

測定・評価部門

わが国の教育測定・評価・統計に関するこの1年間の研究動向

柴 山 直

(新潟大学)

本稿の目的は、1993年7月から1994年6月の期間内に発表された教育測定・教育評価及びそれらの方法論的基礎としての統計学的手法に関するわが国の研究動向を展望することにある。その範囲には「教育心理学研究」、「心理学研究」、「行動計量学」、「Behaviormetrika」の各専門誌、「日本教育心理学会第36回総会」および「日本心理学会第57回大会」の各論文集、ならびに上記の期間中に刊行された著作が含まれる。なるべく包括的に取り上げようとはしたが、結果として筆者の専門領域である教育測定論および教育統計学に関連した論文の選択的な展望となったことは否めない。諒とされたい。

展望の骨子としては、まず教育測定に関する研究論文を取り上げる。ここでは古典的テスト理論をベースとした研究、項目反応理論の応用的あるいは発展的な研究、最近とみにその社会的必要性が増してきた小論文テストに関する研究、そしてテストカタログならびにテストスタンダードに関する著作を概観する。

次に教育評価についての研究論文といくつかの刊行物を見た後、教育心理学における統計学的手法の展望を行う。ここでは方法論をめぐる話題、教育測定の伝統的手法とも言える因子分析的研究に関する論文、またその発展としての共分散構造分析、さらに教育心理学的な研究をより実りあるものにしてくれる可能性を秘めたいいくつかの手法を使った研究のうち、主として多変量解析を中心とした適用例を眺めていく。なお、教育心理学における統計的手法、特に統計的検定の有効性と限界に関しては、本年報の別稿で南風原朝和氏が踏み込んだ考察をされているので、そちらを参考にされたい。

1. 教育測定論

1.1 古典的テスト理論

後述する項目反応理論の登場で、その陰に隠れた観のある古典的テスト理論であるが、現在使われている検査の多くが依って立つ理論であるという点で、まだまだ研究すべき課題が残されている。本年度は、特に実用性との兼ね合いの中での研究がなされたようである。

池田(1994)は測定の信頼性係数を推定するための従来の方法を、推定に必要とされる情報の多寡と前提条件の

強弱との観点から整理し、各下位テストと総得点の分散、それらの間の相関(共分散)を利用して求められる信頼性係数の推定方法を提案している。たとえば、この推定方式をCronbachの α 係数と比較してみると、総得点と各下位テスト得点間の相関(共分散)が必要になる一方で、 α 係数の前提とされる、各下位テストが互いに τ 等価であるという条件は満足していなくても構わない。したがって、より緩やかな前提条件のもとで信頼性係数を推定できるため、 α 係数よりも広範性および簡便性の点ですぐれているのではないかと指摘している。

小笠原(1993)は、被験者により検査間隔が異なる状況下における測定モデルを導入することで、個人の安定性と測定の信頼性を分離することを試みている。これは古典的テスト理論で扱われてきた信頼性係数の推定方法のひとつである再検査法の条件、すなわちすべての被験者が同一の検査を同じ時点で受けるという条件を緩和したモデルとなっている。提案されたモデルには、単一観測変量のモデルと潜在変数(因子)モデルのそれぞれにつき3つのバリエーションがある。各々のモデル自体も、やや複雑なものとなっている。しかし、内田クレベリン検査と注意配分検査(各検査ともその検査間隔は平均8年以上)とによって得られたデータを、いかに分析するかという実用上の強いモチベーションに支えられているだけに、これからの適用例の蓄積がたのしみである。

1.2 項目反応理論の応用

本年度も項目反応理論に関する研究は多かった。しかし従来と違った視点からの研究が増加したように思える。そのひとつとして野口(1994a, b)や藤森(1994a, b)に代表される項目反応理論の実用化のための研究が挙げられる。もちろん、例えば印東他(1962)や芝(1978)など、これまで実用化を主眼にした研究がなかったわけではない。しかしいずれもいわゆる検査方式として紙と鉛筆を使うものが主であった。また柴山他(1987)が芝(1978)の語彙理解力検査をパーソナルコンピュータで行えるようにした研究があるが、これはむしろ項目反応理論の実用化に際しての基礎資料を収集するという意味合いが強かった。これに対して野口(1994a, b)は最初からパーソナルコンピュータを利用することを前提に、適性検査のひとつで

ある識別性検査の開発を行っている。まだ開発中のものであるため、項目反応理論の特徴を活かした、被験者の特性に応じた効率のよい測定を実現した適応型検査には至っていないようである。しかし、各検査項目の図版がコンピュータ画面上でカラー表示されるなど、ここで培われたノウハウの教育現場への応用が期待される研究である。

一方、藤森(1994a)は項目反応理論に基づくコンピュータを利用した適応型検査において、項目情報量の最大化方式で呈示される項目の困難度が、被験者にとっては心理的に難しすぎるため、正答確率が0.7から0.8となるような困難度の項目を使うことを提案している。確かに被験者にとっては難しすぎると感じられる項目を避けることで、不安感など、余計な負担をかけないという意図はよく理解できる。しかし、項目情報量の最大化方式自体が、なるべく少ない項目数で効率よく測定を行うことを目的としているため、それと矛盾する規準(この場合、正答確率が0.7から0.8となるような困難度の項目)を導入した場合、意図とは逆の結果を招きかねないのではないだろうか。この矛盾をどう解決するのか、今後の研究の進展を待ちたい。また別の視点から、藤森(1994b)はコンピュータ化されたテストにおける無答欄が、被験者の反応行動に及ぼす影響を調べ、紙筆検査と大きな差はないことを、実験的手法によって見いだしている。

このように項目反応理論は、様々な心理検査をコンピュータ化する上で、非常に強力な理論的基礎を与えてくれるものであり、技術的には昨今のパーソナルコンピュータの機能の向上が、それを実現可能なものとしてくれている。その意味で、たとえば、村上(1993)で報告されている、MMPI-1の短縮・改訂版である MINI のコンピュータ方式化などにも十分役立つはずである。なお項目反応理論には含まれないが、ニューラルネットワークによる WAIS-R の短縮版を作成する服部他(1994)の試みも、新しい研究動向として見逃せないものである。

さて、項目反応理論においても、これまで、そのパラメタ推定のために、同時推定法、周辺最尤推定法、ベイズ推定法、MMAP 法など理論的に精緻な方法が提案されてきた。しかし、テスト開発およびそのデータ解析に携わってきた筆者自身の経験から言えば、実用上は、いわゆる簡便法によるパラメタの推定値で十分であったことも少なくはない。このような観点から、服部(1994a)は、項目反応モデル、特に2パラメタ・ロジスティック・モデルのための簡便なパラメタ推定方法を、入念なシミュレーション実験のもとに、上記の理論的に洗練された様々な方法と比較検討した研究を行っている。一見地味だが、しかし現場の要求に即したこの種の研究が、モ

デルの地についた普及という意味で、今後も「教育心理学研究」誌に掲載されることを切望する。

項目反応理論関係の研究動向として、もう一つ見逃せないのが、繁樹(1994)に代表される教育工学的な観点からの一連のアプローチである。このアプローチは、テストにおけるコンピュータの活用を念頭に置いているという意味では、野口(1994a, b)や藤森(1994a, b)など同一の方向を目指しているものと言える。しかし、後者があくまでも効率的な測定に重点を置いているのに対し、前者は測定と同時に最適な教授法を提示するシステムの構築を目指しているところに特色がある。

項目反応モデルに限らず、計量心理学的なモデルの特徴として、学習者の心理的特性を少数個のパラメタで表わし、あえてその内的なプロセスまで踏み込むことを避ける傾向がある。これに対し、Embleton(1985, 1991)などは、項目内容を、その項目に正答するために必要と考えられる認知要素にまで分解し、さらに、その要素に対応する潜在変数をパラメタとして、項目反応モデルの中に組み込むことで、学習による能力変化をとらえようとするなどの工夫を行っている。海外のこうした研究を踏まえ、Shigemasu & Ueno (1993)は、学習者個人の理解の程度と通常の正誤パターンに加えて、それとは別の外的な情報を考慮したモデルを提案した。それによれば、まず、各学習者はある知識を1)完全に理解しているか、2)部分的に理解しているか、あるいは、3)全く理解していないかの、いずれかの状態にあると考える。次に、そうした状態が反応結果の正しさへの自信の有無に影響すると仮定するのである。また、植野他(1993)は、学習結果の形成的評価を目指し、学習者の解答過程と解答方略を評価するためのパラメタを新たに組み入れた項目反応モデルの提案を行い、確率計算のテスト問題によって得られた実際のデータへの適用例を示している。

この一連のアプローチの基礎となっている項目反応モデルそのものが、すでに、局所独立などの厳しい条件や、付随パラメタの問題を抱えている。そのため、かなり限定された状況下でないと、パラメタの推定がうまく行えないことが経験的に知られている。そうした項目反応モデルの限界を超え、学習者の動的な学習プロセスに即応しながら評価と測定を一体化したシステムを構築しようとする繁樹(1994)らの試みは野心的ですらある。ただ一つ、あえて批判的なことを述べるなら、「モデルの複雑化は諸刃の剣」ということにつきるのはないだろうか。複雑な現象を記述するために導入した複雑なモデルは、局所的に精緻な記述あるいは予測が行えたとしても、それが他の状況でも有効に機能する保証はない。それは、丁度、マクロ計量経済モデルにおける同時方程式体系の

大規模化が、必ずしも経済予測の精度の向上に結びつかなかったという佐和(1994)の指摘と軌を一にしているのではないだろうか。勿論、繁樹(1994)もこのことを承知の上で、敢えて「複雑な統計モデルの利用」に踏み切っているのだから、今後の研究の発展を待つべきであることは言うまでもない。

なお、項目反応理論の範疇に含まれるわけではないが、テストデータ解析のためのモデルの提案として、小笠原(1994b)の研究がある。これは、ある一人の被験者の誤答数などの分布を問題にするのではなく、集団における事象数の分布を問題にするとき、単一のポアソン分布ではなく、それを特殊例として含む一般化ポアソン分布を使うべきであるという主張のもとになされた研究である。ある検査に関する集団の誤答数の分布が年齢とともにどのように変化するかを適用例として示している。その際、一般化ポアソン分布のパラメタの変換を行うことで、平均と標準偏差や、分散と平均の比の経年変化が解釈しやすくなるような工夫を凝らしている。一般化ポアソン分布が事象数の分布として取り扱いやすいものなので、これから実用に供される場面も多くなるかもしれない。

1.3 小論文テスト

1991年度の日本教育心理学会総会で持たれた研究委員会企画シンポジウム「教育評価の将来像を考える」で、池田(1991)はこれまで教育評価が歩んできた道を、いくつかの対立する概念のせめぎあい、たとえば、相対評価 vs. 絶対評価あるいは学力試験 vs. 内申書・推薦書などの観点からとらえ、現代はテスト方式における対峙概念でいえば、どちらかというお客観式テストから記述式テストへの回帰の時代であるとしている。当然のことながら、単純に記述式テストのみがよいという楽観的な見方を池田(1991)がしているわけではない。しかし、確かに時代の要請として採点者の主観が少なからず影響を持たざるを得ない記述式テスト、ここでは小論文テスト、をいかに処理するのかは大きな問題であると言えよう。

この問題に関し、池田(1994)は、採点者の相違による採点結果への影響を調べ、解答様式と採点基準がよく構造化されているならば、統計的に有意な採点者のバイアスを見いだせないことを報告している。この研究は、項目反応モデルにおいて、もともと被験者集団の違いによる項目の不公平さを検出する方法として発展してきた Differential Item Functioning (たとえば Holland & Wainer (1993)) の考え方を応用したものである。

また、平(1994)は、その問題意識を、1)作文評価において高い信頼性をいかに確保するのか、2)作文能力の測定を行うための課題をいかに工夫するのか、3)作文の出来不出来を説明する要因は何かという点に絞り、基礎的

言語能力や教科・科目の学力などに左右されにくい課題として、絵画から簡単な物語を作成させる方法を提案している。これは純粋に測定方法に関する研究というよりも、作文教育に関する研究と言った方が適當かもしれない。しかし、教育測定にまつわる様々な利点・欠点を踏まえた上で、実践上どうするのかに正面から取り組んだ研究として注目できるものである。

なお、大野木(1994)は、小論文テストへの「過大な期待について、むしろ、思い切って立ち止まって考えよう」という立場から、学校現場で使われるさまざまな教育テストについての問題を検討している。いろいろな問題が交錯する複雑な現場に立って「論文体テスト」を考察している点で、一読に値する著書であると思われる。

1.4 テストカタログとテストスタンダード

教育心理学の研究をする上で、当面の問題に関連した心理検査が必要になることが間々ある。米国では'Tests'(Sweetland & Keyser (1983))などといった心理検査のカタログ類が刊行されているのだが、日本ではこれまでこの種の刊行物は、筆者の知る限りなかった。ところが、本年度においては、上里(1993)と堀他(1994)が相次いで刊行された。上里(1993)では、WISC-RやWAISあるいはMMPIやTATなどの比較的定評のある心理検査を約40個近く取り上げ、それらの特徴や関連研究の動向、具体的なケース例、あるいは効用と限界を丁寧に解説している。また、付録の心理検査一覧は研究に必要な検査を探すときに重宝なものとなろう。一方、堀他(1994)によるものは、日本で開発されたにもかかわらず、いわば、埋もれたままになっているものを中心に、実に104篇にも上る尺度を収録している。

堀他(1994)の編集方針で注目すべきは、すべての尺度について、妥当性と信頼性に関する項を共通に設けている点である。測定論的に言っても、妥当性と信頼性の検討は尺度を使う上で、必須であるにも関わらず、看過され勝ちであった。実際、このカタログによって、妥当性や信頼性にまったく言及していない心理尺度の存在が確認できたことは、予期していたとはいえ、些か驚きであった。ある程度の因子分析や項目分析の素養を持つ研究者であれば、質問紙形式の尺度を「形式的に」作成することは、さほど困難ではない。そのため、極めて狭い関心事に基づいて「作ってみただけ」の尺度も多く、それが、また、淘汰されずに残ってしまっていた。しかし、このようなハンドブックが2冊刊行されたことは、作成された心理尺度が、様々な研究者の厳しい目にさらされ、既存のあるいはこれから作成される心理検査の品質を向上させることを意味する。また、同じ尺度を使った、いくつかの別のデータの分析結果を統合して、メタ分析を行

い、さらに尺度を洗練していくことも可能となるなど、大きな意義を持つ出版である。以上のような機能を、今後も担っていけるよう、定期的に上記の刊行物の内容が改訂されることも、また、あわせてお願いするものである。

さて、高品質の心理検査・教育検査を作成し、それを有効に活用するためには、関心下の領域の深い洞察と、それを実現させるのに必要な心理測定論あるいは教育測定論の知識が不可欠である。さらに、そのテストが社会的な役割を果たして行く上では、何らかのガイドラインが必要となる。良心的に作成された教育・心理テストが存在する一方で、そうでないテストもまた世の中に氾濫している昨今の状況を見るに付けて、その感は深まるばかりである。

このような状況への改善策の1つとして、学部レベルの専門教育の中に、基礎としての測定論と実践としての心理検査法とが密接に関連したコースを設けることが考えられる。筆者の主観的な判断に過ぎないが、これまで心理検査を使う機会の多いカウンセラーをはじめとする専門家と、そうした検査の測定論的基礎を扱う専門家とのあいだには、なぜか、深く越えがたい溝があったように思う。おそらく、それが遠因となって、両者がチームを組んで担当するようなコースがほとんど存在しなかったのではないだろうか。渡部(1993)によるテキストは、そのようなコースにおける橋渡しをしてくれる有力な候補のひとつとなろう。その構成は、前半で、信頼性や妥当性の概念、因子分析や項目反応理論の基礎知識を整理した後、後半では、知能検査・性格検査・職業興味検査・不安検査・発達検査、さらには投影検査にいたるまでの代表的な検査の特徴と実践方法について解説するという内容となっている。

改善策の第2として、社会の中で心理検査・教育検査を使っていく上でのガイドラインを実際に設けることが考えられる。赤木他(1993)による教育・心理テストの規準開発委員会(1985)の翻訳は、まさにそうしたガイドラインを懇切丁寧に示してくれるものである。これは、以前、斎藤他(1973)によって翻訳されたAPA・AERA・NCME合同委員会(1966)の最新版である。内容は大きく4部から構成され、「第I部テストの作成と評価のための専門的規準」では測定論的な観点から見てテストが満たすべき条件が扱われ、「第II部テスト利用に関する専門的規準」では、例えば、臨床場面や学校、あるいはカウンセリングでの心理テストや教育テストの利用に関しての規準などが述べられている。また「第III部特定の適用のための規準」では障害となる条件を持つ人たちなどを対象とする場合の規準、そして「第IV部実施手続に関する

規準」ではテストの実施手続から受検(験)者の権利の保護までカバーされている。勿論、文化的背景の異なる米国でのガイドラインなので、そのままわが国に当てはめるわけにはいかないにしろ、玉石混淆の現状を少しでも改善するのに大きな役割を果たす翻訳と言えるであろう。

2. 評価

これまでの社会科と理科を合併する形で、3年前に発足した、小学校低学年における「生活科」は学校現場においてかなりの混乱をもたらしていると聞く。「生活科」は名ばかりで結局单元ごとにもとの「社会科」であったり「理科」であったり、あるいは単に「お遊び」の時間になってしまったりというのが実状であり、ましてや「生活科」の学力を評価するなど、難しくてできないというのが多くの教師の正直な感想であるともいう。そこで、全国のいわゆる「生活科」研究推進校とその担当教諭を対象に、現場では「生活科」をいかに評価しようとしているのかをアンケートにより調査した結果が、藤岡(1994)によって報告されている。詳細は藤岡(1994)をご覧いただきたいが、指導要録の観点を評価を行うに際して教師が難しいと感じているものが「気付き」、「関心・意欲・態度」であったことが報告されているのは注目すべきことであろう。藤岡(1994)はその理由として、前者が「概念の曖昧さ」、後者が「主観的になりやすいこと」を挙げているが、正にその通りである。評価すべき学力の定義、もしくは概念構成がはっきりしないまま行う評価を、真に評価と呼べるのかという本質的な問いかけを改めて投げかけた報告と言える。

研究論文以外では、梶田(1994a, b, c)の著作集が刊行されたことを忘れるわけにはいかない。その内容を簡単に紹介しておく。梶田(1994a)では「学力観・評価観の転換」と題して、学校現場での教育実践に直接関わる問題についての論考が収録されている。また梶田(1994b)では「学校学習とブルーム理論」として、その著作のうち理論的な論文が中心となっている。さらに梶田(1994c)では教育評価と社会とのかかわり合いを扱った論考が集められ、「社会・学歴・教育評価」という副題が付けられている。もちろん、その時々における論考を集めたものであるため、体系的に、たとえば教育評価のテキストとして使うわけにはいかないかもしれないが、この分野で活躍してきた研究者の集大成であるだけに、理論的実践的課題に対してさまざまな示唆を与えてくれるものとなろう。

もうひとつ取り上げたいのは、天野(1993)による著作である。教育測定論や教育統計学といった、どちらかというと技術的な側面の強い分野を専攻していると、技術自体の発展にエネルギーを集中させてしまい、それが社

会の中で果たしてきた歴史的な役割や、その背後にある教育観や学力観などを（少なくとも筆者は）忘れがちになる。それを補完してくれるのがこの著作である。そこではわが国の初等教育（義務教育）段階における教育評価活動が、近代教育史の流れの中で、教授・訓育活動などの実践の改善にいかにつぎつぎと結びついてきたのかの観点から考察されている。それだけに、たとえば、上でも取り上げた観点別評価の問題、とくに測定論的にみて概念構成のあいまいな「関心・意欲・態度」の評価が、先人達が残した遺産からみてどのような意味を持つのかといった視点で読めば、考えさせられることの多い書物となるはずである。

3. 教育心理学研究における統計的手法

3.1 方法論

本年度の総会においても、教育心理学的研究の方法論としての統計的手法に関する自主シンポジウム（繁樹他（1994））が開かれ、かなりの盛況であった。それだけこの問題に関心を持つ研究者が多いということなのだろう。シンポジウム自体も、「教育心理学の研究者が、自分の研究を遂行するため抱くことの多い疑問」を15個取り上げ、教育心理学と統計学の接点で活躍している5人の話題提供者が、それぞれの立場から答えていく斬新な形態をとり、参加者にもかなり有益なものとなったのではないだろうか。

詳細な手法上の有効性と限界の議論はともかくとして、筆者の印象に残った話題を見ていくことにしよう。まず、このシンポジウムの企画・司会者である繁樹算男氏の次の意見、

統計的方法のメーカーとユーザーを画然と分ける考え方には同情的になれない。自分の研究に有効であると信じて、ある方法を使おうとする場合その使用者はその方法に習熟している必要がある。たとえば、流体の運動の研究者が微分方程式を必要とするならば、その研究者は当然それを勉強し、自家薬籠中のものにするであろう。教育心理学でも同様であるはずである。（繁樹他（1994））

には全面的に賛意を表わしたい。統計的手法が必要な研究で、統計的手法だけを切り放した研究は研究にはならない以上、研究者自身が、計算機操作も含めて、必要な統計的手法を学ぶのが当然であると筆者も考える。

次に、目にとまるのが、森敏昭氏の「教育心理学研究誌の掲載論文を見てみると、以前より統計の誤用が増えたような気がする」という主旨の発言である。森氏の指摘によれば、その原因として、1)研究者が増えたこと、2)（統計的手法の使い方を丁寧に説明してくれる）ディーラーが少

ないこと、3)コンピュータが普及してきたこと、4)Don'tsを言う人が少ないことの4点が考えられるという。また、この指摘と関連して、もう一人の指定討論者である楠見孝氏より、統計的手法を利用しようとする研究者にとって、1)統計教育のポイントがどこにあるのか、2)研究者の自己学習のポイントはどこにあるのか、3)教育統計学者の役割は何なのかがよくわからないから、結果として統計の誤用が増えるのではないかとこの指摘があった。発表時間の関係で、両氏の指摘についての議論が行われなかったのは残念だったが、教育心理学研究における統計教育を考える上で、重要な指摘であると思われる。

また、疑問12の「因子分析を行い解釈して、因子得点による実験群と対照群の比較を行いました…」に関しては、フロアーからの質問に応じて、多変量解析一般の文脈の中で、「多元的にデータを見るためには様々な手法を試みることをすすめる」という主旨の、柳井晴夫氏による発言があった。確かに多次元データの性質を探るといふ点ではその通りなのだが、この疑問12に限っていえば、方法論上の矛盾が生じてしまう。すなわち、因子分析を使わなければならない場面というのは、それが仮説探索的であれ仮説検証的であれ、サンプル数にして数百程度の大標本を前提としている。その一方で、「実験群と対照群の比較」とある以上、いわゆる平均差の検定も行うというのなら、サンプル数の多さが原因となって検出力が高まり、両群の間に統計的有意差がみられて当然の状況となる。したがって、疑問12は、因子得点を使うか使わないかという問題以前に、実験群と対照群の比較に、因子得点を使わざるを得ない方法論上の不備が問題となるはずである。

さらに、宮埜寿夫氏による分散分析における事後検定や、豊田秀樹氏による因果推論の文脈のなかでの分散分析に関する話題提供があった。分散分析を含む統計的検定の問題は、ある条件や被験者群についてのデータから得られた情報を、どの方向へ一般化するのかという問題としてとらえることもできる。この問題を考えるには、「日常的な意味での一般化可能性」と「有意性検定における一般化可能性」を対比的に論じた、南風原（1994）が、統計的推定の本質を考える上で、非常に有益となろう。また、教育心理学の分野からは少し逸脱するかも知れないが、藤越他（1993）による Rao（1993）の訳も、また、方法論としての統計学を考える上で極めて刺激にとむ著作であろう。

3.2 適用例

ここでは教育心理学的な研究における統計的手法の適用例を、新しい使い方という点で特徴的であると思われるものをトピックス的に簡単に見ておこう。

相関分析：これまで「教育心理学研究」誌で推測統計的な手法が使われるときには、有意性検定の結果のみが報告されることが多かった。それに対して、丹藤(1994)は、小学2年生から小学4年生および小学5年生から中学3年生にわたる追跡調査の結果に基づき、学業成績におけるきょうだい間相関の学年変化等を分析する際に、相関係数の信頼区間を併せて報告しているのは、研究報告の新しいスタイルとして注目できる。すこし煩雑なようだが、このようにデータから得られた相関係数の値と信頼区間が同時にわかると、単に母集団分布において相関が無いという帰無仮説を棄却した結果だけからはわからない情報が、読者にとっても読み取れるようになって有益である。ただ、丹藤(1994)の研究全体から言えば些細なことだが、相関係数の値や信頼区間を示す際の工夫、たとえば小数点以下第2位程度に丸めておくとか、文章中には相関係数だけ書いておいて、信頼区間の方は表にまとめておくとかの工夫があればさらに読みやすかった。いずれにせよ、この研究が、データ分析によって得られた相関係数の報告に関する、新しいスタイルの起点として位置づけられることに変わりはない。

また、中村他(1994)はベイズ的方法による四分相関係数の区間推定を行う方法を提案している。ベイズ流のアプローチをとった場合、母集団パラメタがその区間に入る確率によって、その確信区間の幅を直接解釈できるので、伝統的な標本理論による信頼区間の解釈よりも自然である。事前分布の客観性の問題や数値計算上の問題など解決すべきことがらがあるとは言え、研究者が直観的に理解しやすい相関情報の利用方法を提案した点で、注目すべき研究のひとつであろう。

正準判別分析：前田他(1994a, b)は正準判別分析を使った、職業適性の予測に関する妥当性の検討を行っている。ただ、正準判別分析自体が外れ値に極めて敏感な手法であること、また、さまざまな要因が錯綜する現場での応用が未知数なことなどの問題点が残されていることには注意が必要である。

因子分析的研究：因子分析を使った研究はさすがに多い。ここでは、異なる被験者集団から得られた因子を比較するという点で、共通のアプローチをとっている、次の2つの研究を挙げるにとどめておく。1つは、日本とアメリカの大学生の創造的態度を因子分析により比較した繁樹他(1994b)の研究であり、もうひとつが、親子関係の構造を明らかにするために、養育態度の認知が親子でどのように異なるかを、プロクラステス法によって比較検討した小高(1994)の研究である。なお、柏木他(1994)は、因子分析の直交ジェオマックス基準群の収束に関するシミュレーション実験の結果を報告している。

3相因子分析：多変量解析の多くは、変量×被験者のいわゆる2相データの分析を対象としている。しかし、縦断的データや継時的データでは、この他に時間軸の入った、変量×被験者×測定時期の3相データとなることが多い。こうした種類のデータをそのまま分析する目的で開発された手法のひとつが3相因子分析である。田中他(1993)は服部(1993)がパーソナル・コンピュータ用に開発したアルゴリズムを用いながら、継続的に行われた臨床事例を記述するというユニークな分析を試みている。PARAFACモデルをはじめとして、3相データの分析手法は、まだまだ適用例が少ないので、こうした研究がこれからもなされることが必要であろう。

共分散構造分析：共分散構造モデルについては、豊田(1994)が同時方程式モデルのパラメタを推定できる十分条件を導いている。適用例としては、鈴木他(1993)による高校生の進路決定にいたる因果モデルの構築があげられる。さらに、服部(1994b)は共分散構造分析の学習を目的とした、パーソナルコンピュータ用の教育ソフトを開発している。

MDS：多次元尺度法(MDS)は、1970年代から1980年代にかけて非常な発展を見た。この方法は、そもそも人間の認知構造を空間的に表現しようという目的で開発されてきたものだけに、教育心理学的な研究に資することの多い解析手法であると思われる。実際、今井(1993)は非計量多次元尺度法とクラスター分析を有効に用い、英語を外国語とする日本人学生の意味表象がネイティブスピーカーのそれと比較して、空間表現上は拡散している、いいかえれば各々の語の意味が構造化されないままになっていることを見いだしている。このような言語や概念の表現のみならず、たとえば、他者認知や自己認知の構造を明らかにする上でもMDSの果たす役割は、まだまだ大きいと言える。

一方、パラメタの推定方法の発展から見ればMDSもまた、他の手法と同様、最小2乗法から最尤推定法、そしてベイズ推定法へと発展してきた、あるいは発展しつつある方法である。この文脈でみれば、最尤推定法に基づくMDSの一方法であるPROSCALの解に一意性をもたせる土屋(1994)の提案は、刺激座標間に独立性の仮定を先験的に導入するという点で、最尤法からベイズ推定法への過渡期にあるものと位置づけることができるであろう。この研究の今後の発展を見守りたい。

欠測値：教育心理学的な研究を進めていると、調査項目への無解答やテスト項目のやり残し、あるいはデータ収集上の制約などの理由から、一部のデータが得られないまま分析を行わなければならない場面に出くわすことも多い。このような欠測値を含む、不完全データの分析

法に関しては、小笠原(1994a)が、成人用知能検査のデータ分析に対する応用を念頭に、ラッシュの乗法的ポアソンモデルを欠測値がある場合に拡張することを試みている。また、柴山(1994)は、欠測値を含む不完全データ行列の主成分分析のための最適化アルゴリズムの改良結果を報告している。

多変量解析全般：以上、多変量解析の教育心理学における適用例を中心にトピック的に展望してきたが、最後に多変量解析全般の展望論文として、藤越他(1993)と柳井(1994)を挙げておく。前者は、多変量推測統計学と記述的多変量解析の最新の発展を展望している。また、後者は、前者の記述的多変量解析手法の部分を、さらに簡潔に整理しなおしたものである。両者とも多変量解析の最新の研究動向を知る上で貴重な論文となろう。

引用文献^{*)}

- 上里一郎(監修) 1993 心理アセスメントハンドブック 西村書店
- 天野正輝 1993 教育評価史研究 —教育実践における評価論の系譜— 東信堂
- Committee to Develop Standards for Educational and Psychological Testing of The American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education 1985 *Standards for Educational and Psychological Testings American Psychological Association*. (赤木愛和・池田央(監訳) 1993 教育・心理検査法のスタンダード 図書文化社)
- Embleton, S.E. 1985 Multicomponent Latent Trait Models. In S.E. Embleton (Ed.) *Test Design : Developments in Psychology and Psychometrics*. Academic Press.
- Embleton, S.E. 1991 A Multidimensional Latent Trait Model for Measuring Learning and Change. *Psychometrika*, 56, 3, 495-515.
- 藤越康祝・柳井晴夫 1993 多変量解析の現状と展望 日本統計学会誌 22, 3, 313-356.
- 藤森進 1994a 心理的適切さを考慮したCAT 教心 478.
- 藤森進 1994b CATにおける無答欄表示の影響 日心 444.
- 藤岡秀樹 1994 「生活科」の評価についての研究 —「生活科」研究推進校の通知表の現状と担当教諭に対する調査結果を中心に— 教心 479.
- 南風原朝和 1994 一般化可能性の評価と有意性検定 日心S11.
- 服部兼敏・小林重雄・藤田和宏・前川久男 1994 ニューラルネットワークによるWAIS-R短縮版の作成 教心 461.
- 服部環 1993 HPARAFAC.EXE: PARAFACモデルを解くパソコン用プログラム 教育心理学フォーラム・レポート 日本教育心理学会
- 服部環 1994a 簡易推定法の推定精度に関する検討 —MMAP法, 周辺最尤推定法, 疑似ベイズ推定法との比較 教育心理学研究 42, 1, 38-47.
- 服部環 1994b 共分散構造分析の学習を主なねらいとするパソコン用プログラムの作成 教心 462.
- Holland, P.W. and H. Wainer 1993 *Differential Item Functioning*, Lawrence Erlbaum Associates.
- 堀洋道・山本真理子・松井豊(編集) 1994 心理尺度ファイル：人間と社会を測る 垣内出版
- 梶田叡一 1994a 教育における評価の理論I 学力観・評価観の転換 金子書房
- 梶田叡一 1994b 教育における評価の理論II 学校学習とブルーム理論 金子書房
- 梶田叡一 1994c 教育における評価の理論III 社会・学歴・教育改革 金子書房
- 池田英子 1994 主観的判断がテスト採点に及ぼす影響の研究 教心 463.
- 池田央 1991 技術革新と教育評価 教心S11.
- 池田央 1994 信頼性係数再考— α 係数に代わるもの— 教心 459.
- 今井むつみ 1993 外国語学習者の語彙学習における問題点—言葉の意味表象の見地から— 教育心理学研究, 41, 3, 243-253.
- 印東太郎・鮫島史子 1962 LIS推因子測定法 日本文化科学社
- Joint Committee of the American Psychological Association, American Educational Research Association and National Council on Measurement in Education 1996 *Standards for Educational and Psychological Tests and Manuals American Psychological Association*. (斉藤耕二・高橋良幸(共訳) 1973 テスト・スタンダード〜教育・心理検査および手引のスタンダード〜 日本図書文化協会)
- 柏木繁男・青木孝悦 1994 直交ジェオマックス基準群の収束について 心理学研究, 65, 1, 42-46.
- 小高恵 1994 親子間の認知構造の因子分析的研究 心理学研究, 65, 2, 95-102.
- 前田忠彦・和田さゆり・佐藤典子・柳井晴夫 1994a 正準判別分析に基づく適性診断法に関する検討 教心 460.

- 前田忠彦・和田さゆり・佐藤典子・柳井晴夫 1994b 職業適性に関する基礎的研究 (III) —検査バッテリーによる適性予測の増分妥当性— 日心 370.
- 村上宣寛 1993 MINI 性格検査の冊子方式とコンピュータ方式の違いについて 心理学研究, **64**, 4, 279-283.
- 中村和靖・前川眞一 1994 四分相関係数の区間推定 日心 436.
- 野口裕之 1994a パーソナル・コンピュータをベースにした識別性検査に対する受験者の反応 教心 476.
- 野口裕之 1994b パーソナル・コンピュータをベースとする識別性検査A—1001用項目プールの作成 日心 371.
- 小笠原春彦 1993 検査再検査間隔が個体毎に異なるデータにおける信頼性と安定性の推定 心理学研究, **64**, 5, 369-377.
- 小笠原春彦 1994a 欠測値を含むデータにおけるラッシュの乗法的ポアソンモデル 教心 456.
- 小笠原春彦 1994b 一般化ポアソン分布におけるパラメータの変換 日心 435.
- 大野木裕明 1994 テストの心理学 ナカニシヤ出版
- Rao, C.R. 1993 *Statistics and Truth : Putting Chance to Work*, 2nd Edition. CSIR. (藤越康祝・柳井晴夫・田栗正章(共訳) 1993 統計学とは何か～偶然を生かす～ 丸善)
- 佐和隆光 1994 計量経済分析の意義と有効性—過去30年の経過と展望— (竹内啓・竹村彰通編 1994 数理統計学の理論と応用 第11章) 東京大学出版会.
- 芝祐順 1978 語彙理解力尺度作成の試み 東京大学教育学部紀要, **17**, 47-58.
- 柴山直・野口裕之・芝祐順・鎌原雅彦 1987 最適化テスト方式による語彙理解力の測定 教育心理学研究, **35**, 363-367.
- 柴山直 1994 欠測値を含む多変量データのための主成分分析法～最適化アルゴリズムの改良とその結果～ 教心 457.
- 繁桝算男(研究代表者) 1994 多次元測定空間において学習者の軌跡を表現する方法の開発～効果的な教授的介入のために～ 平成5年度科学研究費補助金研究成果報告書.
- Shigemasu, K. and Ueno, M. 1993 A New Item Response Model with Parameters Reflecting State of Knowledge, *Behaviormetrika*, **20**, 2, 161-169.
- 繁桝算男他 1994 教育心理学における統計的方法: DOs and DON'Ts 教心 46-47.
- 繁桝算男・横山明子・スターン・駒崎久明 1994b 日米学生の創造的態度の因子分析による比較 心理学研究, **64**, 3, 181-190.
- 鈴木則夫・柳井晴夫 1993 因果モデルによる高校生の進路意識の分析 教育心理学研究 **41**, 3, 324-331.
- Sweetland, R.C. & D.J. Keyser 1983 *Tests : a comprehensive reference for assessments in psychology, education and business*. Test Corporation of America.
- 平直樹 1994 物語作成課題に基づく作文能力の分析 教心 475.
- 田中堅一郎・篠竹利和 1993 3相因子分析法を用いた心理臨床事例の分析 心理学研究, **64**, 4, 284-288.
- 丹藤進 1994 学業成績及び知能のきょうだい間類似性についての縦断的研究 教育心理学研究, **42**, 1, 29-37.
- 豊田秀樹 1994 An identification rule for the simultaneous equation model 教心 458.
- 土屋隆裕 1994 一意的な軸を持つ PROSCAL の提案 日心 438.
- 植野真臣・久保琢朗・繁桝算男 1993 学習者モデルとしての項目応答モデルの提案 行動計量学, **20**, 2, 1-8.
- 渡部洋(編著) 1993 心理検査法入門～正確な診断と評価のために～ 福村出版
- 柳井晴夫 1994 多変量解析の動向—主に記述統計的立場から— (竹内啓・竹村彰通編 1994 数理統計学の理論と応用 第6章) 東京大学出版会
- 注) 文献末尾の教心 p 及び日心 p は, それぞれ, その年度の日本教育心理学会総会発表論文集と日本心理学会大会発表論文集の p 頁を意味する。

付 記

この論文の草稿に対して数多くの貴重なコメントを頂きました新潟大学教育学部の松井仁助教授に, 厚くお礼申し上げます。もちろん, この論文の中で扱った研究に対する誤解等があれば, その責任はすべて筆者にあることは言うまでもありません。