

公募型 Web 調査における回答時間と回答中断行動

Response Time and Respondent Interruptions in Volunteer Panel Web Surveys

山田 一成
Kazunari YAMADA

問題

Web 調査は安価で機動性が高く、多様なカスタマイズが可能であり、各種パラデータも入手可能であるため、既に広く利用される方法となっている。しかし同時に、近年の方法論的基礎研究によって、Web 調査が従来の調査方法の安価な代替ではなく、パラデータによる検票を必要とする特殊な方法であることも明らかになりつつある（増田, 2017; 増田・坂上・森井, 2017; 埴淵・村中・安藤, 2015）。

本研究でもこのような認識に基づき、Web 調査の検票作業に利用可能なパラデータである回答時間と回答中断行動に注目し、まず、その実態を明らかにすることを試みる。そして、回答中断者の回答が回答非中断者の回答と比較して異質であるかどうかを検討し、そのうえで回答中断行動を管理する必要があるかどうか議論する^(注1)。

なお、本研究で取り上げる回答中断行動は無効票となる脱落（breakoff）ではなく、有効票における一時的な回答中断である^(注2)。こうした回答者の行動については、早くから休止（inactivities）や中断（interruption, distraction）といった言葉による言及があったものの、その実態や発生理由、および、それが回答の質に及ぼす影響などについては、実証研究が十分に行われなままとなっていた（Callegaro, Manfreda, & Vehovar, 2015）^(注3)。

そうしたなか、大規模な Web 調査データに基づき、回答中断行動とその影響について詳細な分析を行ったのが Ansolabehere & Schaffner（2015）である。Ansolabehere らが分析したのは政治を主題とする 4 つの Web 調査で、そのうちひとつは Knowledge Networks 社が 2010 年に実施した調査（パネルからの確率標本）であった。また、それ以外の 3 つは YouGov 社が 2011 年から 2012 年にかけて実施した調査（オプトイン・パネルからのサンプル・マッチング）で、そのうちの 2 つは CCES の一

部であり、残りの1つはANESの一部(モード比較のためのオンライン調査)であった^(註4)。

こうした調査のデータを用い、Ansolabehereらは、家事や電話などの回答中断行動9項目について分析を行い、その結果、(1)所要時間30分程度の調査で少なくとも1種類以上の回答中断行動があった者は約5割にも及んでおり、回答中断行動が一般的であること、(2)調査所要時間が長くなるほど回答中断行動が増える傾向があること、(3)回答中断行動を増加させるのは調査の前半・後半といった質問の位置ではなく、回答が認知的努力を要するかどうかであると考えられること、そして、(4)そうした回答中断行動は回答時間増につながるものの、回答の質を低下させることはない、などといった貴重な知見を報告している。

ただし、Ansolabehereらが分析に用いたデータはアメリカにおいて収集されたデータであり、サンプリングの方法も現在日本で多用されている公募型Web調査における割当法(回答者パネルからの先着順方式)とは異なっている。そのため、回答中断に関する彼らの調査結果がそのまま現在の日本の公募型Web調査に当てはまるという保証はない。そこで本研究では、Ansolabehereらの研究方法を参考に、日本の公募型Web調査における回答中断行動と回答時間の実態を明らかにすることを試みる。以下では、まず、分析1で回答中断行動の分析に必要な回答時間の特性を明らかにし、そのうえで、分析2で回答中断行動の実態について、また、分析3と分析4で回答中断行動と回答パターンの関連について検討する。

方法

調査概要 本研究では下記の2回の公募型Web調査(調査会社委託・ポイント報酬制)のデータを分析に用いる。なお、各回の調査の概要は下記のとおりである。

第1回調査: 一都三県在住の男女20~69歳が対象。調査会社の登録モニターから事前調査で回答者を抽出し、総数1,400人程度を目途に人口比例させた性年代人数を本調査に割り当てた(人口比例は平成27年度の国勢調査に基づく)。事前調査(スクリーニング調査)は2017年1月17日(火)~18日(水)に実施し、69,810人に配信して5,000人から回答を得た。ここから、無効回答、回答に利害の影響が懸念される特定業種の従事者(従事者が家族にいる者を含む)と、通信速度が著しく遅いダイヤルアップ接続者を除外して4,137人を抽出。そこから2,325人をランダムに抽出した後、本調査を2017年1月20日(金)~1月22日(日)に配信し、割当人数分の回答回収時点で調査を終了して、1,444人の有効回答を得た。調査では1画面に1質問を表示し、未回答存在時の警告表示により無回答を許容しない仕様とした。

なお、本調査の質問数はスプリット4群とも18問であった(A票~C票:総クリック数55+数値入力2箇所、D票:総クリック数53+数値入力2箇所+Visual Analogue Scale回答2箇所)。また、以下の分析では、スマートフォンや携帯電話からの回答者を除き、回答デバイスがPCであった

1. 401名を分析対象とする。

第2回調査：一都三県在住の男女20～69歳が対象。調査会社の登録モニターから事前調査で回答者を抽出し、総数600人程度を目途に人口比例させた性年代人数を本調査に割り当てた（平成27年度国勢調査に基づく）。事前調査（スクリーニング調査）は2018年1月17日（水）～19日（金）に実施し、89,253人に配信して5,000人から回答を得た。ここから無効回答、回答への利害の影響が懸念される特定業種の従事者（従事者が家族にいる者を含む）と、通信速度が著しく遅いダイヤルアップ接続者を除外して4,118人を抽出した。そこから1,149人をランダムに抽出した後、本調査を2018年1月19日（金）～1月21日（日）に配信し、割当人数分の回答回収時点で調査を終了して、622人の有効回答を得た。

なお、本調査の質問数はスプリット3群とも17問であった（A票～B票：総クリック数75+数値入力2箇所、C票：総クリック数73+数値入力2箇所+ Visual Analogue Scale 回答2箇所）。また、以下の分析では、スマートフォンや携帯電話からの回答者を除き、回答デバイスがPCであった594名を分析対象とする。

分析1：回答時間の特性

問題と方法 本研究では、最初に、公募型 Web 調査の回答時間の特性と実態を明らかにするために、以下のような時間計測を行った。まず、公募型 Web 調査の回答時間としては、①調査所要時間、および、②質問画面提示時間を測定した（以下では適宜「秒」を「s」と表記する）。これらのうち、①は調査会社の仕様に従いデフォルトで測定される変数であり、②は発注者の指定によって測定される変数である。なお、②は各質問ごとに測定されるため、以下では②の総和を「総質問画面提示時間」または「 Σ ②」と表記する。また、本研究の公募型 Web 調査はスプリット調査として行われたため、分析は、適宜、スプリットの各票別に行った（第1回調査の調査票はA票～D票の4種類、第2回調査の調査票はA票～C票の3種類）。なお、回答時間の分析に際しては、適宜、自然対数化した値を用いた。

結果 まず、第1回調査の結果を見ると、調査所要時間（①）と総質問画面提示時間（ Σ ②）の分布には著しい違いが認められた（①：最小値90s、最大値56,936s、中央値284s、 Σ ②：最小値71s、最大値2,067s、中央値266s）。そこで、こうした結果になる理由を調査会社に問い合わせたところ、両者の間に「① = Σ ② + ③調査依頼画面提示時間」という関係があることが判明した。また、調査依頼画面に表示された回答者へのメッセージは、調査タイトルを含め176字（ボタン操作指示を除く）であったが、③（=① - Σ ②）の分布を調べたところ、調査の冒頭で調査依頼画面が提示された後、長時間「質問1（最初の質問）」に進まない回答者がいることが明らかとなった（③：最小値3s、最大値56,641s、中央値10s、第3四分位15s）。なお、回答者の94.7%は60秒以内に「質問1」に進んで

いたが、1時間以上「質問1」に進まなかった回答者も20名存在した。また、そのうち5名は10時間以上「質問1」に進んでおらず、③の最大値は16時間弱であった。

次に、第2回調査の回答時間について検討したところ、上述の結果とほぼ同様の結果が得られた(①：最小値101s、最大値35,617s、中央値335s、 Σ ②：最小値92s、最大値2,050s、中央値316s)。また、③についても同様に、長時間「質問1」に進まない回答者が存在した(③：最小値3s、最大値35,291s、中央値10s、第3四分位15s)。なお、回答者の94.8%は60秒以内に「質問1」に進んでいたが、1時間以上「質問1」に進まなかった回答者も9名存在した。また、そのうち3名は8時間半以上「質問1」に進んでおらず、③の最大値は10時間弱であった。

考察 こうした結果から、公募型 Web 調査で異常に長い回答時間が測定された場合には、当該の変数が上記①であるかどうか確認する必要があると考えられる^(注5)。また、 Σ ②の最大値が35分弱であることから、数時間にも及ぶ異常に長い回答時間は質問への回答が開始されてからは発生しない可能性が高いと考えられる。ただし、両調査とも1ジャッジ・1クリックで回答可能な質問が5～7割を占めており、 Σ ②の値が低めである点には注意が必要である。なお、一部の回答者が長時間「質問1」に進まなかったことについては、そうした行動の前提に、PCの常時稼働と常時 Web 接続、および、同時並行作業 (multitasking) の習慣化があると考えられる。

以下では、以上のような検討結果に基づき、まず、回答中断行動と回答時間の関係について検討する。なお、以下では特に断りの無い限り、便宜的に質問画面提示時間を回答時間と呼称する。

分析2：回答中断行動の実態

問題と方法 本研究では公募型 Web 調査における回答中断行動の実態を明らかにするために、調査の最後で「あなたは、今回の調査に回答する間に、次にあげるような理由で、一時的に回答を中断したことがありますか」と尋ね、Table 1に挙げた12項目のそれぞれについて、該当するかどうか2件法で回答を求めた(質問画面には「事前アンケートを含めずに、本アンケートの開始後についてお知らせください」と注記した)。また、回答者は、こうした質問について1項目以上該当項目があれば「回答中断群」(以下、中断群)、全項目非該当であれば「回答非中断群」(以下、非中断群)に分類された。

ただし、回答中断行動の分析に先立ち、回答中断行動を尋ねる質問への不正回答の有無について検討した。不正回答の指標となるのは回答パターン(連続同一回答 [straight lining : SL] と準連続同一回答 [1項目以外は同一回答である near straight lining : NSL])、および、回答時間である。以下では、まず、第1回調査について、回答中断行動に関する「該当」のSL・NSLパターン、「非該当」のSL・NSLパターンの順に検討し、その後、第2回調査の結果について同様の検討を行う。そして、そのうえで、回答中断理由、および、回答中断行動と回答時間の関係について検討する。

結果 まず、第1回調査については、「該当」パターンのうち、全項目該当（12点：SL）は0.3%で、これに1項目非該当（11点：NSL）の0.3%を加えても、0.6%（8名）と希少であった（Table 2）。また、これら8名の回答中断行動質問への回答時間の範囲は9s～56sで、 $\pm 2SD$ 基準で除外対象となるのは8名中1名であり、最小値の9秒についても、適切な回答行動が不可能な値ではないと考えられた。そのため、「該当」に関するSL・NSLについては、明らかな不正回答はないと仮定した。

次に、「非該当」パターンのうち、全項目非該当（0点：SL）は57.2%で、1項目該当（1点：NSL）も17.1%と多く（Table 2）、それらは経験的に十分ありうるパターンであると考えられた。ただし、全項目非該当という回答パターンに、虚偽回答としての「中断なし」が混在している可能性がないわけではない。

しかし、そのような虚偽回答は、以下のような理由により、回答時間に注目しても発見・除外することが困難であると考えられる。まず、この質問が回答中断行動に関する質問であることは、質問画面提示後に短時間で理解可能であり、「中断なし」（全項目非該当）と回答する作業も短時間で可能である。そのため、「中断者が全項目非該当と回答する」という虚偽回答（F）も短時間で可能となるが、同時に、「非中断者が全項目非該当と回答する」という適切な回答（T）も短時間で可能となる。このように、今回の分析に使用した質問の形式では、FとTの混在が予想されるとしても、回答時間による虚偽回答の識別は極めて困難であると考えられる。

ただし、非中断群において異常に長い回答時間が測定された場合には、虚偽回答の識別も不可能ではない。そこで、スプリット4群ごとに、非中断群における回答時間の分布を調べた。なお、ここで分析に用いた回答時間は、「質問1」から「回答中断行動質問の直前の質問」までの質問画面提示時間の総和である（以下、便宜的に④分析用回答時間、または、④と呼称する）。その結果、外れ値がA票に1つ、D票に2つ認められたが、それらの値は、短時間の回答中断が疑われる値ではなかったものの、虚偽回答が疑われるような著しく大きな値ではなかった。

次に、第2回調査の結果について検討する。第2回調査の結果も第1回調査とほぼ同様で、まず、「該当」パターンについては、SLとNSLの合計は0.7%（4人）と希少であった（Table 2）。また、これら4名の回答中断行動質問への回答時間の範囲は9s～57sで、最小値の9秒についても、適切な回答行動が不可能な値ではないと考えられた。そのため、第2回調査についても、「該当」に関するSL・NSLについては、明らかな不正回答はないと仮定した。

なお、「非該当」パターンについても、SLが53.4%、NSLが21.0%と、第1回調査とほぼ同様の結果であった（Table 2）。また、非中断群における回答時間については、外れ値がA票に1つ、C票に1つ、対数化回答時間の平均値から $+2SD$ を超える外れ値がB票に5つ認められた。しかし、それらの値も、短時間の回答中断が疑われる値ではなかったものの、虚偽回答が疑われるような著しく大きな値ではなかった。

以上のような検討の結果、上記の質問によって測定される各回答中断行動の割合は、中断者が非中断群に混在している可能性を排除しきれないため、厳密に言えば、一種の下限值として扱うべきであ

Table 1 回答中断行動の頻度 (理由別)

中断理由	第1回調査		第2回調査	
	頻度1	頻度2	頻度1	頻度2
電話	6.3	14.7	4.9	10.5
電子メール	9.4	22.0	8.6	18.4
家の中での家族とのやり取り	18.4	43.0	17.3	37.2
仕事や家事	16.5	38.5	18.0	38.6
食事	10.8	25.3	12.3	26.4
入浴	5.8	13.5	5.6	11.9
睡眠	7.9	18.3	7.2	15.5
休憩	10.2	23.8	8.4	18.1
テレビ視聴	12.6	29.5	14.3	30.7
新聞・雑誌・コミックなどの閲読	2.8	6.5	3.4	7.2
ネットやSNSの閲覧	7.7	18.0	6.7	14.4
その他の理由	11.9	27.8	12.8	27.4
N	1,401	600	594	277

注：数字は列%。頻度2のNは回答中断者。質問項目では各中断理由の後に「による回答中断」と記述。

Table 2 回答中断理由の該当個数

個数	第1回調査		第2回調査	
	%	累積%	%	累積%
0	57.2	57.2	53.4	53.4
1	17.1	74.3	21.0	74.4
2	7.4	81.7	10.3	84.7
3	6.9	88.7	5.1	89.7
4	4.4	93.1	3.4	93.1
5	1.9	94.9	1.9	94.9
6	1.6	96.5	1.9	96.8
7	0.7	97.2	0.8	97.6
8	0.9	98.1	0.3	98.0
9	0.8	98.9	0.7	98.7
10	0.6	99.4	0.7	99.3
11	0.3	99.7	0.5	99.8
12	0.3	100.0	0.2	100.0
N	1,401		594	

ると判断された。なお、一般に、回答中断は好ましくないことと見なされる。また、登録モニターにとって、回答中断行動を問う質問は調査会社のパネル管理作業を連想させる可能性がある。さらには、回答中断者のほうが「中断」かどうかの判断基準が緩いことも十分考えられる。こうした点も、上記の回答中断行動の割合を下限値と見なす根拠になるかもしれない。

以上の検討を踏まえ、以下では、回答時間 $\pm 2SD$ 基準等によるサンプルの除外を行わずに、回答

Table 3 分析用回答時間の記述統計量（第1回調査・回答中断行動の有無別：秒）

	A 票 (Split 1)		B 票 (Split 2)		C 票 (Split 3)		D 票 (Split 4)		
	非中断	中断	非中断	中断	非中断	中断	非中断	中断	
最小値	86	82	86	92	71	62	87	77	
最大値	845	1872	813	1906	981	1713	861	1781	
四分位	25%	185	188	172	185	189	182	175	173
	50%	236	251	240	244	226	246	239	244
	75%	320	349	312	337	295	336	329	356
平均	268.0	309.2	267.0	311.3	259.0	303.7	270.6	317.3	
標準偏差	129.6	221.6	137.2	261.7	126.1	223.0	139.3	274.5	
歪度	1.78	3.68	1.69	4.01	2.02	3.24	1.65	3.50	
N	196	161	213	139	211	153	178	147	

注：分析用回答時間は「質問1」から「回答中断行動質問の直前の質問」までの質問画面提示時間の総和。
非中断群については分析用回答時間の明らかな外れ値をA票で1つ、D票で2つ除外した。

中断理由の頻度を示す。なお、両調査とも、回答中断理由の全項目について、スプリットされた群間に差が認められなかったため、以下では群を併合した場合の結果を示す（Table 1、Table 2）。

まず、第1回調査における回答中断理由の選択個数については、上述のとおり、「0個」（回答中断なし）が57.2%で、回答中断者は4割を超えていた。また、回答中断理由を1つだけ選択した者は17.1%に留まり、理由を複数挙げた者のほうが25.7%と多かった。なお、回答中断理由は単一の項目に集中せず、多くの項目に分散していたが、多かったのは「家の中での家族とのやり取り」と「仕事や家事」で、それらに「テレビ視聴」と「食事」が続いていた。なお、第2回調査についても、ほぼ同様の結果であった（Table 1、Table 2）^(注6)。

次に、回答中断行動の有無と④分析用回答時間の関係についてはTable 3とTable 4に示すとおりである。まず、第1回調査の結果を見ると（Table 3）、スプリット4群ともほぼ同様の結果となっており、非中断群と中断群の間に第3四分位で25s～41sの差があり、中断群のほうで④が長くなっていた。また、④の最大値も、中断群で非中断群よりも742s～1,086s大きくなっていた。ただし、このような増加は、調査所要時間（①）について計測された異常に長い回答時間とは異なり、最大値で見ても最大18分程度の増加に過ぎなかった。

次に、第2回調査の結果を見ると（Table 4）、スプリット3群ともほぼ同様の結果となっており、非中断群と中断群の間に第3四分位で51s～89sの差があり、中断群のほうで④が長くなっていた。また、④の最大値も、中断群で非中断群よりも115s～390s大きくなっていた。ただし、このような増加は最大値で見ても最大6分半程度の増加に過ぎなかった。

考察 回答中断行動によって回答時間が増加するのは当然のことである。しかし、問題はその程度である。もしも、多くの回答者において、回答中断行動により数時間から十数時間もの時間増が生じ

Table 4 分析用回答時間の記述統計量 (第2回調査・回答中断行動の有無別:秒)

	A票 (Split 1)		B票 (Split 2)		C票 (Split 3)		
	非中断	中断	非中断	中断	非中断	中断	
最小値	77	99	84	95	87	101	
最大値	850	1098	675	1065	882	997	
四分位	25%	220	232	191	215	184	232
	50%	292	313	261	313	262	301
	75%	378	419	341	430	354	405
平均	312.1	361.6	272.0	357.5	288.3	348.0	
標準偏差	141.1	202.5	112.4	206.2	146.4	172.0	
歪度	1.18	1.60	0.62	1.39	1.42	1.69	
N	101	92	113	96	97	85	

注:分析用回答時間は「質問1」から「回答中断行動質問の直前の質問」までの質問画面提示時間の総和。非中断群については、分析用回答時間の明らかな外れ値をA票で1つ、B票で4つ、C票で1つ除外した。中断群については、同様の外れ値をA票で1つ、B票で1つ、C票で2つ除外した。

ているようであれば、調査の目的によっては、回答者間でも回答者内でも同じ条件下で回答したとは仮定できなくなる可能性がある。

しかし、本研究の結果からは、回答中断行動によって、そのような大幅な時間増は生じていないことが明らかとなった。中断群と非中断群の④分析用回答時間の差は第3四分位で数十秒程度、明らかな外れ値を除外した場合の最大値で数分または十数分程度である^(註7)。こうした結果が示唆するのは、回答時間増のみを理由に回答中断行動を積極的に管理する必要があるとは言えない、ということである。

ただし、それだけを理由に、直ちに管理不要との結論を下せるわけではない。もしも、中断群と非中断群の間で回答の質に著しい差があるようであれば、管理の要不要について更なる議論が必要になるからである。そこで次に、回答中断行動の有無による回答パターンの異同について検討した。

分析3:回答中断行動と連続同一回答

問題と方法 回答パターンについては、まず、連続同一回答(SL)と準連続同一回答(NSL)について検討した。SL・NSLは不正回答が疑われる回答パターンであり、中断群に不正回答者が多く含まれるのであれば、SL・NSLは非中断群よりも中断群で多くなると予測される。なお、本研究では、そうした検討のために認知欲求尺度(神山・藤原, 1991)が利用可能であった。この尺度は7件法の

Table 5 心理尺度項目15項目における連続同一回答パターンの割合（第1回調査：％）

		肯定		← 回答選択肢 →					否定		
SL・NSL		1	2	3	4	5	6	7	N		
中断群	4.5	SL	0.0	0.0	0.0	0.3	0.2	0.0	0.0	600	
		NSL	0.0	0.0	0.2	3.8	0.0	0.0	0.0		
非中断群	9.2	SL	0.1	0.0	0.0	2.4	0.1	0.0	0.2	801	
		NSL	0.0	0.0	0.1	5.6	0.1	0.0	0.5		

注：SLは連続同一回答、NSLは準連続同一回答（1項目以外同一回答）。

Table 6 心理尺度項目15項目における連続同一回答パターンの割合（第2回調査：％）

		肯定		← 回答選択肢 →					否定		
SL・NSL		1	2	3	4	5	6	7	N		
中断群	7.9	SL	0.0	0.0	0.0	1.1	0.0	0.0	0.0	277	
		NSL	0.0	0.0	0.4	6.1	0.4	0.0	0.0		
非中断群	14.2	SL	0.0	0.0	0.0	3.5	0.0	0.0	0.0	317	
		NSL	0.0	0.0	0.6	9.5	0.0	0.3	0.3		

注：SLは連続同一回答、NSLは準連続同一回答（1項目以外同一回答）。

Table 7 心理尺度項目15項目への回答時間（第1回調査：秒）

			四分位							
		最小	最大	25%	50%	75%	平均	標準偏差	歪度	N
SL・NSL	中 断	9	53	15	21	30	24.0	12.7	1.12	27
	非中断	5	143	16	21	30	26.5	19.6	4.05	74
非 SL・NSL	中 断	7	1481	46	67	95	90.0	119.8	6.20	573
	非中断	9	449	48	67	90	77.7	52.9	2.77	727

Table 8 心理尺度項目15項目への回答時間（第2回調査：秒）

			四分位							
		最小	最大	25%	50%	75%	平均	標準偏差	歪度	N
SL・NSL	中 断	9	200	16	23	28	31.0	38.8	4.33	22
	非中断	9	80	13	19	31	23.9	15.9	1.80	45
非 SL・NSL	中 断	8	774	42	63	86	79.0	79.2	4.83	255
	非中断	7	760	38	56	77	68.3	67.9	5.58	272

Likert 尺度15項目から成り、逆転項目を8項目含んでいるため、中央および中央付近の選択肢以外では、回答がSL・NSLとなる可能性は極めて低いと考えられる(項目の提示順序はプログラム管理により回答者ごとにランダム化された)。

結果 まず、第1回調査の結果はTable 5のとおりで、認知欲求尺度15項目への回答におけるSL・NSLの割合は、中断群で4.5%、非中断群で9.2%と、中断群で低かった($\chi^2(1) = 11.5, p < .01$)。また、SL・NSLの割合を選択肢ごとに見ると、1・2・6・7といった選択肢ではSL・NSLがほとんど認められず(これらの選択肢でのSL・NSLは不正回答の疑いが濃厚である)、希少なながら認められたSL・NSLも非中断群での回答であった。なお、こうした傾向は第2回調査でもまったく同様で(Table 6)、SL・NSLの割合は、中断群で7.9%、非中断群で14.2%と、中断群で低かった($\chi^2(1) = 5.17, p < .05$)。また、選択肢ごとのSL・NSLの頻度についても、第1回調査とほぼ同様の結果であった。こうした結果は中断群で不正回答者が多いとは言えないことを示唆している。

ただし、回答パターンについてはSL・NSLの回答時間の長短も問題となる。もしも、どちらかの群で回答時間が極端に短いSL・NSLが多ければ、群間で回答の質に違いがあることになる。そこで、次に、中断群と非中断群の間でSL・NSLの回答時間に差があるかどうか検討した。

第1回調査の結果はTable 7のとおりで、まず、SL・NSL以外の回答パターン(下段「非SL・NSL」)に比べ、SL・NSLの回答時間(上段「SL・NSL」)は著しく短かった(中央値で約45秒の差があり、この差が全てカーソル移動時間の差によるものとは考えにくい)。また、SL・NSLの回答時間について(上段「SL・NSL」)、中断群と非中断群の四分位範囲を比較すると、両者はほぼ重なり合っていた(「非SL・NSL」についてもまったく同様であった)。

次に、第2回調査の結果はTable 8のとおりで、SL・NSL以外の回答パターンに比べ、SL・NSLの回答時間は著しく短かった(中央値で約40秒の差)。また、SL・NSLの回答時間についても、上記と同様に四分位範囲はほぼ重なり合っていた(「非SL・NSL」についてもまったく同様であった)。

考察 以上の結果から、まず、公募型Web調査においては、Likert型の心理尺度項目群に一定の割合で不正回答が含まれている可能性があることが示唆された。しかし、そうした不正回答が疑われる回答パターンが中断群で多くなる傾向は認められず、むしろ、SL・NSLの割合は中断群のほうで低いことが示された。

認知欲求尺度は一定の認知的負荷を伴う項目群であり、不正回答としてのSL・NSLが生じやすい質問であったと考えられる。そのため、中断群が不正回答を頻繁に行う回答者を多く含むのであれば、中断群でSL・NSLが多くなるはずであるが、そうした傾向は認められなかった。こうした結果は、回答中断行動が直ちに回答の質を低下させるわけではないことを示唆している。

ただし、非中断群のほうでSL・NSLが多くなっていた点については、不正回答を行う回答者が、どの質問でも短時間で回答を終わらせようとし、SL・NSLという回答パターンを採用した可能性があることを示唆している。そこで次に、各質問への回答時間の質に注目し、回答中断行動との関連について検討する。

分析 4：回答中断行動と超短時間回答

問題と方法 最後に、回答中断行動の有無によって回答時間の質に違いがあるかどうか検討した。回答時間の質の指標となったのは超短時間回答数と超長時間回答数である。なお、ここで言う超短時間回答（extremely short response time：ESRT）とは、不正回答の疑いが濃厚な短時間回答のことで、一定の長さの質問文や複数の選択肢が並んでいる質問画面で、わずか数秒の回答時間、すなわち、質問文や選択肢の読解、および、回答のための思考がなされたとは言いがたい短時間の回答を指す。他方、ここで言う超長時間回答（extremely long response time：ELRT）とは、回答中断を想定しないと了解不能な長時間回答のことで、回答時間の値が明らかに外れ値であるような長時間の回答を指す。

本研究では回答者ごとに調査における ESRT 数と ELRT 数を算出し、それらの値が中断群と非中断群の間で異なるかどうか検討した。なお、ESRT 数の値は高くなるほど不正回答数が多いことを示唆し、当該回答者の回答の質が低い可能性があることを意味する。また、ELRT 数の値は、回答中断が頻繁に行われることが回答や調査への非専念性を示唆するため、値が高くなるほど当該回答者の回答の質が低い可能性があることを意味する。従って、非中断群よりも中断群で ESRT 数や ELRT 数が高い回答者の割合が多ければ、回答中断行動は不正回答や非専念回答の生起と関連すると考えざるを得ず、そのため、回答中断行動を管理する必要が生じることになる。

なお、分析にあたっては、各質問画面ごとに、回答時間が2.5パーセンタイル値未満の回答を ESRT、97.5パーセンタイル値超の回答を ELRT と操作的に定義した。また、ESRT 数と ELRT 数の値は、調査末尾の回答中断行動質問を除く全質問についてカウントしているため、両指標とも、第1回調査（質問数18）では0～17、第2回調査（質問数17）では0～16の値を取る。

結果 中断群と非中断群の ESRT 数と ELRT 数の割合は Table 9 に示すとおりである。まず、第1回調査では、ESRT 数の割合は中断群も非中断群もほぼ同様の分布となっている。他方、ELRT 数については、予測どおり中断群で高い値の割合が多くなる傾向がうかがえるが、中断群で高い値の割合が著しく多くなつてはいなかった。また、こうした結果は第2回調査についてもまったく同様であった。

なお、ESRT 数と ELRT 数のそれぞれについて、回答中断の有無と関連するかどうかカイ二乗検定を行った。カテゴリー数は、ESRT 数と ELRT 数については2以上の値を併合して「0」「1」「2以上」の3つとし、回答中断については「中断群」と「非中断群」の2つとした。その結果、第1回調査において、回答中断と ESRT 数の割合との間には関連が認められなかったが、ELRT 数の割合との間には有意な関連が認められ（ $\chi^2(2) = 21.0, p < .001$ ）、中断群で高い ELRT 数の割合が多くなっていた。また、第2回調査においては、回答時間の有無と ESRT 数との関連について有意傾向が示されたが（ $\chi^2(2) = 5.4, p < .1$ ）、ESRT 数が高い回答者の割合は非中断群で多くなっていた。他方、回答中断の有無と ELRT 数との間には有意な関連が認められ（ $\chi^2(2) = 8.3, p < .05$ ）、中断群

Table 9 超短時間回答数と超長時間回答数 (回答中断の有無別)

個数	第1回調査				第2回調査			
	ESRT数		ELRT数		ESRT数		ELRT数	
	非中断	中断	非中断	中断	非中断	中断	非中断	中断
0	81.6	81.2	80.6	70.2	81.1	83.8	83.3	73.6
1	9.6	9.8	12.4	19.8	9.5	11.6	10.1	16.6
2	3.6	3.5	3.1	5.3	2.2	2.5	3.5	4.3
3	1.7	1.8	1.4	2.2	2.8	0.4	1.6	3.6
4	1.4	1.2	1.5	1.0	2.8	0.4	0.6	1.1
5	1.0	0.5	0.5	0.7	0.3	0.7	0.9	0.0
6	0.6	0.8	0.2	0.3	0.3	0.4	0.0	0.0
7	0.0	0.5	0.0	0.2	0.3	0.4	0.0	0.0
8	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.7
9	0.1	0.0	0.1	0.2	0.3	0.0		
10	0.0	0.2	0.1	0.0	0.3	0.0		
11	0.0	0.0	0.0	0.0				
12	0.0	0.3	0.0	0.2				
N	801	600	801	600	317	277	317	277

注：ESRTは超短時間回答、ELRTは超長時間回答。数字は列%で、%は各指標の最大値まで表示。

で高いELRT数の割合が多くなっていた。

考察 まず、ESRT数と回答中断の間には有意な関連が認められなかったため、多くの画面で不正回答をするような回答者が中断群により多く含まれているとは言えない結果となった。また、ELRT数と回答中断の間には有意な関連が認められたものの、多くの画面で非専念回答を行っているような回答者が中断群により多く含まれているとは言えない結果となった。こうした結果は、いずれも、回答中断行動を管理する積極的な理由がないことを示唆している。

総合考察と結論

本研究では、まず、公募型Web調査の登録モニターのなかに、希少ながら、調査冒頭の調査依頼画面を提示したまま長時間にわたり回答を開始しない回答者がいることが明らかとなった。また、そのため、回答時間をパラデータとして利用する際には、調査会社に回答時間の測定方法を確認する必要があることが示唆された。

次に、公募型Web調査については、有効回答者の少なくとも4割超が何らかの理由で一時的に回答を中断していることが明らかとなった。また、そうした回答中断により回答時間は増加するものの、時間増は数十秒程度であることも多く、最大値も20分弱であった。

さらに、連続同一回答、準連続同一回答、超短時間回答数、超長時間回答数などを不正回答や非専念回答の指標とし、回答中断群と回答非中断群の間で回答の質に違いが認められるかどうか検討したところ、両群間に著しい差は認められず、回答中断群に不正回答や非専念回答を行う回答者がより多く含まれることを示す結果は得られなかった。

以上の結果から、公募型 Web 調査においては、現段階では、データ納品時の有効票について、一時的な回答中断のみを理由に、回答中断者や回答非中断者の回答を無効とする必要があるとは言えないことになる。もちろん、公募型 Web 調査に、回答者全員が同じ状況下で回答した、という仮定や、同一回答者の回答は全て同じ状況下で行われた、という仮定が必要となるのであれば話は別である。しかし、公募型 Web 調査にそこまでの厳密さが求められるケースは希であるように思われる。

なお、本研究では回答中断行動を不正回答や非専念回答の指標と見なして分析を進めたが、回答中断行動を別の側面から見れば、回答中断によって中断後の回答の質が高まる可能性があることも否定できない。具体的に言えば、長時間に及ぶ調査では、休息により認知資源の回復が期待できるし、各種の回答中断行動によって、回答に必要な認知情報処理の妨害刺激を除去できる可能性もある。本研究ではこうした点について直接検討しているわけではないが、回答中断行動の有無によって回答の質に差が生じていないのは、そうしたことが原因となっている可能性もあると考えられる。

ただし、以上の議論から、回答中断行動というパラデータの重要性が否定されるわけではない。長時間に及ぶ調査の後半で、回答中断を意味する超長時間回答が頻発するようであれば、回答者の疲労が閾値を超えている可能性が高くなり、そうした情報から、適切な質問量や調査時間に関する貴重な資料が入手できる可能性もある。

そのような意味では、必要なのはパラデータそれ自体ではなく、公募型 Web 調査がパラデータによる検票を必要とする特殊な調査手法であるという認識であり、経験に基づいた適切な調査管理であると言うべきである。そして、そのような意味においてであれば、現時点で問題にすべきなのは、回答中断行動よりも、むしろ、回答中断の有無に関わらず一定数存在する超短時間回答であると考えられる。ただし、現段階では超短時間回答の操作的定義や管理方法について十分な実証研究が行われているとは言えず、そうした点に関する実証研究は当該研究領域において急務となっているように思われる。

なお、最後に、本研究の結果の一般化可能性、および、方法上の制約について述べておきたい。まず、本研究に利用した Web 調査の回答者は一都三県に限定されており、こうした限定がない場合にどのような結果になるかについては未検討である。また、本研究に利用した Web 調査では回答デバイスを PC に限定していたが、スマートフォンからの回答を許容した場合にも同様の結果になるかどうかは不明である^(注8)。さらに、本研究で利用した Web 調査の質問数は20問未満と比較的少なかったため、質問数の多い Web 調査でも同様の結果が得られるとは限らない。

なお、本研究では回答中断行動の有無を自己報告によって測定しているが、分析2で検討したように、そうした自己報告に虚偽回答が含まれていないという保証があるわけではない。もちろん、現状

で、そうした虚偽回答が著しく多いことや、中断者の虚偽回答が非中断者の虚偽回答よりもはるかに多いことなどを示す具体的な知見が得られているわけではない。しかし、もしもそのようなことが起こっているとしたら、本研究の結果の解釈にも少なからず修正が必要となる可能性がある。こうした点もまた、今後の重要な研究課題であるように思われる。

以上のように、本研究の結果の一般化可能性には一定の制限があると言わざるを得ない。しかし、重要なのは結果の一般化であると同時に、公募型 Web 調査の有効利用にはパラデータをを用いた検票が不可欠であるという認識である。そのような注意喚起も含め、公募型 Web 調査については、さらなる方法論的基礎研究が必要であると言わねばならない。

注

- 1) 具体的な管理方法としては、まず、中断者の回答を分析から除外するという方法が挙げられる。例えば、Heerwegh & Loosveldt (2008) は回答時間 3 時間以上の回答者を除外しているが、そうした方法は中断者が少ない場合には一種の安全策として有効であると考えられる。ただし、そうした方法においては無効票判定基準の根拠が不明・曖昧であることも多く、中断者が多い場合にはサンプル数の減少や選択バイアスが問題となる。次に、より積極的な管理方法として冒頭宣言が挙げられる(増田, 2017)。冒頭宣言とは調査の冒頭で回答者に真面目に回答することを約束してもらう方法であるが、増田 (2017) はこの方法が不良回答者を減少させることを報告しており、同様の方法が回答中断行動を減少させることも十分予想される。
- 2) 本研究では、Web 調査の途中で回答が放棄され、特定の質問画面以降の回答が全て無回答となっている状態を脱落と呼ぶ(中止、あるいは、調査自体からの離脱と表現することも可能である)。これに対し、Web 調査の個々の質問画面への回答作業が一時的に中止されても、最終的には回答作業が完了している場合には、そうした中止を回答中断と呼ぶ。なお、脱落が個々の質問画面で生じた場合には無回答が発生するが、本研究で分析に用いた Web 調査は全質問画面で無回答を許容しない仕様であったため、回答中断による無回答は発生していない。
- 3) 同様の問題は訪問留置調査や郵送調査においても潜在していたはずであるが、これまでそうした問題が十分検討されてきたとは言えないようである。他方、Web 調査の場合は、回答時間というパラデータが入手できることで、そうした問題を経験的に検討することが可能となっており、そうした点も Web 調査のメリットのひとつであると考えられる。
- 4) CCES は Cooperative Congressional Election Study の略。ANES は American National Election Studies の略。どちらもアメリカの選挙研究のために実施される大規模調査である。
- 5) 公募型 Web 調査の仕様の詳細については発注者から質問してはじめて入手可能となるが、全ての質問について詳細な回答が得られるとは限らない。発注者の知識・技量・経験が問われる所以である。
- 6) 本研究の目的は回答中断行動が回答の質に与える影響を明らかにすることであり、そのために必要なのは回答中断者を識別するための操作的な指標である。そのため、ここでは個々の回答中断行動についての詳細な検討は行わない。
- 7) 群間比較のために、各条件で、長時間方向にある外れ値を数件除外している。
- 8) Antoun, et al. (2017) の実験によれば、PC 回答者とスマートフォン回答者の間に、回答の質の差は認められていない。ただし、この実験がパネルからの確率標本についてのものである点には注意が必要である。

引用文献

- Ansolabehere, S., & Schaffner, B.F. (2015). Distractions: The incidence and consequences of interruptions for survey respondents. *Journal of Survey Statistics and Methodology*, 3, 216-239.
- Antoun, C., Couper, M.P., & Conrad, F.G. (2017). Effects of mobile versus PC Web on survey response quality: A crossover experiment in a probability Web panel. *Public Opinion Quarterly*, 81, 280-306.
- Callegaro, M., Manfreda, K.L., & Vehovar, V. (2015). *Web survey methodology*. London: Sage.

- 埴淵知哉・村中亮夫・安藤雅登 (2015). インターネット調査によるデータ収集の課題——不良回答、回答時間、および地理的特性に注目した分析—— *E-journal GEO*, 10, 81-98.
- Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality, *Public Opinion Quarterly*, 72, 836-846.
- 神山貴弥・藤原武弘 (1991). 認知欲求尺度に関する基礎的研究 *社会心理学研究*, 6, 184-192.
- 増田真也 (2017). Web 調査における不良回答の検出法の比較と減少のための試み *日本行動計量学会大会発表論文抄録集*, 45, 56-59.
- 増田真也・坂上貴之・森井真広 (2017). Web 調査における不良回答と回答時間 *日本心理学会第81回大会発表論文集*, 10.

付記

本稿は日本行動計量学会第45回大会・特別セッション「Web 調査における回答行動の検討」、および、同学会第46回大会・特別セッション「Web 調査導入に伴う発想の転換」における筆者の発表内容に、大幅な加筆修正を加えたものである。なお、本研究は平成28～30年度日本学術振興会科学研究費補助金・基盤研究 (A)「多肢選択肢における回答行動の統合的研究：質問紙・ウェブ調査法の設計と妥当性の検討」(研究代表者：坂上貴之、課題番号16H02050) の助成を受けて行われた。

【Abstract】

Response Time and Respondent Interruptions in Volunteer Panel Web Surveys

Kazunari YAMADA

One of the benefits of Web surveys is that a variety of paradata regarding how respondents answer questions are generated by the survey process and that these paradata can be used to evaluate response quality of Web surveys. However, few studies have examined whether respondent interruptions alter response quality. To address this discrepancy, this article investigates the interruptions and their consequences for response time and response quality in volunteer panel Web surveys. Data were based on two Web surveys in the Tokyo metropolitan area (one of 1,444 Japanese adults in January 2017, and the other of 622 Japanese adults in January 2018). The results revealed the following: (1) interruptions are common (more than 40% of all respondents) and they add substantially to the duration of surveys, but (2) interruptions do not appreciably alter response quality, and (3) interruptions do not have a significant effect on the number of straight lining answers on the Likert grid containing 15 items, nor do they reduce the number of extremely short response time to questions. Thus, while interruptions may be an issue for survey administration, they do not appear to reduce the quality of data collected by researchers through volunteer panel Web surveys. Finally, the paper concludes by discussing implications and limitations of this research.