

測定・評価部門

教育心理学研究における調査法

土屋 隆 裕

(統計数理研究所)

はじめに

本稿でレビューの対象とした研究は、2003年7月から2004年6月までに発行された『教育心理学研究』および2004年度の『日本教育心理学会第46回総会発表論文集』に掲載された論文が中心である。ただし筆者は、測定・評価部門が本来扱うべきであろうテスト理論や心理統計学、教育評価(日本教育心理学会, 2003; 廣瀬, 2004)に関しては、最近の動向に関する情報はもとより、基本的な知識についても持ち合わせていない。筆者の関心領域はむしろ、誰を対象にどのような道具を用いれば、質の高いデータを得られるだろうかといった統計的な調査法・データの獲得方法にある。したがって、そのような調査データの獲得方法という観点から、今年度発表された研究のいくつかをレビューすることとした。

当然のことながら、調査研究によって得られた結果には、ある程度の一般性・普遍性が求められる。どの程度一般化できるのかという判断基準の一つは、調査結果に含まれる誤差の大きさであり、サンプル調査の場合、誤差にはサンプリング誤差と非サンプリング誤差とがあることはよく知られている。本稿では、サンプリング誤差については、特に集落抽出法あるいは二段抽出法における推定量の分散という視点から、また非サンプリング誤差については、調査不能と質問紙作成という二つの視点から、いくつかの調査研究を取り上げ、論ずることとした。なお、今年度発表された調査研究のうち、上記視点に関連するものについてだけであっても、その数は膨大であり、すべてを取り上げることは不可能である。本稿で取り上げた調査研究は、その中からいくつかを任意に選んだものである。また、結果として、多くの論文では必要な情報の記述が欠けているという指摘をすることになってしまったが、各論文著者の意図の読み違えや、重要な記述の読み飛ばしをしていたり、誤った憶測をしていたとしたら、それらの誤謬はすべて筆者の責任であり、ご容赦いただきたい。

サンプリング誤差

1. 集落抽出法と二段抽出法

学校を通した調査研究 教育心理学の分野では、児

童・生徒やその保護者、あるいは教員を対象とした調査研究が数多く行われている。それらの調査では、学校を通して調査対象者にアプローチすることが多い。調査対象者へのアプローチのしやすさなど、サンプル当たりの回収コストの低さが、その主な理由の一つであろう。実際、TABLE 1に示すとおり、レビュー対象とした『教育心理学研究』4誌に掲載されている37論文のうち、約半数にあたる18論文で、幼稚園・小・中・高等学校を通して、児童・生徒・教員を対象とした調査が実施されている。

TABLE 1 学校を通した調査を実施している論文の数

巻 号	学校を通した 調査論文数/全論文数	
第51巻第3号	7論文	/ 11論文
第51巻第4号	4論文	/ 8論文
第52巻第1号	4論文	/ 8論文
第52巻第2号	3論文	/ 10論文
合 計	18論文	/ 37論文

さらに、学校内で調査対象の一部をサンプルとして抽出する場合には、調査の容易さという利点があることから、学級というまとまり・集落を利用して抽出を行い、選ばれた学級に在籍している全児童・生徒をサンプルとすることが多いであろう。すなわち、学校を通した調査では、サンプルの抽出方法として、調査対象個人を抽出単位とするよりも、学校あるいは学級を最終抽出単位とした一段あるいは二段集落抽出を採用することが多いと言ってよい。

実際の調査場面では、現実の様々な制約条件があるため、厳密な確率抽出を行えないことがほとんどであろう。しかし、理論的に導かれた推定量の分散の式を吟味することで、サンプルが確率抽出法によって得られたものであるか否かを問わず、調査結果の誤差あるいは精度の評価にあたってどのような観点や情報が重要であるか、という示唆が得られる。例えば、サンプルが非復元単純無作為抽出法により得られたのであれば、母集団平均値の不偏推定量 \bar{y} の分散は、 $v(\bar{y})=(1-f)s^2/n$ によって推定

できることはよく知られている。ただし、 f は抽出率、 s^2 はサンプル分散、 n はサンプルサイズである。すなわち、調査結果を、その誤差を勘案しながら見ていくためには、サンプルサイズ n とサンプル分散 s^2 が必要であり、だからこそ、調査結果を報告する際には、両者を提示することが求められているとも言える。そこで以下では、集落抽出法と二段抽出法における推定量の分散を再確認する。

集落抽出法 まず集落抽出法 (SIC) では、集落内の等質性を表す級内相関係数 ρ (繁榊・柳井・森 (1999) などを参照) が正の場合には、同じサンプルサイズの単純無作為抽出法 (SI) に比べ推定量の分散が拡大してしまい、集落の大きさがほぼ同じ (\bar{M}) であるとき、そのデザイン効果 (design effect; 分散の比) を次式で近似できることはよく知られている (Cochran, 1977, p. 242; Särndal, Swesson, & Wretman, 1992, p. 132)。

$$\text{deff}(\text{SIC}, \text{SI}) \approx 1 + (\bar{M} - 1)\rho \quad (1)$$

例えば、学級という集落を考えるとすると、平成16年度学校基本調査速報値によれば、1学級あたりの児童・生徒数の平均は、小学校が26.3人、中学校が31.0人となっている。簡単のため $\bar{M} = 31$ とすると、仮に級内相関係数が $\rho = 0.1$ の場合には、

$$\text{deff}(\text{SIC}, \text{SI}) \approx 1 + (31 - 1) \times 0.1 = 4$$

となって、同じサンプルサイズの単純無作為抽出の場合に比べ、推定量の分散は4倍に拡大してしまう。このことは、逆に言えば、多くの統計学の教科書にあるような単純無作為抽出を仮定した仮説検定などを行おうとするのであれば、サンプルサイズを実際の1/4とみなさなければならない、ということの意味する (Kish, 1965, p. 577)。仮に200人のサンプルを回収したとしても、学級を集落抽出していたのであれば、その誤差の大きさは、個人を単純無作為抽出して回収した50人と同程度、ということである。さらに学級内では児童・生徒の回答がほとんど一致しており、 $\rho = 1$ に近づけば、分散は単純無作為抽出法の場合の30倍近くにまで跳ね上がってしまう。したがって、学校や学級を集落抽出した調査の結果を報告する際には、少なくとも、学校や学級といった集落の大きさ (\bar{M}) と、集落内の等質性 (級内相関係数 ρ) を提示することは必須であり、その上で、集落抽出法というデザインを考慮した分析が必要である。

二段抽出法 また二段抽出法では、簡単のため抽出法は各段ともに単純無作為抽出とし、第一次抽出単位

(PSU)、第二次抽出単位 (SSU) の抽出率をそれぞれ f_p 、 f_s とすると、母集団平均値の推定量 \bar{y} の分散を次式で推定できることはよく知られている (Cochran, 1977, p. 278)。

$$v(\bar{y}) = \frac{1-f_p}{n_p} s_p^2 + f_p \frac{1-f_s}{n_p n_s} s_s^2 \quad (2)$$

ただし、 n_p と n_s はそれぞれサンプルに含まれる PSU の数と (各 PSU における) SSU の数であり、 s_p^2 と s_s^2 はそれぞれサンプルにおける PSU 間の分散と (PSU 内の) SSU 間の分散である。(2)式が意味することは、PSU の抽出率 f_p が非常に小さく 0 に近い場合には、(2)式の第2項は f_p が乗じられているため無視してよく、推定量の分散は PSU 間の分散 s_p^2 とサンプルに含まれる PSU の数 n_p のみによって推定できる、ということである。例えば、平成16年度学校基本調査速報値によれば、全国に小学校は23,420校、中学校は11,102校ある。母集団をこれらに在籍する児童・生徒とし、サンプルとして数校から数十校を選んだのであれば、PSU である学校の抽出率 f_p は非常に小さく $f_p = 0$ とみなしてよい。このとき、調査結果の精度を評価するために必要なのは、回収された児童・生徒の総数 $n_p n_s$ ではない。分散は、サンプルとして選んだ学校の数 n_p と学校間の分散 s_p^2 だけに依存しており、もし、回収された児童・生徒の総数 $n_p n_s$ だけしか報告されていなければ、調査結果の精度は評価できない。さらに、調査を実施したのが仮に1校のみである場合には、PSU である学校間の分散 s_p^2 を計算することができず、調査結果の精度を評価することは不可能となってしまふ。

このような、調査対象の選び方・サンプリングデザインとして、学校や学級といった集落を利用した場合には、データの独立性が失われてしまうため、分析にあたって集落という性質を考慮する必要があることは、吉田(1995)や栗田(1999)、柳井(2000)などによって繰り返し指摘されてきた。また、集落抽出とは異なるが、個人内の反復測定データという文脈では、千野(2003)や、2004年度日本教育心理学会第46回総会の自主シンポジウム「個人内の変化・共変データの解析」において杉澤(2004)が、データの階層構造を考慮した分析を行う必要性を指摘している。しかし、レビューした調査研究の多くでは、分析や結果の記述にあたって、集落をあまり考慮していないような印象を受けた。以下では、学校を通して実施された調査研究のうちのいくつかを概観する。

2. 小学校における調査研究

萩原(2004)は、今回レビューを行った調査研究の中では、級内相関係数に言及している唯一のもののようなのである。すなわち IRT の枠組みにおいて、受験者の特性値を第一次抽出単位での特性値と第二次抽出単位での特性値

とに分解するモデルを、小学校6年生を二段抽出した「基礎学力」調査に適用し、学校内の分散成分と学校間の分散成分とを求めている。その結果、級内相関係数は0.157であったとしている。

藤井(2004)は、学校現場における怒りの経験を調べるため、岩手県内の小学校教師500名を対象とした郵送調査を行っている。分散分析の結果、「同僚や保護者の言動」と「多忙な仕事」については有意な性差が見られ、「多忙な仕事」については年齢の主効果も有意であったとしている。分散分析においてサンプリングデザインを考慮する必要はなかったのであろうか、という疑問が生じるが、それに答えるために、少なくとも、500名のサンプルがどのように選ばれたのかを明記する必要がある。

玉木・山崎・松永(2004)は、児童の攻撃性が社会的情報処理過程の第3段階(目標の明確化)を介してストレス反応にどのような影響を及ぼすかを調べるため、小学校4年生から6年生1,079名を対象とした調査を実施している。1,079名というサンプルサイズは、仮にサンプルが単純無作為抽出法により得られたものであれば、ひとまず少なくはないと言ってよいであろう。しかし、もし学校を集落抽出しており、その抽出率が非常に小さいのであれば、全体のサンプルサイズだけでなく、対象となった学校数が、まず第一に重要な情報であることは先述のとおりである。学年別・性別のサンプル構成を示す前にまず、どのようにしてサンプルが選ばれたのかという記述が必要であろう。

山田・森嶋・岩間・芳賀(2004)、森嶋・岩間・山田・芳賀(2004)、岩間・山田・森嶋・芳賀(2004)は、その一連の研究において、授業不安と級友との関連を調べるため、都内小学生3～6年生425名を対象に質問紙調査を実施し、414名から有効回答を得ている。授業不安に関する項目である「先生に指名されたとき」や級友との関係に関する項目である「学校と一緒に来たり、帰ったりする友達がいる」などは、学校内あるいは学級内の等質性・級内相関係数が高いと考えられ、(どのように選んだのかは分からないが)サンプルの選び方によっては、調査結果の誤差が大きいのではないかと懸念される。

3. 中学校における調査研究

本間(2004)は、いじめ加害者の特徴やいじめ加害の停止と関連する要因を明らかにすることを目的として、中学校2校に所属する中学生1,245名を対象とし、学級単位で質問紙調査を実施している。いじめに関する回答は、学校内あるいは学級内での等質性がかなり高いのではないかと、といったことを多くの人が疑うであろう。1,245名という生徒数から逆算すると、学級数はおそらく40学級前後であろうが、PSUである学校数がはたして2校でよ

かったのだろうかという疑念は残る。そのような疑念を払拭するためには、各質問項目の級内相関係数を検討し、集落を考慮した分析方法を試みる必要があるだろう。

牧・関口・山田・根建(2004)は、中学生における無気力感と主観的随伴経験との関係を調べるため、茨城県内の中学生447名や東京都内の中学生1,148名などを対象とした質問紙調査を実施している。主観的随伴経験尺度の作成やその妥当性検討などのために、異なるサンプルを対象に少なくとも4回の調査を実施している。447名や1,148名といった各調査のサンプルサイズが十分大きなものであるのか否かといった判断は、そのサンプルを選び出した手続きに依存するが、3つの調査では明確な記述が見当たらない。もし在籍する中学校を通して実施していたのであれば、仮に各調査のサンプル校は1校だけであったとしても、すべての調査データを用いることで、学校間の分散を求めることはできよう。「部活で自分なりに頑張っていたのに、先生に非難された」など、学校間の違いが認められるのではないかと考えられる質問項目の級内相関係数が、実際にどの程度であるのか検証し、結果を示すことは、本調査研究が目的の一つとしている尺度の標準化にあたって必要なだけでなく、他の同様の質問項目を用いた尺度の開発にあたっても有益である。

柴橋(2004)は、中学・高校生の友人関係の中での「自己表明」と「他者の表明を望む気持ち」に、どのような心理的要因が関連しているのかを調べるため、私立A中学・高校の中学2年生から高校2年生まで717名を対象に、集合調査を実施している。A校は「入学者の学力、家庭環境にはかなり幅がある」ということであるが、重要なのは、対象校内での調査項目に対する回答の幅(分散)が、母集団におけるそれと比べて十分に大きいか、ということである。A校を選んだ理由である「中学と高校でなるべく学校の風土、環境が大きく変わらないこと」を逆にうがった見方をすれば、母集団における学校間の違いは大きい、ということにもなる。調査は、残念ながら1校でしか実施されていないため、さらに別の学校でも調査を実施し、学校間の違いの有無を検証していく必要がある。

本田・石隈・新井(2004)は、中学生の親・教師・友人との対人関係に対するピリーフ(「私は絶対に～しなければならない」など)が、援助要請行動・適応感に与える影響を調べるため、二つの中学校の生徒380名を対象にした質問紙調査を実施している。その内容は、親・教師・友人それぞれとの関係に関する援助要請行動尺度、ピリーフ尺度、学校生活適応感尺度から成っている。親との関係に関する内容はともかく、教師や友人との関係に関する質問項目については、ある程度の級内相関係数の高さが予

想される。本研究では、いくつかの F 検定を行っているが、生徒が在籍する学校あるいは学級を考慮した場合にも、はたして同じ結論が導かれるかどうか確かめる必要はあるであろう。

丹波・田中・小杉(2004)は、社会的スキルと、ストレスナーとなりうるイベントの体験およびストレスナーに対する負担度との関連を調べるため、中学生312名を対象に集合自記式調査を実施している。本調査に用いた中学生用ストレスナー尺度は、「教師との関係」「友人関係」「部活動」「学業」の4つの下位尺度から成っている。このような学校生活に密着した調査項目は、学校内での等質性が高いと予想されるため、おそらく集落抽出されたであろう312名が、ある一つの中学校から得られたのか、あるいは複数の中学校から得られたのかといった記述は、最低限の情報として必須であろう。逆に、調査対象となった中学生の年齢の平均が13.2歳であることや、その標準偏差が0.89~0.93であるといった情報は、中学生の年齢が普通12歳から15歳であることを知っている一般的な日本人を读者として想定する場合には、紙幅も限られているのであるから、削除してよいのではないだろうか。また、学級担任がホームルーム時間中に集団法にて実施したにもかかわらず、有効回答が78.8%にとどまり、66名もの無効が生じた理由についても探る必要がある。

出口・木下・吉田(2004)は、生徒の社会性を向上させることを目標とした授業の効果を調べるため、中学校における2学級の生徒80名を対象とした質問紙調査を実施している。調査では、学級適応感を測定するために「授業満足度」「学級に対する帰属度」「学級連帯性」「生活・授業態度」といった下位尺度に分類される項目群を用いている。このうち、例えば「学級連帯性」などは学級内の等質性が高い項目の典型と考えられるため、調査対象が2学級のみであるということが分析結果にどのような影響を与えているのか吟味する必要がある。

齋藤(2004)は、英語学習に対する学習動機を探るために、中学2年生131名を対象とした質問紙調査を実施している。その結果、「英語でわからないことはそのままにしておきたくないから」といった充実志向や、「英語は、いずれ仕事や生活の役に立つと思うから」といった実用志向の項目については、「はい」という回答の割合が高く、一方「英語の成績がよければ、こづかいやほうびがもらえるから」といった報酬志向や関係志向の項目については、その割合が低い、という傾向を見出している。こういった傾向は、学級や学校の間で大きな違いがあるとは考えにくいいため、サンプルとなった131名の中学2年生の選び方がどのようなものであったとしても、結論

自体は変わらないかもしれない。しかし、結果数値の信頼性は、その数値を得るための手続きによって判断されるものである。そのため、サンプルの抽出手続きについても明記するのがよいであろう。

濱口(2004)は、攻撃行動のうちの反動的攻撃を測定するための尺度の作成と、その信頼性・妥当性の検討を目的として、中学校5校30学級の生徒1,069名を対象とした自記式調査を実施している。攻撃性に関する調査内容は、学校間あるいは学級間で回答にさほど大きな差が生ずるとは考えにくく、級内相関係数は0に近い値であろうと思われる。5校30学級というある程度の大きさのサンプルを確保しているのであるから、学校間また学級内の分散成分がどの程度の大きさであるのか確認してみるのも興味のあるところである。

4. 高等学校における調査研究

永瀬(2004)は、crowdednessタイプの密度概念の質的理解の発達過程を明らかにするため、中学校3年生を除く小学校5年生から高校2年生までの児童・生徒399名を対象に、学級ごとに調査を実施している。そして例えば、均等分布課題の正答率は、小学校6年生から中学校2年生にかけてはほとんど変化がなく、高校1年生で一時的に低下し、高校2年生になると回復する、といった発達に伴う変化を横断的に調べている。「サンプリングにあたっては、横断的に抽出された各サンプルが等質となるよう十分配慮した」とあるが、高校1年生での正答率の落ち込みは、サンプルとなった国立附属中学校と公立高校との間の学力差ではないかという疑いは残る。さらに、集落抽出であることから、学年ごとの正答率の分散は、単純無作為抽出を仮定した場合よりもかなり大きく、本来であれば統計的に有意な差とはならないのではないかと予想される。サンプルとなった学校の数は、明記されていないためあくまで憶測ではあるが、おそらく学校種ごとに1校ずつであろう。そうであるとするならば、PSUである学校間の分散を推定することができず、残念ながら上記の疑問に答えることはできなくなってしまう。

高木・速水・木野(2004)は、2つの高校の各1学級、計76名を対象に、他者軽視傾向に基づく仮想的有能感尺度の妥当性を検討している。尺度の項目として例示された「自分の周りには気のきかない人が多い」などは、教室内で調査を実施した場合には、同級生のことを尋ねているものと受け取られ、まさに級内相関係数が高くなると考えられる項目の一例であろう。高橋(2004)は、学校生活への適応に影響する人格特性と親子関係との関連を調べるため、高校1年生320名を対象に、集合自記式調査を実施している。調査は、自己愛傾向尺度、学校生活

満足度尺度、日本版 MLAM 承認欲求尺度、親子関係診断尺度、という4つの内容から成るが、このうち特に学校生活満足度尺度は、調査対象校の校風に応じて、学校内の級内相関係数が高くなるのではないかと予想される。もし集落抽出をしていたのであれば、少なくとも、調査対象が何校なのかといった記述は必要であろう。

本多・井上(2004)は、高等学校における学級集団への帰属意識を測定する尺度の構成を目的として、高校生250名を対象に、学級適応診断テストや学校生活適応感尺度などを用いた調査を実施している。学級集団への個人の帰属意識は、学校あるいは学級によって大きく異なると同時に、学級内ではより等質になるものと考えられる。例えば、「クラスの皆が仲が良い」や「クラスの中でケンカが絶えない」などといった項目は、学級内の等質性は明らかに高いであろう。これらの項目について、因子分析の結果得られた因子負荷量だけではなく、級内相関係数の値も提示することは、当該調査結果の精度を評価するために必要なだけでなく、新たな調査の企画にあたって参考となる。

伊田(2004)は、基本的学習観および学習行動と、課題価値との関連を調べるため、5学級の高校生178名を対象とした質問紙調査を実施している。相関係数を用いて関連の程度を調べているが、特に高校では、学業に関する質問項目は学校内の等質性が高いと考えられ、サンプル学級の抽出方法が結果に影響しているのではないかと懸念が生ずる。少なくとも、対象となった5学級が、5校の各1学級なのか、1校の5学級なのかといった情報を明記する必要がある。

調査不能

調査では、計画したサンプル全体から漏れなく有効回答を得られることはほとんどなく、サンプルの一部はしばしば調査不能となる。この調査不能がどのような経緯・理由によって生じたのか、また、調査不能となったサンプルは、回収サンプルと比較してどのような異なる特徴を持つのか、を検討することは、計画サンプルが母集団からの確率抽出によって得られたものであるか否かを問わず、調査結果の一般化ということを考えると重要な作業である。現実には、調査不能サンプルは、調査への協力が得られず、調べられなかったからこそ調査不能なのであって、それが回収サンプルとどのように異なるのかを具体的に明らかにすることはかなり難しい。しかし少なくとも、調査不能がどういった理由で、どの程度の割合で発生しているのか、デモグラフィックな特性など知り得る情報についてはどのような偏りが見られるのか(あるいは見られないのか)、といったことには最低限触れ

るべきであろう。

回収率すら明記されていない論文も多数見受けられる中で、河村・鈴木・岩井(2004)は、調査不能が生じた理由についてまで言及しているという意味では数少ない好例である。調査は、学校現場において、生徒の態度に対して教師が抱く陰性感情と対応行動との関係を明らかにすることを目的として実施された。12の中学校の教師を対象に、不快感喚起場面尺度と教師特有のビリーフ尺度から成る質問紙調査を実施しようとしたところ、3つの中学校長から「質問内容が生徒の態度の良くない場面に偏っているので、回答者が不快になる。教員から校長への反発を招く恐れがある」として、調査を拒否されたという。さらに協力が得られた中学校において回収率が44.3%にとどまった理由も、調査内容に対する不快感によるものではないかと分析している。このような調査不能の検討は、検討結果を数値化してデータの分析に直接反映させることは難しいが、分析結果を眺める上で非常に有用である。例えば、仮に、調査を拒否した理由が不快感を感じたためであるとすれば、調査不能となった教師は不快感を感じやすいのかもしれないし、回答者と調査不能者とは不快感への対処方法が異なるのかもしれない。とすると、著者らが得た「教師の感じる不快感が大きいほど生徒への介入が強いという結果」を、調査不能となった教師も含めた母集団にまでそのまま一般化することには、かなり慎重な姿勢をとる必要があることになってくる。

瀬戸・石隈(2004)は、中学校におけるチーム援助のコーディネート行動の具体的な活動内容を明確にすることなどを目的に、平成10・11年度スクールカウンセラー活用調査研究委託校に指定された中学校148校の学年主任、生徒指導主任、教育相談担当の長、養護教諭、スクールカウンセラーを対象に学校を通じた郵送調査を実施している。その結果、65.5%にあたる97校から回答を得たとしている。調査対象校が、スクールカウンセラー活用調査研究委託校であり、また調査内容が、生徒の問題に対する援助活動であるにもかかわらず、計画サンプルの3校に1校もの割合で調査不能が生じた理由について、論文中では触れられていないが、精査すべきであろう。著者らは、高校において実施した同様の調査では、コーディネート能力・権限が5因子であったのに対し、中学校を対象とした本調査研究では4因子となった理由を、中学校と高校という学校種の違いに求めている。しかし、あくまで例えばの話ではあるが、調査不能となった学校では「状況判断」を行う立場の者がチーム援助には積極的ではなく、そのために調査に協力しなかったのであるとすれば、因子構造の違いは学校種の違いではな

く、調査への協力・非協力の違いを反映したものとなってしまふ。そのような仮説の真偽の判断材料の一つとして、(結果としてあるいは役に立たないかもしれないが)調査不能理由を検討することは不可欠な作業である。

前田・田頭・三浦(2004)は、日本の高校生英語学習者の語彙学習方略使用について、その一般的傾向を明らかにすることなどを目的に、15高等学校の1,177名を対象とした調査を実施している。英語の学習成果を測定するためのC-test得点のサンプル平均や標準偏差が、他の二段抽出された38校1,584人を対象とした調査のそれらと大きく異なることを理由に、調査データが母集団を反映しているとしているが、一般に調査結果の偏りの程度を判断する材料は、サンプルの選び方(とそれに応じた推定量)と、回収者と不能者との間の特性の違いである。サンプルとなった15校は、個人的に依頼を行ったとされているが、なぜどのようにしてその15校を選び、選ばなかった他の高校とはどのような違いがあるのか、他に依頼して拒否された学校はなかったのか、また、大きく記入漏れがあったために無効な回答としたサンプルの割合はどの程度だったのか、どのような記入漏れがなぜあったのか、といった情報がまず必要であろう。回収サンプルの中だけでその性質をいくら調べても、偏りの有無を判断することは難しい。調査内容に鑑みれば、得られた結果に実際に大きな偏りがあるとは考えにくい、結果の信頼度を高めるためには、より詳細な調査手続きの記述が必要であろう。

請川・星(2004)は、小学校においてクラスが機能しないような状態(学級崩壊)があるのかどうか、またあるとしたらそれはどのような要因によると考えているか、を調べるため、北海道内の小学校教師を対象に郵送調査を実施している。有効回答数は792であり、配付した1,489校に対する回収率は53.2%と、半数近い調査不能がある。本調査の内容は、学級崩壊という小学校教師にとっては切実な問題に関するものである。それにもかかわらず調査に応じなかった学校あるいは教師が、調査に協力した学校あるいは教師とどのような点で異なるのかを探ることは、得られた結果の一般化をする上で是非とも必要な作業である。そのためには、例えば、700校近い調査不能のうちの何校かに、調査に協力しなかった、あるいはできなかった理由などを直接尋ねてみるのが有効であろう。調査不能が、調査の実施手続き・手順の不備によるものなのか、あるいは回答したくない調査内容だからなのか、単に忙しさに紛れて忘れてしまっただけなのか、といった手がかりは得られる。こういった調査不能理由の探索結果は、具体的に何らかの数値として表現することは難しいけれども、実際に調査を行った者だけが

知り得る情報として、適切なデータ分析方法の選択や結果の解釈・表現にあたって非常に役立つ。

岩原・八田・川上(2004)は、高齢者の生涯学習に対する意識構造を調べ、なぜ社会教育の機会を利用する人が少ないのかを検討するため、東大阪市民300名を対象に郵送調査を実施している。有効回答数は133件であり、44%という有効回収率は、一般市民を対象としたこの種の調査としては必ずしも低くはない。しかし仮に、調査不能となった半数以上の調査対象者が、調査内容に無関心であるがために調査不能となったのであり、回答者は生涯学習や社会教育に関心のある者に限られていたとすれば、調査目的に鑑みて得られた結果は大きな偏りを持つことになってしまう。調査不能者の特性を調べることは、常に困難なことではあるが、例えば調査票の早期返送者と後期返送者の回答内容を比較するなどして、その特性を探る試みをすることは重要であろう。また調査では、社会教育の機会を利用しようとしめない原因・理由を直接質問している。そういった原因・理由こそ研究者が最も知りたいことではあるが、一般論として、理由を直接尋ねることは避ける方がよいであろう。回答者は、明確な理由に基づき行動しているわけではなく、漠然と考えていることが多かつたりする(西平, 1985, p. 30)からである。さらに、理由を質問すること自体が、回答者のその後の回答行動に影響を与えてしまうという研究結果もある(Wilson, Dunn, Kraft, & Lisle, 1989; Wilson, LaFleur, & Anderson, 1996)。

質問紙作成

今回レビューを行った調査研究では、サンプルサイズが数百を超える場合にはほとんどが、回答者自らが質問紙に回答を記入する自記式調査法を採用していた。自記式調査法では、質問紙の見目なども含めた、あらゆる意味での質問紙の作り方が、回答者の回答行動に影響を与えることがある。

林(2004)は、郵送調査法における返送率という観点から、質問紙や調査実施方法の影響に関する数多くの調査研究を紹介している。サンプルサイズが小さい調査が多いため、統計的には有意な差が見られないものが多いが、それらの結果は、あらゆる細部にわたって質問紙作成の一般的な指針を再度確認する必要性を感じさせる。

また、レビューを行った多くの調査では、4件法や5件法など段階評定に基づく尺度を用いており、我が国における伝統的な形式として、おそらく、選択肢に付せられた数値ラベルを選んでもらう形式をとっていると思われる。Schwarz, Knauper, Hippler, Neumann, & Clark(1991)によるあまりにも有名な実験調査の結果は、各選

択肢に付与する言語ラベルに加え、数値ラベルも回答者が択肢の意味を解釈する上で影響を与えるというものであり、質問紙上での評定尺度の提示方法には十分な配慮が必要である。例えば、城(2004)は、大学での公開授業への参加体験が、高校生の進路形成にどのような影響を与えるのかを調べるため、高校生319名を対象に、大学に進学したい理由の変化を体験授業前後および1年後に調査している。その際、+2を「全くそのとおり」、-2を「全くあてはまらない」とする4件法を用いたとしている。Schwarzらの研究結果に鑑みると、(おそらく用いられたであろう)+2,+1,-1,-2という数値ラベルは、その間隔が一定でないため回答に何らかの影響を与えたのではないかと、また数値ラベルが「全くそのとおり」や「全くあてはまらない」といった言語ラベルとは干渉しなかったのだろうか、という疑念が生ずる。

上地・竹中・鈴木(2004)は、子どもにおける身体活動の行動変容段階と意思決定バランスの関係を検討するため、子ども用身体活動行動変容段階尺度および子ども用身体活動の恩恵・負担尺度を開発している。子ども用身体活動行動変容段階尺度は、調査対象となった小学校4年生から6年生にも回答しやすいようフローチャート形式を採るなど、興味深い工夫がなされている。通常の濾過質問を用いる代わりにフローチャート形式を用いることで、どの程度回答ミスなどの非サンプリング誤差が軽減されているのか興味あるところである。ただし、「いつから毎日30分以上身体を動かす始めようと思っていますか」といった質問項目の表現には、さらなる工夫が必要であろう。「身体を動かす」ことの意味には遊びなどが含まれている、と説明してはいても、「1ヶ月以内」や「2ヶ月以内」といった将来の特定時期に関する内容を尋ねること自体が、「身体を動かす」ことの意味を特定したり、あいまいにしてしまうおそれがあるからである。

おわりに

統計的な調査を実施しデータを得るという作業には、費用はもちろんのこと、時間や労力などを含め膨大なコストが必要であり、その上、確率抽出法の理論に忠実に従ったサンプリングは困難な場合が多いことは、筆者自身よく了解しているつもりである。しかし、膨大なコストをかけ、回答者の善意の協力の下にデータを収集したのであるから、サンプルが非確率抽出された場合であっても、その分析結果は最大限有効活用すべきである。そのためには、当該研究が想定する母集団はどのようなものであるのか、そして、得られた調査結果がその母集団の推定値としてどの程度の誤差を持つものなのか、という判断材料を、調査実施者は提供すべきである。中村

(1999)は、「追試に際して必要な情報は全て掲載すべきである」とし、「データの分析に関する必要な情報」を記述する必要性を述べている。それらに加え、収集したデータの特徴を評価するために必要な情報も記述すべき、ということである。

本稿では、まず、教育心理学研究で多用される集落抽出法や二段抽出法では、PSUの抽出率が小さい場合には、全体のサンプルサイズだけでなく、サンプルにおけるPSUの数やPSU間の分散が、第一に重要な情報であることを述べた。しかし、レビューした調査研究の多くでは、PSUの抽出率がおそらく小さいであろうと思われる場合であっても、それらへの言及が見当たらず、重要性があまり認識されていないような印象を受けた。教育心理学の分野を対象とした統計学の教科書のほとんどでは、無限母集団からの単純無作為抽出のみを前提としており、サンプリングの理論や方法に関する記述は少ないことが、その一因かもしれない。既に収集されたデータを、そのサンプリングデザインに沿って再分析し直してみることは、少なくともサンプルにおけるPSUの数が一つでなければ可能であろうし、ぜひ行ってみるべきであろう。

また、最近の悪化した調査環境の下では、調査における回収率は一般に低くならざるを得ない。しかし、低い回収率が、必ずしも調査結果の偏りの大きさを意味するわけではない。回収サンプルと不能サンプルとの間で、分布に大きな違いがある場合にのみ、回収率の低さが問題となる。そのため、不能サンプルの特徴をできる限り知ろうとすることも、回収サンプルの特徴を評価するためには不可欠である。

さらに、調査に用いた質問紙自体も、収集したデータの特徴を評価するために必要な情報である。調査内容が同一であったとしても、質問紙の作り方が変われば、それによって調査結果が変わるおそれがあるからである。本来であれば、『教育心理学研究』には、用いた質問紙そのままを掲載すべきである。紙幅の都合でそれは不可能な場合がほとんどであろうから、例えばホームページからダウンロードできるようにするなど、質問紙そのものを入手できるようにする何らかの工夫が必要であろう。

引用文献

- 千野直仁 2003 反復測定データの分析とその周辺 教育心理学年報第42集, 107-118.
 Cochran, W. G. 1977 *Sampling techniques* (3rd ed.). New York : John Wiley & Sons.
 出口拓彦・木下雅仁・吉田俊和 2004 生徒の社会性を育む授業の効果1—質問紙による測定— 日本教育心

- 心理学第46回総会発表論文集, 395.
- 藤井義久 2004 小学校教師の怒り経験に関する実証的研究(1)—小学校教師版怒り尺度の開発— 日本教育心理学会第46回総会発表論文集, 380.
- 萩原康仁 2004 「基礎学力」調査の項目困難度と分散成分の推定 日本教育心理学会第46回総会発表論文集, 459.
- 濱口佳和 2004 反応攻撃性 (reactive aggression) 尺度(中学生版)の作成—反応的・道具的攻撃性尺度(RIS 中学生版)の改訂(2)— 日本教育心理学会第46回総会発表論文集, 493.
- 林 英夫 2004 郵送調査法 関西大学出版部.
- 廣瀬英子 2004 測定・評価に関する研究動向と展望—テスト研究と評価研究— 教育心理学年報第43集, 99-106.
- 本多公子・井上祥治 2004 高等学校における学級集団帰属意識尺度作成の試み 日本教育心理学会第46回総会発表論文集, 181.
- 本田真大・石隈利紀・新井邦二郎 2004 中学生の対人関係に対するピリーフが援助要請行動、適応感に与える影響 日本教育心理学会第46回総会発表論文集, 263.
- 本間友巳 2004 中学生におけるいじめの停止に関連する要因といじめ加害者への対応 教育心理学研究, 51, 390-400.
- 伊田勝憲 2004 高校生版・課題価値測定尺度の妥当性検討—基本的学習観および学習行動との関連から— 日本教育心理学会第46回総会発表論文集, 433.
- 岩原昭彦・八田武志・川上綾子 2004 高齢者の生涯学習に対する意識構造 日本教育心理学会第46回総会発表論文集, 431.
- 岩間恵美子・山田睦子・森嶋尚子・芳賀明子 2004 小学生の授業不安に関する調査研究III 日本教育心理学会第46回総会発表論文集, 696.
- 城 仁士 2004 高校生のキャリア発達支援活動の進路意識への影響(2) 日本教育心理学会第46回総会発表論文集, 347.
- 河村夏代・鈴木啓嗣・岩井圭司 2004 教師に生ずる感情と指導の関係についての研究—中校教師を対象として— 教育心理学研究, 52, 111.
- Kish, L. 1965 *Survey sampling*. New York : John Wiley & Sons.
- 請川滋大・星 信子 2004 小学校のクラス機能不全をどうみるか—北海道内の小学校に向けた質問紙調査から— 日本教育心理学会第46回総会発表論文集, 262.
- 栗田佳代子 1999 実際のデータを用いた t 検定および検定力分析の「観測値の独立性」からの逸脱に対する頑健性の検討—人工データによる研究結果との対応および項目の尺度化の影響を中心に— 教育心理学研究, 47, 263-272.
- 前田啓朗・田頭憲二・三浦宏昭 2004 高校生英語学習者の語彙学習方略使用と学習成果 教育心理学研究, 51, 273-280.
- 牧 郁子・関口由香・山田幸恵・根建金男 2004 主観的随伴経験が中学生の無気力感に及ぼす影響—尺度の標準化と随伴性認知のメカニズムの検討— 教育心理学研究, 51, 298-307.
- 森嶋尚子・岩間恵美子・山田睦子・芳賀明子 2004 小学生の授業不安に関する調査研究II 日本教育心理学会第46回総会発表論文集, 695.
- 永瀬美帆 2004 密度概念の質的理解の発達—均等分布理解と関係概念の定性的理解からの検討— 教育心理学研究, 51, 261-272.
- 中村知靖 1999 測定・評価に関する研究の動向 教育心理学年報第38集, 105-119.
- 日本教育心理学会(編) 2003 教育心理学ハンドブック 有斐閣.
- 西平重喜 1985 統計調査法改訂版 培風館.
- 齋藤嘉則 2004 中学生の英語学習に対する学習動機の特徴—「学習動機を測定する質問項目」による調査分析— 日本教育心理学会第46回総会発表論文集, 430.
- Särndal, C. E., Swesson, B., & Wretman, J. 1992 *Model-assisted survey sampling*. New York : Springer-Verlag.
- Schwarz, N., Knauper, B., Hippler, H. J., Neumann, E. N., & Clark, L. 1991 Rating scales : Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570-582.
- 瀬戸美奈子・石隈利紀 2004 中学校におけるチーム援助に関するコーディネーション行動とその基盤となる能力および権限の研究—スクールカウンセラー配置校を対象として— 教育心理学研究, 51, 378-389.
- 柴橋祐子 2004 青年期の友人関係における「自己表明」と「他者の表明を望む気持ち」の心理的要因 教育心理学研究, 52, 12-23.
- 繁栴算男・柳井晴夫・森 敏昭(編) 1999 Q&Aで知る統計データ解析 DOs and DON'Ts サイエンス社.
- 杉澤武俊 2004 階層的線形モデルによる個人内反復測定データの分析 日本教育心理学会第46回総会発表論文集, S 80-S 81.
- 高木邦子・速水敏彦・木野和代 2004 仮想的有能感尺度の妥当性検討 日本教育心理学会第46回総会発表論

- 文集, 33.
- 高橋美知子 2004 高校生の学校生活への適応に影響する自己愛傾向について—親子関係診断尺度 (EICA) との関連から— 日本教育心理学会第46回総会発表論文集, 99.
- 玉木健弘・山崎勝之・松永一郎 2004 教師評価による小学生の攻撃性が社会的情報処理過程ならびにストレス反応に及ぼす影響の検討 日本教育心理学会第46回総会発表論文集, 381.
- 丹波秀夫・田中健吾・小杉正太郎 2004 中学生の社会的スキルとイベント型ストレスラーとの関連 日本教育心理学会第46回総会発表論文集, 394.
- 上地広昭・竹中晃二・鈴木英樹 2004 子どもにおける身体活動の行動変容段階と意思決定バランスの関係 教育心理学研究, **51**, 288-297.
- Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. 1989 Introspection, attitude change, and attitude-behavior consistency : The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology : Vol. 22*. Orlando, FL: Academic Press. Pp.287-343.
- Wilson, T. D., LaFleur, S. J., & Anderson, D. E. 1996 The validity and consequences of verbal reports about attitudes. In N. Schwarz & S. Sudman (Eds.), *Answering questions : Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass Publishers. Pp.91-114.
- 山田睦子・森嶋尚子・岩間恵美子・芳賀明子 2004 小学生の授業不安に関する調査研究 I 日本教育心理学会第46回総会発表論文集, 694.
- 柳井晴夫 2000 因子分析法の利用をめぐる問題点を中心にして 教育心理学年報第39集, 96-108.
- 吉田寿夫 1995 学校教育に関する社会心理学的研究の動向—研究法についての提言を中心として— 教育心理学年報第34集, 74-84.

付 記

草稿を読んで有益なコメントをしていただいた統計数理研究所前田忠彦助教授に感謝いたします。