

授業方法の効果測定に関する方法論的検討

上 笹 恒 *

1. はじめに

この小論の目的は、教師が授業に関して設定する目標に、生徒たちがどの程度達成したか、生徒たちにどんな心理的インパクトを与えることができたか等、授業の与えた諸効果についてそれをどのように把握したらよいのかの方法論について検討することである。

授業の目標を設定するといっても、“社会認識を通して公民的資質を育成することをめざす”といった最も広い意味にとらえた抽象的目標設定から、社会科の教科の特定分野における極めて具体的な知識量の増大を目的とするといった個別的、具体的な目標設定までさまざまなレベルが考えられる。また、実際には授業の効果は単一のものではなく、質的に異なったかなり多面的なものであることが一般的であろう。

いずれにしても、その目標の設定自体は教師の主體的決定に委ねられるべきであり、その目標に到達するために教師はさまざまな授業方法を考案することになる。そして、考案されたいくつかの授業のやり方の中で、どんな方法がいろいろな評価視点からみて最も効果的であったかを知りたいという場面を想定して、本論ではその効果の比較を効率よく科学的に行う方法論について言及しようというのである。

実際には、授業の効果とは何かということは内容的に明確に表現することがあまり簡単ではなく、うまく公式化できないこともあるかもしれない。しかし、ここでは、何らかの観測・測定を伴うものを扱うので、さまざまな検討を経てある種の操作的な定義を必要とする。すなわち、一般的な社会科学の科学論に準拠するとすれば、測定の対象は実体的なものではなく抽象的概念ということになる。その場合、この評価の対象となる構成概念の妥当性を検証しなければならないという大変やっかいな作業が伴う。

しかしながら、ここではそのような検討は別の機会にするとして、次のようなことがらが授業の影響で何らかの変化を示した場合に何らかの効果があったと規定しよう。それらは、情報、知識、学力、諸概念等の獲得、いろいろな社会的対象にたいする認知、イメージ、態度における変容等であり、さらにいえば、ある分野についての理解度を高め、応用力を身につけ、問題解決能力や自己決定能力を身につけるといった、生徒たちのトータルな評価を必要とするものである。従って、最終的には価値観、世界観の変化を測定しなければならないという大きな問題に至る。これらは全て全体的に体制化された構造をもったものであるので多次元の把握が必要となる(ただし、この多次元の構造の中から、特定の一次元の構造を抽出してそこに焦点を当てた見方も可能である)。

こうして考えてくると、授業の目標設定もむずかしい問題となり、それへの生徒たちの到達度やその他の心理的インパクトの把握を行うための測定は一層むずかしい問題になってしまう。従って、このような授業効果の分析の研究には出発点から、あとの測定やデータの解析のプロセ

* 筑波大学哲学・思想学系

スを全体的に事前に考慮した研究のためのデザインが要求される。

2. 授業の効果をとらえる方法

(1) データをどのようにとるか

ここでは一応定量的に扱うデータを想定して話をすすめよう（実際には、最近の統計的データ解析学の発展により、定性的データも十分扱えるようになってきているが）。

最も単純でなじみのあるものは、知識度を測るテストや学力テストなどである。たとえば、あるクラスで一定の授業を受ける前のテストでの正答率を $P_1(\%)$ とし、一定の授業を受けた後の同じテストの正答率を $P_2(\%)$ とすると、その授業の効果は $(P_2 - P_1) / (100 - P_1) \times 100$ とし、算出され、有効度指数(effective index)といわれ、従来から学習の進歩の程度を表すメジャーとされていた（文献〔1〕参照）。同じテスト問題を、前のテストの影響を受けない程度で適当な期間をあけて再度施して、授業のよい効果があれば2度目のテストスコアの方が全体的に高くなることが予想されるということである。この場合は、上記例のようにクラス全体としての正答率といった数値でもよいが、一般には100点満点のスコアが個々人に付与され、クラス全体としてみれば平均値を算出することもできる。

ところが、一方では、このようにはじめからスコアとして扱えない分野も多い。たとえば、何らかの社会的対象にたいするイメージや態度などがその典型である。この場合には、一般に、何らかの言語表現を媒介させないことには測定することができない。そこで、よく用いられるのが尺度(スケール)である。方法としては社会調査などでよく用いられる質問紙調査法がある。

そこで、研究対象として授業の影響を測りたいイメージや態度について、できるだけ事前の情報に基づいて構成された当該の態度領域の内容を表す態度ステートメントをたくさん作成する。イメージについては、関連領域に適合する形容詞などを集める。これらからさまざまなタイプのスケールを構成することができる。この場合、スケールは一次元(unidimensional)と多次元(multidimensional)のものが考えられるが、おそらく現場からの要請で考えれば多次元的スケールの方が一般的であろう。スケールの構成は、アプリアリに評定法スタイルで間隔尺度を仮定した数値を与える場合と、ある方法でデータ解析したあとで、スコアとして与える場合と2通りある。

(2) 測定の信頼性、妥当性

これらのスケールのうち特に態度スケールについては測定の①信頼性と②妥当性の検証が要求される。

①信頼性(reliability)というのは、測定の安定性、一貫性を表わす。これは、物理的な測定、たとえばものさしで何かの長さを測定しても、測る都度必ずしもぴったり一致はしないが、何回か測ったとすれば、ほゞある一定の値の近くに値が集中して、一定の小さい値のばらつきで分布が形成される。このばらつきが小さい方が信頼性が高い。態度ステートメントなどでは、特に内容的にむずかしい表現のものは繰り返しの測定にたいしてばらつきが大きくなる。もっとも、心理的な現象を扱った場合は、履歴効果があるので現実には反復測定を行うことができない。そこ

で、信頼性の評価には種々工夫がなされている。

もし、履歴効果が入らないような測定が考えられるならば、(i) 再テスト法 (test-retest method) といわれる方法で得られる、一定期間を置いて同一個人について2回測定されたスコアの相関係数をもって信頼性係数が算出される。

しかし、反復測定のできない内容の場合には、(ii) 代替形式法 (alternative-form method) が用いられる。これは、全く同じ測定項目が2回使われるのではなく、測定内容に関してほぼ等しいと考えられる2組の測定項目群を意図的に作成し、一定の期間を置いて同一個人にテストを行ない、この2つのスコアの相関から信頼係数を推定する。

上記の(i)と(ii)では、同一集団に2回のテストを実施する必要があるが、これと対照的に、(iii) 折半法 (split-halves method) では次のようにして1回の測定で処理することができる。すなわち、全項目を半分に分けて2種類の代替形式テスト (これを平行テストともいう) を同時に施して、その相関から信頼係数をある方式に従って推定するやり方である。

しかし、全項目を2組に分ける分け方は、項目数が増加すると組合せ計算から算出されるように、大変大きな数になる。そこで、(iv) Cronbachの α 係数という一回の測定でそのテストの信頼性が推定できる内的整合性 (internal consistency) の測度が考案された。そして、後にこの係数のもつ(ii)、(iii)の方法との関連が数学的に明らかにされ、今日ではこのメジャーが信頼性係数の推定値として最もよく用いられている。

②妥当性 (validity) というのは、ミクロにみた態度スケールなどの妥当性で、いくつかの態度スケールである社会的対象にたいする態度といった実体のない抽象的構成概念が本当に測定されているかどうかを判断するという大変むずかしい問題である (妥当性に関しては、測定項目に関するミクロな見方だけではなく、研究全体に関する仮説、理論を含めて大きく見た研究全体の妥当性も当然要求されるが、そのことにはここでは言及しない)。

心理学のテストや態度測定などの分野でよく扱われる妥当性として、(i) 基準関連妥当性が、まず、挙げられる。これは、当面のところ目的にしているある態度ドメイン (領域) と密接に関連するその他の態度ドメインや、その他の何らかの意味で基準になり得るものとの整合的関連の強さで表現される。たとえばテストでいえば、大学の入学試験のスコアが高校3年生の大学入学後の成績と強い相関をもっていて、適確な予測ができるかどうかで評価される。

この場合の大学入学後の成績のように、基準になるものが、本来の測定である入学試験時からみて時間的にあとになる場合を予測的妥当性という。また、基準が当該の測定と同じ時点で相関させられる場合には、同時的妥当性として評価される。

次は、(ii) 内容的妥当性 (content validity) である。これは、基本的には、経験的な測定が当面のところ対象とする、たとえばある態度についての内容の領域をどの程度反映しているかに依存している。研究者は先行研究のレビューやプリテストなどによって、この領域全体を把握していることが必要となる。すなわち、事前にある程度仮説された次元が内容領域を適切に反映しているかどうか依存している。たとえば、算数の計算力をテストしたい時、計算テストが加算問題しか含んでおらず、乗・除・減算の問題が入っていない場合は、明らかにこのテストは内容的妥当性を持たないことになる。場合によっては尺度の因子分析によって結果的に確認される

こともある。

さらに、(iii) 構成概念妥当性 (construct validity) がある。例を挙げると、たとえば Rosenberg の自尊心尺度によって、自尊心という構成概念の妥当性を検証しようとする時、Rosenberg は理論的に生徒たちの自尊心の水準が学校活動への参加と正の相関があるとした。これら2つの測定の相関が正で、ある程度大きいことがこのチェックとなる。

この一例のように、構成概念妥当性は、測定される諸概念に関して理論的に導出される仮説と整合的に合致する他の測定とどの程度関連しているかという観点からチェックされる。

まずはじめに、諸概念間の関係を理論的に検討し、次に諸概念の測定間の経験的關係を吟味し、データが示すそれらの結果から特定の測定の構成概念妥当性をどのように明らかにしているかを検討する、というプロセスを経る。これは厳密に考えるとなかなかむずかしい問題である。

妥当性の問題をもっと系統だった方式で考える時、参考になるのは Campbell と Fiske の多重特性一多重法 (multitrait-multimethod) によるアプローチである。もし、測度が妥当なものであれば、ある抽象概念をたとえ異なった方法で測定したとしても、その結果はよく類似したものになるはずだという論理である。この方法は、研究全体の妥当性評価に関わる根本的な部分をおさえているが、実例がないとわかりにくいので、文献 [2], [3] などを参照されたい。

また、最近ではもっと直接的に、構成概念を用いた原因—結果モデルを、共分散構造分析という統計解析法を用いて確認する方法も開発されているので、入門編として文献 [4] を参照するとよいであろう。

(3) 実験計画法の考え方

すでに記述してきたように、授業の効果としてどのような側面を捉えたいかを決定し、その概念的公式化を測定の変数のレベルまで分解していく操作化のプロセスを経て、実際の観測を行うことになるのであるが、その際事前に準備しなければならない実験計画法について次に取り上げよう。それは、手続きの一般論としていえば、ある授業を受ける前の生徒たちの状況をまず最初に測定しておき、授業を実施した後で再び同じ測定を繰り返すことによって事前—事後の比較を行ない、その変化から効果を抽出することである。

しかしながら、この時に考慮しなければならない諸要因が多く絡むので、事はあまり簡単ではない。そのために、以下に記すような各種の実験計画法が考案されてきている。

① *ex post facto* research

この方法は、*ex post facto* というラテン語の語源が示すように、事後において過去に遡って現象を研究する方法をとる。たとえば、いくつかの授業のやり方、学習法のトータルな効果として異なったいくつかの特性を持ったグループが結果的に発生したとしよう。その時、これらグループの差異をもたらした要因を過去に遡って (retrospectively) 追究し、原因—結果の関係を推測しようとする。この場合は事前の測定は全く行われていない。ある集団で、ある実験的操作を受けた事後にのみ測定が行なわれ、その結果にのみ基づいて何が差異をもたらしたのか源泉を探っていくのである。この方法は、実験的に状況を操作できない場合に適用される。

従って、この方法は、研究主体が被験者をランダム化したり、操作 (独立) 変数を統制する

ことができない、原因—結果の関係を同定、確認できない、いろいろな要因が絡んで来てしまう、といった制約をもつ。

② The One Group Pretest-Post-test Design

今ある一つの新しい教授法の効果をみるために、事前にある一つの集団に、たとえば何かに対する態度を従属変数 (O_1 と表す) として測定したとする。そのあと実験的操作 (独立変数として X と表す、すなわち、ある方法による授業など) を施して、再びその集団の態度 (O_2) を測定する。そして、事前と事後の比較を行う。このデザインを図示すると次のようになる。

グループ	事 前	実験的操作	事 後
(Experimental)	O_1	X	O_2

この ($O_1 - O_2$) の差が、もし何らかの外生的変数、研究主題と関係ない変数の影響などがないとすれば、実験的操作の効果を表現するはずである。

③ The Pretest-post-test Control Group Design

②で記述した方法にはさまざまな統制されない要因が入って純粋な授業の効果だけを捉えられない可能性があるので、さらに、次のようなデザインを考える。

グループ	事 前	実験的操作	事 後
(Experimental)	RO_1	X	O_2
(Control)	RO_3		O_4

①と異なるのは無作為化(randomization) によって、等質に構成された2つの集団 (実験群と統制群) を用いる点である。上図の中の記号Rはランダムにわりあてられたグループを意味し、これらに O_1 と O_3 という同じ事前の測定を行うという意味である。そして、この場合、 ($O_1 - O_2$) の差と ($O_3 - O_4$) の差の差が全体的差異となる。これは、古典的実験デザインともいわれる。実験群と統制群が全くランダムにわりあてられていたとすれば、 O_1 と O_3 は殆ど等しく、 O_2 と O_4 の差が X のおかげであるということになる。

しかし、厳密に言えば事前の測定 (O_1, O_3) によって受ける事後への影響が入ってしまう。そこで、このことを考慮した次のデザインが考案された。

④ Solomon's Three-Group Design

グループ	事 前	実験的操作	事 後
(Experimental)	RO_1	X	O_2
(1st Control)	RO_3		O_4
(2nd Control)		X	O_5

これによって、 O_2 と O_4 の単純な比較から X の効果を推定するだけでなく、プリテストがポストテスト状況に何らかの有意な影響を及ぼしたかどうかをチェックすることも可能になる。たとえば、もし、2nd Control Group の平均値 O_5 の値が 1st Control Group の O_4 の値より大きかったとすれば、プリテストそれ自身は実験群の被験者には何ら影響を与えていないと言ってよいであろう。また、

$$(O_2 - O_1) - [(O_4 - O_3) + \{O_5 - (O_1 + O_3) / 2\}]$$

のようにして、最初の測定と実験的操作の交互作用を検出することもできる。

⑤ Solomon's Four-Group Design

上記デザインについて、さらに第4のグループを追加しないと統制できない要因の効果を同定するために、次のような Solomon の4グループデザインが考えられた。

グループ	事前	実験的操作	事後
(Experimental)	RO_1	X	O_2
(1st Control)	RO_3		O_4
(2nd Control)		X	O_5
(3rd Control)			O_6

事前に関しては、4つのグループとも当然等質であると仮定されるが、事前の測定は Experimental と 1st Control のグループだけに行われる。附加されたグループについても事前の測定はない上に、実験的操作も受けず、ただ事後の測定だけが行われる。この場合、 O_1 と O_3 の平均値で代表される事前の推定値と、事後の O_6 の差が統計的に有意であれば、統制されなかった何らかの変数の効果が混入したことになる。これは、最初の測定、実験的操作、それらの間の交互作用のいずれにも起因しない変動が生じたということを示す。

⑥ Factorial Design

今までの説明は、実験的操作に関して一つの独立変数の原因—結果の効果を調べる方式をとっていたのであるが、現実の場面では同時に2つ又はそれ以上の独立変数を扱い、それらの操作の結果としての従属変数を測定することも必要となる。

はじめに話を簡単にするために2つの要因、変数だけを操作する場合を例にとる。たとえば、いま、ある方式の授業の効果を比較したいと考え、授業の方式として、(X) ディベートやディスカッションの有無（「有」の場合1、「無」の場合0と添字で表す）、(Y) 授業材料としてビデオ教材の有無（「有」は1、「無」は0）の2要因を取上げる。この組合せで次のような表を作成する。そして、各セルに $E_1 \sim E_4$ の等質な実験群をランダムにわりあてる。

このような配置で考えた時、授業方式（討議やビデオ教材、およびその組合せ）で最も効果的である方式

		ビデオ教材 (Y)	
		有 (1)	無 (0)
討 論 (X)	有 (1)	E_1	E_2
	無 (0)	E_3	E_4

は何か、それぞれの要因の効果はどのように評価されるかを明らかにすることができる。そこで、これに事前・事後を入れたデザインで考えると、統制群を1つ加えて次の5グループが構成される。

グループ	事前	実験的操作	事後
Experimental 1 (E_1)	Be_1	$X_1 Y_1$	Ae_1
Experimental 2 (E_2)	Be_2	$X_1 Y_0$	Ae_2
Experimental 3 (E_3)	Be_3	$X_0 Y_1$	Ae_3
Experimental 4 (E_4)	Be_4	$X_0 Y_0$	Ae_4
Control (C)	Bc	なし	Ac

被験者の無作為化がうまく行われ、統制群の事前・事後に差がなければ、事後の平均値の差のみで、どの方式の組合せがよかったかなどが明らかにされる。なお、要因は3つ以上にふやすこともできる。

以上、レビューしてきた方法以外にもまだ何種類かの手法を挙げることができる。たとえば、ある刺激を提示する実験的段階を中に入れて事前・事後に複数の観測時点を取り、同一被験者の変化のプロセスを時系列的に追跡することもできる。従って、うまく適用できれば比較的長期に亘る態度変容なども把握できるかもしれない。(以上の記述内容については文献 [5], [6] を参照されたい)。

(4) データの解析・分析

データの解析は、実は、(3)の実験計画法と分けて考えることはできない。実験のデザインとそれによって規定される統計的検定の方法が一体化されたものとなる。

たとえば、同一被験者に実験的操作をはさんで2回の測定を行ない、前後の変化を比較するという場合、仮りにグループ平均値で差をみるにしても、その平均値の統計的差の検定法は、全く独立に抽出された2つの被験者集団の間の平均値の差の検定とは異なる。同一個人の繰返し測定は、測定値間の相関があると見做し、独立の2つの集団の場合は被験者個人の対応はついていないので相関がないと見做されて、それぞれ異なった平均値の差の検定法が用いられる。

また、いくつかの要因を組合わせて一度に複数の要因効果を抽出したい時に用いる factorial design についていえば、同一の被験者が全ての要因の組合せを実験的操作として受ける場合、要因の組合せの1つ1つに何名かの被験者をランダムにわりあてる場合ではデータの扱いが異なる。

一般に、実験計画法は分散分析 (Analysis of Variance) というデータ分析法、統計的検定 (必要に応じて推定も含める) と一体化されて適用される。分散分析法の統計的モデルは諸要因の主効果、それら間の交互作用、誤差などの諸項目が加算的であるという形をとる。そして、この統計モデルはさまざまな実験計画法に対応して作られる。たとえば、単一の要因 (A) につ

いて、 A_1, A_2, \dots, A_n といくつかの水準を設定する。水準というのはその要因に関して異なった実験的操作を被験者に与えるその内容の違いを表す。従って、いくつかの教授法を一定期間実施し、テストの得点によって測定した時、 A_1 が最もよかったとか、 A_2 はあまりよくなかったといった評価を平均値の差で検定する。この単一の要因の場合は一元配置分散分析という。授業方式で第1の要因A（たとえば、集団討議・ディベートの有無の2水準）と第2の要因B（たとえば、ビデオ教材と印刷教材の2水準）の組合せで効果を検討するタイプは2元配置分散分析という。一般に、factorial designで用いられるのは、2元配置、3元配置、…、多元配置分散分析である。

その他、どのような効果を抽出したいのかによっていろいろな実験計画法があり、それに対応した分散分析法がとられる。

このように、ある要因の効果を抽出したい時には、前以って特定の実験計画を立て、それに則って実験的操作を施すのであるが、もし、事前に準備されなかったある要因がどうも関連しているらしいということが後からわかった場合に、時には事後の解析の段階で、いわば統計的コントロールによってそれを考慮することが可能な場合もある。それが、共分散分析 (Analysis of Covariance) といわれる方法である。たとえば、授業方式の差にどのような効果の差異がもたらされるかを検討したい時に、学力水準の差が混入していることが後からわかった場合に、本来の実験的操作変数以外に学力という補助変数をモデルの中に組み込んでそれをコントロールできる。

なお、今までのデータ解析の話では、測定されたデータとして得られる値はある要因のある水準について、あるいはいくつかの要因とその水準の組合せで得られるある条件のもとで、1つの値（実験的操作の結果得られる従属変数が1つ）であったが、これを複数の従属変数の1セットとして、そのまま全体的に取扱うこともできる。これは多変量分散分析 (Multivariate Analysis of Variance) といわれる。たとえば、もし、授業の効果のために変容すると思われる、生徒たちの何かにたいする認識やイメージ、態度の変化を捉えようとする場合などは、同一の被験者に多くの項目の測定を同時に課すので、その1セットの測定値を全体的に扱うことを考えた時などに適用される。各項目を1つ1つ別個に扱うならば、通常の分散分析でよいのであるが、それを1セット全体としてみる時に多変量分散分析が用いられる。

これには、たとえば学習の進展状況と対応して、時系列的に変化していくある特性に着目した場合のように、 t_1, t_2, \dots, t_n の時点という n 個の時点での測定値の1セットを対象とした場合にも応用できる。

実験計画法と分散分析に関しては、多くの文献があるが、Kirk (1968) あるいはその改定版である Kirk (1982) に体系的によく整理された説明がなされている (文献 [7], [8] 参照)。また、既存のプログラムを利用する場合は、SAS の利用を解説した高橋・大橋・芳賀 (1989) がある (文献 [9] 参照)。

3. まとめ

以上で、授業の効果を捉えるための適確で効率のよい、科学的方法論としての実験計画法と分散分析などの統計的解析への概略を記述してきた。まずはじめに、授業の効果をどのような観点

から捉えるかを検討し、目標を明確にした上で、それらを測定レベルにまでパラフレイズしていくことから始まる。この時、特に入念な検討を要するのは抽象的概念の測定である。最初に全体的概念規定を明確に行ない、それを操作的概念にさげていき、さらに測定変数に具体化して表現していく。このプロセスは研究全般の妥当性が検討される必要の中で最も中心的な位置を占める。さまざまな工夫がなされた測定の結果、当初想定された目標が特定の授業という実験的操作を経て、直接関連しない諸変数、諸要因の影響をのぞいて、純粹に授業の効果として抽出される。

授業の効果は、記憶現象が短期記憶から長期記憶へ、そして体制化された記憶から知識体系に codify されていくように、時間をかけてより複合度の高い、抽象度の高い、態度や世界観の体系に繰り込まれていくと考えられる。従って、あまり短期的な効果測定では見かけ上の効果しか測定されないかもしれない。時間経過に伴ってどのように変化するのかを追跡することも必要となるであろう。ただし、あまり長期に亘る場合には、また別の統制されない諸要因の影響が入ってしまうので、純粹な効果の持続性、定着性を検討することは簡単ではないかもしれない。

2の「授業の効果をとらえる方法」で記述したことは、測定変数の個々の項目について、事前・事後でどのような変化を示したかを主として検討するものであった。これは、効果測定というものの概念的理解を容易にするための単純化であった。しかしながら、上述のようにわれわれが測定したいと思うことは生徒たちの統合化され、システム化された多次元の構造をもつ認識や態度、価値観などである。

こういう観点からすると、現状で要求されるこの種の実験は必ずしもテキストどおりにいかないことであろう。その場合でも、今までに記述された諸方法の趣旨は多くの統計的データ解析の援用により、よく生かされると思う。

たとえば、何らかの社会的態度変化を例にとれば、これは多くの場合明らかに多次元的なものである。多次元尺度構成法 (Multidimensional Scaling, MDS と省略される)、因子分析、その他の多次元解析法から、事前・事後の構造の変化を抽出して効果の検討を行うことができる。この場合、特に、多次元尺度構成法にはかなり多様な統計的モデルがあるので、研究のデザインによっては大変有効な方法論を提供してくれるはずである。被験者×態度項目×時点、といった3相データの変数を同時にとらえることができる INDSICALモデルやそれをさらに一般化した、ウェイト付き多次元展開法モデルなどはその一例である。

MDS の概念の入門書としては、林・鮑戸 (1976)、クラスカル・ウィッシュ (1980) (文献[10], [11]) などが最適であり、理論的には齊藤 (1980)、高根 (1980) (文献[12], [13]) などがよい (応用については文献[14]を参照されたい)。

いずれにしても、研究の事前のデザインが非常に重要になる。全ての計画はデータを収集することからその解析までの全体的展望の中に明確に位置づけられなければならない。このようなケース・スタディーが多く蓄積されていくならば、将来より望ましい授業法の研究開発に大きく貢献できるようになるであろう。

なお、このような実験計画法の考え方を適用したわれわれの身近な例として中切論文 (文献[15]参照) を挙げておきたい。

参考文献

- [1] 辰野千寿・高野清純・加藤隆勝・福沢周亮編(1986) 多項目 教育心理学辞典, 教育出版。
P. 387
- [2] Sullivan, J. L. and Feldman, S. (1979)
Multiple indicators : an introduction. Beverly Hills, Calif : Sage Publications.
広瀬弘忠・石塚智一(訳)(1983) 多重指標分析 朝倉書店。
- [3] Carmines, E. G. and Zeller, R. A. (1979)
Reliability and validity assessment. Beverly Hills, Calif. : Sage Publications. 水
野欽司・野島栄一郎(訳)(1983) テストの信頼性と妥当性 朝倉書店。
- [4] 豊田秀樹・前田忠彦・柳井晴夫(1992) 原因をさぐる統計学—共分散構造分析入門 ブルー
ボックス, 講談社。
- [5] Cohen, L. and Manion, L. (1989)
Research methods in education. 3rd ed. London : Routledge.
- [6] Sedlack, R. G. and Stanley, J. (1992)
Social research : theory and methods. Boston : Allyn & Bacon.
- [7] Kirk, R. E. (1968) *Experimental design : procedures for the behavioral sciences.*
Belmont : Brooks / Cole.
- [8] Kirk, R. E. (1982) *Experimental design : procedures for the behavioral sciences.*
2nd ed. Belmont : Brooks / Cole.
- [9] 高橋行雄・大橋靖雄・芳賀敏郎(1989) SASによる実験データの解析。東京大学出版会。
- [10] 林 知己夫・鮑戸 弘編(1976) 多次元尺度解析法—その有効性と問題点—。サイエン
ス社。
- [11] Kruskal, J. B. and Wish, M. (1978)
Multidimensional scaling. Beverly Hills, Calif. : Sage Publications. 高根芳雄(訳)
(1980) 多次元尺度法 朝倉書店。
- [12] 齊藤堯幸(1980) 多次元尺度構成法 朝倉書店。
- [13] 高根芳雄(1980) 多次元尺度法 東京大学出版会。
- [14] 林 知己夫編(1984) 多次元尺度解析法の実際 サイエンス社。
- [15] 中切正人(1994) ファシズム理解を深めるシミュレーション教材の開発とその実証的研
究—高等学校世界史における「アイヒマン実験」の導入—平成5年度 筑波大学大学院教育
研究科修士論文