

広島大学 高等教育研究開発センター 大学論集  
第42集 (2010年度) 2011年3月発行：245-263

## 大規模継続学生調査の可能性と課題

山田 礼子

# 大規模継続学生調査の可能性と課題

山田 礼子\*

## はじめに

筆者等は、2004年以来継続的な学生調査研究を行っている。現在まで、①学生の成長を測定する学生調査の意味、②学生の成長を支える高等教育機関の効果の測定の開発という研究目的を掲げ、UCLA 高等教育研究所（以下 HERI = Higher Education Research Institute）の大学生調査（CSS = College Student Survey）と新入生調査（TFS = The Freshman Survey）とそれぞれ互換性のある日本版大学生調査（以下 JCSS = Japanese College Student Survey）、日本版新入生調査（以下 JFS = Japanese Freshman Survey）、および日本独自の短期大学生調査（以下 JJCSS = Japanese Junior College Student Survey）を開発してきた。3種類の学生調査を総称して JCIRP（Japanese Cooperative Institutional Research Program）と呼称している。

高等教育がユニバーサル化するなかで、学生の変容も著しいと指摘され、従来とは異なる教育方法や教員の関わりが求められつつある。以前とは異なる時代環境の中で、模索される新しい大学像と教育を具現化するためにも、これまで日本の大学と学生研究では限界があるとされてきたカレッジ・インパクトモデルの再考を通じて、大学生への教育効果・成果について検討してきた。

カレッジ・インパクトモデルを検証するためには、当初開発した JCSS の継続調査に加えて、複数の時点での調査が不可欠であることから、JFS は、JCSS のインプット部分として利用できるように設計した。JFS2008には、163大学から19,661人（うち新入生は19,332人）の新入生が参加した。同時に開発した JJCSS2008には、9短期大学1,996人の新入生が参加した。2009年度は、JFS には46大学から8,534人、JCSS には24大学から4,183人、JJCSS には30短期大学から約8,000人が参加した。2004年以降から現在までの3種類の継続調査への累積参加学生数は54,000人を超えている<sup>1)</sup>。既に JFS と JCSS、JFS 同士、JCSS 同士、JJCSS 同士による機関パネルとしてのデータ利用が整備されつつあるが、機関によっては、学生の個人パネル調査も可能である。JFS と今まで実施してきた上級生を対象とした JCSS、JJCSS を組み合わせることで、カレッジ・インパクトによる大学生への教育効果の測定研究の展開の可能性が見えてきているのが、プロジェクト型大学生調査研究 JCIRP の現状である。しかしながら、間接評価の代表的な方法である学生調査を通じて、カレッジ・インパクト理論の検証を積み上げながら、教育効果の測定研究を展開していくには、克服しなければならない課題も少なくない。本稿では、最初に直接評価と間接評価という測定概念を高等教育の文脈に照らし合わせて整理する。次に、間接評価である学生調査と大規模な継続調査が克服すべき問題

\*同志社大学社会学研究科教授

を明らかにし、そうした課題に対して何をしなければならないかをデータを通じて探索的に探ることを目的としたい。

## 直接評価と間接評価

2008年の中教審答申で提示された「教育の課程の修了に係る知識・能力の証明」である学位の本質に鑑みて、到達目標や獲得すべき「ラーニング・アウトカム（学習成果）」の提示が俎上に上っているが、本節では、学習成果を教育効果とあわせて教育評価として幅広く捉え、教育評価の測定としての学生調査の意義と限界について検討する。

バンタは、教育評価方法は成果に対する直接評価の一種である科目試験やレポート、プロジェクト、卒業試験、卒業研究や卒業論文あるいは標準試験による検証と学生の学習行動、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を評価する学生調査に代表される間接評価に分類できるとしている (Banta, 2004, pp.4-5)。直接評価は、学習成果を直接に測定する方法として受けとめられ、直接評価を組み入れた学習成果や教育効果を探る先行研究が米国を中心に積み重ねられている (Shavelson, 2010)。直接評価のなかでも特に標準試験をめぐる先行研究が積み重ねられているが、標準試験の信頼性と妥当性を巡る議論が活発に行われ、とりわけ妥当性についての方向性は一定ではない。標準試験の多くは、一般教育の成果測定や文章力や批判的思考力 (クリティカル・シンキング) の測定を意図して開発されている。(Palomba & Banta, 1999, pp.98-99)。

1998年に出版された *The Mental Measurements Yearbook* には、369種類にもものぼる標準試験の内容と700人から成る評者からの標準試験の検証と評価が掲載されている。信頼性と妥当性を巡る議論が顕著であると先述したが、*The Mental Measurements Yearbook* の標準試験の評価においても、多くの学生に適用されることを目的として開発されたテストに、教授内容とその結果として学生が習得した能力やスキルが偏りなく反映されているかという妥当性の検証に頁が多く割かれている。一般教育の成果測定として開発された CAAP (Collegiate Assessment of Academic Proficiency) の妥当性については、テスト理論を用いて得点の等化が複数回検証されてから汎用化されるようになっている。CAAP や MAPP (Measure of Academic Proficiency and Progress)<sup>2)</sup> は、多肢選択方式であることから、等化によって妥当性を高めることができると理論的には考えられるが、新たに開発された CLA (Collegiate Learning Assessment) は、一般教育の成果測定のために開発された CAAP や MAPP とは異なり、大学で学んだ成果全般を標準的に測定し、大学間での比較を可能にするような測定ツールとして開発された標準試験である。また、多肢選択ではなく、「クリティカル・シンキング」「分析的な理由づけ」「問題解決」「文章表現」を包摂した包括的な能力を測定することを目的として、「make-an-argument task」「break-an-argument task」というパフォーマンス・タスクを基本とした問題構成から成り立ち、採点方法も機械的ではなく、記述に基づいて採点を人が行うという方式である。テスト理論による得点の等化を CLA に適応することは容易ではないだけでなく、妥当性の問題がクリアされていないという批判や大学間での比較の信頼性に関する論争は収斂していない (Borden

& Young, 2007) が、年々 CLA を利用する大学も増加しており、大規模な学生数と継続的参加により、CLA の妥当性と信頼性に関する研究が積み重ねられ、改善も期待できよう。米国では、一般教育やクリティカル・シンキング等を測定する直接評価の開発が理論にもとづいて実施されていることから、米国の一般教育の成果として掲げられている能力・スキルを学習成果として測定することは多種多様に開発されてきた標準試験で可能という論理に結び付くのは整合的である。

カレッジ・インパクトの基本となる I-E-O モデルでは、アウトカムもしくはアウトプットは成果、具体的には学生の成績や学習成果、学位取得に相当する。インプットは学生の既得情報であり、学生の高校時代の成績や在籍していた高校の種別、難易度、家庭の所得や親の学歴等に相当する。環境は学生が教育課程内外のなかで経験することを意味する。このモデルに直接評価・間接評価を当てはめ、学習成果を O として見ると直接評価は O を測定するために妥当な測定方法であるといえる。I として入学時に標準試験を実施し、3～4年次終了時に同一標準試験を実施することで、シェイベルソンが開発した「付加価値」方式により、一定の期間での得点の伸張（あると仮定して）である O を検証することも論理的には可能である。しかし、その間の得点の伸びあるいは停滞を測定する場合、直接評価は適切な測定方法であるとはいえない。つまり、O である成果としての直接評価（この場合は標準試験と仮定する）にいたる過程には、学生の大学での経験や関与、その基本となる自己の認識や価値観等が深く関連しているが、直接評価はこの過程を把握することができないからである。そこで、学習成果が提示する部分だけでは、教育評価としては十分ではないという視点から見ると、学生調査やインタビュー等あるいは授業評価で実施される間接評価は、学生の期待度や満足度、学習行動の把握、関与（エンゲージメント、インボルブメント）や経験を把握することができ、成果につながる教育の過程を評価するという機能を伴っている。アナヤは直接評価である GPA、GRE（大学院入学適性試験）と学生の自己評価による成長度合が整合的であることを知見として示したが（Anaya, 1999）、それ以外にも学生調査における学生の自己評価結果が同一学生の直接評価による成果測定結果と整合性があることを検証した研究も積み重ねられている（Astin, 1993; Pascarella & Terenzini, 2005）。直接評価と間接評価の差異について表1にまとめている。

表1 直接評価と間接評価の差異

	測定対象	測定方法	測定分野
直接評価	学習成果	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、標準試験、卒業研究	一般（共通）教育、専門分野、語学
間接評価	学習の過程（学習行動、自己認識、満足度、価値観、経験等）	学生調査、卒業生調査	

先行研究により、直接評価とプロセス評価としての間接評価結果の組み合わせによって、大学生生活を通じての学生の成長に関する精緻な結果測定が可能になると纏められるが、実際に米国では、直接評価、間接評価である学生調査各々が精査され、改良され、その妥当性や信頼性も高められているのが現状である<sup>3)</sup>。我々は、主にプロセス評価として、継続的に学生調査を実施することにより間接評価の精緻化を目指してきているが、残念ながら直接評価との組み合わせまでは実現できていない。しかし、既に米国では、学生調査は、現在では単体で利用される教育評価というよりは、

直接評価と組み合わせて利用される評価としての合意が形成されていることも付加しておく必要がある (Gonyea, 2005)。

## 教育評価の測定における学生調査の意義と限界

JCIRP を通じて、学年、学生の関与、教員の関与という大学という環境の中での学生の活動や教員との相互関係等々を意味するカレッジ・インパクトがもたらす環境面での教育効果が検証された。また、学年が上昇するにつれて、学生が獲得する知は上昇し、学生が様々な大学内での環境に関与すること、それを支える教員の関与が強い教育効果をもたらすことが確認され、学習成果の上昇には、学生のエンゲージメントの充実と質の高い授業の提供が不可欠であることも検証された (山田, 2007, 2008)。個別大学ごとに検討した場合、入学難易度とは関係なく、カレッジ・インパクトが学生の教育効果に影響を及ぼしている大学の存在が確認され (山田, 2007)、大学での経験を通じて大学生生活を充実していると捉えるなどのポジティブな評価をしている学生と、入学時の進学理由が自発的であったにもかかわらず大学での経験を通じて大学生生活を前向きに捉えなくなっているネガティブ学生が存在していることが得られた。このようなネガティブ学生に見られる効果の落ち込みは深く、小さい大学間格差が明らかになり、大学での経験を通じて変化する学生類型といった行動様式も検証された (杉谷, 2007)。さらには、多くの先行研究同様に高校時代までの学習等のインプットの部分の大きさを再確認しつつも、高校入学前の学習経験と大学での学習の適応との関係性を JFS2008 から検証した (木村・西郡・山田, 2009)。JCIRP 以外の大規模学生調査を通じて大学教育のアウトカムを検証する研究も積み重ねられてきている (小方, 2008)。

このように、カレッジ・インパクト理論の日本の大学での適応性についての検証が積み重ねられてきている一方で、大きな課題が立ちはだかっている。複数の調査から成り立つ JCIRP は、毎年複数調査を通じての大規模データが集まるが、参加大学が異なり、サンプルサイズが年度によって、かつ参加大学によっては対象学年も統一性が取れていないという性格を伴っている。同時に、年度毎に、調査項目が追加あるいは改定されるといった状況に置かれている。したがって、(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ことから、学生の成長を統計的に安定的に汲み取る複数の方法論を提示することが不可欠である。(2) 単純集計だと共通項目のみしか比較できないことから、共通項目以外の項目も含んだ総合的測定をおこない、質問項目の変化を勘案しながら、標準的な共通尺度を作成していくことも視野にいれなければならない。換言すれば、多様なサンプルから成り立つデータという性格を前提として、複数の検証方法を通じて、カレッジ・インパクト理論の安定的な検証を行い、同時に、頻繁に入れ替わる項目や大規模データの集積という性格を持つデータから、質問項目の安定性と精度の向上を図り、質問項目の妥当性を提示することが大規模継続的學生調査の課題である。そこで以下の節では、データを複数の検証方法によるデータの検討に焦点を当て、探索的研究を試みる<sup>4)</sup>。大規模継続調査が詰め重ねられてきている米国においては、データについての様々な検証方法を探索的に試みることで、カレッジ・インパクトに関する理論研究の積み重ねや新しい知見などを生み出してきている。加えて、新しい検

証方法の探索と開発も米国の大規模継続を通じて積極的に実践されている。翻って、日本では継続的な大量のデータの利用法としての分析手法の開発については、継続調査があまり実施されてこなかった状況を反映して、それほど主眼が置かれてこなかった。そこで、本稿では、継続調査という性格を伴う本調査のデータを用いて、近年異なる機関や異なる学生像を持つデータ解析に合致しているという手法を探索的に試みる<sup>5)</sup>。

## マルチレベル・モデル分析によるカレッジ・インパクトの検証

村澤は、近年社会科学系分野においても普及しつつあったマルチレベル・モデル分析を高等教育分野において適用を試み、その特徴を明らかにし、高等教育研究への適応可能性を提示した(2006)。その際、単層的データを想定した従来の分析方法が対応し得なかった階層構造を持つデータの分析に適している分析手法としてのマルチレベル・モデル分析の特質を明らかにしたが、その分析モデルについては村澤の論考あるいは詳細な解説については Kreft & Leeuw (1998), Raundenbush & Bryk (2002) を参照されたい。米国のカレッジ・インパクト研究においては、2010年にボルクワインとロビックが学生のエンゲージメントと機関の特性の関係性についてマルチレベル・モデル分析を用いて検証するなど本モデルによる分析の緒を開いている (Lovic & Volkweine, 2010)。

本稿では、JFS2008とJFS2009のデータを用いてマルチレベル・モデル分析を行う。JCIRP 研究では、例えば、これまで JCSS2005, JCSS2007, JFS2008を通じて、回帰分析による大学間比較や学系別比較を実践してきた。しかし、これらのデータは、大学や学部あるいは学系という集団から成り立っており、集団内には均質性や類似性がみられる。こうしたデータは、集団毎にネストされ、集団内で類似しているという階層性のあるデータである。これまで階層データを分析する際には、集団内のサンプルを独立性があることを前提として、集団毎に平均して集団数をサンプル数としてきたが、この方法ではサンプルは集団内で類似しているという性格から独立ではなく、平均値は純粋な集団の性質を反映していない。したがって、JCSS2005, JCSS2007, JFS2008で実施してきた大学間比較、学系別比較は、大学や学系毎の類似性を反映しないまま回帰分析を遂行してきたことになる。すなわち、通常回帰分析(単回帰分析、重回帰分析も含む)だと大学をひとくくりにした分析と大学別の分析結果は必ずしも一致するとは限らず、逆の結果を導くこともあり得ることから、大学間の違いと大学内(学生間)の違いを区別した分析を行うことが推奨される。また、通常回帰分析では、切片と傾きは単一の値しかとらないために、切片や傾きが学校間で異なるような散布図に対しては通常回帰分析は適切な分析方法ではないのに対し、マルチレベル・モデル分析では大学間で異なる切片や傾きを推定することが可能となる。つまり、変数間の関係を、1次抽出単位内(大学内、学生間)の違いと2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するモデルであるが、式で表せば以下ようになる。

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad \text{---①} \quad (i \text{ は個人})$$

$Y$  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

次に、各母数が集団ごとで違っていると仮定すると

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad \text{---②} \quad (j \text{ は集団})$$

となる。②がマルチレベルの基本式であるが、本式では切片  $\beta_{0j}$  と係数  $\beta_{1j}$  が「ある値を中心に大学ごとに分布している」とみなすことから、切片あるいは係数にランダム効果を設定したことになる。

これを数式であらわすと

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (u_{0j} \text{ は誤差項}) \quad \text{---③}$$

切片  $\beta_{0j}$  がある値  $\gamma_{00}$  を中心に個々の大学によって異なり、

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (u_{1j} \text{ は誤差項}) \quad \text{---④}$$

切片  $\beta_{1j}$  がある値  $\gamma_{10}$  を中心に個々の大学によって異なることを示している。

「効果のある学校」研究にマルチレベル・モデル分析を適用した川口（2009）は、学力生活実態調査データを用いて「学校の効果」について検証した。本調査は、国語・算数（数学）等の教科を対象として、CRT（教研式学力テスト）と生活実態調査から成り立っている。川口の研究では、学校の効果について小4時点の国語、算数のテスト得点と小5時点の国語、算数のテスト得点を効果測定の変数として、性別、通塾割合、社会的経済的変数を統制変数として扱った。地域（ある県のある市）の学校を横断する直接評価としてのテストと実態調査から明らかになる間接評価を組み合わせた好事例であるといえる。

筆者は、直接評価とプロセス評価としての間接評価結果の組み合わせによって、大学生活を通じての学生の成長に関する精緻な結果測定が可能であるという立場を取るが、日本の高等教育においては大学間の横断的な直接評価による先行研究はほぼないに等しい。村澤の先駆的なマルチレベル・モデル分析による研究においても、「大学教育への投資値頃感」と「教養教育の充実度」を被説明変数と説明変数として扱っている。学習指導要領によって学ぶ範囲が規定されている初等・中等教育では、測定の妥当性が担保されているのに対し、学ぶ範囲の合意が一律的に見出されていない高等教育においては横断的なテストを開発したとしても、そのテストが測定したいものを測定することが容易ではない高等教育の特性でもある<sup>6)</sup>。したがって、本稿においても、「大学での充実度」を被説明変数として設定し、山田が、「全体的な授業の質に対して満足している」が「大学での経験満足度」の最も大きな要素であることを決定木分析により明らかにしたことを勘案し（山田，2008）、「全体的な授業の質の満足度」を説明変数として関係を検討してみる。アスティンも I-E-O モデルのアウトプットあるいはアウトカムの指標として情緒的指標である満足度変数を提示している（Astin, 1993）。シューレ等は学士課程プログラムの満足度が学業へのモチベーションを高めるといふ知見を示すなど（Suhre, Jansen & Harskamp, 2007）、海外のカレッジ・インパクト研究においても満足度変数を被説明変数あるいは説明変数として利用していることも少なくないことから、本稿においてもこの2つの変数を使用する。

「大学での充実度」と「大学教育の質の満足度」にマルチレベル・モデル分析を適用すると切片にランダム効果を設定するランダム切片モデル、傾きにランダム効果を設定するランダム傾きモデル、切片と傾きの両方にランダム効果を設定するランダム切片+傾きモデルの3つのパターンが設定できる。特に、切片と傾きの両方にランダム効果を設定するランダム切片+傾きモデルが現実的

には、多様な高等教育機関や学生像を反映していると推察されるが、本稿では探索的にマルチレベル・モデル分析を試みるという理由により、切片にのみランダム効果を設定するランダム切片モデルを採用する。その場合、学生レベルは、大学での充実度  $ij = \text{切片 } j + \text{傾き} * \text{大学教育の質の満足度 } ij + \text{誤差 } ij$  となり、大学レベルでは、傾き  $j = (\text{傾きの}) \text{切片} + \text{誤差 } j$  となる。

そこで②の式

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (j \text{ は集団})$$

に③の式  $\beta_{0j} = \gamma_{00} + u_{0j}$  ( $u_{0j}$  は誤差項) を代入すると

$$Y_{ij} = (\gamma_{00} + u_{0j}) + \beta_{1j}X_{ij} + e_{ij} \quad (j \text{ は集団}) \quad \text{⑤になる。}$$

⑤の式は、 $u_{0j}$  は大学レベルの誤差項 (固有効果) 切片の平均  $\gamma_{00}$  からのずれを示していることから、たとえばA大学の学生はA大学以外の大学の学生に比して切片すなわち「大学での充実度」の平均  $\gamma_{00}$  から  $u_{0j}$  だけ異なる。ただし、回帰係数すなわち、充実度と質への満足度の関係はグループ間では変わらないと解釈できる。

## データの概要とモデルの設定

分析に用いる JFS2008と JFS2009の基本情報を表2に記す。JFS2008と JFS2009は、参加大学・学部数も参加人数も異なるが、本稿の目的のひとつに、多様な分析手法を用いて、カレッジ・インパクト理論を安定的に説明し、信頼性を高めることを掲げていることから、参加人数や規模の違い、参加大学の違いが分析結果にどう反映されているかを検証していくことが継続的データを収集する上では不可欠であると考える。

表2 JFS2008とJFS2009の基本情報

	専門分野	比率	偏差値分類	比率	設置別	比率	学年	比率	性別	比率
JFS2008 19,661名	人文系	12.6	39未満	19.0	国立	18.3	1年生	98.6	男性	53.9
	社会科学系	28.7	39~44	20.2	公立	6.2	2・3年生	1.3	女性	46.1
	理工農生系	25.5	44~49	20.7	私立	75.5	無回答・その他	0.1	合計	100.0
	医療系	12.2	49~52.5	20.2	合計	100.0	合計	100.0		
	教育系	9.0	52.5以上	20.0						
	家政系	2.5	合計	100.0						
	芸術系	4.1								
	その他	5.4								
	合計	100.0								
JFS2009 8,534名	人文系	18.5	39未満	33.0	国立	11.8	1年生	96.9	男性	51.7
	社会科学系	28.2	39~44	17.5	公立	4.6	2・3年生	3.0	女性	48.3
	理工農生系	16.6	44~49	18.5	私立	83.7	無回答・その他	0.1	合計	100.0
	医療系	9.2	49~52.5	20.6	合計	100.0	合計	100.0		
	教育系	9.1	52.5以上	10.4						
	家政系	4.4	合計	100.0						
	芸術系	5.3								
	その他	8.8								
	合計	100.0								

JFS2008には132大学から163学部が参加した。本調査は新入生を対象とした調査であるが、一部上級生が回答していたため、上級生を除くと対象者は19,332人となった。また、マルチレベル・モ

デル分析を実施する際の適切なデータ数については、様々な議論がある<sup>7)</sup>。本稿では第一次水準である学生数の最低数を20名として設定した結果、3大学44名を対象外とし、対象となる大学数は129大学、学生数は19,288名となった。JFS2009には46大学から56学部が参加し、上級生を除いた分析の対象者は8,229人となった。最低数20名を第一次水準の最低数と設定した結果、1大学13名を対象外とし、対象となる大学数は45大学、学生数は8,216名となった。JFS2008とJFS2009の差異として、JFS2009が偏差値分類において、より下方にシフトしていること、人文系の比率が高めであるのに対し、理工系、農・生物系、医学・医療系の比率が下がり、設置形態では国立が減少していることが挙げられる。

先述したように、「大学での充実度」と「大学教育の質全般の満足度」の関係を探ることを主な目的としているため、従属変数は「大学での充実度」と設定し、「大学教育の質全般の満足度」を説明変数としているが、JCIRP 先行研究からインプット要件である、「入試形態」、「大学への志望度」が学生の大学での適応や満足度に大きな影響を与えていたことから、今回の分析でも、入学後日の浅い新入生にも影響を与える可能性があると考え、変数としてこれらを組み入れている。

分析の際には、設置形態については、国立・公立=1、それ以外に0のダミー変数とし、入試形態については、一般入試・センター入試=1 推薦、AO等=0のダミー、志望度については、第一志望=1 それ以外=0のダミー変数とした。第一水準は学生個人であり、上位集団である第二水準は大学とした。学部を上位集団としてマルチレベル・モデル分析をすることも可能であるが、今回は学部が多様で分類することが複雑なため、学部を水準として扱わないことを記しておく<sup>8)</sup>。

マルチレベル・モデル分析では、説明変数を含まないモデル0をNullモデルと呼ぶが、このモデルを基本に、その後変数を付け加えたモデルと、逸脱度やAICの値の変化を見ながら、ICCを含めて適合度の高いモデルを選択する。モデル1は、教育への「大学教育の質全般の満足度」を加えたモデル、モデル2にはモデル1に設置形態ダミーを加え、モデル3には、志望ダミーと入試ダミーを加えた。

Yは大学での充実度を示す。

モデル0. Nullモデル

$$Y_{ij} = (\gamma_{00} + u_{0j}) + \varepsilon_{ij}$$

モデル1. ランダム切片+傾きモデル

$$Y_{ij} = (\gamma_{00} + u_{0j}) + \beta_{ij} \text{大学教育の質全般の満足度}_{ij} + \varepsilon_{ij}$$

モデル2. ランダム切片+傾きモデル

$$Y_{ij} = (\gamma_{00} + u_{0j}) + \beta_{ij} \text{大学教育の質全般の満足度}_{ij} + \gamma_{20} \text{設置形態ダミー}_{ij} + \varepsilon_{ij}$$

モデル3. ランダム切片+傾きモデル

$$Y_{ij} = (\varepsilon_{00} + u_{0j}) + \beta_{ij} \text{大学教育の質全般の満足度}_{ij} + \gamma_{20} \text{設置形態ダミー}_{ij} + \gamma_{30} \text{志望ダミー}_{ij} + \gamma_{40} \text{入試ダミー}_{ij} + \varepsilon_{ij}$$

## 分析結果

表3に分析結果を示している。逸脱度とAICの値が小さくなればなるほどモデルの適合度が高いことを意味するので、JFS2008とJFS2009ともにモデル3すなわち、「大学教育の質全般の満足度」、設置形態ダミー、志望ダミー、入試ダミー変数が加わったモデルの当てはまりが良いといえる。

JFS2008, JFS2009ともにNullモデルとモデル1の間で最も逸脱度, AICの値が減少していることから、大学教育の質全般の満足度が大学の充実度を左右する重要な変数であると解釈できる。JFS2008, JFS2009ともに投入した説明変数はすべて有意な結果を示していることから、「大学教育の質全般の満足度」、設置形態としては、「国立・公立大学」、志望度、入試方式としては、「一般入試・センター入試」が大学での充実度に効果的に働いているといえる。

ICCは、大学間の分散を学生間分散と大学間分散の合計である全体の分散で割った値であり、本モデルにどれだけ機関である大学が影響しているか、説明できるかを示す値であるが、JFS2008では、Nullモデルが4%、モデル3が3.4%である。JFS2009では、モデル3が5.4%とJFS2008よりもICCの値は高くなっている。いずれにしても、「大学での充実度」には機関である大学ごとの特性が影響を与えている率は低く、おそらく学生間の要因の方が大きいと推察できる。

NSSEの開発者であり実施責任者であるマコーミックは、NSSEのこれまでの研究蓄積から大学毎の要因で教育の効果を説明できる値は高く10%程度であり、残りは個別機関の学生要因であると論じているが(2009)、この主張とも本分析結果のICC値は若干低い、整合性がある。JFS2009のICC値が上昇しているが、JFS2009では、国立大学の参加率が低下し、私立大学の参加率が高まっていることから、機関の類似性が強くなっているとも考えられ、継続データで機関による特性のベンチマークを積み重ねて検討することが不可欠であろう。

表3 JFS2008, JFS2009のマルチレベル・モデル分析結果

	JFS2008 大学間				JFS2009 大学間			
	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				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***	3.124***	2.23***	2.23***	2.202***
大学教育の質全般の満足度		0.388***	0.391***	0.386***		0.274***	0.273***	0.269***
設置形態ダミー			0.168***	0.148***			0.04***	0.034***
志望ダミー				0.130***				0.06***
入試ダミー				0.047**				0.01**
学生間分散	1.162	1.045	0.149	0.139	0.556	0.49	0.108	0.106
大学間分散	0.048	0.037	0.005	0.005	0.015	0.01	0.006	0.006
共分散			-0.029	-0.027			-0.025	-0.025
ICC	4.00%	3.40%	3.20%	3.40%	2.60%	2.00%	5.30%	5.40%
逸脱度	58168	55355	55279	53488	18916	17106	17069	17058
AIC	58176	55363	55293	53506	18922	17114	17083	17076

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

## 誤差項に見られる大学ごとの差異の推定

式③で示した大学ごとの誤差項  $u_{0j}$  の値に注目して、切片を中心とした大学毎の充実度の値の差

異を見てみる。初期条件だけの Null モデル上での  $u_{0j}$  の値と初期条件に新たに変数を加えたモデル 3 上での  $u_{0j}$  の値を比べてみると大学毎の充実度の伸びがわかる。図1にはモデル3により充実度の高い上位10位までと下位10位までの大学の結果をグラフ化してみた。

上位10大学に注目すると、C 大学は見かけの充実度は最も高いが、初期条件を統制すると充実度は若干低くなること、初期条件を統制すると最も充実度が高くなっている、すなわち  $u_{0j}$  という誤差項の変化により、その変化が明確に何かと説明することはできないものの何らかの大学での特性や影響などが高いのは、D 大学であることがわかる。また、下位集団に目を向けると見かけも初期条件を統制しても充実度は低い、それでも何らかの大学の影響による効果によって上昇しているのが T 大と U 大であることが示されている。もちろん、前述したように、 $u_{0j}$  は学生か機関の差か明確に説明できない目に見えない誤差項のため、一概にカレッジ・インパクトとは言い切れないが、あえて、こうした説明しきれない部分を何らかの大学の影響からもたらされる効果としてみると、こうした効果を示している大学内での環境はどのようなになっているのだろうか。JCIRP は各機関がベンチマークをできるようにという視点から設計されていることから、探索的に本結果を利用してベンチマークを試みる。ここでは、上位5大学 (A ~ E) にはいかなる共通性や差異があるのかに着目してデータを比較する。

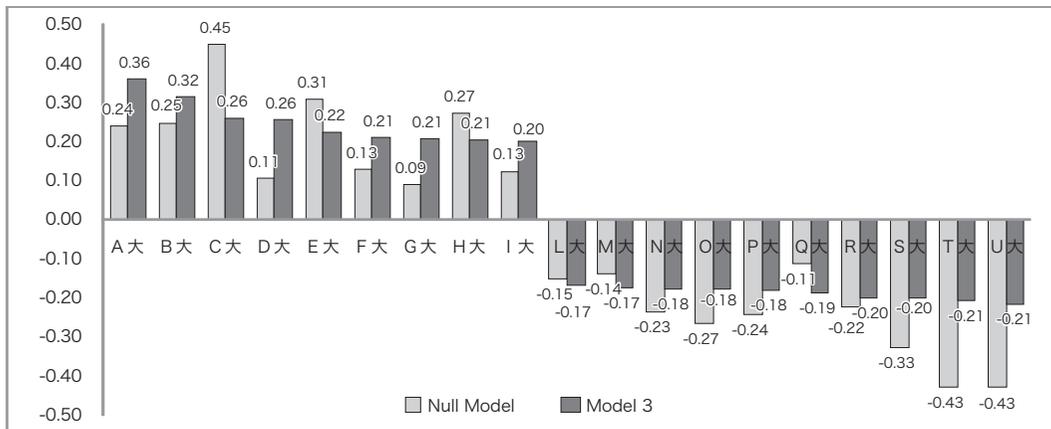


図1 モデル3における上位10大学と下位10大学の $u_{0j}$ の値

新入生対象の調査であることから、「大学志望理由」と「満足度」に注目して5大学の共通性を見てみる。まず、5大学は全て私立大学であり、複数参加学部を考慮した偏差値の分布は、39未満17%、44~49 41.6%、49~52.5 29.1%、52.5以上12.3%と分布はそれほど偏りが無い。専門分野は社会科学(社会福祉)系29.1%、医学・医療・薬学24%、教育・教員養成17%、その他(スポーツ科学)29.9%である。

下の表4は5大学の学生の志望動機の高い項目について上位4位までの各大学の比率を提示している。D 大学はマルチレベル・モデル分析での条件統制後の固定効果(誤差項)が高いが、志望動機項目については、「オープンキャンパスに参加した」という項目以外は、重要と答えている割合が

低い。しかし、オープンキャンパスに参加したことが志望度を高めていることに効果を示しているようだ。一方、B大学とC大学は4つの項目の回答比率に類似したパターンが見受けられることから、「学術」「社会活動」「卒業生の就職」に注力しており、それが一定の広報として効果をあらわしていると推察される。

次に満足度にどのような特徴がみられるのかを見てみよう<sup>9)</sup>。満足度の平均値の高い上位5項目を表5に示した。表には、設備に関する満足度と学生同士の交流関連の満足度が高いことが示されている。入学後それほど月日が経っていない状況で実施されていることから、教員との関係や個々の

表4 大学の志望動機上位項目

社会活動に対する評価が高い	まったく重要でない	あまり重要でない	ある程度重要	とても重要	合計%
C大学	13.0	28.7	40.1	18.2	100
A大学	41.5	29.6	24.6	4.2	100
E大学	21.0	24.0	39.0	16.0	100
D大学	40.2	41.2	17.5	1.0	100
B大学	8.3	19.8	45.5	26.4	100
学術レベルに対する評価が高い					
C大学	10.5	19.8	48.3	21.4	100
A大学	46.5	29.6	20.4	3.5	100
E大学	14.7	18.6	47.1	19.6	100
D大学	37.1	38.1	23.7	1.1	100
B大学	14.4	23.1	46.5	16.0	100
卒業生の就職が良い					
C大学	15.3	10.8	45.4	28.5	100
A大学	57.4	17.0	16.4	9.2	100
E大学	58.0	10.0	27.0	5.0	100
D大学	62.8	18.6	12.4	6.2	100
B大学	14.2	13.8	43.5	28.5	100
オープンキャンパスに参加した					
C大学	41.1	11.3	26.2	21.4	100
A大学	56.7	10.6	20.6	12.1	100
E大学	83.7	7.1	8.2	1.0	100
D大学	58.8	11.3	18.6	11.3	100
B大学	31.4	12.4	33.5	22.7	100

表5 満足度上位項目

コンピュータの施設や設備	大学名	とても不満	不満	どちらでもない	満足	とても満足	合計%
C大学		3.2	11.7	32.9	34.9	17.3	100
A大学		3.5	3.6	21.3	45.4	26.2	100
E大学		4.0	7.9	34.6	41.6	11.9	100
D大学		3.1	3.1	39.6	40.6	13.6	100
B大学		1.7	11.3	23.1	45.8	18.1	100
図書館の設備							
C大学		2.0	6.4	28.9	47.0	15.7	100
A大学		9.9	12.7	28.2	37.9	11.3	100
E大学		1.0	6.9	27.7	50.5	13.9	100
D大学		2.2	6.2	45.8	35.4	10.4	100
B大学		2.1	8.4	21.3	49.0	19.2	100
他の学生と話をする機会							
C大学		2.4	8.0	32.0	38.4	19.2	100
A大学		7.0	11.3	36.6	18.2	16.9	90
E大学		2.0	6.0	38.0	32.0	22.0	100
D大学		2.1	18.6	41.2	30.9	7.2	100
B大学		3.8	10.0	44.1	34.6	7.5	100
実験室の設備や器具							
C大学		0.8	2.4	42.8	35.6	18.4	100
A大学		5.7	4.3	67.3	17.0	5.7	100
E大学		4.0	6.0	38.0	46.0	6.0	100
D大学		0.0	0.0	10.4	54.2	35.4	100
B大学		3.0	7.6	72.5	13.9	3.0	100
学生交流の機会							
C大学		1.6	10.0	29.6	33.6	25.2	100
A大学		4.3	10.6	34.8	34.0	16.3	100
E大学		2.0	6.0	39.0	36.0	17.0	100
D大学		1.0	16.5	59.8	19.6	3.1	100
B大学		3.8	10.0	53.7	26.7	5.8	100

授業の内容よりは、設備関連や学生間での親睦が深まる機会が提供されていることが重視されていると推察できる。今回のマルチレベル・モデル分析により、カレッジ・インパクトとは言い切れないものの大学の影響が強い大学での学生の分野が医学・医療・薬学、社会福祉、スポーツ科学が多いことから、教育と深い関連性がある施設・設備の充実が満足度に影響を及ぼしていると解釈できよう。

## 本研究のインプリケーションと課題

2008年の中教審答申『学士課程教育の構築に向けて』では、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させることが強く求められ、より具体的な学士課程共通の学習成果に関する参考指針として、実際に、各専攻分野を通じて培う「学士力」が提示されたのも新しい動向である。2008年の答申を高等教育の今後の方向性のある程度規定づけるとすれば、高等教育の教育成果や効果の提示への圧力が強くなっているとみなすことができよう。高等教育の成果や効果測定の方法についての見解は統一されているわけではなく、特に直接評価によって成果を測定するには、国家資格と直結した医学・医療・看護等に代表される領域以外では容易ではない。本稿では間接評価としての学生調査の教育成果や効果を測定する上での可能性と課題について論じてきたが、その際の学生調査の特性は大規模継続調査である。カレッジ・インパクト理論を検証するには、普遍性を担保するデータの安定性と方法の信頼性を提示しなければならない。とりわけ、大規模継続調査には、異なる機関や学生像からなる回答者が参加することになる。機関の特性や学生個々の差異を超えてデータの性格が安定していることを提示する方法のひとつとして、今回探索的マルチレベル・モデル分析を試みた。マルチレベル・モデル分析を通じて、参加大学や学生の特性が異なっていたとしても「大学教育の質全般の満足度」、「設置形態」、「志望度」、「入試形態」を組み入れたモデルが、新入生という入口の時点での学生の「大学での充実度」に有効であり、本モデルは異なる機関や学生像から構成されているJFS2008、JFS2009両方に適合していることが判明した。ただし、ランダム切片モデルのみの使用という点で限定的であることは否めない。

一方ICCというこのモデルを機関が説明する比率は3.4%~5.4%と決して高くはない<sup>10)</sup>。むしろ、学生の差異によって説明できる可能性が高いとも捉えられる。しかし、アメリカでのNSSEを始めとする継続的大規模学生調査においても、多様な統計的手法を用いても大学機関が説明する比率は高く10%程度であるとされ(McCormik, 2009)、日本の機関の説明力が特に低いというわけでもなさそうであるが、いずれにしても、大学間によるベンチマークの意味がないと捉えるべきなのだろうか。筆者はそうではないと考える。大学ごとの効果結果が示しているように、充実にむけての効果を示している大学には共通点が見出された。本稿では紙面の都合から効果の高い大学の教育環境や学生の特徴の詳細分析を実施していないが、大学間だけでなく、大学内での分析を通じて高等教育全体もしくは個々の高等教育機関の教育改善の土台にもなる。換言すれば、共通性や差異をもたらす要因や環境を詳細に研究者や個々の機関が検討することで、カレッジ・インパクト研究の普遍性の検証のみならず、学生の認知的・情緒的成長をもたらす環境と機関特性との関係性の解明と

効果的な教育環境の形成のための機関ベンチマークの可能性を拓くことにもつながる。

本研究は、しかしながら、大学教育の成果や効果を間接評価として厳密に測定しているとはいえない。新入生調査という言葉が象徴的に示しているように、入学後の月日がそれほど経過していない学生が参加しているために、いわゆる高校時代の既得要因の影響は強いと推察できる。新入生調査をインプットとして大学生調査と合わせて縦断的に成果を検証しなければならない。更に、機関間の横断的検証、学年進行による縦断的検証、異なる学年間での縦断的検証など継続的大規模調査により可能な研究に着手することが次のステップとして不可欠である。

## 【注】

- 1) 2010年にも JCSS と JCSS を実施しており、現時点で回収中である。
- 2) MAPP も一般教育の成果測定のために開発されたテストである。
- 3) アスティンが開発した CIRP やインディアナ大学が中心となって開発した NSSE (National Survey of Student Engagement) は、プロセス評価としてあるいはベースラインデータとして利用するという意図で、設計や項目の精査も並行して行われてきた。CIRP が依拠している理論はアスティンの Involvement であり、NSSE はクールの Engagement 理論である。クールが開発にかかわっている NSSE は Engagement という理論が質問全体の基本思想となっている。
- 4) 質問項目の安定性と妥当性を向上するためには、木村が項目反応理論 (IRT) を用いて、質問項目の精査を実施している。
- 5) 古田は JCIRP のなかの JCSS2007 のデータを用いて学生の学習成果を測定するために、多重対応分析の手法を試みている。
- 6) 職業資格に直結した医学、看護学等の領域では一定の学ぶ範囲や到達目標の合意がなされていることを付加したい。医学分野の卒業生を対象とした標準試験は、学ぶ範囲についての妥当性の検証や試験の信頼性の検証が行われたうえで、TOEIC や TOEFL のようにコンピュータ上での試験の受験が可能になっている。ただし、そのためには予備項目も含めて約60,000題が準備されている。残念ながら、日本の一般教育においては学ぶ範囲の合意について、すなわち妥当性の研究がほとんどないことから、標準試験はないに等しい。
- 7) 村澤は様々な適切なデータ数についての議論を提示したうえで、一集団あたりのサンプルが十分に多く、集団の数自体も比較的多く、さらに欠損値がほとんど無いか皆無であるデータが適切であるとしている。
- 8) 参加大学の基本情報として専門分野も分類しているが、多様な学部を専門分野として分類しており、その結果は学生が所属している学部とは必ずしも一致していない。
- 9) マルチレベル・モデル分析に投入した「大学教育の質全般の満足度」変数は本満足度項目の平均値比較には組み入れていない。
- 10) 学部の特性が結果には反映されていると推察できるが、今回の分析では機関別に主眼を置いたため学部を水準としては扱っていないが、次回は学部を水準として扱いたい。

## 【参考文献】

- 小方直幸 (2008) 「学生のエンゲージメントと大学教育のアウトカム」『高等教育研究』第11集, 45-64頁。
- 川口俊明 (2009) 「マルチレベルモデルを用いた「学校の効果」の分析—「効果的な学校」に社会的な不平等の救済はできるのか—」『教育社会学研究』第84集, 165-182頁。
- 木村拓也・西郡大・山田礼子 (2009) 「高大接続情報を踏まえた「大学教育効果」の測定—潜在クラス分析を用いた追跡調査モデルの提案—」『高等教育研究』第12集, 189-214頁。
- 杉谷祐美子 (2007) 「大学内における教育効果の学生間比較」『転換期の高等教育における学生の教育評価の開発に関する国際比較研究』平成年度科学研究費補助金基盤研究 (B) 研究成果報告書, 49-72頁。
- 日本テスト学会編 (2010) 『見直そう, テストを支える基本の技術と教育』金子書房。
- 村澤昌崇 (2006) 「高等教育研究における計量分析手法の応用 (その1)」『大学論集』第37集, 309-327頁。
- 山田礼子 (2007) 『転換期の高等教育における学生の教育評価の開発に関する国際比較研究』平成16-18年度科学研究費補助金基盤研究 (B) 研究成果報告書。
- 山田礼子 (2009) 「学生の情緒的側面の充実と教育成果—CSS と JCSS 結果分析から—」『大学論集』第40集, 181-198頁。
- 山田礼子・木村拓也・古田和久・吉田文・杉谷祐美子 (2009) 「JCIRP に見る大学生の諸相—プロジェクト型大学生調査の目的・方法・課題—」『日本教育社会学会第61回大会要旨集』, 早稲田大学, 285-290頁。
- Anaya, G. (1999). College Impact on Student Learning: Comparing the Use of Self-Reported Gains, Standardized Test Scores, and College Grades. *Research in Higher Education*, 40(5), 499-526.
- Astin, A. W. (1993). *What Matters in College?: Four Critical Years Revisited*. San Francisco, Calif: Jossey-Bass.
- Banta, T. W. (Ed.) (2004). *Hallmarks of Effective Outcomes Assessment*. San Francisco, Calif: Jossey-Bass. A Wiley Company.
- Borden, V. M. & Young, J. M. (2008). Measurement Validity and Accountability for Student Learning. *New Directions for Institutional Research, Assessment Supplement 2007*. Jossey-Bass, 19-37.
- Chen, P. D., Gonyea, R. M., Sarraf, S. A., Brckalorenz, A., Korkmaz, A., Lambert A, D., Shoup, R. & Williams, J. M. (2009). Analyzing and Interpreting NSSE Data. *New Directions for Institutional Research, Survey Research Emerging Issues*, 141, 35-54.
- Ewell, P. T. (2008). Assessment and Accountability in America Today: Background and Context. *New Directions for Institutional Research, Assessment Supplement 2007*, 7-17.
- Gonyea, R. M. (2005). Self-Reported Data in Institutional Research: Review and Recommendations. *New Directions for Institutional Research, Survey Research Emerging Issues*, 127, 73-89.

- Impara, J. C. & Plake, B. S. (Eds.) (1998). *The Thirteenth Mental Measurements Yearbook*. Lincoln, Nebraska: The Buros Institute of Mental Measurements, the University of Nebraska-Lincoln.
- Kreft, I. & Leeuw, J. D. (1998). *Introducing Multilevel Modeling*. Sage (小野寺孝義訳 (2006), 『基礎から学ぶマルチレベルモデル—入り組んだ文脈から新たな理論を創出するための統計手法—』ナカニシヤ出版).
- Lovik, E.G. & Volkwein, J.F. (2010). The Effects of Institutional Characteristics and Student Engagement on College Student Spiritual Development. Paper presented at AIR Forum 2010, Chicago, Ill.
- McCormik, A. C. (2009). Toward a Nuanced View of Institutional Quality. *NSSE Promoting Engagement for all Students: The Imperative to Look Within 2008 Results*. 6-8.
- Palomba, C. A. & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. San Francisco, Calif: Jossey-Bass. A Wiley Imprint.
- Pascarella, E. T. & Terenzini, P. T. (2005). *How College Affects Students*. San Francisco, Calif: Jossey-Bass. A Wiley Company.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Second Edition, London: Sage Publications.
- Rhodes, Tl. L. (2008). VALUE: Valid Assessment of Learning I Undergraduate Education. *New Directions for Institutional Research, Assessment Supplement 2007*, 59-70.
- Shavelson, R. J. (2010). *Measuring College Learning Responsibly: Accountability in a New Era*. San Francisco, Calif: Stanford University Press.
- Suhre, C.J.M., Jansen, E.P.W.A. & Harskamp, E.G. (2007). Impact of Degree Program Satisfaction on the Persistence of College Students. *Higher Education*, 54, 207-226.

## The Possibility and Challenges of a Large Scale Continuous Student Survey System: through analysis of JCIRP data

Reiko YAMADA \*

Universities and colleges in Japan have experienced sweeping and sudden changes during the past ten years. Many factors have contributed to the changes: universalization of HE, obligatory accreditation systems, and so on. For example, “quality assurance” has rushed us into reform of the curriculum, program, and pedagogy. Many universities and colleges have now put an emphasis on teaching and learning rather than research. So, what is happening in Japanese higher education and what are the impacts of the emergence of the accountability issue and the progression of massification? One of the new trends is exemplified as the rising movement for learning outcome assessment. However, in Japan, we have little historical background to assess learning outcomes and students’ growth through standardized tests and student self-reported surveys.

Our research group has developed a student surveys systems for learning improvement, the so-called JCIRP (Japanese Cooperative Institutional Research Program) since 2004. Our program consists of three student surveys, the Japanese Freshman Survey (JFS), the Japanese College Student Survey (JCSS), and the Japanese Junior College Student Survey (JJCSS). Up to 2010, more than 54,000 students from 310 four-year and two-year colleges have participated in our program. It seems that our student survey system has institutionalized a culture for assessment of student learning outcomes based on the self-reported data that has been widely accepted in Japanese higher education institutions as an indirect assessment system.

The learning outcomes issue can be approached through diversified perspectives. Learning outcomes can be examined as knowledge acquired through university education. It can also be examined by measurement of outcomes assessment tools. Definition of learning outcomes is wide and complicated. It covers the field of general education, academic disciplines and the method of assessment. In the United States, various tests have been developed in order to assess the learning outcomes of general education, of the academic disciplines and of high-school learning. However, assessment toward students’ learning is not well developed in Japan.

In the United States, the Spellings’ Commission on the Future of Higher Education issued a report that considered introduction of national tests for assessment. However, as is mentioned by Banta, the context for considering a national test for college students was actually much fuller in 1990 than in 2006, and approaches for designing the test was more careful and deliberate (Banta, 2007). Various arguments about the assessment issue have been developed and research on assessment has accumulated. In other words, many researchers have engaged in measurement, testing and assessment study and thus have developed various tools for

---

\* Professor, Graduate School of Social Studies, Doshisha University

assessing cognitive skills, learning outcomes and the affective aspects of students.

In contrast, there is little research on assessment in Japan. Thus, it is not easy to define what assessment is and what the assessment of students' learning is in Japanese higher education. However, it seems that there are two kinds of outcome assessment. These are divided into direct assessment and indirect assessment. Direct assessment is assessment of the direct learning outcome of students through tests, paper, portfolio, graduation examination, graduation research paper or standardized tests in the fields of general and discipline-based education. Indirect assessment can be utilized to assess the learning process through student surveys asking about learning behavior, student experience, self-perception and satisfaction. These surveys are practiced when students are entered the college, at the end of the first-year, at senior-year level and after graduation. Direct assessment and indirect assessment can be used to complement each other. Thus, our JCIRP is expected to function as the indirect assessment for student learning outcomes.

This paper first demonstrates how we have tackled the development of student survey systems in Japan, what problems are to be focused on in terms of assessment of learning outcomes and what we have to do in the next stage for development of our survey system. Second, by using the JFS2008 data, the method used to explore the complex structure of college student analysis is indicated, that is, by analysis of JFS2008 data by a multi-level model. Multi-level model analysis is identified as a mixed-effects model or Hierarchical Linear Model (HLM). This model can be applied to nested or hierarchical data in which individuals belong to various types of groups. In this case, college students belong to different gender groups, different school year groups, and various colleges at different levels, and so on. Using multi-level model analysis, the study shows that college student satisfaction is caused by personal factors rather than college factors.