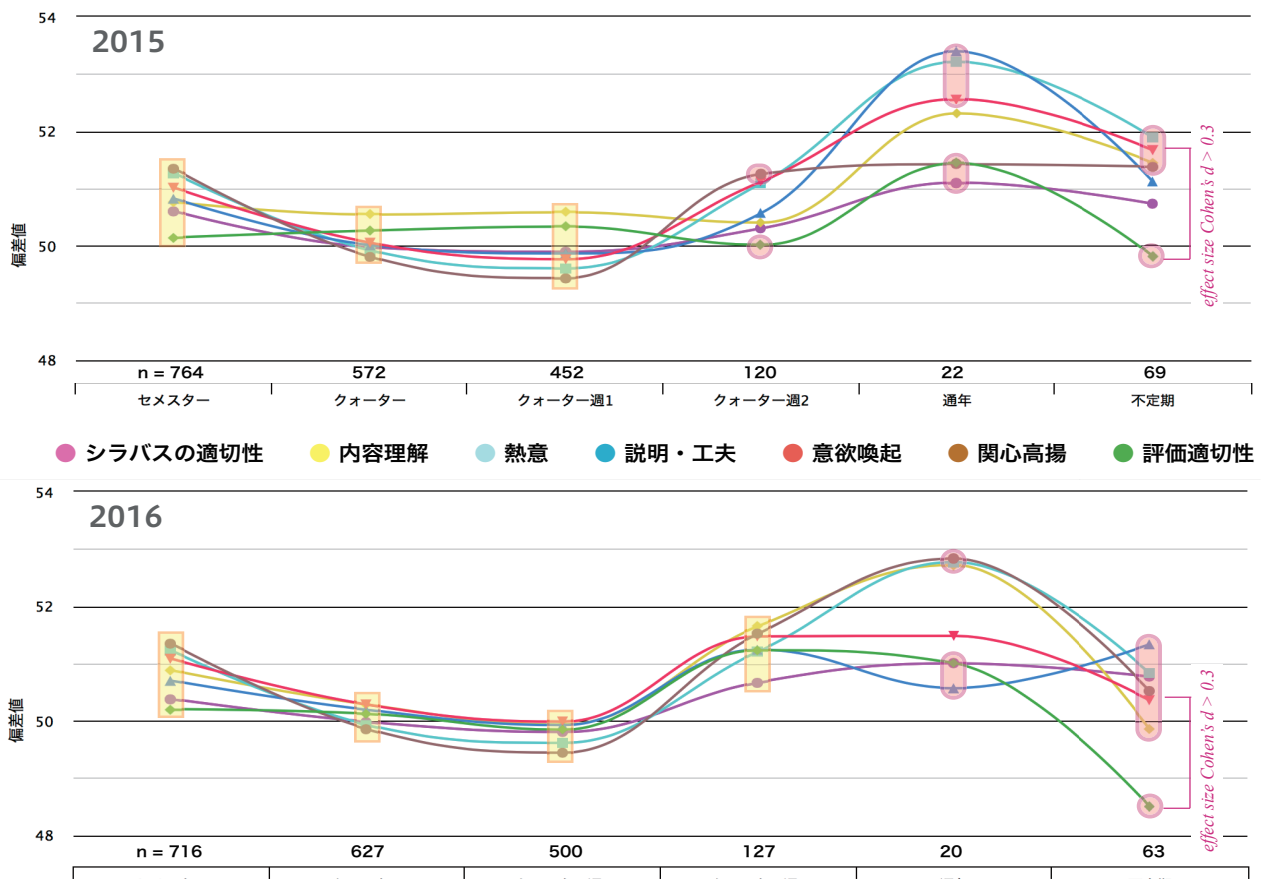


Figure 8 年度別、開講期間分類別の授業アンケート 7 項目の偏差値平均値の比較

みればあきらかである。だが、ときに「履修生の少ない授業を実施していく意義はあるか」といった調子の言明がなされることはある。そうした際には、「いやいやこの大学でそれをいうなら、真逆で、履修生の「多い」授業をしていく意義はあるのか、ですよ」ということになる。少なくとも事実データは前者よりも後者の問いかけに正当性があることを伝えている。

開講学期間の差異

お茶の水女子大学では 2014 年度に 4 学期制を取り入れ、クォータ科目と従前の 2 学期制セメスター科目が併存するハイブリッド式での学期運用がなされている。これに全科目数の 1% 程度だが卒業論文/研究などの通年開講科目と 4~5% 程度の不定期開講科目があり、そのほか集中開講科目が存在している。クォータ科目には週 1 回開講と週 2 回開講の 2 タイプがある。前者は多くの場合、年 2 回内容が異なる連続性をもった授業で構成されている。したがってこの場合、授業実施そのものはセメスター科目と同様である。ただ、ある年次に前半を履修し、年度を変えて後半を履修することを可能にしている点でセメスター科目と異なっている。

4 学期制の導入は大きく 2 つの目的で実施された。第一は、セメスター制を分割した履修を可能にすることで海外の多くの大学の学事歴にあわせ、留学の出入りを柔軟にすることであった。週 1 回開講のクォータ科目とは主としてその対応を図る目的で設定された。その授業運営はセメスター科目とほとんど変わらない。そのため、教員にとっては 4 学期制への移行障壁が少なく、学事歴の変更という施策の導入と推進の上でも効果があった。その効果のほどは 2016 年度で全科目のおよそ 1/3 が週 1 回開講のクォータ科目になっていることからあきらかである。ただし、その眼目のひとつであった柔軟な履修を可能にすることで留学、とりわ海外大学のサマープログラムを中心とした短期留学を促進することについては少なくとも開始から数年の限りでは必ずしも目に見える効果を見ていない。その詳細はここでの目的とは異なるのでこの施策の背景にあったグローバル人材育成支援事業の報告書等を参照されたい。

4 学期制導入の目的のもうひとつは、多様な学修スタイルを可能にすることでそれぞれの授業内容にあわせた柔軟な授業運営を可能にし、学修効果を高めることであった。そのなかには授業時間外の学修量を多く

して単位制度の実質化を図るといふ狙いも含まれていた。たとえば、週1回開講のクォータ科目は仕組みとしては連続した内容のセメスター分の授業を異なる学期（たとえば、1学期と4学期）で実施してそのあいの間に単位制度の実質化にそくした授業外学修を明示的に課すといったあらたなタイプの授業運営も可能にしている。また、一般にクォータ科目は週2回実施による集中学習による学びの効果を得ようとする狙いがある。その一方で、週1回開講のクォータ科目には授業実施のあいだに一定期間を設けることで分散学習による学びの定着を図るといふ効果を企図することもできる。こうした4学期制の導入による学習効果の測定は当然、第一には学修成果の高揚から測る必要がある。しかし、たとえば外国語科目において外部の検定試験などをとおしてある程度客観的にスキルの発達を比較できる場合は別として、一般には大学の諸授業において端的な学修成果の比較検証はむずかしい。ただ授業アンケートをつうじて学習者自身の授業に対する振り返りによって、理解の促進や関心の高揚等の自己評価を指標に、間接的に授業運営の相違による効果の比較検証はできる。

4学期制導入から2、3年目という時期にあつた

当アンケートの結果はまさにその検証機会に相当した。その結果は Figure 8 に示したとおりであった。図中横軸群の「クォータ」はその右隣の2件、クォータ週1回開講と2回開講を足し合わせた結果である。両年度つうじてセメスターと事実上それと同様のクォータ週1には項目間に評価差異は認められなかった。クォータ週2回については2015年度で関心、意欲喚起、教員の熱意が評価適切性よりもあきらかに高く評価された。これらで項目間の差異が目立たなくなるのはサンプル数が大きくなって集計により回答が平準化する傾向にあつたためである。したがって、クォータ週2ですこし項目間差異が出がちになることも他の2者に比してサンプル数が1/5以下と少ないことによるところがある。それよりも注目すべきは両年度ともクォータ科目がセメスター科目に比べてとくに高い評価を得ている事実は認められなかったことである。さらに、クォータ週1の場合は授業実施上、セメスターと同様であるにもかかわらず、やや低い評価を得た。

手続きの項でも述べたように、この比較分析では原則的に横軸の科目群間での項目平均値の比較検証はおこなわない。だが、ここは4学期制の効果検証とい

Table 1 セメスター科目群と2種のクォータ科目群間の各設問における回答偏差値平均に関する検定結果

	left	right	Zscore left	Zscore right	ES Cohen's d	ES Hedges' g	Welch's test
2015	シラバスの適切性	セメスター vs. クォーター週2	50.598	50.296	0.072	0.072	n.s.
	内容理解	セメスター vs. クォーター週2	50.762	50.402	0.075	0.075	n.s.
	教員の熱意	セメスター vs. クォーター週2	51.261	51.088	0.037	0.037	n.s.
	説明・工夫	セメスター vs. クォーター週2	50.816	50.574	0.050	0.050	n.s.
	意欲喚起	セメスター vs. クォーター週2	51.012	51.123	0.023	0.023	n.s.
	関心高揚	セメスター vs. クォーター週2	51.342	51.247	0.022	0.022	n.s.
	評価適切性	セメスター vs. クォーター週2	50.137	50.007	0.030	0.030	n.s.
	シラバスの適切性	セメスター vs. クォーター週1	50.598	49.885	0.180	0.180	n.s.
	内容理解	セメスター vs. クォーター週1	50.762	50.579	0.039	0.039	n.s.
	教員の熱意	セメスター vs. クォーター週1	51.261	49.604	0.371	0.371	p < .001 significant
	説明・工夫	セメスター vs. クォーター週1	50.816	49.858	0.201	0.201	p < .001 significant
	意欲喚起	セメスター vs. クォーター週1	51.012	49.763	0.268	0.268	p < .001 significant
	関心高揚	セメスター vs. クォーター週1	51.342	49.433	0.438	0.437	p < .001 significant
	評価適切性	セメスター vs. クォーター週1	50.137	50.333	0.048	0.048	n.s.
2016	シラバスの適切性	セメスタ vs. クォーター週2	50.376	50.668	0.071	0.071	n.s.
	内容理解	セメスタ vs. クォーター週2	50.884	51.649	0.163	0.163	n.s.
	教員の熱意	セメスタ vs. クォーター週2	51.222	51.210	0.003	0.003	n.s.
	説明・工夫	セメスタ vs. クォーター週2	50.700	51.238	0.115	0.115	n.s.
	意欲喚起	セメスタ vs. クォーター週2	51.077	51.473	0.082	0.081	n.s.
	関心高揚	セメスタ vs. クォーター週2	51.338	51.515	0.039	0.039	n.s.
	評価適切性	セメスタ vs. クォーター週2	50.204	51.230	0.231	0.230	n.s.
	シラバスの適切性	セメスタ vs. クォーター週1	50.376	49.803	0.138	0.138	n.s.
	内容理解	セメスタ vs. クォーター週1	50.884	49.932	0.208	0.208	p < .001 significant
	教員の熱意	セメスタ vs. クォーター週1	51.222	49.607	0.357	0.356	p < .001 significant
	説明・工夫	セメスタ vs. クォーター週1	50.700	49.921	0.164	0.164	n.s.
	意欲喚起	セメスタ vs. クォーター週1	51.077	49.977	0.229	0.228	p < .001 significant
	関心高揚	セメスタ vs. クォーター週1	51.338	49.440	0.408	0.407	p < .001 significant
	評価適切性	セメスタ vs. クォーター週1	50.204	49.843	0.082	0.082	n.s.

う全学的な教育の質保証にかかわる必要性もあることから、例外としてこれら3群の結果について比較検定をおこなった。結果はTable 1に示したとおりで、きわめて明快であった。両年度とも一貫してセメスターとクォータ週2回開講のあいだにはすべての評価項目においてWelchのt検定で有意差なし、Cohenのd、Hedgesのgによる効果量も0.25未満の低い値であった。つまり、当初理論上、期待されていた週2回開講による集中学習等による学修促進効果は少なくとも学生自身からの回答には認められなかった。

他方、授業運営上は同等であるセメスターとクォータ週1回開講のあいだには両年度とも「教員の熱意」「関心高揚」の両項目について効果量0.3以上、かつWelchのt検定で有意差（有意水準 $p < .001$ ）が認められた。両年度ともそれに加えて2項目に有意差が認められたが効果量は小さかった。

事実上、同じ開講スタイルをとったこの両群において2項に明確な差異が生じた原因を考えれば、結果的にクォータ週1回開講というスタイルを選び取った授業は、クォータ科目ではありながら、実際はほぼセメスター科目のままで看板だけ掛け替えたという面がなきにしもあらずであったから、その対応をとった科目においてとくに「教員の熱意」と「授業内容への関心高揚」という点においてセメスター科目より相対的に弱い性格が感じられたのかもしれない。両群に授業運営上、ただひとつ異なる点があるとすれば、クォータ週1回開講の科目ではセメスター科目と比較してクォータ期間末に成績評価のための試験やレポートが課されることである。そのため学生にとっては負荷が高まる。だが、このことがクォータ週1回開講の授業に対する評価を劣化させたとすれば、そのことは「成績評価の方法や基準はあらかじめ明確で適切であった」の評価に差異をもたらしたことだろう。しかし、その項目の結果をみると両年度とも効果量は0.1未満で十分に小さく、いわゆるセメスター科目に中間試験が入ったような科目であったことから授業評価を落としたという気配は認められなかった。

以上から、4学期制の導入は授業アンケートの2年間をみるかぎり、最も注目された週2回開講においてその評価はセメスターと変わらず、逆に週1回開講科目はセメスターより評価を低下させたという結果になり、授業の質を劣化させているということになった。もしこれがその一方で、柔軟な学事歴運用によって目論見どおり、留学数が飛躍的に増大し、国際流動性が高まったといった事実を得ているのなら、課題

は劣化した週1回開講クォータ科目の授業実施努力ということになろう。だが、4学期制導入に伴う留学増はこのアンケート期間においては認められなかった。このかぎりでは、この施策展開は180度の方向転換が改善への道ということになる。

ただ、引き返すことを潔しとしないとすれば、この結果は4学期制を徹底せず、セメスターとの混淆という中途半端な方法をとっていることに難があり、残余する過去の習慣に引っ張られて新しさの良さが引き出されていないのだという冒険的判断をして4学期制の徹底を断行するというのも一案であろう。その際の覚悟は冒険が失敗すれば全学期において授業の質に減退を招くことになる。少なくとも現在の授業アンケートの結果はその招来を予測させるものとなっている。だから、その断行はまさに賭けになる。

セメスター制への帰還をするまえに、週1回開講クォータ科目も半数をセメスターに戻し、その戻した科目に戻さずにそのまま週1回開講クォータ科目を続ける科目とのあいだで授業アンケートの結果を比較検証してみるという実験的施策をとることも一案であろう。週1回開講クォータ科目に対する評価はそれがクォータだからではなく、そもそもそういう評価の科目であったからかもしれない。そのことが示せば、4学期制の導入は未だ効果は明確ではないが今後への期待の余地があることになる。だから、このまましばらく継続し、効果検証をつづけるという判断もできる。

また、この結果は4学期制導入当初のものである。だから、クォータへの転換科目は授業実施に不慣れな点もあり、個々の授業での実施経験に基づく改善効果はこれからあらわれ、そこに理論上語られていた効果は現れ出すという可能性はある。したがって、この結果をもって効果検証とするには早すぎることも確かである。少なくとも数年先までは継続的に運営し同様の検証をつづけることが穏当なところだろう。

最後に4学期制を離れて、通年開講科目と不定期開講科目の結果に触れる。まず、通年科目は項目間に評価差異があるものの、全項で平均以上の評価、「教員の熱意」「内容理解」という2項について、両年度ともいずれの他群よりも高いベスト・プラクティスの評価を得た。ただしこの群の科目数は他群よりも大幅に少ない20件程度であった。だから結果の安定性は低い。その点を抜きにしてこの結果からいえば、授業アンケートが伝えていることは学期の構造を変えて授業の質改善の方途を開くなら、それはセメスターを細

分化するのではなく、実は反対にまとめる方向にこそあるということのようである。エビデンス・ベースの施策とは流行や思い込みではなく、実際のベスト・プラクティスに従うことである。むしろ、通年開講に倣うにあたってはその授業が特殊な一群、たとえば卒業研究・論文のようなものでは端から質が異なってしまう。だが、ここで検証対象になっている通年開講の授業はいずれも文教育学部の専門科目であったが、少なくとも授業題目において同一種のもので並んでいるといった特殊性はなかった。この事実からしてまた、このことが他の学部においても同様にいえるかどうかはわからない。

不定期開講科目は60～70科目あった。「教員の熱意」「内容理解」「意欲喚起」について「評価の適切性と明示」よりもあきらかに高い評価がなされ、この両者には項目間に明確な差異があらわれ通年に近い高評価を得ている。この群には実際のところ前期、後期、通年の違いがある。だが、開講が不定期であることが明示されている場合、それら3期の群とは分離して不定期開講科目として群化している。その開講期内容をあらためてみると不定期開講群の約6割は通年で開講されている。だから、ここでの通年開講群とサンプルの重なりはないのだが、通年開講としての評価特性がここにも反映されている可能性はある。

成績評価の仕方の相違と授業アンケート結果の関係

お茶の水女子大学では成績評価は授業ごとに教員の判断で素点またはS～D、5段階のレターグレードで評定する。素点のレンジと不合格点は授業毎に任意である。これを学務処理においてレターグレードでの評定はS=95、A=85、B=75、C=65、D=0に素点化する。素点評価は合格レンジを60～100に線形変換し、不合格点は0とする。これらのスコアXを用いて、 $GP = (X - 55) / 10$ によってグレード・ポイントを算定している。さらに素点評価はこのように100点満点に変換した値を用いて100～90をS、90未満～80をA、80未満～70をB、70未満～60をC、60未満をDと画定している。ここではそのレターグレード評定の授業毎の割合のパターンの相違と、授業アンケート結果とのあいだに何らかの関係性が見出されるか否かを、2016年度の結果をもとに検討した。

大学の授業には昔からその成績評価をもとに「楽勝科目」とか「鬼畜科目」と呼ばれるような授業があった。今世紀に入ってから大学教育に厳格な成績評価が課題として掲げられたこともあり、さすがにかつて

ほど学生の会話に飛び交うことばではなくなってきた。それでも厳格・厳正に成績を評価したがゆえに、結果的に楽勝になったり、鬼畜になることはありうる。だから、今でも結果的にそれに該当する授業科目がなくなったわけではない。

これらの名称には楽勝科目とは授業内容の良し悪しや魅力、あるいは授業における教員の熱意などとは無関係に、極端には履修して試験さえ受ければその成果とは比較的無関係にかなり良好な成績が得られ、不合格になることなどまずない、というニュアンス、鬼畜科目にはその反対の意味合いが含まれているという通念がある。とすれば、成績が判明する以前に、授業実施のあれこれについて細かく尋ねるアンケートには、楽勝にも鬼にもさほど関係のない評価、あるいはことによると、楽勝科目では授業評価そのものは手厳しく、鬼畜科目では教員の熱意などにおいては高評価を得ているのではないかと、といった推察もできる。

今や日本の大学の約8割が成績指標にGPとGPAを用いるようになった。伴って学生の会話にも日常的にGPAのことばが交わされ、それを気にする声もよく聞くようになった。授業毎にGP範囲やその平均値が学生に示されるようにもなった、そのため教員側もかつてときにみられたような成績評価に対するおおらかな裁定には自己規制がかかり、履修者全般にわたって弁別性の高い評価を出す傾向も一般化してきた。とはいえ、成績評価のパターンは授業のポリシーや目標設定の仕方はもちろん、そうした事前の基準とはほとんど無関係に履修してくるその都度の学生により、まさに生(なま)ものとしての性質が入り込んでくる。したがって、必然的にそれぞれの学期によって成績評価のパターンは変化する。よってどのような年度、学期においても成績をたとえばレターグレードで分画してその分類をとれば、Table 2に示したような授業が見出されることになる。ただ、それぞれの分類に該当

Table 2 成績のレターグレード配分とその授業の数 (2016年度)

レターグレード配分	授業数	履修生数レンジ*
SかAだけの授業	107	4 - 169
SがAより多い授業	129	4 - 50
S～Dの各LGがある授業	307	8 - 191
S・Aがない授業	8	4 - 40

*この算定では履修生数が4以上の授業を対象にしている。

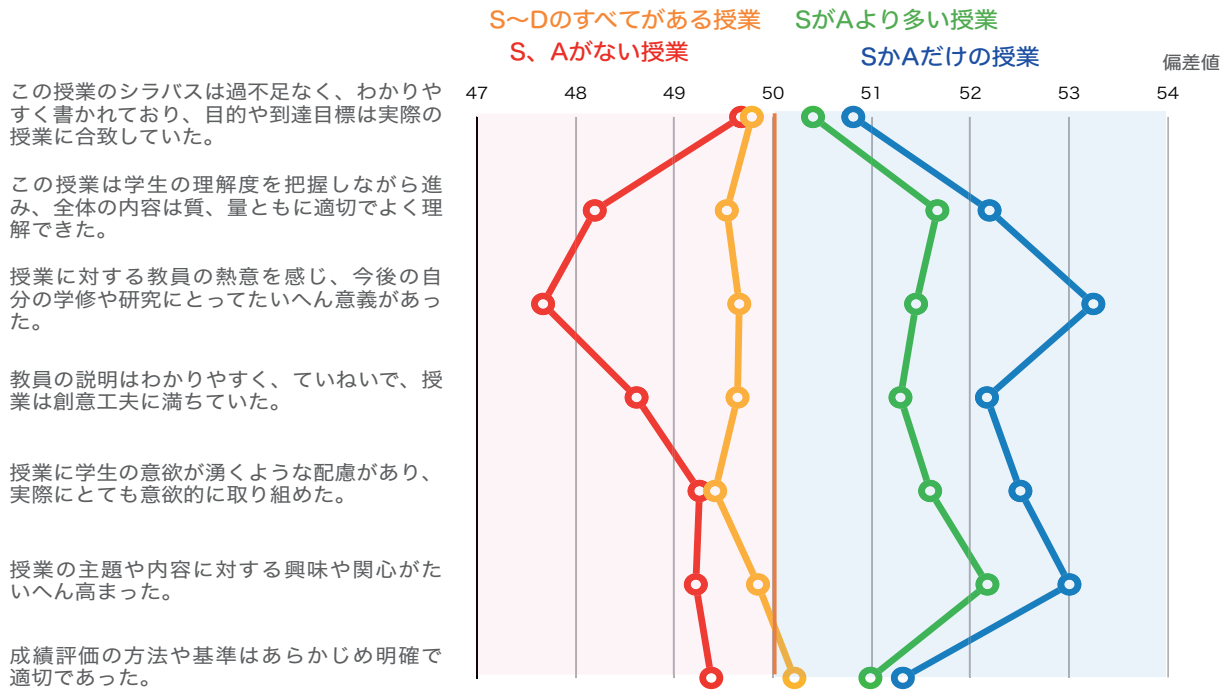


Figure 9 レターグレード成績評定の仕方（一授業内のレターグレードの配分）における4つのパターン別の授業アンケート7項目に対する回答結果偏差値平均値の比較（お茶の水女子大学 2016年度）

する授業や数そのものはその内容を追い求めてもあまり意味はないはずである。確かに見た目には楽勝とか鬼と形容されるような授業があることは確かである。

ここでとりわけ注目したいことは以下にみるような、これらに分類された授業に対する授業アンケート結果の比較である。なお、Table 2は2016年度のお茶の水女子大学における結果だが、これらは全授業のうち偏差値が算定できた授業（全授業のおよそ2/3）で履修生が4名以上であった授業を対象にした数値である。したがって、この表の数値は比較上の相対量として参考にできる値であり、その数量自体を云々するためのものではない。

Figure 9はこれらの評価パターン群の各々について授業アンケートのリニア・ルーブリック評価7項目それぞれの偏差値平均を求め、それを群毎に折れ線グラフにして示した結果である。一目瞭然、レターグレードのパターン群間においてすべての評価項目にわたり相互に重なり交わることなく差別化された結果が表出した。成績にSかAしかない授業の評価が全項目にわたり最も高く、つぎにSがAよりも多い授業（B～Dもありうる）、S～D（不合格）5段階の成績すべてがある授業、SとAがない（B以下だけの）授業の順で授業アンケートのスコアは低下した。後二者は1項目以外が偏差値50以下、つまり全授業の平均値

を下回る評価であった。

これら成績パターン間の差異が7項目すべてに認められたということは授業アンケート総体でみたとき、あきらかに差異のある結果があらわれているといわざるをえない。しかも4つの成績パターンのあいだで最も差異が小さかった設問は「この授業のシラバスは過不足なく、わかりやすく書かれており、目的や到達目標は実際の授業に合致していた」であった。

つぎがとりわけ重要だが「成績評価の方法や基準はあらかじめ明確で適切であった」であった。一方でSとAしかつけていない授業があり、他方それらがひとつもない授業もある。そのいわば両極端を比較しているこの結果において「成績評価の方法や基準」の適切性については相対的に大きな違いがなかった。このことはこの授業アンケートが成績評価を知る前に実施されているから、この設問に対する評価は十分にはなしえず、差異が生じにくかったということ、それと同時に、このアンケートの回答結果は成績評価の結果を織り込んでおらず、授業実施内容そのものにもとづくものになっていることを示している。

反対に最も差異が大きかった設問が「授業に対する教員の熱意を感じ、今後の自分の学修や研究にとってたいへん意義があった。」であった。成績にSかAしかつけない授業は、この設問において7項目中最も高く評価され、偏差値平均で53.2、それと正反対に

SもAもないすべてがB以下の成績を出している授業は、この設問に対する回答が最も低く、偏差値平均で47.7、全設問中、最大の差異を示した。後者は前者に比べて構成サンプル数が8と少なく結果には不安定さが否めない。だが、他の成績パターン群の結果とも併せて、ここに示した結果にはことによると授業の成績評価には授業実施ないし運営上の正否が反映されるのではないかと、といわざるをえないかたちがあらわれている。いわゆる仏の楽勝科目というのは、まさに成績評価が象徴しているように、授業運営そのものも楽勝的？で、その内実も成績同様に概括的、平板なものになっているのではないかと、憶測されがちだろう。また、厳格厳正な成績評価の仕方に気を配り、生真面目にA～Dのすべての成績評価を出している授業〈Table 2にあるようにサンプル数はこのパターンの授業数が最も多い〉は平均的かそれをやや下回る授業アンケート結果になったのである。

鬼のように成績が厳しい授業もわずかに認められたが、それは成績評価においてそうであるだけでなく、その厳しさが必ずしも熱意や意欲喚起としては受け取られておらず、アンケート結果から推察するに授業内容が相対的に学生に伝わっていないがゆえのB以下と考えられるのであった。しかし、それはその意味では必然ともいえる。仮に成績が授業の達成目標の到達具合をあらわすとして、他の授業よりも授業内容が伝わっていない授業の到達度が低く抑えられることに矛盾はない。まさに学生に対してなしているはずの成績評価とは自分の授業実施に対する自己評価になっているという気配を色濃く感じさせるところである。

これらの結果を補強する他の観点からの成績評価パターンの結果も併せて呈示しておく。Figure 10には不合格(D評価)がなかった授業〈n=843〉とあった授業〈n=610〉を、Figure 9に重ね描きした結果

である。

この両者間についても設問7項すべてについて正負同方向での差異が認められた。不合格がある授業はほぼ偏差値50ないしそれを下回る全学平均的な評価であった。これに対し不合格がなかった授業、換言すれば履修生全員がお茶の水女子大学が明示している成績評価基準の表現でいえば、基本的な目標を達成した授業は授業アンケート5項において偏差値平均51程度とやや高評価になった。不合格者がいなかった授業はいうまでもなくおめでたい授業だが、これは繰り返すことになるが結果がおめでたかったからアンケート評価も高かったというわけではなく、時間順序としては授業のあれこれに対する評価が差別的にやや高かった授業は全員がその授業の基本的な目標をクリアし、合格したということなのである。

最後に、より極端な成績評定パターンにも触れておく。レターグレード評定がSかAしかない状態でSのほうがAよりも多かった授業である。お茶の水女子大学の成績評価基準ではS評定にかぎっては履修者の15%以内、履修者数が10人未満の場合は2名以下にするという「目安」が掲げられている。したがって、これは履修生数が3名以下で起きている特例にあたりそうだが、この分析では履修生数4名以上のケースを扱っているもので、いずれも目安を超えてそのようにならざるを得なかった授業ということになる。当然ながらそうした授業は数も少ない(n=14、ちなみに履修生数レンジは4～41)。そのため結果は不安定だが、Figure 11に示したように、Figure 9で示した授業アンケートの結果が最も高評価であった「SかAだけの授業」をさらに上回る高評価を示した。

先の成績評価基準によれば、A評定とは「基本的な目標を十分に達成している」でS評定ではこれに加えて「かつきわめて優秀な成果をおさめている」であ

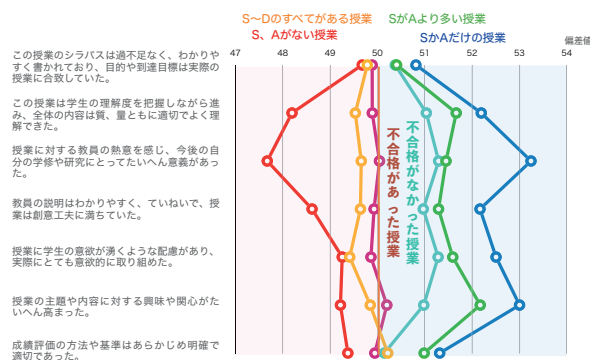


Figure 10 不合格がある授業とない授業を分けて Figure 9の偏差値グラフに重ねて表示した。

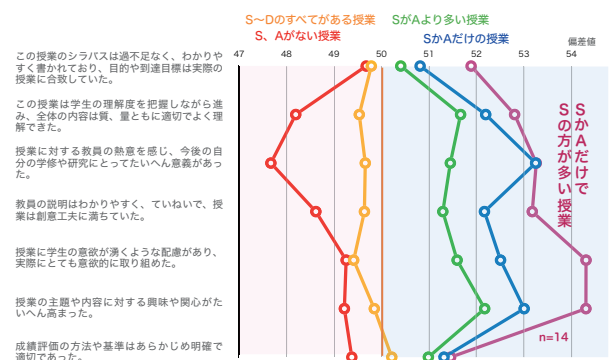


Figure 11 レターグレード評定にSかAしかなく、かつSの方が多かった授業を Figure9の偏差値グラフに重ねて表示した。

る。履修生全員が基本的な目標を「十分に」達成できているような授業がたいへんよろしいことはいままでもない。そのよろしきは学生からの諸観点での授業アンケートでも同様に写しだされており、しかもそのなかできわめて優秀な成果をおさめた学生がマジョリティになるような授業が「意欲喚起」や「授業主題や内容に対する関心高揚」において偏差値平均で54を上回るという特段の評価を得ている。このことはそれ自体、目安を超えても出さざるを得なかった学修成果であったことを推察させる。

ここではつきりとあらわれた結果は授業アンケートでの評価が履修生全体につけている成績評価の仕方と明白な正の相関関係にあるということである。当然これは因果関係ではないから、おしなべて全員に高い成績をつければ授業アンケートの評価も高くなるということではない。評価自体の時間関係からもそれはいえない。しかし、そこに因果ではなくあきらかな相関が見出されたということがもつ意味は大きい。

ひとつには授業の成果が全般に思わしくなかった場合、まちまちであった場合、総じてよかった場合、並外れてよかった場合、といったそれぞれが学生の側から授業を振り返り評価した際のその高低に素直に写し返されるというあまりにも無矛盾で自然な結果が、この授業アンケートそのものの回答妥当性を示しているという点で意義深い。

もうひとつはよりだいたいな再確認的な発見だが、今世紀に入って大学界にかまびすしく反響した「厳格な成績評価」ということが、もとより成績評価の営みをひとえに厳しく、辛くするという語りではなく、決しておろそかにあつてはならない厳格厳正さという点において、とりわけ教員の評価行為そのもの以上に、制度における厳格厳正さのそれに焦点があつたことをあらためて確認させる結果であつた点で大きな意味がある。

お茶の水女子大学ではその「制度の」厳格・厳正さという点に焦点を合わせ、授業それぞれの裁量下での素点成績をそのままグレードポイントに直列させる functional GPA を成績評価の基本指標にする制度改革を十年ほど前におこなつた。それにより教員が厳格厳正におこなう成果評価を制度において丸め損ねてしまうようなこと（多くの GPA 制で生じていること）がないよう教育の質保証を実現した。当結果はそれが果たされたうえでの発見と確認であつた。

くだんの厳格な成績評価の履行の話題のなかでは、すべての授業において評定するレターグレードの割合

をあらかじめ規定するといったほとんど無理な機械的封建主義のようなモデルが呈示されたこともあつた。その余波として現在でも成績評価の分布の妥当性をめぐって、レターグレードのおおのがそれぞれに然るべき比率をもってつけられていることが正統であるかの観点が示されたりする。だが、今回のこの結果はそうした机上の理想観念モデルは学生からの授業への評価を低下させ、結果的には教育の質保証を損ねる方向に向かわせるものであることを暗示している。あらためて成績の評価基準に掲げた文言を基準に考えてどのような授業を実施することが学生にとっても教員にとっても、ひいては大学全体にとっても望ましい結果を得ることになるのか、根底のところから再考を促すうえで参考になるデータを手にできたといえるだろう。

注

- 1) スマート・ウェブ・サーベイ以前のお茶の水女子大学の授業アンケートは質問紙配布と回収による匿名調査で実施されていた。その結果としてときに自由記述には誹謗中傷にあたる記述もあつたようだが、その経験を経たためか、たとえば nigala に転換したばかりの時点では、本人認証の手続きを経ているにもかかわらず、自由記述欄に「バナナ」といった回答をわざわざしてきた例もあつた。それに対して特に対応したわけではないが、そうした遊びはしないほうが賢明、ということをお学んでもらうためにも、ウェブを介した調査には一種の電子パノプティコンとしての特性があることを学生に伝えていくことは同時代的に生きる方途を確認したり実践的に学ぶ機会となる。したがって、結果からみえてくるいくつかの個人的な行動特性に関する結果のフィードバックを学生にしていくこともこの今の時代の「教育機関内における」調査の役割としてたいせつなことだろう。
- 2) 大学院での授業アンケートが nigala システムで実施されるようになるのは 2018 年度からである。この報告の範囲（2015-16 年度）においては大学院では従前の質問紙配布回収による方法で実施されていた。そのため大学院生が学士課程科目を履修した場合は自主的にこのアンケートにアクセスしておこなう必要があつた。このことからそもそもその種の履修は稀なケースであつたが回答も稀となつた。
- 3) ここでいう標準化とは各値から平均値を引き、標準偏差で除した Z スコアのことである。偏差値はその値を 10 倍して 50 を加えた値である。よって平均 50、標準偏差 10 になる。

- 4) 高知工科大学の例がそのようである(2016年高知県公立大学法人総務部長福田直史氏との談話)。
- 5) 本文で述べたように、偏差値はその標準化算定における適切性の観点から、一学生において履修科目数が9以上あった場合に求めている。したがって偏差値が求められた科目数は減少した。アンケートそのものは9割の回収率を得ても偏差値が求められた科目数が3~4割減少することはスケール値で得られる回答傾向との齟齬を免れ得ず、結果読み取りにあたり、スケール値と偏差値のどちらを優先すべきか迷う状況が生じることもわかってきた。それを踏まえて2018年度からは偏差値算定はおこなわないことにした。
- 6) 上の注に記したように2018年度からは偏差値の算定を過去に遡及して取りやめた。しかし、ここで述べたことは有効回収全数のなかでの各授業科目の各設問についての評価相対値によって代替し充たされている。
- 7) 教学比較IR コモンズが実施しているALCS学修行動比較調査では2017年度の調査などで英語科目に関する複数のスキルの成長感について、ある大学の学生が他大学に比して統計的な有意差をもって高い自己評価

をなしていることが見いだされている。ただし、同大のこの調査での回収率は比較他大学に比して20~40%低い状態であったから、その後の調査での再検討を要する留保つきの結果である。

参考文献

- 半田智久 2015「サード・リアリティ時代に生きる持続可能な内部質保証基盤：スマート・ウェブ・サーベイによる授業アンケートの概要とその駆動原理：PDCAより自己創出 (autopoiesis) へ」高等教育と学生支援,6,75-93.
- 半田智久 2016「教学 IR: 2013~16年度学修行動調査：お茶の水女子大学の学内分析報告」高等教育と学生支援,7,6-29.
- Melzack,R. 1975 The McGill Pain Questionnaire: major properties and scoring methods. Pain, 1, 275-299.

2018年2月23日 受稿