

センシティブな社会調査における使い捨て項目の 効果について

吉村治正*・船木昭夫**

Influences of Throwaway Questions for Sensitive Social Survey.

Harumasa YOSHIMURA and Akio FUNAKI

要 旨

本稿では、社会調査において「使い捨て質問 (throwaway question)」を導入することで、回収率や偏りにどのような影響が見られるかを、実際の社会調査に付随して行った実験によって検証した。調査票の構成によって調査票にコンテキストを持たせること、これは従来バイアスの源泉であるとして否定的に扱われてきた。ところが近年、調査票にコンテキストを持たせることによって回答者の回答行動が促され、回答拒否や欠損の発生を抑制するという議論が現われている。本稿では、この新しい見解を立証すべく、一般人口を対象としたセンシティブなトピックの社会調査において二種類の調査票を作成、配布して回収状況を比較した。結果は、「使い捨て質問」を用いてコンテキスト形成を行った調査票の方が、コンテキスト形成を行わない調査票に比べ、有効回収率で有意に高い回収率を示した。他方、回収率が上がることによって心配された回答の偏りについては、明確なものは見つからなかった。

【キーワード】 社会調査、コンテキスト効果、非回答誤差

1 問題の所在

本稿では、社会調査 (social survey)¹において「使い捨て質問」を導入することで、回収率や偏りにどのような影響が現れるかを論じる。

調査票の作成にあたって個々の質問項目をどのように並べるべきか。社会調査の実施に携わる際に必ず直面する、この極めて素朴な疑問に対する回答は、実は社会調査に精通した専門家たちの間では対極に分かれる。一方には、調査項目の順序は考慮する必要がないとして、調査項目をどう並べるかという問いかけそのものを否定する立場がある。そして他方には、質問項目はその内容に応じて、一定のパターンで配列し、全体としてのコンテキストを作成しなければならないとする立場がある。この後者の立場に立つSheatsleyは、次のように述べる。

最初の質問は簡単で当たり障りのないものにする。最初の質問は、回答者がインタ

ビューアーに接する最初の機会であり、これによって、これから行われる課題（インタビュー）の中身のトーンが決まる。もしも回答者が最初の質問を難しい、自分の知識や経験では答えられない、あるいは何らかの形で困惑したり圧迫されていると感じるのであれば、彼らはそこで回答をやめてしまうことだろう。他方、もしも最初の質問が簡単で答えることに不快感を感じなければ、そのまま続けることに前向きになる。このため、調査者は、難しい質問やセンシティブな項目に進める前に、回答者に準備をさせ、自分の役割に慣れさせることを目的に、使い捨ての質問（throwaway question）を用意することも珍しくない。こうしたルールの示すものは簡単である。つまり、回答者がインタビューにのめり込んでこなければ、難しい質問やセンシティブな項目を聞くことはできない、ということである。（Sheatsley 1983, 220-221）

この言明は、二つの重要な論点を含んでいる。第一に、Sheatsleyは調査票をコンテキストを持つ総体（entity）としてとらえ、そのコンテキストに対する反応として回答行動を把握している。これは、調査票を個々の調査項目の集合体（aggregate）とみなし、コンテキストの存在を認めない行動科学の基本発想と相いれないものである。第二に、Sheatsleyはコンテキストを作成するという目的のために、集計の必要性が想定されていない項目、実際にデータ分析の場面で使用される予定のない項目を加える事に肯定的である。「使い捨て質問」というのは、要するに集計も分析も想定されていない、調査の目的に照らして直接的に必要な項目、端的に言えば無駄な項目ということである。無駄な質問項目を含めると調査票は冗長になり、回収率は低下し測定誤差も増大する。したがって調査票は可能な限り短くすべしというのが、一般に知られている調査票作成の鉄則である。この常識に従えば、データとして分析する予定のない項目は、調査票の作成段階でまっさきに削除されなければならない。ところがSheatsleyは、この分析に使われる予定のない無駄な調査項目を導入することで、かえって回答者の回答行動が促されるというのである。これは、明らかに従来の社会調査の常識に反する。

社会調査は、ある行動科学上の命題が成立することを前提としている。それは、調査対象となる個人の行動や態度、価値観などは、調査が実施される状況や環境に依存することなく比較的安定した状態で存在し測定することが可能である、というものである。回答は回答者によってのみ決まり、調査実施の状況的要因、例えば前後の質問の内容や調査主体、調査員（訪問調査の場合）の印象などの影響を受けることはないし、また受けてはならない。もしも調査員によって回答者の回答が変わるのであれば、同じ調査員が担当した回答者の間でしか比較分析ができなくなるし、調査主体によって回答が変わる事があれば、複数の社会調査の結果を対比することは無意味になる。同様に、もしも前後の質問の内容によって回答が変わるのであれば、前後の質問が同じ順序で並べられた調査同士でしか、個々の集計結果を比較することはできなくなる（Schuman et. al. 1981, 122）。つまり、この行動科学上の命題が成立すると前提することで、初めて社会調査のデータを項目ごとに集計し、比較・分析を行うことが可能になる。

冒頭で触れた、調査票のなかでどのように質問項目を並べるかという問いかけの存在そのもの

を否定する立場は、この社会調査の前提となる命題に従うことから生じる。質問項目が互いに前後の項目の影響を受けないということは、項目の順序はランダムであるということである。もしも質問項目を何らかの意図に基づいて一定の順序に配列すると、前後の質問のつながりがコンテキストを形成し、これによって回答が偏り (bias) を示すようになる。したがって、偏りを防ぐために質問項目の順序はランダムでなければならない、ということになる。

だが、こうした立場は、原理としては理解されるが、実務的な観点から見ると深刻な困難さを伴う。質問の順序によって回答が変わる、いわゆるコンテキスト効果については、すでに1940年代からその存在が知られ、問題視されてきた。そのため、このコンテキスト効果を排除するために、コンピューターを用いて回答者ごとに質問項目をランダムに並べ直して調査、集計するという試みが70年代に行われたこともあったらしい。しかし、実際に調査の現場に関わった人々からは、質問項目をランダムに配列した状態でインタビューを行うには多大な困難が伴うという強い批判がなされ、こうした試みが定着することはなかった (McFarland 1981, 86)。むしろ、調査実施の現場では、回答者から正確な回答を得ようとすれば、調査項目はある一定のパターンで並んでいる必要があるという認識が、比較的早くから広まっていた。問題は、この「一定のパターン」がどのようなパターンであり、それはどのような根拠によるのか、という点にある。

今日、調査票の構成は、簡単で当たり障りのない質問を前の方に、センシティブな質問や複雑な質問を後ろの方に配置するとしている例が多い (Biemer et. al. 2004; De Leeuw et. al. 2008; De Vaus 1990; Dillman 2009; Groves 2004; Marsden & Wright 2010; Trochim & Donnelly 2008; Weisberg 2005)。これは、何よりも調査実施の積み重ねの中で生まれてきた経験則である。したがって、研究レベルにおいては、その経験則がなぜ成立するかという問いかけが必要となる。

質問項目の順序について第一に言われるのが、簡単な質問を前に配置することで Sudman & Bradburn (1974) のいう信頼感効果 (rapport effect) が働き、これによって後半での回答拒否やごまかしの回答、いい加減な回答が減少する事が期待できるという主張である。面識のないインタビュアー (あるいは調査者) に対して、回答者は多少とも警戒感を持っており、調査開始直後にセンシティブな内容の質問を聞いても、回答拒否やごまかしの回答が多数発生する。これを防止するためには、まず比較的簡単で当たり障りのない質問と回答のやりとりを続け、インタビュアーに対する回答者の警戒感が次第に薄れていき、回答者の調査内容への関心が高まるのを待ってから、センシティブな質問を聞けばよい。つまり、最初に簡単で当たり障りのない質問を並べるのは、回答者の警戒心を解くためのストラテジーということになる。これは極めて社会的な発想で、日常場面におけるコミュニケーション過程において、初対面で相手の内面に深く立ち入るようなことを禁忌する社会的規範が存在するために、社会調査もこの規範に従う必要がある、とするものである。

簡単な質問を前に、複雑な質問やセンシティブな質問を後に持ってくる第二の根拠としてあげられるのが、前の質問への回答が、後に続く質問への回答を思い出すきっかけ (cue) となるという指摘である。この具体例として、Biemerらは、世帯所得を聞く場合には、世帯内に有所得成員が何人いるかとか、就労所得以外の所得の有無などについて先に聞くことで、より正確な回答が得られると主張する (Biemer et. al. 2004, 34)。これは実際にインタビューを行った経験のある

人であれば、すぐに得心が行く。世帯所得というのは世帯構成員各人の就労所得に加え、銀行預金の利息や投資・家賃収入などの非就労所得、住宅等の売却による一時的な所得など、構成項目が極めて多く、したがって、とっさに計算して答えるのは難しい場合が多い。そのような場合、記憶を喚起するキーワードや計算に含めるべき項目などを直前の質問で聞くことで、回答に要する心理的コストを低減する。これは複雑な質問や正確に思い出すのが難しいような調査項目を聞く場合に用いられる、認知心理学的なストラテジーといえる。

今日、こうした議論をさらに進め、調査票にコンテキストを形成させる事を積極的に評価する立場があらわれている。この代表例が、冒頭のSheatsley (1983)やDillman (2009)らである。Dillmanは次のように述べる。

無秩序に並べられた調査項目のリストと、これを回答者に対して効果的に提示する調査票との違いは、とてつもなく大きい。調査票をデザインする際には、これを受け取った人に対して、どうやったら回答者として調査へ回答してもらえるように動機づけるかという問題を考えねばならない。また、意図せざる順序効果が現れたり、特定の項目で異常に高い欠損率が現れたりといった、多岐にわたる測定上の問題が発生することも避けねばならない (Dillman 2009, 151)

Dillmanによれば、社会調査への回答は、それ自体で社会的行為であり、回答者は、回答することによって得られる効用とこれに要するコスト、心理的・社会的なリスクなどを見積もりながら、回答するかしないか、あるいは回答するとしても正確に回答するか、いい加減に回答するかといった判断を下していく。調査者 (researcher) から提示された質問に回答者が答え、その発話に対する反応として次の質問が提示される。この時に、回答者は両者の間に期待の相補性が成立していると想定する。つまり相手 (調査者) は自分 (回答者) の発話内容を理解した上で次の質問を発していると想定することで、次の質問の内容を予測しようとする。もしも次に続く質問の内容が回答者の予測から大きく外れれば、回答者は調査者が自分 (回答者) の先ほどの発話を理解しなかったか、あるいは調査者の期待を自分 (回答者) が誤って理解していたのではないかと考え、調査者が何を期待しているかを推測しなおそうとする。こうした推測のやり直しが続くと、回答者は調査者とのコミュニケーションを継続することに次第に困難さを感じるようになり、その結果、無回答や不正確な回答が発生する。つまり、回答者から正確な情報を得ようとするれば、回答者が個々の質問にどのように答えるか、そしてそれに伴ってどのように思考を展開するかをあらかじめ予想し、その思考の展開パターンに合致するように質問項目を配列し、調査票を構成していく必要がある。この点、Dillmanは「調査票は、社会規範にしたがって派生する会話のように作られねばならない」(Dillman 2009, 157)と主張するが、その「会話」は話のコンテキストも内容も全て調査者によって事前に計算されつくしたものでなければならない。

Dillmanは、調査票のコンテキストによって無回答や回答の正確さに影響が出る可能性を指摘しているが、これだけでなく、コンテキストの与え方によって個々の質問項目の内容に対する認知が変容する可能性があることも指摘されている。Leeによれば、ある特定の質問がどの程度セン

シティブであると認知されるかは、調査票全体のコンテキストに依存するという (Lee 1993)。例えば飲酒行動などは、消費行動の一部として飲酒頻度を聞かれる場合と、薬物依存に関する調査項目の中に飲酒頻度を聞く質問がおかれる場合とで、回答することに対する心理的な負担感が全く異なる (Lee 1993, 78)。したがって、センシティブな質問項目を調査する場合は、質問内容がセンシティブであるということを回答者に認知されないように調査票全体のコンテキストを形成し、また前後の質問を並べる必要がある、というのがLeeの主張である。

しかしながら、このLeeの主張は、冒頭にあげた行動科学の前提と矛盾する性質を持つ。人間の行動も態度の表出も、すべてそれが行われた状況、コンテキストに依存する、したがってコンテキストから切り離れたデータを分析・解釈しても意味がないとする原則主義的な批判は、すでに1930年代から行われているし、今日でも繰り返しとりあげられる (Bogardus 1934; Cicourel 1964; Suchman & Jordan 1990)。しかしながら、もしもこうした批判が究極的な意味において正しいとしても、これを受け入れてしまうと、社会調査 (social survey) という概念自体が成立しなくなる。

この点、Wanke & Schwarz (1997) の議論は一考に値する。彼らは、そもそもなぜコンテキスト効果が表れるのかを、Tourangeauの回答行動に関する認知心理学的モデルにしたがって整理した。Tourangeauによれば、ある人がある質問を聞かれた時、それに対する回答を発するまでに、以下の5段階の心理的過程を必要とするという (Tourangeau 2000; Wanke & Schwarz 1997; Weisberg 2005)。

- ①何を聞かれたのか、言葉どおりでの理解 (semantic understanding)
- ②その人は実際に何を知りたいのかという理解 (pragmatic understanding)
- ③回答に必要な情報を記憶層から抽出 (retrieval)
- ④抽出した情報から判断もしくは態度を形成 (judgment formation)
- ⑤態度表出のおよぼす影響を考慮し回答を再編成する (editing)

つまり、回答という行為は極めて複雑な心的過程であり、多大な心的エネルギーの動員を必要とする。そのため、質問の背後にある調査者の意図が把握できない、あるいは正確に思い出すのが難しいなど、質問の内容によって回答に大量の心的エネルギーを必要とする場合、心的エネルギーの浪費を抑えようと、様々な方策がとられる (Weisberg 2005; Tourangeau 2000)。

この一つの方法として、直近の質問を手がかりとして上記の心的過程を効率化しようとする現象があげられる。つまり、ある質問がなされたとき、上記の認知的過程を順番に逐一改めて行うのではなく、それが直近の別の質問と同じ内容であると認知することで、調査者が何を聞きたいのかという意図を推測することを容易にし、あるいは記憶層から新たに呼び出す情報量を抑制する。こうすることで、態度形成を効率化する (同化効果: assimilation effect)。あるいは、直近の質問と相反する内容であると認知することで、同様に態度形成を効率化する場合もある (対照効果: contrast effect)。これがいわゆるコンテキスト効果である。したがって、コンテキスト効果は、回答者の心的エネルギーの浪費を抑制するために常に生じているものである。反対に、もしもコンテキスト効果が働かなければ、一つの質問ごとに上記の認知心理的過程を逐一最初から

やり直すことになり、回答に要する心的エネルギーは莫大なものとなる。その結果、回答者は回答することに疲れ、たとえばKahnemanのいうヒューリスティックな判断 (Kahneman et. al. 1982) など、他の方法によって心的エネルギーの抑制を図るか、あるいは回答拒否という選択をすることになる。つまりTourangeauの認知心理学的モデルに従えば、ある程度までコンテキスト効果が働くことで正確な回答が期待できるということになる。

Wanke & Schwarz (1997) が指摘したのは、原則論を議論するよりも、回答者の行動を正確に把握することによって、様々な制約のある実際場面で効果的にコンテキスト効果をコントロールする方法を検討すべき、という点であった。つまり、もしも二つの質問の間にコンテキスト効果が生じては困るというのであれば、二つの質問が関連しているという認知を低下させればよい。具体的には、その二つの質問の間に他の質問項目 (緩衝項目: buffering item) をはさむことで、最初の質問への回答で記憶層から引き出した情報へのアクセシビリティを低下させ、新たな認知過程を行わせるように仕向ければよい。

このWanke & Schwarzの論理を延長すれば、ある質問への回答が多大な心的エネルギーを必要とする場合、そのエネルギーを効率化するために、その前にその質問と認知面で関連するような別の質問をおいてやるという方法が成立する。つまり、まず比較的認知心理学的に負担の少ない質問 (つまり質問の実際上の意図が理解しやすく、また、これに関して記憶層から情報を引き出す時も、比較的容易に行い得るような質問) を先に聞き、それに続いて本来聞きたかった質問を行うと、回答者は、その前に聞かれた質問への回答を参照する (二つの質問を関連させる) ことで、認知心理学的なプロセスを効率化することができ、結果として回答しやすくなる。つまり、これがSheatsleyのいう「使い捨て質問」の効用であり、DillmanやLeeのいうコンテキストの作成による回答の効率化ということになる。したがって、設定される使い捨て質問は、回答に要する心的エネルギーが比較的小さく (つまり答えやすく)、なおかつ次の質問と関連していると回答者に認知されるような内容でなければならない。

使い捨て質問によるコンテキスト形成の理論的根拠づけは、上記のようになされる。しかしながら、コンテキスト効果についての研究事例は極めて乏しく、これに加え、その希少な事例においても期待されていたような効果が現われる例はさらに少ないと言われている (Groves 2004, 482)。以下、筆者らが遭遇した調査の機会において、これをどのように検証したかを論じることにする。

2 実験の設定

今回の実験は、青森県自殺対策検証研究会が2013年1月に実施した、青森県居住者を対象とした郵送調査に際して行われた²。この調査は純然たる学術目的の調査というよりは、政策提言を目的として実施されたものであり、そのために調査計画の立案の時点で基本的なデザインはおおむね決まっていた³。すなわち、この調査の核心は、精神衛生と自殺の関係、もう少し具体的に言えば、精神疾患としての「うつ」を抱えることで自殺のポテンシャルが高まるかどうかにあった。日常生活における困難さに直面した個人々が、精神的に追い込まれることで次第に躁うつ病ないしは抑うつ症状を示すようになり、やがて自殺という行動へと至る。したがって、抑うつ

スクリーニングと適切な治療、つまり精神衛生を改善することで、自殺を防止できる。これは、今日の日本の自殺研究の基調をなす考え方であり、内閣府の自殺予防対策も基本的にこの考え方を踏襲したものとなっている。

しかしながら、調査のデザイン段階で問題となったのは、調査項目のセンシティブリティであった。抑うつ傾向を測定する項目として、CES-Dと呼ばれるうつ病自己評価尺度（20項目）を、自殺のポテンシャルを計る項目として、一般に希死念慮と呼ばれている項目（3項目）を用いる。他の調査との比較を考え、ここまでは異論なく決まった。ところが、この希死念慮についての質問項目は、「死について何度も考える事がありますか」、「気分がひどく落ち込んで、自殺について考える事がありますか」という二つの質問と、さらに「自殺について考えた事はない」「自殺について考えた事はあるが実行した事はない」「自殺したいと思った」「もし機会があったら自殺するだろう」という四つの選択肢の中から最近の気持ちをもっともよく表しているものを一つ選ばせる質問の、合計三問から構成されている。つまり、自殺したいかどうかを極めて直截的に聞いている。また、CES-Dにしても、元来は心療内科やカウンセリング機関などでの問診の場面を主に想定して作成されており（島 1975）、そのために、例えば「急に泣き出すことがある」「皆が自分をきらっていると感じる」など、一般人口に対して聞くには、ためらいを感じる項目も少なくない。つまり、現在の日本の社会通念では、抑うつや自殺は極めてセンシティブな事柄であり、これを果たして一般的な社会調査によって調べる事は可能なのか、という疑問が調査のデザインの段階で相次いで提示された⁴。

この点、この調査はLeeやDillman、Schwarzらの仮説、すなわち特にセンシティブな調査の場合に調査票の構成によってコンテキストを形成する事で回答者の回答行動を助長できるという議論を立証するいい機会であったと言える。そこで、まず前半に、基本属性（性別、年齢、配偶者の有無、就労か非就労か、など）に続き、飲酒・喫煙の習慣、そして健康状態に関する質問（疾病と定期的通院の有無）という順序で項目を配列した。これらはいずれも、抑うつ傾向と希死念慮の先行変数となることが予定された変数である。

実験として操作したのは、調査票の後半である。実験では、後半部の異なる二種類の調査票を作成した（図1）。第一の調査票は、定期的通院の理由についての質問の後に抑うつ傾向を問うCES-D項目（20項目）、そしてこれに続いて希死念慮（3項目）を並べた（B票）。これはコンテキストに配慮せず、単純に質問を並べたものである。これに対して、第二の調査票（A票）は、コンテキストを重視し、質問の流れが自然に受け取られるようにいくつかの質問を新たに付け加えたものである。具体的には、まず定期的通院の理由の後に生活ストレスを感じるかという質問14項目を加えた。この14項目は、直前の質問と関連するような、体調不良や疾病によるストレスから始まり、やがて日常生活の困難さから対人関係のトラブルによるストレスへと、回答者の関心を次第に自身の精神状態へと向けていくように並べてある。そして、このストレスの質問の直後に、抑うつ傾向を測定するCES-D項目（20項目）が置かれた。また、このCES-D項目の直後に、希死念慮へと回答者の関心を向けるために、新しい項目（8項目）を加えた。これは、自殺という単語が唐突にならないように、病死と事故死というキーワードを一緒に与えておくことが目的だった。さらに、希死念慮の質問（3問）で調査票を終えると自殺という単語の印象が強く残ってし

まうという共同研究者からの指摘があったため、A票では気分転換の方法や困ったときに頼りになる友人などの質問を最後に加えた。これは、今回の調査が郵送によるものであるため、回答者は後の質問を見てから回答する事ができるという特性を利用して、自殺という単語を目立たない

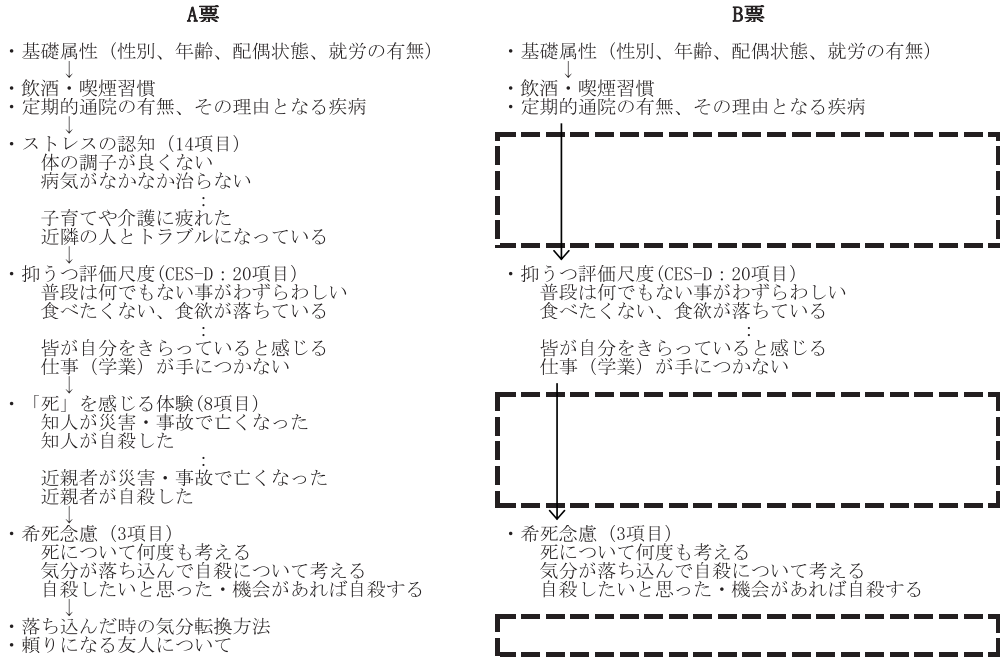


図1：調査票の構成（フロー）

ようにする工夫であった。

このA票で加えられた「使い捨て項目」は全部で27項目にのぼる。B票の項目数、すなわち「使い捨て項目」を加えていない状態での調査項目総数は40項目なので、コンテキスト形成のために加えられた使い捨て項目は、A票全体の40%に達していることになる。

調査では、コンテキスト形成を行った調査票（A票）を正式の調査票とし、使い捨て項目を追加していない調査票（B票）は実験用の調査票、テスト票とした。実際に配布された調査票は総数で3000部、このうち2700部はA票で、残る300部がB票となった。調査票の配布は2013年の1月、配布および回収は郵送によった。配布地域すなわち調査対象地域は青森県内全域、標本抽出は電話帳より無作為に送付先世帯を抽出し、宛名の個人名を特定せず「居住者の皆様へ」として送付した。なお、A票とB票はランダムに混ぜ合わせて発送した。

3 結 果

まず本調査票発送の前に行ったプレテストの結果に触れておきたい。調査票の構成段階で、筆者の担当する授業の時間を使ってプレテストを行った。手順としては、授業の出席者31人をラン

ダム（たまたまその日の授業で机の右端に座っていた者と左端に座っていた者）に分け、一つのグループにはA票、もう一つのグループにはB票を配り、その場で回答してもらった。そして回答後に全員、目を閉じさせ、このアンケート調査が嫌いである、答えたくないという者に挙手をさせた。その結果、A票（コンテキスト形成のための使い捨て項目を追加してあるもの）に回答した16人のうち、3人が答えたくないと挙手をした。これに対して、B票（使い捨て項目を足していないもの）に回答した15人のうち、答えたくないと挙手したのは7人にのぼった。授業後に学生に聞いてみると、やはりB票の場合は自殺というキーワードが頭の中に強く残ってしまうらしい。ここから、コンテキスト形成とそのための使い捨て項目の導入は、かなりの程度、効果が期待できるといふ見込みが立った。本調査でA票を正式の調査票としたのは、この結果を踏まえたもの

表1：回収率および有効回収率

	A票 (コンテキストあり)	B票 (コンテキストなし)	有意差確率 (p値)
送付完了数	2228	240	
回収数	775	71	
(粗回収率)	34.8%	29.6%	0.1154
有効回収数 ^{1*}	769	69	
(有効回収率 ¹)	34.5%	28.8%	0.0734
有効回収率 ^{2*}	736	65	
(有効回収率 ²)	33.0%	27.1%	0.0695

*1 項目欠損率50%超を無効回答としたもの

*2 項目欠損率30%超を無効回答としたもの

である。

本調査票の回収については、まず2700部発送したA票のうち472件が、300部発送したB票のうち60件が、宛先不明として返送されてきた。20%近い宛先不明というのは極めて高い。これは、いくつかの原因が重なったことによる。第一に、送付先住所を電話帳から学生が一件ごとに手書きで転写しているため、転写のミスが出たこと。作業を行っているのが人間である以上、このミスがある程度発生するのは避けられない。第二に、電話帳に記載された住所が略式で、建物名などが抜けている例が多かったこと。これは事前に予想された事態ではあったが、標本抽出のフレームとして他の選択肢がなかったため、やむを得ないと言えよう。第三に、宛先の個人名がないため、住所が不完全でも郵便局員の経験で正しい住所に転送することができなかったこと。少なくとも、この3つの事情が確認できている。

送付数から宛先不明数をのぞいた送付完了数A票2228部、B票240部のうち、回収できたのは、A票が775件、B票で71件となった。ここから回収率を計算すると、A票が34.8%、B票が29.6%となる（表1）。すなわち項目数の多いA票の方が回収率が高い。ただし、この二つの回収率に差があるかどうかをフィッシャーの正確性検定⁵で検証したところ、 $p=0.1154$ となった。つまり、この差は有意ではない。しかしながら、回収票のデータ入力を始めると、白票つまり回答が一切ない

回収票や項目欠損が著しく多い票が含まれていることに気づいた。そこで、項目欠損率50%超⁶のケースを無効回答として除外すると、有効回答数がA票で769件、B票で69件となり、回収率（有

表2：抑うつ得点および希死念慮の差

	A票 (コンテキストあり)	B票 (コンテキストなし)	有意差確率 (p値)
CES-D得点平均値	11.93	12.7	0.456
希死念慮 1 ^{*1}	29.70%	30.90%	0.890
希死念慮 2 ^{*2}	8.8%	11.8%	0.379
希死念慮 3 ^{*3}	2.5%	1.6%	1.000

*1 「死について何度も考えることがあるか」という質問に「ある」と回答した割合

*2 「気分がひどく落ち込んで、自殺について考えることがあるか」という質問に「ある」と回答した割合

*3 「自殺したいと思った」「もし機会があったら自殺するだろう」と回答した割合（合計）

*4 CES-D得点平均値については一元配置分散分析で、他はフィッシャーの正確性検定を用いて算出した結果

効回収率) にするとA票34.5%、B票28.8%という結果が得られた。この有効回収率の有意差検定の結果は、 $p=0.0734$ となった。つまり、 α 水準を10%とやや甘めに見積もれば、A票つまり項目数の多い方が回収率が高いということになる。なお、確認のために、項目欠損率30%ではどうなるかを調べたところ、A票の有効回答数が736件（33.0%）、B票が65件（27.1%）、有意差検定の結果は $p=0.0695$ となり、項目欠損率50%の場合と同様な結果となった。

コンテキスト形成のもう一つの問題、つまり偏り (bias) の発生について、まず抑うつ傾向を測定するCES-D項目の総合得点を比較してみた⁷。すると、A票の平均値が11.93点、B票の平均値が12.7点となり、事実上、差が認められなかった (表2)。さらに、希死念慮の3項目についても、それぞれ自殺のポテンシャルが高いと判断される回答をした者の割合はほとんど差がみられず、誤差の範囲に収まることが明らかとなった。つまり、コンテキストの形成をうまく行って回答率をあげても、個々の質問に対する回答は変わっていない。言い方を換えれば、偏りが発生していることは観察されなかった。

4 考 察

今回の実験結果をまとめると、調査票の構成において、使い捨て質問を用いてコンテキストの形成を行うことで、有効回収率が高くなっている。この点、SheatsleyやDillmanの指摘したとおりの結果が得られている。だが、コンテキスト形成によって回答に偏りが生じていることは観察されていない。当初、筆者らは使い捨て質問を用いてコンテキスト形成を行うことで、中核的質問項目への認知が変わる、つまり抑うつや希死念慮の各質問項目をもっと日常的なものとしてとらえる回答者が現われ、その結果として、抑うつ傾向や希死念慮を高め回答する傾向が生まれるのではないかと予想していた。しかしながら、今回の実験に関する限り、こうした結果は現れていない。これは、コンテキストが存在しても、個々の質問項目に対する認知そのものは変わっ

ていないことを示している。いわば、今回の実験で、使い捨て項目は直前の質問項目から次の質問項目へと回答者の意識を向けさせるための「橋渡し」ないしは「リエゾン」の機能を果たしていたと言える。

今回の実験結果は、従来の社会調査の常識とは合致しない。しかしながら、一見常識外に見えるこれらの傾向は、回答者のとりアクションを理論的に検討していくと、極めて合理的に理解できることがわかる。従来の調査方法論の研究では、個々人は属性の集合体としてとらえられていた。これに対して、近年注目を集めているTourangeauやDillman、Schwarzらの研究の特徴は、回答者を意思を持った個人としてとらえ、社会学・心理学的な立場から、その行動（respondent behavior）を分析しようとする。つまり、回答を求められた調査対象者が、どのような状況においてどのような思考を展開したか、その思考の展開パターンを正確に把握することで、非回答なり様々な測定効果が生じることを予測する。それを可能にするのは、心理学なり社会学なりの専門知識に加え、調査の現場において何が起きているかを正確に把握する現場経験の豊富さである。この点、今日の社会調査研究は、新たな局面に向かっているのではないだろうか。

註一覧

- 1) 本稿で社会調査という場合、それらは全てsocial surveyを示すものとする。これは、エスノグラフィが社会調査でないとか、歴史資料の分析を調査と認めないとか、そういった議論を念頭におくものではない。単純に、訳語の問題を本文中で逐一触れる必要性を認めないだけである。本稿は技術的な論点を扱うものであり、筆者としては、日本で社会調査という言葉をどのように定義すべきかという哲学的な議論には関心がない。
- 2) 調査結果および基礎集計表に関しては、青森県自殺対策検証研究会『平成24年度青森県自殺防止対策取り組み検証事業報告書』（2013年3月刊行、非売品）を参照されたい。
- 3) 近年、日本における自殺率の高さに注目が集まり、社会問題として認識されるに至った。政策側の対応も比較的積極的で、2006年に「自殺対策基本法」が施行されて以来、2007年から『自殺対策白書』が刊行されるなど、内閣府を中心に自殺問題に関する調査研究の促進と統計の整備が進められ、また厚生労働省は各自治体での自殺予防対策の振興を指示してきた。こうした中で、都道府県別の自殺率で全国第三位（2007年）および第二位（2008年）という高い水準にあった青森県は、自殺予防の実務プログラムだけでなく自殺行動の調査研究にも極めて高い関心を示していた。筆者らが今回の調査に関わるようになった背景には、こうした事情があった。
- 4) この点については、内閣府の行っている調査も同様であり、2007年に内閣府の行った「こころの健康（自殺対策）に関する世論調査」および「自殺予防対策に関する意識調査」という二つの意識調査では、回答者を自殺行動の当事者とは見なさず、第三者として自殺のメディアでの取り扱いや自殺未遂者に対するケアの必要性などを調査しており、回答者自身の自殺へのポテンシャルを聞く事は避けている。これに対し、2008年度および2011年度に実施された「自殺対策に関する意識調査」では、回答者を自殺行動の当事者と位置づけ、抑うつ傾向や自殺へのポテンシャルを聞いているが、この調査に際しては、前年度と異なり、インタビューアからも回答がわからないように面接法から自記式に回答方式を変更したり、無回答を明示的に認めるなど、様々に配慮を施している。内閣府の調査については、<http://www8.cao.go.jp/jisatsutaisaku/survey/index.html> を参照されたい。
- 5) フィッシャーの正確性検定を用いた理由は、標本数がA票とB票で大きく異なるためである。ちなみに、

一般的な二標本の比率の差の検定を行った結果は、以下のとおりである。

粗回収率	p=0.1067
項目欠損率50%以上を無効	p=0.0731
項目欠損率30%以上を無効	p=0.0613

つまり、二つの検定方法で異なる結果は出していない。

- 6) 項目欠損率の計算にあたっては、コンテキスト形成のために加えられた使い捨て項目は除外してある。
- 7) CES-Dは保健医療現場での使用を目的としているため、得点化の手続きもあらかじめ決められている。CES-Dは全20項目のうち、第4, 8, 12, 16項目が逆転項目となっている。これらの逆転項目については、「ほとんどなかった」= 3、「少しはあった」= 2、「時々あった」= 1、「たいていそうだった」= 0として得点化する。他の16項目については、「ほとんどなかった」= 0、「少しはあった」= 1、「時々あった」= 2、「たいていそうだった」= 3として、同様に各項目を得点化し、最後に全20項目の得点を合計し、これを抑うつ傾向得点とする。つまり、点数が高いほど抑うつ度が高いという事になる。

文献一覧

- Biemer, Paul P. et.al. 2004. *Measurement Errors in Surveys*. Wiley-Interscience.
- Bogardus, Emory. 1934. Interviewing as a Social Process. *Sociology and Social Research*, 19.
- Bradburn, Norman & William Mason. 1964. The Effect of Question Order on Responses. *Journal of Marketing Research*, 1. Reprinted in Martin Bulmer ed. 2004. *Questionnaires*, 2.
- Cicourel, Aaron. 1964. *Method and Measurement in Sociology*. Free Press.
- De Leeuw, Edith et. al. 2008. *International Handbook of Survey Methodology*. Psychology Press.
- De Vaus, David A. 1990. *Surveys in Social Research, Second Edition*. Unwin.
- Dillman, Don A. 2009. *Internet, Mail and Mixed-mode Surveys*. Wiley.
- Groves, Robert. 2004. *Survey Errors and Survey Costs*. Wiley-Interscience.
- Kahneman, Daniel, Paul Slovic & Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Lee, Raymond. 1993. *Doing Research on Sensitive Topics*. Sage.
- Marsden, Peter & James Wright ed. 2010. *Handbook of Survey Research, Second Edition*. Emerald Group Publishing.
- McFarland, Sam. 1981. Effects of Question Order on Survey Responses. *Public Opinion Quarterly*, 45-2. Reprinted in Martin Bulmer ed. 2004. *Questionnaires*, 2.
- Schuman, Howard, Stanley Presser & Jacob Ludwig. 1981. Context Effects on Survey Responses to Questions about Abortion. *Public Opinion Quarterly*, 45-2. Reprinted in Martin Bulmer ed. 2004. *Questionnaires*, 3.
- Sheatsley, Paul B. 1983. Questionnaire Construction and Item Writing. In Peter H. Rossi et al. *Handbook of Survey Research*. Academic Press.
- 島悟・他 1985. 「新しい抑うつ性自己評価尺度について」, 『精神医学』, 27.
- Suchman, Lucy & Brigitte Jordan. 1990. Interactional Troubles in Face-to-Face Survey Interview. *The Journal of American Statistical Association*, 85.
- Sudman, Seymour & Norman M. Bradburn. 1974. *Response Effects in Surveys*. Aldine.
- Tarnai, John & Don. Dillman. 1992. Questionnaire Context as a Source of Response Differences in Mail and Telephone Surveys. In Norbert Schwarz & Seymour Sudman ed. *Context Effects in Social and*

Psychological Research. Springer.

Tourangeau, Roger et al. 2000. *The Psychology of Survey Response*. Cambridge University Press.

Trochim, William & James Donnelly. 2008. *The Research Methods Knowledge Base, Third Edition*. Atomic Dog.

Wanke, Michaela & Norbert Schwarz. 1997. Reducing Question Order Effects: the Operation of Buffer Items.

Lars Lyberg et al. ed. *Survey Measurement and Process Quality*. Wiley.

Weisberg, Herbert F. 2005. *Total Survey Error Approach*. University of Chicago Press.

Summary

This paper examines the influences of imposing throwaway questions for sensitive social survey. Making a context in questionnaire through the introduction of throwaway items is a relatively new idea in survey methodology fields, and has still very few researches reported. Authors, thus, examined this at a sensitive social survey, whose major topics were mental depression and suicide propensity among general population in Aomori Prefecture. We composed two questionnaires: one with throwaway items that made the flow of questions more smooth, and another without them. The valid response rate of the questionnaire with additional items indicated 34.5% whereas another showed 28.8%, whose difference was significant at 10% confidence level. Response bias between the two questionnaires, however, was not observed.

[Key words] social survey, context effect, nonresponse error