

Title	第五回高等教育質保証研究会兼・大学教育実践センター「教育実践公開研究会」「学生調査の活用：新入生調査、大学生調査から見えてくる教育成果への影響」
Author(s)	山田, 礼子
Citation	大阪大学大学教育実践センター紀要. 2011, 7, p. 111-127
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/4081">https://hdl.handle.net/11094/4081</a>
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

## 第五回高等教育質保証研究会 兼・大学教育実践センター「教育実践公開研究会」 「学生調査の活用：新入生調査、大学生調査から見えてくる教育成果への影響」

山田 礼子

Utilization of Student Survey: How We Can Analyze the Data of Student Survey

Reiko YAMADA

2011年1月7日(金)、16時より、同志社大学の山田礼子教授を講師に迎え、第5回高等教育質保証研究会(兼・第9回教育実践公開研究会)「学生調査の活用：新入生調査、大学生調査から見えてくる教育成果への影響」を、ステューデントコモンズ1階開放型セミナー室で開催した。

山田先生はカリフォルニア大学ロサンゼルス校(UCLA)の調査をベースとして、日本版の新入生調査、大学生調査を開発し、複数の大学の協力を得て調査研究を行っており、その開発の趣旨と実データを用いた分析結果を中心に説明をされた。

現在、わが国では学士課程の充実という重要な課題に取り組んでいる。すなわち、大学とは単に教育を授ける場として存在するのではなく、学生が真に成長する場となるべきであり、そのためには学業の達成や態度の変容に至るまで教育のアウトカムを測定し、それをマネジメントに活かすことが求められている。山田先生の行う調査が明らかにしようとするアウトカムの測定は、当初は学習プロセスから評価する間接指標による間接評価であったが、今後の展開として、他大学との連携のもとで、これまでの間接指標と大学のIRシステム(Institutional Research System)とを関連づけることによって、間接指標がどのように直接指標とリンクするかが明らかになりつつある。

従来、把握することが困難とされてきた教育のアウトカムに対し、学生調査やIRの機能から可能な限り有用なデータを収集し、大学の教学マネジメントに活かしていくのが現在のトレンドである。ディプロマポリシーの明確化と合わせ、教育の質保証に向けてのデータ収集の重要性が再確認された。

参加者数は16名であり、外部の教職員も多く参加して活発な質疑応答が行われた。ここではセミナーの記録のうち講演部分について掲載する。(齊藤貴浩)

### アウトカム・アセスメントを機能させるには(#1)

私の今日のタイトルは「学生調査の活用：新入生調査、大学生調査から見えてくる教育成果への影響」ということになります。

さて、なぜこのようなことを今日お話しさせていただくかということですが、今現在、世界の高等教育はアウトカム、つまり学習成果(=ラーニング・アウトカム)をしっかりと証明するというのが求められるようになってきています。しかし、そうしたアウトカム・アセスメント、つまり学習成果の評価を機能させるためには、いろいろなクリアしていくべき課題もありますし、いろいろな方法があります。そういうことを、学生調査

を通じてどのように把握していくかを、今日のテーマとして見ていければと思っています。

さて今申し上げましたように、基本的にはこの学習成果ということ置き換えますと、学士課程を充実させるという大きな課題になってきます。つまり1年生から4年生までの学士課程教育を考えた場合に、単に一般教育や専門教育と考えるのではなく、学士課程としての教育、そしてその学士課程の教育の成果として学士という学位を与えるということを考えなければいけません。すなわち、各大学が自らの教育理念と学生の成長を実現する学習の場として学士課程教育を充実させていくということが求められています。

しかしそうした場合に、学習成果というものが最終的なアウトカムであるとしても、学習成果に至るまでに学生が常に成長していくとすれば、どうやってその成果に至る成長を測定するのか、そして学生が成長していく結果としての成果と学士課程教育とにどのような関係があるのかについては、残念ながら日本国内の研究の蓄積は意外に少ないと言えます。

そこで、私どもは研究グループとして、2004年からこの研究を行ってきております。学生の成長に関するそういった理論を構築していくことと、その間接的な評価測定法である継続的な学生調査を開発すること、そして、その中でデータを継続的に収集し、データの中身が安定することによって、学生調査そのものが一般的にどの大学でも使えるようにするという意図で行って来ました。

現在、日本における学生調査研究は新たな段階に入ってきています。先ほど申し上げましたように、アウトカム・アセスメントの要求というのが世界的な傾向でして、欧米やアジアなどでもそうです。最も先鞭を切っているのは成果を測るという文脈では、OECD諸国の高等教育版の能力調査であるAHELOというのがありますし、アメリカなどでも大学がア krediteーションを受ける際には、学習成果のアセスメントについて、どんな方法でもいいからとにかく行い、そしてその結果をエビデンスとして示すということが求められています。アジアなどでも、理工系を中心に、そうしたアセスメントというものが進展していることは間違いありません。

アウトカムをアセスメントする、または評価するにはいろいろな種類があります。例えば学習成果や学業達成のような直接アセスメントです。詳細は後で申し上げますが、現在、日本学術会議などで分野別の標準規定を作り上げていこうということが言われています。そうした動きに関連した直接的なアセスメントとしましては、医学部の卒業に際しての一般的な認定の資格などもあります。あるいはJABEEなどが要求しているようなアセスメントもここに関連してくるかと思えます。もう一つは情緒的側面です。つまり学生を中心とした間接的なアセスメントがあります。

いずれにしても、大学の教育力への期待と要望という政策的な大きな転換がこの背後にあります。大学は従来からありました研究の場であるだけでなく、大学という環境が学生を成長させるための装置であり、したがって、そのためのカレッジ・インパクトというのは一体ど

うあるべきかという理論が重要であるということが再認識されていると言えるかと思えます。

先ほども申し上げましたが、学習成果を測定するには大きく分けると二つあります。

直接評価。これはダイレクト・エビデンスと申しまして、学習成果を直接に評価します。直接評価は私ども教員であれば日々行っていますし、1月になって、そろそろそのような時期に入ってきています。例えば学期末の科目試験やレポートなど、そして卒論や卒業試験などもこの時期に行っていることです。一般的なTOEICやTOEFLなどの標準試験や資格試験などもどちらかといふところの直接評価にあたります。ですから、分野としては一般教育であったり、専門分野別であったり、また資格に関連した分野であるといふところは、直接評価で実際に測定していくことが可能なおところですよ。

直接評価というのは大学の基本となるような評価ですが、直接評価がそれだけで効果的かというクエスチョンマークが付くところです。といいますのは、ダイレクト・エビデンスは確かに学習の成果ですが、成果に至るまでの過程というものが見えてきません。もちろんその成果として、例えばGPAでいうならば4.0とは言いませんが3.8など非常に高いGPAを修めている学生が、「プロセスが見えない」といふのはおかしいと思われるかもしれません。しかし、彼ら彼女らが、一体その成果に至るまでにどれぐらいの学習時間を確保してきたか、大学教育への満足度をどのぐらい持っているか、あるいは価値観はどうであるか、つまり成果に至るまでのプロセスというものは見えてきません。

他方で、間接評価は今申し上げたようなプロセスを見る評価です。私はインダイレクト・エビデンスと呼んでいます。学習行動、生活行動、自己認識、大学の教育プログラムへの満足度など、成果に至るまでの過程であるプロセスを測定することで、非常に効果的な評価となります。一般的に学生調査や卒業生調査で行うわけですが、卒業後15年ぐらいたって企業の中で働いている卒業生を対象に調査を行う場合には、それは間接的な評価と言えるかもしれません。評価を行う時期としては、入学時、1年次終了時、上級学年時、卒業後というように分けられるものであると思います。

そういたしますと、今申し上げた二つの評価方法がありますが、実際にこれらをばらばらで行うことに意味があるのかということが問題になります。私どもが考えているのは、一番有効な使い方として、直接評価と間接評

価を組み合わせるということ。それは後ほど、私どもが行っている戦略的連携支援プログラムに関してご説明する際に、その評価方法についてご紹介したいと思います。

それでは、最初にアメリカで行われている直接評価としての標準試験について説明したいと思います。おそらくこの中には理工系の先生もたくさんおられると思いますが、私がいろいろところで今日の内容を発表した経験から言えば、大概理工系の先生方は直接評価というものを非常に重視される傾向があります。JABEEの基準もありますし、それは当然なのです。そして、はっきりとした成果がなければ分からないということをよくおっしゃってください。実際そのとおりだと思います。そこで、直接的評価による成果測定の代表的な標準試験を作るとはということなのかをお話ししたいと思います。

標準試験は直接的な評価です。例えばアメリカでいえばSATやACTという大学入学に関する一般的な標準試験があります。大学院であればGRE (The Graduate Record Exam) という標準試験があります。それから一般教育の成果を測定する方法、標準試験としまして、CAAP (Collegiate Assessment of Academic Proficiency) やMAAP (Measure of Academic Proficiency and Progress) など、いわゆるマークシート式の選択方式の試験があります。また、TOEFLを作っていることで著名な機関であるETS (Educational Testing Service) は、専門分野別の試験なども作っています。実際に、これらの標準試験は、アメリカの大学でよく使われています。大学院に入学を志望している者は必ずGREを受けなければいけませんし、高校生が大学進学を目指す場合にはSATやACTが必要となってきます。

これらの標準試験を実用化するためには、いくつかの保証すべき性質があります。これが非常に大事です。まず妥当性、つまり測ろうとする性質に対して実際にテストしている内容が適切であるか、妥当であるかを保証する必要があります。そして、信頼性、その得点傾向や回答傾向などを継続的に測定していき、安定した結果が得られることを保証する必要があります。長期間での大規模な検証を行いつつ、理論に基づいてテストを作成することでこれらが成り立ちます。私はテスト理論の専門家ではありませんので、このあたりはテストの専門家しかご説明いただけないと思いますが、古典的なテスト理論や最近ではIRT (項目反応理論) というような理論を

使って、標準試験の作成と検証がなされるようになってきています。こうした経緯を経て、ラーニング・アウトカムの測定方法が開発されるわけです。

例えばGREなどは、実際に60~70年の歴史を持っています。その実施の過程の中で改良されて改良されて使われてきていますように、標準的な試験になるには時間がかかるということです。もう一つの課題としては、例えばアメリカの一般教育の成果を測るテストとしてCAAPやMAAPを挙げましたが、これが日本でも同じような一般教育の成果を図るテストとして成立するかということです。アメリカの一般教育のカリキュラムというのは基本的にどの大学に行っても汎用性があるように作られています。つまり、ある意味ではコアカリキュラムというものが、一般教育において人文系、自然科学系、あるいは社会科学系の中で作られていますので、標準的なテストがしやすいという状況にあります。それでは、日本の大学で教養教育を行う共通教育段階において、コアカリキュラムができていくかということ、まだまだそこまで到達していません。一般教育の成果を測る標準的なテストが、非常に作りにくいという現状があります。

それでは、次に間接評価としての学生調査について申し上げたいと思います。これは先ほど申し上げたように、プロセスを測定するのに非常に適したものだと思います。ただし、実用化に向けて保証すべきところはやはり妥当性と信頼性です。つまり、長期間での大規模な検証が必要であるということ、そして理論に基づいた項目の作成が必要であるということです。

間接評価の問題点をわかりやすくご説明いたしますと、私は教育社会学の研究者なので、その中の領域として高等教育を扱ってきました。そうすると、私どもが質問紙調査を作るときには、比較的自分たちの研究に関心のある項目で作ることになります。そのため、関心が変わっていくとまた違う質問紙ということで、調査項目が長期間継続することがないのです。それからもう一つの問題は、心理学などでは行われていることなのですが、それ以外の学問分野ではなかなか理論に基づいた調査項目を作成しないということもあります。言い換えれば、研究者の問題意識や関心に基づいて、質問項目が頻繁に変えられてきたという状況があります。しかし、それではどの大学もが使える標準的な学生調査は作れません。

そこで、アメリカの私どもが依拠してきたカレッジ・インパクト理論に基づいた質問紙調査のCIRP (Cooperative Institutional Research Program) というプロ

グラムを用いることといたしました。これはUCLAの教授でもあった先生がお作りになったものですが、1966年に初めてCIRPの新入生調査を作られて、今現在2010年ですから、40年以上これを改良しながら使っています。もう一つ、より学生がかかわる、エンゲージするということに焦点を絞った理論があります。これに基づいて作られているのがインディアナ大学を中心としたNSSE (National Survey of Student Engagement) という質問紙調査です。こうしたものを長期的に、そして幅広くいろいろな大学が参加することで、大学生であればこのようなことが測れるというような標準化したデータが取れるようになってきています。

そうした中で、間接評価はいわゆるアスティン先生が言うインプット (I)・エンバイラメント (E)、あるいはアウトプット・アウトカム (O) という学生が成長するモデルです。インプットは前提となる要因ですから、例えば学生の高校時代あるいは家庭背景などです。Eはそれぞれ大学という環境、Oは成果です。成績、就職状況、大学院への進学などいろいろあります。そうしたものを検証する方法に使われます。

#### JCIRP (Japanese Cooperative Institutional Research Program) 開発の状況 (#7)

それではJCIRP開発の状況について説明したいと思います。

今現在、高等教育の質の保証というのが話題になっていますが、それに向けてJCIRPを開発してきました。例えばシラバスやGPA、CAP、FDといったものがあるいろいろな大学で導入され実践されるようになってきました。

その次のステップとして、大学内にはいろいろなデータがあり、大学はデータの宝庫だと思いますがしかし、宝庫であるけれども散在しています。一つの場所にありません。入試のデータだったら入試担当部署にあるし、教育のデータだったら教務担当部署にあります。あるいはものによっては学生の担当部署にもあるなど、いろいろです。それらのデータをある意味で一つの集まりとして使いながら教育改善へと結び付けるIR (Institutional Research) という機能がこれからの大学には必要になってきますし、またそれが充実してくることが必要です。そうした中で、先ほどの学生調査をどう使うかということが問題です。

ですから、質保証に向けての教育改善モデルを考えますと、この図 (#9) のような形で回ると思うのです。標準的の学生調査を実施して開発する、データを収集して分析する、それを教育の自己点検や評価活動に使っていく、その結果をカリキュラムや教員などの実際の教授法などに生かすという教育改善につなげていくというようにPDCAのサイクルを回していきます。これが学生調査を使った質保証に向けての教育改善モデルを、どのように構築するかということになります。

私どものこれまでの研究の特徴について申し上げます。それは、先ほどから何度も申し上げているような、学生の成長を支える高等教育機関の効果の測定の開発をしてきたということです。

もう一つは国際比較ができるという特徴があるのですが、一応アメリカで蓄積されてきたカレッジ・インパクトの研究の流れをベースに、日本的に変えているということがあります。

それでは、どのようなことをこれから測るかということです。後で申し上げますが、今までの調査は、2009年現在で5万4000件のデータを持っています。2010年調査も今現在行っていますので、大学生と短期大学生調査を含めると7万件ぐらいのデータが蓄積されるようになっていきます。

そういう中で今まで分かってきたことは何かというと、やはり学生が成長していく上で大学という環境は非常に重要な位置を占めているということです。

また学年効果ですが、これは学年が上がるにつれて、あるいは学生がいろいろなことに積極的に関与するということが、そして同時に教員が学生に関与していくというようなカレッジ・インパクトが実際に教育効果に意義をもたらしているということも分かりました。

もう一つは、もしもここに心理学系のカウンセラーの先生がおられるとお分かりになることだと思いますが、ネガティブ学生、つまり前向きになれない、言葉が悪いですが積極的になれずにひきこもりがちで、ある意味で退学予備軍とか退学のリスクが非常に高い学生の存在などは、比較的大学間格差が小さいです。ですから、第一志望で入った学生が多い大学は、仮説としてはネガティブ学生は少ないのだろうと思っていたのですが、必ずしもそうではありませんでした。第一志望で入学してきたとしてもそれが精いっぱい頑張ってきた場合、入ってからの学習が難しく付いていくことができ

ないということがあったり、あるいは非常に成績が優秀であったとしても、今度は人間関係を構築しづらい学生も中にはたくさんいるので、そうした学生はやはりネガティブ学生になっていきます。そこには成績や高校時代、あるいは入試というものとは関係なく、あるいは大学の威信度にも関係がないとは言いませんが、そのような差に比して小さい格差であるということがあります。

もう一つは、大学の経験を通じて変化する学生類型が何かと言いますと、基本的にネガティブ学生がポジティブに変わっていくということも確認できました。その反対もちろんありますが、比較的ネガティブからポジティブに変わっていくことの方がよく見られます。

ですから、そういう意味で大学の環境というものは学生に対して非常にポジティブに後押しするような力を持っているということも言えるかと思えます。しかし一方で、いわゆるI（インプット）の部分というのはどうしても否定できません。やはり小学校あるいは家庭の背景、高校までに作り上げられてきた価値観や学習習慣が大きな影響を及ぼしています。しかしそうは言っても、そのIの部分で全部決まってしまうということになってしまうと大学の意味がありません。影響が大きいとしても、こうした規定されてきた行動能力を、大学教育がどのように伸長させるかということが、おそらく現在の大学に求められていることではないかということです。

### 国際比較ができる（#13）

では、国際比較ができるというところを見てみましょう。

ここで使うデータは3種類です。CSS2005はアメリカのUCLAが行っているもので、2005年度のデータです。サンプルは3万人程度で、私どもと比べると大きすぎるのですが、使えないことはありません。2005年の大学生調査、2007年の大学調査です。

この2005年と2007年の参加校の一部は一緒ですが、違うところもたくさんあります。どちらかと言いますと2005年調査は比較的難易度を統一しました。都市型の大学で、難易度が同じとは言いませんが、非常に似た大学を選んでいきます。それから2007年調査はどちらかというとばらばらになっています。

このグラフ（#15）は何を意味しているかという、自己評価なのですが、自分の大学での成績はどの程度かということを開いたものです。中位というところを見て

いただくと、アメリカ、そして日本は2005年度、2007年度と、そんなに変わりはありません。ですから非常に妥当な回答になっていると思うのですが、違う点は上位です。上位と回答した者はアメリカの方が多いです。これはやはりGPAシステムが機能していますし、よい成績を確保して継続しない限り奨学金がもらえないということがありますから、上位と答える学生も多いということになります。

そして、もう一つの違う点は、一番右の欄です。アメリカでは「わからない」という学生がいませんでした。日本では2005年は15.5%、つまり自分の成績が相対的にどれぐらいかというのを把握していないのです。2007年には若干減少しています。このあたり、2005年から2007年で何が原因かと考えたときに、少しはGPAというのが取り入れられる大学が増えてきましたので、それを意識する学生が増えたのかもしれない。しかし、それでも日本の学生は成績に無関心だということです。これは大学だけの問題ではありません。基本的に産業界があまり成績などを重要視しないという慣行がずっと続いてきて、現在は若干変わりつつあるかもしれませんが、そういうものが影響しています。

次の図（#16）は、大学での経験の日米比較です。学際的な授業を履修したか、自主的な学習プロジェクトへの参加や、授業の内容について他の学生と議論したとかというようなことを聞いています。

特徴としましては、日米そして日本の継続データでもインターネットなどの利用度はほとんど変わりませんので、このあたりは日米で全く差はありません。一方で日本の学生のよい点としては、授業内容についての議論の度合いというのは日本の学生の方が高いです。これはどうしてかと考えたときに、やはり日本の学生たちは帰属場所がはっきりしていますから、例えば大阪大学でも法学部に入学してきて語学のクラスで一緒だということがあると思います。ですから、お互いが顔見知りで議論することも結構あります。しかし、アメリカの場合は学生の帰属場所というものがなく、専門に行かないとなかなか一緒に授業を取れませんので、人為的にそのようなクラスを作ることもせざるを得ないので、疎外感があるということもあるかもしれません。また、アメリカの方が課題の提出度は低いのです。これは課題をたくさん課すからということも言えるかもしれませんが、日本の学生の方が課題はきっちり提出するということがありません。

一方で日本の学生の大きな特徴は、学際的な授業や自主的なプロジェクトに参加する率が低いということです。そしてもう一つは、学生は授業につまらなさを感じているということです。これは教員にとっては反省すべき材料かもしれません。あるいは、今の学生にとって授業の中身というものが、私どもが学生にこれだけは身に付けてほしいということと、学生の高校時代までの学習とを考えたときに、合っていない可能性もあります。

この図 (#17) が何を意味しているかということ、授業や実験への出席時間を見たときに、日本の学生は20時間以上という人が非常に多くなっていることです。アメリカは7.7%です。一方で、授業時間以外で勉強するといったところはどうかということ、長時間学習する人たちの割合はあまり変わらないのですが、それ以外の回答で随分短い時間しか学習していないのは日本の方です。このことは、単位の実質化を考えたときに、大きな課題です。1週間で取得する単位が多いにもかかわらず、自主学習の時間が非常に少ないというのが日本の学生です。それは2005年、2007年のデータを比較しても授業時間は短縮していませんし、学習時間も伸びていません。これだけ世界の中で学習成果ということが言われ、試験の成果だけではなく、学生が自ら学習していくということが必要とされているにもかかわらず、日本の学生の自主的な学習時間が短いままであるということが気になるところです。

この図 (#18) がラーニング・アウトカムの自己評価です。一本だけ上にあるグラフがアメリカです。日本の方が低いのは一目瞭然です。ただこれは大きく増えたという比率だけをピックアップしていますので、日本とアメリカの学生の自己評価の位置付けが違っていることも考慮しなければいけません。つまり日本の学生は控えめに自己評価し、アメリカの学生は高めに評価するということです。これは先行研究でもはっきり言われていて、アメリカの中で人種民族的に自己評価の傾向を調べた研究者がいます。そうするとアジア系の学生は非常にモデレートに謙遜して自己評価するのですが、ヨーロッパ系の学生、つまりホワイトアングロサクソンの学生は高めに評価するという結果が出ています。ただし、それで20~30%割り引いたとしても、まだまだ日本の学生による成果に関する自己評価は低くなっています。

とりわけ、グローバルな問題の理解や、文章表現の能力、プレゼンテーションの能力や、数理的な能力という

のは分野にもかかわってくるのですが、異文化の人々と協力する能力や、批判的に考える能力、分析や問題解決能力といったような能力は今現在世界中で必要とされているような汎用的なスキルという考え方からすれば、もう少し評価が上がっていかねばいけない点でもあるかもしれません。

それでは、カレッジ・インパクト理論を検証するためにはということで、今までのデータをお見せしましたが、これまで継続的調査をどれぐらい実施してきたかということをご説明します。

JCSSは大学生調査で、今お見せしたのはこの調査です。大学2年、3年、4年生を対象とするのが適切で、私は比較的3年生が一番適切な対象者になるだろうと思います。2004年、2005年、2007年、2009年に実施しています。JFSは新入生調査で、入学してきて6月ぐらいに行うもので、2008年、2009年と行っています。JJCSSは短期大学生調査で、累積約5万4000人です。この調査は、いわゆる大学間での機関ベンチマークに使えるということになります。

複数の時点での調査を使ってみると、例えばJFSとJCSSを組み合わせると、入学時、そして2年後の3年生に機関内パネル調査として使った場合には、個別学生のパネル調査にもできます。もちろんそこには学籍番号を入れるなどの工夫が必要で、私どもは各学生に暗号化したコードを作り、それでデータを一致させています。それから4年制大学機関と2年制の短期大学機関を比べることもできます。ということで、ある程度理論の部分ではできるようになってきたのですが、まだまだ課題はあります。

プロジェクトの課題は、この一覧 (#20) のようになっています。毎年これらの複数調査の大規模データが集まります。参加大学が異なり、サンプルサイズも年度によってまちまちです。実は大阪大学も参加して下さったことがありまして、800人ぐらいのデータが大阪大学から集まったことがあります。そのデータがどこにあるかは知りませんが、お返ししていますので、お使いになっていただければこのような機関内の分析ができるのではないかと思います。

新入生用調査と上級生用調査では目的が若干違います。そして入れ替わりがある、追加項目が増えるということで、ここにあるような課題が見えます。では、それをどのようにしていかなければいけないかというのが次

の第2ステージになります。

まず、さまざまな検証方法を開発して、多様な手法を通じてカレッジ・インパクトを検証することが必要です。これはアメリカのCIRPに見られるような理論サポートの安定性の要因です。アメリカではCIRPに関して、あるいはNSSEに関して、それらを用いた研究発表の蓄積が豊富です。しかし、日本では、このような分野の研究者が少ないこと、そして、例えば心理学と高等教育あるいは教育社会学というような領域の研究者が交流して、共同研究をすることはそれほど多くないという事情があることから、残念ながらアメリカに比べて理論や実証研究の蓄積がまだまだ少ないというのが現実です。アメリカの場合はそこが非常に面白くて、なぜこういう理論がしっかりと作られてきたかというときに、一つは社会学系の領域の先生方と心理学系の先生方が一緒になって高等教育研究を構築してきたということがあります。それでいろいろな理論面での検証ができるようになってきたということです。

大規模データの集積に関しては、質問項目自体が安定して精度の高いものにしていかなければなりません。言い換えれば、これは標準試験の項目と同じで、一つ一つの質問項目に意味付けをして、精度を高めていかなければならない、それができて質問項目の精査に入り、標準化調査の作成ということになっていって、そして初めて妥当性や信頼性が検証されるということになっていきます。まだここまでは行っていない状況です。

それを何とかしようとしているのが、こうした分析の方法(#22)です。

### 新入生調査から見た日本の大学1年生の学び(#23)

例として、JFS(新入生調査)から見た日本の大学1年生の学びを見てみたいと思います。

この図(#24)で用いられているのは2008年のデータで、1万9000人ぐらい集まっています。国立、公立、私立の比率も学校基本調査とほとんど変わりありませんので、ある程度信頼できるものと思われまます。

この図(#25)は、専攻分野別で見た高校時代の向学態度を見たものですが、この中で授業中に質問をしたかどうかということを見ました。授業中の質問の傾向には分野ごとの大きな違いは見られません。ただ教育学、これはどちらかというと言っても教員養

成系ですが、教員を志望する高校生にはやや積極的な態度が見られるということが分かります。

次(#26)は意見を論理的に主張したかどうかということですが、その経験を高校時代にどれくらいしたかということを知っています。全くしなかったという学生が全体の50%を占めています。私も大学教員が現在の高校生に期待すること、あるいはこうあるべきという高校生像になっているのかということを考えさせるデータではないかと思うところです。

それから、この図(#27)はこちらにもおられる理系の先生方にとっては大事な点だと思うのですが、科学論文、これはものすごく難しい論文というのではなくて、新聞や雑誌の記事なども含めてですが、それを読んだかどうかということ考えた場合に、読まなかった、全く読まなかったという学生は全体の70%近くになります。しかし、これはいろいろな分野がありますので、よく読んだ学生に視点を当てますと、理学で11%、農学9%、生物学9%程度、その一方で全く読まなかった学生が理学で61%、農学で57%、生物学54%というようにやはり過半数となっています。

次には補習授業、いわゆるリメディアル教育を受けている比率を見たものです(#28)。この横軸のタイプは、学生が高校時代にどのような基準で学習をしたかということ類型化したものです。その類型として、まずは目的型、そして探究学習型Aというのは探究的な学習を経験していただいてもそれを理解しないまま受けていた学生、高校指導従順型は教師あるいは予備校などの指導に本当に素直に従っていた学生、受験勉強型Aは取りあえず受験をして進学しようというタイプの学生、探究学習型Bはそういう問題背景や問題解決型の授業を受けてその意味を分かりながら学習していったというタイプです。受験勉強型Bははっきりと大学で学びたい分野が決まっていたという学生です。

そうして見ますと、補習教育を受けている学生が一番多いタイプは、高校指導従順型です。これは少し分かりづらいかと思いますが、学生が自ら補習教育を受けるといってもなく、大学が受けろということでもありません。ですから、高校指導従順型の学生は、自発的に選ぶというよりは、とにかくメニューにあるもの、つまり時間割にあるものを履修しようという傾向が強いということです。また、補習授業の内容では英語の割合が一番高



くなっていますので、このあたりは、英語教育に携わっている先生方にとって、設定している水準に到達していない学生が増えているのかもしれない。

この図（#29）は、専攻分野によって、補習授業を受けたい、あるいは受けたくないという希望を聞いたものの対応分析です。この図ではプロットが近いところを見ていただきたいので、右側の輪で囲んでいるところを囲まりとして見ることになります。言い換えれば、「現在受講していないし、希望もない」という学生です。これは文章表現、すなわち日本語表現の補習授業を意味しています。つまり、医学系、工学系、情報科学系、農学系、理学系、生物学系の分野の学生たちは、文章表現の補習授業を受けたくないという人たちが比較的多いということになります。初年次教育の中で全学で共通教育で提供しているときに、学生たちは比較的ばらばらに文章表現など、レポートの書き方などの補習授業を受けると聞いています。そういうときに、理系の学生たちにとっては文系の先生方が教えるレポートの書き方というのは、もしかすると必要ないのかもしれない。むしろ理系の書き方というのを教える必要があるとも取れるし、逆にこれを教養として考えたときに理系の学生たちも文系のレポートの書き方が必要と見ていくという見方もできます。

一方で、現在受講していないけれども希望が高いという分野は、社会科学、そして人文科学です。このあたりは文章表現というのは、今の学生たちにはもしかするともう少し丁寧に教えていかなければならないという意味を持っているのかもしれない。

次に、この図（#30）は初年次教育の満足度です。教員養成系あるいは教育学系の学部、農学系などの初年次教育は満足度が高い学生たちが比較的多く、理学、家政学の方は不満足派が比較的多いということが見てとれます。中間派は経済学、医学、文学という領域になります。家政学のあたりなどでは、初年度教育プログラムがまだ浸透していないということを示しているのかもしれない。

こちらは学生の類型と専攻分野で、ネガティブな学生、ややネガティブ、ややポジティブ、ポジティブな学生というのはどのような分野に多いのかというのを見たものです。私立大学ではそのような傾向があるのですが、国立でも比較的経済や法学などでは大教室で行われ

ている大きな単位のクラスが多いのではないかと思います。そこにネガティブ学生が比較的多くいます。これは実践されている先生方から見ると「違う」というご認識もあるかもしれませんが、データ上ではこのような結果が出てきました。

ややネガティブな学生が家政学、心理学、理学などに少し固まっています。ややポジティブは医学、生物学のあたりです。ポジティブは教育学、農学、そして芸術学といったところに多いということが見えます。

### 新入生・大学生調査における複数の統計的手法を用いての分析（#32）

さて、それでは少し違う手法を用いて見てみたいと思います。

ここではマルチレベル分析というのを使います。なぜこの分析をここで用いるかというと、基本的には先ほど申し上げたように、私どもは将来、この調査によって大学間、あるいは学系間の比較ができるように、つまりベンチマークができるようにしていけたら、どの大学もこのデータが使えると思っています。しかし、実際には大学間での階層性があります。つまりA大学、B大学、N大学というようなところで、大学単位のデータはそれぞれが独立したサンプルではないかもしれない。似た学生たちという形で、実際には階層性のあるデータから成り立っているのではないかと。

それを今までは集団内のサンプルを独立なものとして扱っていたり、集団ごとに平均して集団数をサンプル数として扱ってきました。

しかしこの問題点は、サンプルは実際に独立ではありません。大阪大学には大阪大学の学生文化があるので、分野は違っていても実際集団内で類似しているということが多分あるだろうと思います。

ですから、平均値というのは純粋な集団の性質を反映しません。そういう意味でいうと、私どもの今までのデータを使った多変量解析などでは、こうした大学や学部ごとの類似性を反映しないまま分析してきたという反省があります。

それをマルチレベルという方法で分析すれば、サンプルの類似性というものを前提として分析することができるということです。

そしてここでは大学での充実度と大学教育の質の満足度というものを、マルチレベル分析を使ってどうなるかということを見ていきたいということです。

モデルとしては設置形態を国立・公立、それ以外としています。そして入試の変数、志望の変数を第一志望かそれ以外というように分けてみました。第一水準を大学とし、類似性があるとしています。第二水準は個人です。もちろんこれは、本来は大学だけの類似性というよりは、おそらく学系とか学部ごとの類似性の方があられるかもしれません。しかし、今回は学部が非常に多様で複雑であったために、学部を水準として扱わない方法で分析を行いました。従属変数は大学での充実度です。

この図(#38)が結果です。この結果でいうと、モデル1というのは大学教育の質の満足度と充実度だけで表したモデルで、そしてあとは説明変数を増やしていくのですが、モデル3の全部入れたモデルが基本的には有効であると見ることができます。ただ、ICCの値が意味するのは、分散を説明できる割合が3.4%ということで、その値は低いと思われるのではないかと思います。しかし、先行研究でいえば、アメリカで長期間にわたって開発されてきたNSSEやCIRPの例を申し上げましたが、あれらの標準試験も大学間ベンチマークとして使われている歴史が非常に長いのですが、その研究結果でも高くてもせいぜい10%です。ですから、あとは個人間の要因として説明していくということしかできないのかなというところですね。

つまり大学間の要因はそれほど大きくありません。いずれも3~4%程度です。これはマコーミックなどの先行研究と整合的です。やはり大学教育の質全般の満足度が大学での充実度を左右していくということになりますし、設置形態、志望度が大学での充実度にプラスの影響を及ぼしていることが明らかになっています。

こちら(#40)はまた別の方法です。多重対応分析を用いて見たもので、学習生活と大学教育の関係性を、大学在学中にどのような知識・技能を獲得するかということと多重対応分析で見たものです。

その分析結果から見ると、学業成績と学習時間との対応関係というのは明確です。個々の知識・技能について専攻分野によって異なります。

当たり前のことですが、例えば多くの項目に文科系の分野の学生が肯定的に回答する傾向が見えます。これは学問分野によって知識内容の違いということを反映しているのかもしれませんが。先ほど申し上げた中で、ネガティブ学生ということをお願いしたと思います。実は多くの学生の中で、一生懸命高校まで学習してきた第一志望の大学に入ってネガティブに変更する学生が若干いるのですが、そのような学生は理系に多いのです。つまり高校の学習成果として大学入学ができた、それがその時点で精一杯だったかもしれないのです。文系の方はそれだけの差が出てこないの、非常に肯定的に回答する傾向があります。逆に言うと、はっきりと学習成果を測ることが語学以外ではあまりないのかもしれない。

こちらは別の研究者が行っている分析です。これは私はあまり説明できませんので、グラフをご覧ください。いわゆるIRTで等化した後の大学間比較ということになります。ですからこれで見ると、F大学はJCSS2005では初年次の満足度は高かったのですが、2007年では低くなっています。この間で等化した結果なので、F大学の初年次教育プログラムに何らかの変更があったのではないかとということです。どちらかというD大学が若干下がってはいるけれども、満足度が上に向いているという結果を見ると、まあいいのかなというところですね。しかし、これはデータが少ないので、つまり私どもの抱えている課題ですが、1年、2年では分かりません。何年もかけて等化していかないと安定的なデータが得られないということを、この結果が示していると思います。

今までのまとめがこちらです。項目反応理論(IRT)で得られた見解としては、大学満足度は学年ごとに変化します。2年次に満足度がいったん下がるということがいわれています。こうしたものを研究から分かったこととして、個別大学で改善のためのツールとして機能するためにはどうしなければいけないかということが、あくまでも研究としては分かるのですが、実際にこれを使ってツールとして大学の中でどうしていくかということが、第3ステージとして非常に重要になってくるかと思います。

そこで課題です。教育改善・自己点検というPDCAの3番目の図を思い出してください。それをするためには、例えばこの標準的に開発された学生調査ができた

して、大学機関としてどう使用しているか、どう活用したいかということがあります。例えば調査する主体が研究者のグループであるのか、その一方で改善していく主体はどこであるかという問題です。ここでは調査主体と改善主体が一致しているのか、別個であるのかで全然違ってきます。あるいは、教育改善につなげるための説得力は誰が担うか。当然、そこでその意味を執行部が理解しているかということも大事になってきます。学生調査をこのIRとして機能させるためには、直接評価とどう結び付けていくのか、学内の学生データとのリンクと分析は誰が担うかといった大きな問題もあります。あくまでも先ほどの評価は間接評価ですから、直接評価の結果は分かりません。プロセスは見えますが、実際にAという学生の成績はどうであるか、単位の取得状況はどうであるか、あるいはあまり積極的に大学には来ていないのかということは見えません。学生調査はそういった学習成果の直接評価と結び付けていくことが必要になりますが、それを一体誰が担っていくのかということも考えていかなければならない課題になります。

#### 4大学の戦略的支援プログラム (#45)

今までお見せしてきたのは2004年からずっと行っている科研の部分で、あくまでも純粋な研究として行ってきたものです。これを少し応用したもので、4大学の戦略的支援プログラムというものが、2009（平成21）年度に採択されました。来年度が最終年度となりますが、こちらをご紹介します。こちらをご紹介します。

これは私ども同志社大学が代表大学となっていますが、北海道大学、大阪府立大学、甲南大学という設置母体も全く違うところが4大学で取り組んでいるものです。学生調査は共通の学生調査の1年生調査を作り上げて、それを軸とするIR機能の開発を基盤として、IRを通じた連携大学間での相互評価の結果を学士課程教育の向上に結び付ける、質保証システムとして作り上げてきました。その中でどのようなことをしているかということ、4大学のIR学生調査ネットワークで共通の学生調査を使っています。ただそれだけではありません。それだけだったら分析だけで終わってしまいます。

そこで、実は、新たにネットワークのシステムを今作り上げていて、学生調査の結果と大学の中にある学生のデータとをリンクさせるということを行っています。つまり、各大学の暗号化した教務データとして、単位の累

積状況、所属学部の平均GPA、高校の場所、入試の種別などのデータをもらい、それらをこの学生調査と結び付けます。ただ、結び付けるときには、当然ながら学生のセキュリティを大事にしなければいけませんので、各大学の中で暗号化したものを別の人がキーで結び付けてくれます。分析者が私どもの研究センターの中にいて、それを分析するという形になっており、その結果を4大学の間で相互評価していくということになります。

そしてベンチマーキングしていくということですが、例えば北海道大学は非常に熱心に単位の実質化を進めてこられています。つまり、学生の自発的な自習時間、学習時間を確保することを行ってきていますので、そういうことがもし結果に表れていたら、データを分析して何がそうさせているのかということ、われわれは学んでいくというか、応用していきたいと考えています。

もう一つはそのシステムですが、これはどの大学にも分析できる人がいるとは限りません。ですから、ボタン一つで学生調査の結果がグラフ化されるようなシステムを今開発しています。大阪大学の人間科学のセンターにはあると思うのですが、例えば慣れていない教務の方や職員の方でも使えるようにと考えています。

それから、最終的にこのプログラムはアウトカムを非常に重要視しています。そのアウトカムの部分は英語教育の成果としました。北海道大学、大阪府立大学、同志社大学、甲南大学は全部違います。例えば北海道大学や大阪府立大学では理系にどちらかという特徴がありますし、私ども同志社も甲南も理系よりはどちらかという文系・社会科学系の学生が多く、そういう分野の方が多いといった特徴があります。そこで共通点は何かということ議論した上で、やはり英語教育だろうということで英語の指標を使いました。

その英語の指標は何かということ、CEFR (Common European Framework of Reference for Languages) というヨーロッパ共通言語フレームワークです。ご存じの方もおられるかと思いますが、それを使いまして、4大学でベンチマークしていこうということです。この戦略連携のプログラムは10年を考えるとと言われましたので、その結果を10年を単位として何かに活かしていかなければいけないとしています。しかし、皆さんもご存じのように文部科学省のプログラム全体が事業仕分けの対象になりまして、その結果「廃止」ということになっていますから、どうなるか分かりません。

そこで、それでは大学の英語教育のカリキュラムに生かそうということで、主に甲南大学が考えてくれていま

すが、英語教育というものを、ナンバリングなりで100番台から400番台というように階層性のあるものにしていく予定です。そして英語の教育が積み上げでできるようなカリキュラムになることを考えていますが、そういうようにこの相互評価とベンチマーキングの結果を最終的には反映したいということを考えています。

最後のIR人材の育成のためのワークショップ事業というのは、当然これはそういう分析をする人を育てなければいけませんので、分析の手法なりそういうシステムなりを使ってワークショップを行っています。

形にすればこういうことになります (#47)。これは実際の大学の中にあるものですが、イメージとしてはデータを集計してこのシステムの中に入れて分析して、ベンチマークと相互評価の結果が出力されます。これは参加大学であれば各大学の担当者がすぐにボタン一つで表示

できるということを考えています。

このプログラムが今年で2年経過しましたので、2年間のデータが集まっています、それを初めて実際の各大学の中にある教務データ・学生データあるいは機関データと組み合わせていくということで相互評価ができるようになり、それと同時に、各大学の中でも分析する人がいれば非常に細かい分析ができて教育改善に生かすことができるという内容です。

現在は4大学ですが、パイロット校を増やすということも言っています、今年度からA大学がパイロット校として参加してくださいました。将来的にはこれをIRコンソーシアムにしていかなければならないようにしていますので、もしご関心があればご参加していただければということで、最後は宣伝で終わらせていただきます。

**学生調査の活用:**  
**新入生調査、大学生調査から見える**  
**教育成果への影響**

大阪大学・大学教育実践センター  
 第4回 高等教育質保証研究会

2011年1月7日  
 山田礼子(同志社大学)

**アウトカム・アセスメントを機能**  
**させるには**

**学士課程の充実という大きな課題**

✕ 学士課程教育の構築に向けて各大学が自らの教育理念と学生の成長を実現する学習の場として学士課程を充実することが求められている

↓

But

✕ 学生の成長をどう測定するのか?  
 ✕ 学生の成長の成果と学士課程教育の関係性についての研究蓄積の少なさ

↓

How

✕ 学生の成長に関する理論構築と間接評価測定法である継続的學生調査の開発 データの継続性と安定性

**日本における学生調査研究の新たな段階**

✕ アウトカム・アセスメントへの要求  
 世界的な傾向 欧米、アジア等

➢ アウトカム・アセスメントの多様な種類

- ・ 学習成果や学業達成の直接アセスメント
- ・ 分野別直接アセスメント
- ・ 情緒的側面の間接アセスメント

✕ 政策的転換 大学の教育力への期待と要望  
 ✕ 大学という環境への注目

カレッジ・インパクト理論の重要性の再認識

**アウトカム・アセスメントの方法**

□ アウトカム・アセスメント＝ラーニング・アウトカムを測定するには？

直接評価＝ダイレクト・エビデンス＝学習成果の評価  
 内容＝科目試験、レポート、プロジェクト、ポートフォリオ、卒業試験、卒業研究や卒業論文、標準試験  
 分野＝一般教育、専門分野別

間接評価＝インダイレクト・エビデンス＝学習プロセスの評価＝学習行動、生活行動、自己認識、大学の教育プログラムへの満足度等成果にいたるまでの過程  
 内容＝学生調査、卒業生調査等  
 時期＝入学時、1年次終了時、上級学年時、卒業後

**アメリカの直接評価としての標準試験**

- ・ 標準試験は代表的直接評価
  - 例 SAT, ACT, GRE
  - 一般教育の成果を測定するには、CAAP, MAAP等 (多項選択式)
  - 専門分野別試験(ETS)

実用化に向けて保証すべきところは？

- 妥当性 テストしている内容が適切か、妥当であるか
- 信頼性 得点傾向、回答傾向等

○ 長期間での大規模な検証 ○ 理論にもとづいたテストの作成  
 古典的テスト理論  
 IRT (Item Response Theory)

↓

ラーニング・アウトカム測定の方法のひとつ

**間接評価としての学生調査**

- ・ 直接評価で測定できないプロセス  
 学習成果に導くための学生像や成長度合の把握

実用化に向けて保証すべきところは？

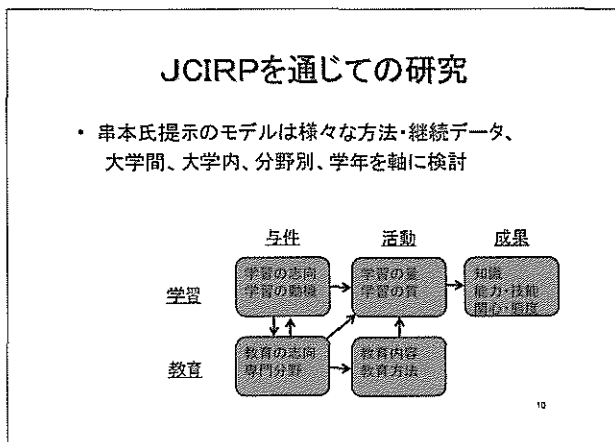
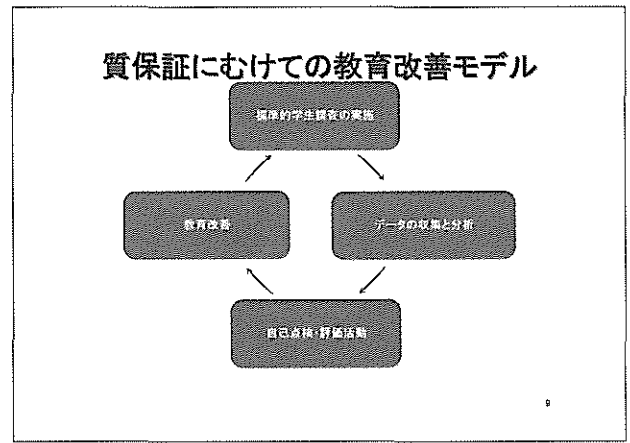
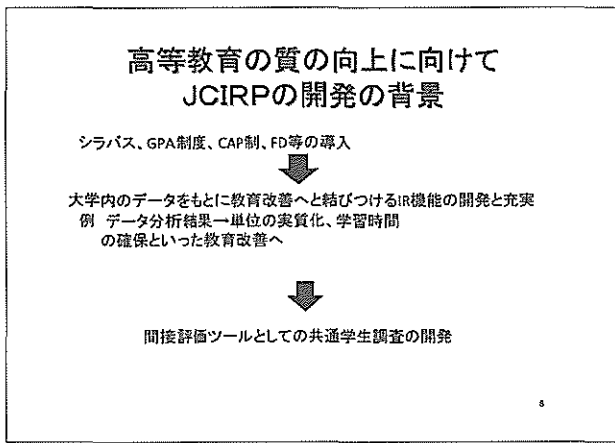
- 妥当性と信頼性
- 長期間での大規模な検証
- 理論にもとづいた項目の作成
  - 例 総合的なカレッジ・インパクト理論 → CIRP
  - Engagement 理論 → NSSE

間接評価(学生調査)はI-E-O モデルのE(環境)を検証する方法のひとつ

↓

直接評価と間接評価を組み合わせるとより効果が期待

**JCIRP開発の状況**



- ### 研究のこれまでの目的と特徴
- 学生の成長を測定する学生調査の意味、学生の成長を支える高等教育機関の効果の測定の開発
  - 研究の特徴: アメリカで蓄積されてきたカレッジ・インパクト研究の流れをベースにアスティンのI-E-O(既得-環境-成果)モデルに依拠しつつ、大学生への教育効果・成果について検討  
カレッジ・インパクトモデルの検証と発展
  - アメリカの学生調査と互換性のある調査の開発と実施
- 11

### 大学生調査研究の目的と成果

**問題設定**

- 一 大学の影響・教育力(カレッジ・インパクト)と学生の成長の関係性はいかなるものか
- 一 大学間および大学内での差や共通性はどのようなものなのか
- 一 環境要因と教育成果との関係性はどのようなものであるか

**調査からの検証**

- ・ 大学という環境が学生の成長に及ぼす重要性の存在
- ・ 学年、学生の関与、教員の関与というカレッジ・インパクトがもたらす教育効果への意義
- ・ ネガティブ学生に見られる落ち込みと小さい大学間格差
- ・ 大学の経験を通じて変化する学生類型

**課題**

アスティンのI-E-OモデルのIの部分はかなり大きい  
大学入学以前に規定されている行動・能力をどう大学教育が伸長する

12

### 国際比較ができる

13

### データについて

■ CSS2005  
調査対象 米国の4年制大学に在学している3年から4年になる学生30188人  
(男性11,367人、女性18,821人)

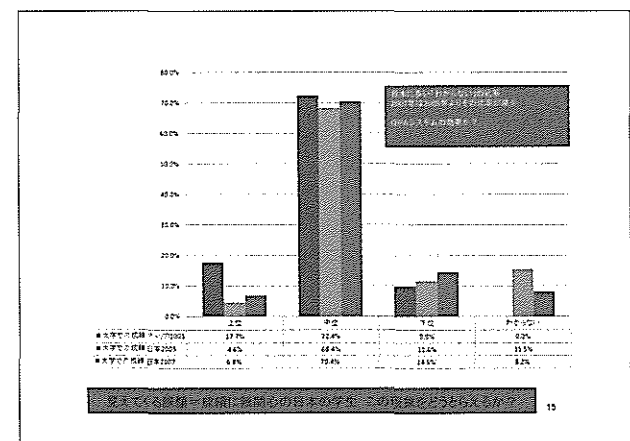
■ JCSS2005  
調査対象 国公立大学8校3961名  
調査時期 2005年10月-2006年1月

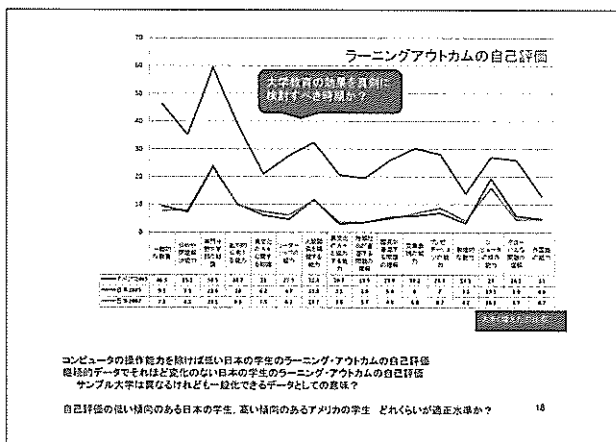
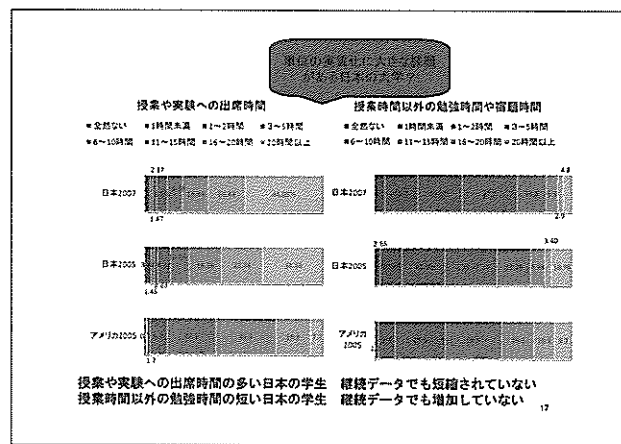
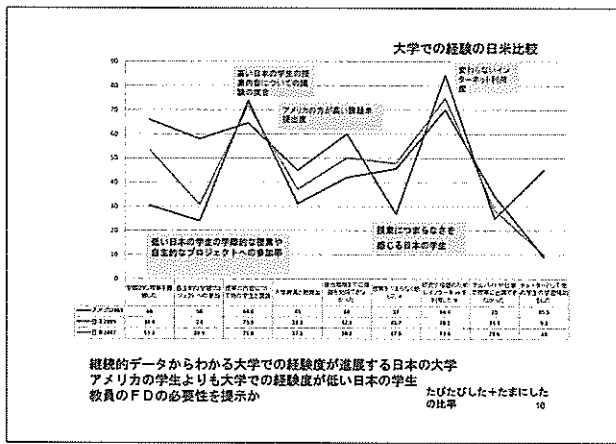
大学	国立	私立	合計
A(理)	680	0	680
B(理)	239	0	239
C(理)	97	0	97
A(文)	384	0	384
C(文)	211	0	211
D(文)	91	237	327
E(文)	0	811	811
F(文)	0	237	237
G(文)	0	644	644
H(文)	0	386	386
合計	1611	2094	3705

■ JCSS2007  
調査対象 国公立大学16校6228名  
調査時期 2007年12月-2008年1月

大学名	国立	私立	国立	私立	合計
A	621	0	0	0	621
B	51	0	0	0	51
C	51	0	0	0	51
D	453	0	0	0	453
E	313	0	0	0	313
F	0	497	0	0	497
G	0	2145	0	0	2145
H	0	0	23	0	23
I	0	0	83	0	83
J	0	0	357	0	357
K	0	0	157	0	157
L	0	0	322	0	322
M	0	0	473	0	473
N	0	0	227	0	227
O	0	0	112	0	112
P	0	0	167	0	167
合計	1543	2559	1724	319	4251

14





### 成果からみえる次の課題に向けて ：第一ステージ

- カレッジ・インパクト理論を検証するためには：
  - 一継続的調査 → JCSS(2004, 2005, 2007, 2009) JFS (2008, 2009)
  - JCSS(2008, 2009) 累積約54000人 → 機関ベンチマークの進展
  - 一複数の時点での調査 JFSとJCSS 機関内パネル調査に  
機関によっては個別学生パネル調査として利用
  - 一4年制大学機関と2年制短期大学機関での検証  
JCSSの開発と実施
- 一成果の検証
- カレッジ・インパクト理論の検証はある程度可能に  
しかし、立ちちはかかる大きな課題

### JCIRPプロジェクトの課題

- JCIRPの置かれた状況
  1. 毎年集まる複数調査の大規模データ
  2. 参加大学が異なり、サンプルサイズが年度によってまちまち
  3. 参加大学によって、対象学年もまちまち
  4. JFS(新入生用調査)とJCSS(上級生用調査)とで目的が違う調査の比較
  5. 調査項目がJFSとJCSS間でも年度間でも入れ替わる
  6. 今後も時代の要請・変化によって追加項目が増える可能性
- JCIRPの抱える課題
  1. 経年変化・学年変化に際して、単純集計結果では統計的因果関係を説明しにくい
  2. 学生の成長を統計的に汲み取る方法論が必要
  3. 単純集計だと共通項目のみしか比較できない → 例えば、共通項目以外の項目も含んだ総合的測定が必要
  4. 学年進級による質問項目の質的変化の勘案が重要
  5. 共通尺度の必要性 → JCIRPの統計的標準する項目も → IRTによる等化

共同研究者木村作成(長崎大学) 20

### 成果からみえる次の課題に向けて ：第二ステージ

- 様々な検証方法の開発 → 多様な手法を通じて  
カレッジ・インパクトが検証されねばならない → アメリカのCIRPに見られる理論サポートの安定性の要因
- 大規模データの集積 → 質問項目の安定性と精度の向上 → 質問項目の精査 → 標準化調査の作成 →  
カレッジ・インパクト理論と標準質問項目の妥当性、信頼性の検証

多様な分析方法による検証：クロス集計分析、対応分析、マルチレベル分析、多重対応分析、IRT

新入生調査からみる日本の大学一年生の学び

### JFS(日本版新入生調査)の概要

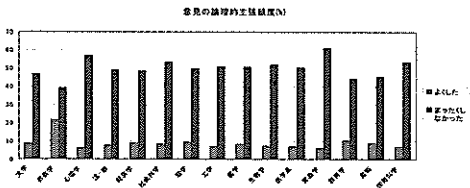
設置形態	度数	比率	高校成績	度数	比率
国立	3,523	17.9	上位	3,536	18.3
公立	1,568	8	中位の上	5,287	27.3
私立	14,570	74.1	中位	4,583	23.7
合計	19,332	100	中位の下	2,753	14.3
			下位	2,556	13.2
			その他	565	2.9
性別	度数	比率	小計	19,274	99.7
男性	10,103	52.3	無回答	59	0.3
女性	8,710	45.1	合計	19,332	100
無回答	519	2.7			
小計	19,332	100			

### 専攻別高校時代の向学態度(1)



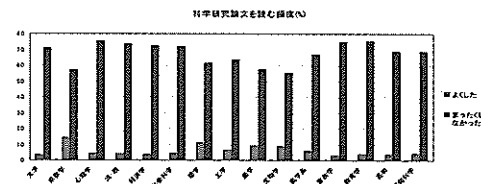
- 1年生(短期大学の1年生を含む)。宗教学は実数28(以下同)
- 授業中の質問の傾向に分野ごとの大きな違いは見られない。
- 将来「教育学」を専攻する高校生にやや積極的態度?

### 専攻別高校時代の向学態度(2)



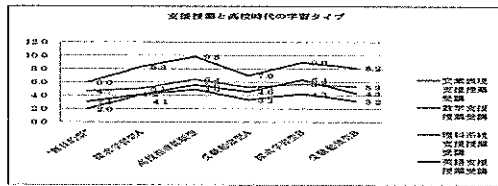
- 「まったくしなかった」: 全体の50.3%
- 将来「教育学」を専攻する高校生にやや積極的態度?
- 「家政学」: 意見の論理的主張の経験のない学生が61%

### 専攻別高校時代の向学態度(3)



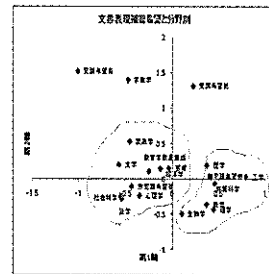
- 高校生は原則として科学研究論文を読まない。「まったくしなかった」: 全体の69.7%
- よく読んだ学生: 「理学」11.4%、「農学」9.0%、「生物学」8.7%
- まったく読まなかった学生: 「理学」61.4%、「農学」57.3%、「生物学」54%

### 補習(支援)授業と学生類型



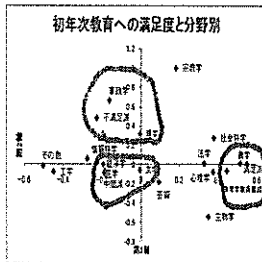
- 英語の補習授業を受講している学生の比率が最も高い
- 高校指導従属型学生がいずれの科目においても補習授業を受講する比率が高い
- 自主的学習の基礎が確立されていない?

### 専攻分野による補習(支援)授業受講希望の対応分析



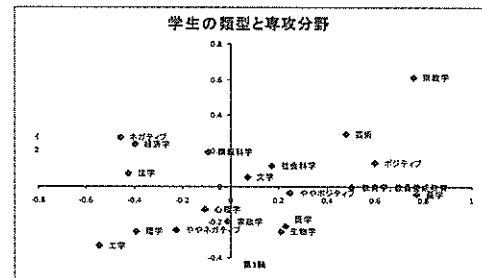
- ・文章表現の補習(支援)授業を受講していないが受講希望者が多い専攻分野はレポートや記述式試験の多い領域である文系に集中
- ・受講しないのか受講希望も無い専攻分野は理系に集中形式的レポート作成だけで十分か?

### 専攻分野別に見られる授業の満足度の対応分析



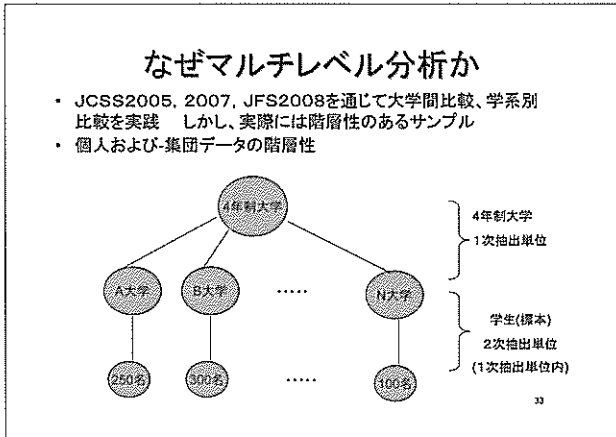
- ・初年次教育プログラムへの満足度の高い専門分野は教育学系と農学系
- ・不満足群が多い領域は家政学と理学系統
- ・初年次教育プログラムの浸透度が低い? プログラム構築に課題か?

### 学生の類型と専攻分野





新入生・大学生調査における複数の統計的手法を用いた分析



### データの階層性とその扱い

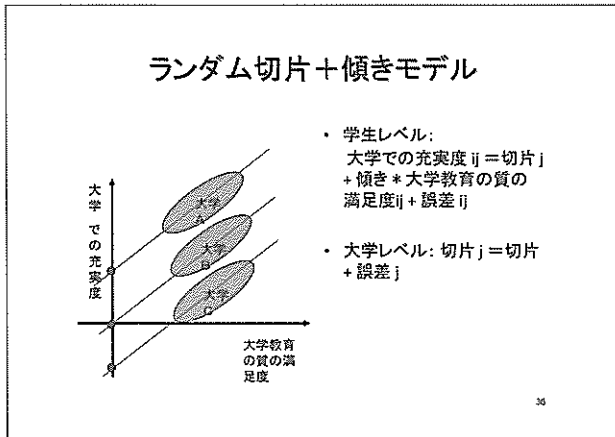
- データの階層性
  - ー 集団ごとにネストされたデータ
  - ー 集団内で類似したデータ
- これまでの階層データの扱い
  - ー 集団内のサンプルを独立なものとして扱う
  - ー 集団ごとに平均して集団数をサンプル数とする
- 上記の方法の問題点
  - ー サンプルは独立ではないー集団内で類似している
  - ー 平均値は純粋な集団の性質を反映しない

↓

JCSS2005, 2007, JFS2008の大学間比較、学部間比較、学系別比較は大学や学部ごとの類似性を反映しないまま分析

### マルチレベル分析による利点

- 変数間の関係を、1次抽出単位内(大学内、学生間)の違いと1次抽出単位間(大学間)の違いに分解し、階層的データを適切に分析
- 「集団内類似性」を評価し、それに合わせてモデリング
  - ー  $Y_i = \beta_0 + \beta_1 X_i + e_i$  (iは個人)
  - Yは従属変数、Xは個人レベル変数、eは誤差変数
  - ・  $\beta_0$ は切片、 $\beta_1$ は回帰係数
- ー 次に、各母数が集団ごとで違うと仮定
  - ー  $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$  (jは集団)



### モデルの設定

- ランダム切片と傾きモデルを基本に設定
- 設置形態ダミーー国立・公立=1 それ以外0
- 入試ダミーー一般入試・センター入試=1 推薦、AO等=0
- 志望ダミーー 第一志望=1 それ以外=0
- 第一水準 大学(類似性があるとする)、第二水準 個人 (今回は学部が多様で複雑なため、学部を水準として扱わない)

大学での充実度と大学教育全般の質の満足度の関係を探る

- 従属変数=大学での充実度

	大学教育の質全般の満足度と大学での充実度			
	マルチレベルモデル		ランダム切片+傾きモデル	
	Nullモデル	モデル1	モデル2	モデル3
固定効果				
切片 (Y00)	3.727***	2.53***	2.48***	2.393***
大学教育の質全般の満足度		0.388***	0.391***	0.386***
設置形態ダミー			0.168***	0.148***
志望ダミー				0.130***
入試ダミー				0.047**
学生間分散	1.162	1.045	0.149	0.139
大学間分散	0.048	0.037	0.005	0.005
共分散			-0.029	-0.027
ICC	4.10%	3.40%	3.20%	3.40%
AIC	59176	55363	55293	53506

\*\*\*: p<.01 \*\* : p<.05

### モデルが意味することと今後の検証課題

- ー Nullモデル、モデル1、モデル2と比べた場合、AIC値、逸脱度からみてモデル3が適切なモデル
- ー 大学教育の質全般の満足度が大学での充実度を左右
- ー 設置形態(国・公立)、志望度が大学での充実度にプラスの影響
- ー 大学間の要因は大きくはない。いずれも3~4%程度、個人間の要因の方が大きい(マコーミックの先行研究と整合的)

今後の課題

- ー 充実度という尺度は個人間の違いによるものが大きいのが、大学間比較は、継続的データを使い、ベンチマーキングをしていくためには不可欠。それゆえ、項目の精査と尺度の開発が課題

### 学習成果と間接評価

- 1) 大学教育に対して学習成果を求める動き
  - 2) 在学中の経験が職業的地位や所得にもたらす効果  
松繁編(2004)、矢野(2005, 2009)など。
  - 3) 大学在学中にどのような知識・技能を獲得するか？  
村澤(2003)、葛城(2006)、小方(2008)など。  
→ 学生の学習活動や専攻分野による違いを確認。  
・ただし、獲得された知識・技能を集約して捉えているために、専攻分野や学習時間がどんな知識・技能の形成に寄与しているかは十分にわかっていない。  
・比較の視点の必要性(専攻分野間の比較の可能性)。
- 多重対応分析を用いた探索的分析

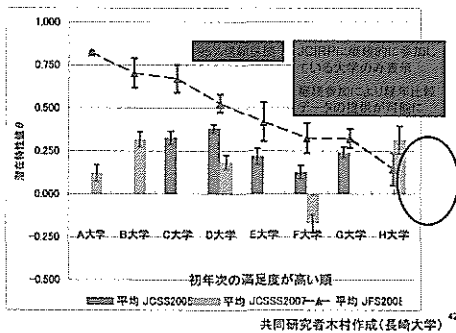
共同研究者古田作成(同志社大学) 40

### 探索的多重対応分析からの知見

- 学業成績と学習時間との対応関係が明確。
- 個々の知識・技能について専攻分野によって異なる。
- 多くの項目に、文科系の分野の学生が肯定的に回答する傾向。  
... 学問分野により知識内容の違い。  
・特定の項目では、理科系の分野で自己評価が高いものも確認された(「数理的な能力」)。  
... 文科系、理科系の分化。
- 客観的変数と意識変数の対応から、知識・技能の自己評価でもある程度の妥当性が認められる。
- ただし、理科系の分野の特徴をつかめていない可能性がある。→ 質問項目の再検討。

共同研究者古田作成(同志社大学) 41

### IRTで等化後の大学間比較



### 研究からわかったこと

- 大規模継続調査の利点：複数の時点、複数の大学での 幅広い調査データの蓄積が可能になってきた
  - 1. マルチレベル分析で得られた見解  
→ 大学の充実は、大学教育の質全般に関係するが、大学間の要因よりも個人間の要因の方が大きい。しかし、ベンチマーク(類似大学のグループ化)により、大学間の要因は大きくなる可能性  
→ 学生群の分類がIRモデルの効果測定には役立つ
  - 2. 多重対応分析で得られた見解  
→ 学習時間と学業成績、専攻分野と獲得した知識・技能の対応関係。  
→ 調査項目のある程度の妥当性。ただし再検討の必要性。
  - 3. 項目反応理論で得られた見解  
→ 大学満足度が、学年ごとに変化。特に、2年次の落ち込みが激しい  
→ 各大学の継続的な参加によって、各大学の教育状況をモニタリングできる機能が、JCIRPIには備わっていない
- ↓
- 本当に個別大学で改善のためのツールとして機能しているのか？

43

### 教育改善・自己点検のために

- ① 標準的 学生調査を大学機関としてどう使用しているか、どう活用したいか  
調査主体と改善主体は一致？ 別個？
- ② 教育改善につなげるための説得力は？  
誰が担うか？ 執行部が理解？
- ③ 学生調査をIRとして機能させるためには  
直接評価とどう結びつけるのか？  
学内学生データとのリンクと分析は誰が担うか？

44

新入生調査・大学生調査の応用編：4大学戦略的支援プログラム

### 連携取組の目的と内容

#### 目的

国公立4大学の連携により、学生調査分析を軸とするIR機能の開発を基盤として、IRを通じた連携大学間での相互評価の結果を学士課程教育の質的向上に結びつける質保証システムの創出と全国規模のIRコミュニティの育成を目指す

#### 内容

- ・4大学IR学生調査ネットワークの構築と活用
- ・連携大学間におけるIRネットワークシステムの構築
- ・IRネットワークを活用した相互評価とベンチマーキング
- ・IR人材育成のためのワークショップ事業

45

### IRシステムの概要

複数大学の調査結果を集計・分析し、レポートを自動生成

