

公募型 Web 調査の回答の質に回答デバイスが及ぼす影響

Effects of Response Devices on Response Quality in a Volunteer Panel Web Survey

山田 一成

Kazunari YAMADA

江利川 滋*

Shigeru ERIKAWA

問 題

回答デバイスの変化

2010年以降のスマートフォン（以下SP）の急速な普及にともない、近年、Web調査におけるSP回答の割合が急増している。日本マーケティング・リサーチ協会によれば、日本の主要調査会社のWeb調査におけるSP回答の割合は、2013年には15.7%であったが、2019年には56.1%と過半数を占めるに至っている。また、そうしたSP回答の割合は年代が若いほど高く、2019年時点では20代で84.6%、10代では94.7%に達している（注1）。

こうしたWeb調査環境の変化は、一方においては、それまで調査が困難だった層からの回答を期待させるものであったが（Couper et al., 2017 ; Wells, 2015）（注2）、同時にそれは、SP回答における「回答の質」の低下を懸念させるものでもあった（Tourangeau et al., 2017 ; Antoun et al., 2017 ; Sommer et al., 2017）。そして、そうした懸念は調査設計における重大なジレンマにも関わっていた。もしも、回答デバイス間に回答の質の著しい差があれば、SP回答者の除外を検討しなければならないが、SP回答者の属性に顕著な特徴がある場合には、SP回答者の除外は致命的なセレクション・バイアスになってしまうからである（以下では、このジレンマをスマホ・ジレンマと呼ぶ）。

こうした状況を背景に、海外では2010年代の前半から回答デバイスと回答の質に関する実証研究が盛んに行われてきた。また、そうした先行研究において回答の質を低下させる原因だと考えられたのは、SPの操作性と携帯可能性であった（Antoun et al., 2017）。まず、操作性については、SPの画面が小さく狭いことや、ズームやスクロールなどのタッチスクリーン操作によって回答負担が増すこと、および、そうしたことによって誤入力が増えることが懸念された。さらに、携帯可能性については、家庭外での回答において様々なディストラクション（刺激による回答への注意の途切れ）が生じ、最小限化回答が増えることが懸念された。また、周囲に回答画面を覗くことができる人がいるとプライ

*株式会社TBSテレビ

バシーが守られているという意識が低下し、微妙な内容が回答されなくなることも危惧された。

では、そうした事柄が原因となって、実際に、SP回答とPC回答の間に回答の質の差が生じていたのだろうか。以下では、まず、海外での研究結果を概観する。

2017年以前の先行研究の概観

回答デバイスが回答の質に及ぼす影響については多くの研究が行われているが、2010年から2016年までの実証研究についてはTourangeau et al. (2017) が網羅的なレビューを行っている。彼らは当該の問題を扱った査読付き論文15点について、その調査結果を次のように要約している。

まず、レビューの対象となった複数の実証研究の結果は一貫しており、SP回答はPC回答に非常によく似ていた。測定誤差については、概して、回答デバイス間にほとんど差が認められなかったのである。もちろん、SP回答は他のデバイスからの回答に比べ、回答時間が長く(注3)、非回収や脱落や無回答が多く、自由記述が短い、といった特徴を有してはいたが、それらも、あまり大きな違いとは言えないと判断された(注4)。こうした結果は当初の予想に反するものであり、「SPは画面サイズが小さくデストラクションが生じやすいにも関わらず、回答の質を低下させていない」という逆説を示唆するものであった(以下では、この逆説をスマホ・パラドックスと呼ぶ)。

もっとも、Tourangeauたちも認めているように、レビューの対象となった実証研究には欧州での研究が多く、ロシアでの研究には携帯電話が含まれているなど、一般化に関する制限がないわけではない。また、レビュー対象となった研究にはLISSやGESIS Panelなどの確率標本を用いた研究が多く(注5)、公募型Web調査においても同様の結果が得られるという保証はない。そうしたこともあってか、彼らの中間総括以降も、こうした問題を扱った実証研究の結果は次々に発表されることとなった。

2017年以降の先行研究の概観

そうした後続研究としてまず挙げられるのは、Tourangeau et al. (2017) による2015年の米国でのフィールド実験である。この実験では、回答者は面接者から回答デバイスを渡され、それを用いて調査に回答する方式が採用された。回答デバイスはSP、タブレット端末(以下TB)、ノートPCの3種類で、回答の質の指標として採用されたのは、単一選択における順序効果、多件法における順序効果、2質問の提示方式(分離/結合)の影響、複数項目の提示方式(分離/グリッド)の影響、回答参照情報の提示方式の影響であった。

こうした指標について、当初Tourangeauたちは、質問全てを見渡すためにスクロールが必要となるSPでは、視認性の高い選択肢が選ばれやすくなるなど、他のデバイスとは異なる回答傾向が確認されると予想していた。しかし、調査の結果、各指標の主効果が認められることはあっても、回答デバイスが関わる交互作用はほとんど認められなかった。また、この調査データについては、回答時間、無回答、ストレートライニング、尺度の信頼性と妥当性などを指標とする分析も行われたが、回答時間がSPで長いこと以外には、回答デバイスによる差は認められなかった(Tourangeau et al., 2018)。

こうした結果は過去の研究結果とも整合的であり、SP回答への懸念を払拭するもののように思わ

れた。ただし、こうした結果は調査員が身近に存在する場合のものであったため、そのまま公募型 Web 調査に当てはまるわけではなかった。

次に、Tourangeau et al. (2017) と同時期の研究報告として、Antoun et al. (2017) が2013年にオランダの LISS パネルを用いて行った実験的調査が挙げられる。この調査では参加要請に応じた者が SP 群と PC 群に分けられ、各群から無作為抽出された者が回答者となった。また、回答者は、まず、どちらかのデバイス条件に回答し、その 1 ヶ月後に、同じ調査票にもう一方のデバイスから回答した (クロスオーバー・デザイン)。なお、回答の質の指標となったのは、非識別化回答 (nondifferentiation)、端数処理 (rounded numerical responses)、自由記述量、認知反射テスト (CRT)、マルチタスキング (ディストラクション)、社会的望ましさ、誤入力、などであった。

こうした調査の結果、まず、回答者のほとんどが家庭で回答していることが明らかとなった。ただし、SP 群ではマルチタスキングが多く、回答時に他者がいる場合が多く、回答時間も長くなっていたため、SP 回答の質の低下が予想された。しかし、上述の指標を調べると、SP 群と PC 群の間にはほとんど差が認められず、自由記述量は予想に反し PC 群で少ない傾向が認められた。また、社会的望ましさにはデバイス間に差がなかったが、誤入力は SP 群で多くなっていた。

こうした結果について Antoun たちは、SP の画面サイズの小ささが誤入力を増やすと指摘しながらも、SP の携帯可能性は回答の質に影響しないため、SP 回答者は Web 調査に含めるべきであり、SP の調査票は SP に最適化すべきである、と述べている。ただし、Antoun たちも認めているように、回答者は確率パネルであり、回収率は低く、1 ヶ月後とはいえ同じ調査票に 2 回答しているなど、一般化のための制限は多く、上述の研究結果も、そのまま日本の公募型 Web 調査において再現されるとは限らない。

さらに、回答の信頼性・妥当性についての研究としては Sommer et al. (2017) が挙げられる。彼女たちは2013年のドイツの選挙調査 (非商業パネル) を用いて、デスクトップ PC 群とモバイルデバイス群 (SP と TB) を比較している。回答の質の指標は、政治知識項目の信頼性係数、投票意図と政党支持・連立支持・前回投票先との異同などであったが、そうした指標について、上記 2 群の間には有意な差がまったく認められなかった。

こうした結果について Sommer たちは、モバイルデバイス群ではデスクトップ PC 群よりも低年齢層や女性が多く、また、脱落が多く回答時間が長いが、データの妥当性は損なわれていない、と結論づけている。ただし、彼女たちは、この研究は長期的に安定した政治的態度の研究であるため一般化には制限があるとし、回答時間が重要な意味をもつ質問や認知的負荷を必要とする質問については、さらなる研究が必要であると述べている。

なお、公募型 Web 調査についての研究としては Zou et al. (2021) が挙げられる。彼女たちは米国 (2017 年) と中国 (2018 年) のそれぞれで、観光をテーマとする公募型 Web 調査を行っている (調査会社委託、性・年代による割当法)。この調査では、モバイル回答者は米国で 48.7%、中国で 66.2% であり、そうしたデータについて各国ごとに、モバイル群と PC 群の間に回答の質の差があるかどうか検討さ

れた。指標となったのは、回答時間、心理尺度の内的一貫性、自由記述量、「その他」回答であった。

結果を見ると、回答時間については、中国では差がなく米国ではモバイル群で長くなっていた。心理尺度の一貫性については、両国ともモバイル群でいくつかの尺度の α 係数が低くなっており、自由記述量については、中国では差がなく米国では調査末尾の質問でのみモバイル群で自由記述量が少なくなっていた。なお、「その他」回答については、両国とも回答デバイスによる影響は認められなかった。こうした結果についてZouたちは、まず、米中ともモバイル回答とPC回答に大きな差はなく、モバイル回答の質が低いとは言えない、と結論づけている。また、彼女たちは2国間の差について、中国は「PC時代」を経験せずにモバイル機器の利用が始まったため、モバイル機器への親近度が高いことが原因である、と主張している。

こうしたZouたちの研究は、SP回答の割合が明示されていない点に議論の余地を残すものの、回答デバイス間に回答の質の差がないことを示している点では、従来の研究結果と概ね整合的であると考えられる。また、この研究では近年の公募型Web調査が分析対象となっており、その点でも注目に値する。

本研究の目的

以上のように、回答デバイスの影響に関する実証研究はTourangeau et al. (2017) のレビュー以降も盛んに行われている。また、そうした後続研究の結果を概観すると、Tourangeau et al. (2017) の中間総括と同様に、「SP回答はPC回答に比べ、長時間化するものの、回答の質が低下するわけではない」との総括が得られるようである。

もちろん、後続研究においては、個別具体的な質問形式や回答行動に特化した研究も行われており(注6)、そうした個々の分析を詳しく検討すれば、SP回答とPC回答の差異が部分的に観察されることもある。しかし、そうした詳細な結果報告も含め、海外では既に、回答デバイスの影響について基本的な情報の共有と多岐にわたる検討が行われており、今後の調査設計においても先行研究の結果を参照することが可能となっている。

そうした状況と比較すると、日本においては回答デバイスに関する実証研究は今なお希少であると言わざるを得ない。特に、公募型Web調査における回答デバイスの影響については実証的な知見に乏しく、現状では調査設計時に必要な情報も十分ではないように思われる(注7)。そこで本研究では、公募型Web調査を実施し、回答デバイスごとに回答者属性が異なるかどうか検討するとともに、回答デバイス間に回答の質の差がないという逆説的状况(スマホ・パラドックス)が生じているかどうか検証する。そして、そのうえで、回答デバイスに関わる問題状況(スマホ・ジレンマ)をどう捉えるべきか検討する。

方 法

分析方法

以下、分析1では、回答デバイスを指定しない公募型Web調査において、回答デバイス種別の割合を明らかにする。また、回答デバイス各群の回答者属性を比較し、回答デバイス各群が主要属性について等質かどうか検証する。次に、分析2では、回答デバイス各群間で回答ストレス感、回答中断行動、回答時間に違いがあるかどうか検討する。最後に、分析3では、回答の質の指標としてDKNA回答、非識別化回答（nondifferentiation）、ストレートライニング（straight-lining）、尺度項目の内定一貫性という4種類のパラデータを取り上げ、回答デバイス間に回答の質の差が認められるかどうか検証する（注8）。

調査概要

公募型Web調査（調査会社委託、ポイント報酬制、スクリーニング調査と本調査を一体化した調査）（注9）。一都三県在住の男女20歳～69歳が対象。調査は2017年3月30日（木）～31日（金）に実施。本調査はスプリット調査であり調査票種別は4種類。各調査票ごとに、直近の一都三県の性年代別人口比に基づいて、性年代別の10区分に対し1,000人を割り付け、総計4,000人を目処に実査を行った。実査では、まず、調査会社の登録モニターを対象に、スクリーニング調査を68,555人に配信して5,239人から回答を得た。次に、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（家族に従業者がいる者を含む）、および、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、4,317人を本調査の対象者とし、最終的に4,140人の有効回答を得た。なお、調査では1画面に1質問を表示し、警告表示により無回答を許容しない仕様とした。また、SPの調査票はSPに最適化される仕様であった。

結 果

分析1：回答デバイスと回答者属性

回答デバイスの割合 まず、各回答デバイスの割合を調査会社の調査システム上で調べたところ、PCが56.4%、SPが39.7%、タブレット端末（TB）が4.0%で（ $N=4,140$ ）、SPの割合は日本マーケティング・リサーチ協会の2017年度統計とほぼ一致する結果であった。なお、今回の調査には回答デバイス種別を尋ねる質問が設けられていたため、調査システム上の回答デバイス記録と自己報告の一致度を調べた（Table 1）。その結果、記録と自己報告は概ね一致するものの、一部に誤解や虚偽と思われる回答パターンが認められた。そのため、以下の分析では回答デバイスの識別に調査システム上の記録を用いることとした。また、こうした結果から、今後の実証研究における回答デバイス種別の識別には、十分な注意と事前検討が必要であると考えられた。

Table 1 回答デバイスの自己報告と調査システム上の記録

記録	自己報告						N
	DT-PC	N-PC	SP	FP	TB	Other	
PC	39.7	57.4	1.5	0.4	0.6	0.3	2,333
SP	1.5	3.3	92.4	0.7	1.6	0.5	1,643
TB	6.1	7.3	4.9	0.0	81.7	0.0	164
計	23.2	34.0	37.7	0.5	4.3	0.4	4,140

注：数字は行%。記録は調査システム上の記録。DT-PCはデスクトップPC、N-PCはノートPC、SPはスマートフォン、FPは携帯電話（SP以外）、TBはタブレット端末、Otherは「その他」。

Table 2 回答デバイス各群の回答者属性

Device	性別		年代					N
	男性	女性	20代	30代	40代	50代	60代	
PC	55.8	44.2	10.7	13.2	23.1	22.7	30.3	2,333
SP	44.1	55.9	25.7	31.6	26.2	11.4	5.0	1,643
TB	48.2	51.8	12.8	20.1	25.6	23.2	18.3	164

注：数字は行%。Deviceは回答デバイス種別。SPはスマートフォン、TBはタブレット端末。

回答デバイス各群の回答者属性 次に、回答デバイス各群（PC、SP、TB）の間で回答者属性に違いがあるかどうか検討した（Table 2）。その結果、性別については、SP群ではPC群より女性が10ポイントほど多くなっていた（ $\chi^2(2)=53.3, p<.001, \text{Cramer's } V = .11$ ）。また、年代についてはPC群よりもTB群、TB群よりもSP群で若い年代が多く、特にSP群では30代以下の割合が6割弱と、PC群の2.5倍近い割合となっていた（ $\chi^2(8)=676.7, p<.001, \text{Cramer's } V = .29$ ）。なお、以上の結果は海外での先行研究と一致するものであった。

一方、教育水準、世帯年収、未既婚、子供の有無、一都三県の別については3群間に顕著な差は認められなかった。米国では低教育水準層や低収入層でSP回答が多いとのレビューもあるが（Wells,2015）、本研究ではそうした傾向は認められず、属性との関連については社会状況や標本抽出法の影響が大きいと考えられた。

分析2：回答デバイスと回答負荷

回答デバイスと回答ストレス感 回答ストレス感の質問位置は全13問中12問目で（注10）、Table 3に挙げた10項目について、それぞれ該当するかどうか2件法で回答を求めた。分析に際しては、不正回答の疑いがある超短時間回答者を除外するために、各回答デバイス群ごとに、回答ストレス感質問の質問画面提示時間を自然対数化した値を用いて、それぞれ-2SD未満の回答者を除外した（SDは回答時間の長時間方向の外れ値を除外した後に算出した）（Smyth et al., 2006）。そのうえでクロス集計を行い、回答デバイス各群間で回答ストレス感を比較した。その結果、8項目で有意な関連が認められ（Bonferroni法）、そのうち6項目でSP群の該当率が最も高率となっていた（Table 3）。特に、項目8と項目9は回答行動への影響が強く懸念される内容であり、SP群の該当率も4割前後と高率

Table3 回答デバイス各群の回答ストレス感

回答ストレス感	回答デバイス			V	p
	PC	SP	TB		
1 処理速度が遅いので、ストレスを感じた	13.6	16.5	12.3	.04	ns
2 ディスプレイ画面が狭いので、ストレスを感じた	8.0	30.1	12.9	.29	***
3 ブラウザの表示が遅いので、ストレスを感じた	13.4	15.8	10.4	.04	ns
4 カーソル操作（タッチ操作）が思うようにならず、ストレスを感じた	10.7	21.5	21.5	.15	***
5 キーボードが使いにくいので、ストレスを感じた	5.7	12.8	13.5	.13	***
6 画面に表示される文字が小さいので、ストレスを感じた	16.9	24.4	27.6	.10	***
7 ときどきブラウザが反応しなくなるので、ストレスを感じた	14.1	18.8	16.6	.06	**
8 文章や選択肢が画面からはみ出して、見づらい質問があった	9.8	43.0	12.9	.38	***
9 ディスプレイ画面の表示が小さいので、選択肢が選びにくかった質問があった	12.4	37.0	27.0	.29	***
10 長い文章の表示が続いて、答えるのが面倒になった質問があった	24.7	31.5	28.2	.07	***
N	2,283	1,597	163		

注：数字は列%（超短時間回答者を除外）。SPはスマートフォン、TBはタブレット端末。VはCramer's V。有意水準は表全体をBonferroni法で調整。* $p < .05$, ** $p < .01$, *** $p < .001$

Table4 回答デバイス各群の回答中断行動

回答中断行動	回答デバイス			V	p
	PC	SP	TB		
1 電話	4.2	5.8	8.0	.05	ns
2 電子メール	3.5	6.9	4.9	.07	***
3 家の中での家族とのやり取り	7.7	12.9	17.3	.10	***
4 仕事や家事	7.1	10.7	4.9	.07	**
5 食事	3.9	7.0	1.9	.08	***
6 入浴	2.7	3.9	3.1	.03	ns
7 睡眠	2.9	5.8	3.1	.07	***
8 休憩	4.2	6.7	4.9	.05	*
9 テレビ視聴	7.9	9.9	6.8	.04	ns
10 新聞・雑誌・コミックなどの閲読	2.1	2.8	2.5	.02	ns
11 ネットやSNSの閲覧	2.9	6.2	3.7	.08	***
12 その他	6.4	10.5	8.6	.07	***
N	2,316	1,632	162		

注：数字は列%（超短時間回答者を除外）。SPはスマートフォン、TBはタブレット端末。VはCramer's V。有意水準は表全体をBonferroni法で調整。* $p < .05$, ** $p < .01$, *** $p < .001$

であったことから、SPの画面の小ささ・狭さに起因する悪影響が生じている可能性が高いことが示唆された。

回答デバイスと回答中断行動 回答中断行動の質問位置は最後（13問中13問目）で、Table 4に挙げた12項目について、それぞれ該当するかどうか2件法で回答を求めた。なお、分析に際しては、上述の回答ストレス感と同様に、超短時間回答者を除外してクロス集計を行い、回答デバイス各群間で回答中断行動を比較した。その結果、8項目で有意な関連が認められ（Bonferroni法）、そのうち7項目でSP群の該当率が最も高率となっていた（Table 4）。また、回答中断行動の該当項目数を算出したところ、回答中断行動が1項目以上あった者の割合は、PC群の25.3%に対し、TB群で30.2%、

Table5 回答デバイス各群の調査回答時間

Split	Device	最小	最大	percentile				歪度	N	除外
				25%	50%	75%	95%			
A	PC	72	1863	188	253	337	616	3.5	1,162	1
	SP	64	1869	183	251	353	648	3.1	824	5
	TB	93	529	165	223	318	460	0.7	77	1
B	PC	74	2161	216	286	387	781	3.2	1,167	3
	SP	76	2531	208	280	398	684	4.4	809	5
	TB	135	895	218	279	378	678	1.6	83	3

注：数字は秒。調査回答時間は同一調査票種別の回答者全員に共通する質問の回答時間の合計。Splitは調査票種別、Deviceは回答デバイス。SPはスマートフォン、TBはタブレット端末。除外は除外された回答時間の外れ値数（長時間方向）。

SP群で34.7%と、モバイルデバイスでやや多くなっていた ($\chi^2(2)=40.9, p<.001$, Cramer's $V = .10$)。こうした結果は、回答デバイスの携帯可能性により、SP（およびTB）ではデストラクションが起りやすいことを示唆していると考えられた。

回答デバイスと回答時間 本調査の質問のなかには、他の質問への回答内容に従って、質問表示の有無や表示される選択肢の種類が異なる仕様の質問があった。そのため、回答時間の算出にあたっては、全13問のなかからそうした2質問を除外し、残る11問の回答時間の合計値（以下、調査回答時間）を比較した。ただし、そうした11問のなかには、スプリット法により回答形式が2種類用意されていた質問がひとつあったため、調査票の種別は2種類となり（Table 5のSplitのAとB）、分析もそれらの種別ごとに行われた。

結果はTable 5に挙げたとおりで、調査回答時間のパーセンタイルを比較すると、両調査票種別に共通して、SP群とPC群の間に大きな差はみられなかった。また、各調査票種別ごとに、調査回答時間を自然対数化した値について一元配置分散分析を行ったが、回答デバイス間に平均値の有意差は認められなかった（注11）。

しかし、回答時間は年代が若いほど短いことが予想され、SP回答の割合は若い年代ほど多かったため、調査回答時間の比較を年代別に行ったところ（Table 6）、両調査票種別に共通して、全ての年代で、SP群ではPC群より調査回答時間が長くなる傾向がうかがわれた（TB群はサンプル数が少ないため表記を省略）。また、そうした調査回答時間の差は年代が上がるほど大きく、50代で最大となっていたが、60代ではPC群の調査回答時間が長時間化することにより差が小さくなっていた。なお、差の大きさについては、A票では、第2四分位での最大は86秒、第3四分位での最大は164秒であった。また、B票では、第2四分位での最大は65秒、第3四分位での最大は127秒であった。こうした値は質問数（11問）を考慮すると、平均で1質問あたり、第2四分位で最大8秒弱、第3四分位で最大15秒の長時間化であった。

ただし、上述の対数化回答時間について回答デバイスと年代を要因とする2要因分散分析（平方和

Table 6 回答デバイス各群の調査回答時間（年代別）

Split	年代	Device	最小	最大	percentile				歪度	N
					25%	50%	75%	95%		
A	20	PC	72	1843	144	212	290	576	4.4	125
		SP	64	1801	161	220	301	593	3.8	208
A	30	PC	77	1327	152	204	284	709	3.0	157
		SP	77	1869	174	237	310	606	3.9	262
A	40	PC	89	1863	169	230	306	560	4.1	268
		SP	105	1353	194	256	356	638	2.5	215
A	50	PC	101	1704	196	247	334	637	3.4	266
		SP	93	1500	251	333	498	970	2.3	91
A	60	PC	111	1663	242	310	392	629	3.4	346
		SP	148	1291	259	340	492	894	1.9	48
B	20	PC	74	2586	165	238	318	1073	4.2	124
		SP	85	2302	176	246	362	699	4.3	213
B	30	PC	84	1263	178	228	323	832	2.5	148
		SP	90	2531	204	275	378	573	5.3	253
B	40	PC	90	2140	197	242	332	801	3.6	271
		SP	76	2206	218	293	406	700	4.2	213
B	50	PC	114	1226	227	281	369	702	2.6	264
		SP	119	2449	254	346	496	834	4.4	96
B	60	PC	95	2111	295	363	461	764	3.2	361
		SP	174	1300	272	373	499	946	2.3	34

注：数字は秒。調査回答時間は同一調査票種別の回答者全員に共通する質問の回答時間の合計。Splitは調査票種別、Deviceは回答デバイス。SPはスマートフォン。

はタイプ II) を行ったところ、まず、A 票については、回答デバイスの主効果 ($F(2, 2048) = 16.82, p < .001, \eta^2 = .01$) と年代の主効果 ($F(4, 2048) = 47.84, p < .001, \eta^2 = .08$) が有意で、交互作用は有意ではなかった。また、多重比較 (Tukey の HSD) の結果、回答デバイス間には有意差が認められなかったが、年代については有意差が認められ、若い年代ほど回答時間が短かった (20代と 30代、30代と 40代の間を除く)。次に、B 票については、回答デバイスの主効果 ($F(2, 2044) = 9.93, p < .001, \eta^2 = .01$) と年代の主効果 ($F(4, 2044) = 44.07, p < .001, \eta^2 = .08$) が有意で、交互作用は有意ではなかった。また、多重比較 (Tukey の HSD) の結果は A 票の結果とまったく同様であった。

以上のように、本研究では海外の先行研究とは異なり、SP 回答の回答時間が PC 回答のそれより明らかに長いという結果は得られなかった。

分析 3：回答デバイスと回答の質

DKNA 回答 教育水準と世帯年収を尋ねる質問の DKNA 回答率を調べた。まず、教育水準の「答

えたくない」という選択肢の回答率は、PC群1.2%、SP群1.0%、TB群1.8%と各群とも希少で、両変数の間に有意な関連は認められなかった ($N=4,140$)。また、世帯年収の「わからない・答えない」という1 選択肢の回答率は、PC群22.3%、SP群18.7%、TB群22.6%とSP群でやや低くなっていた ($\chi^2(2)=7.9, p<.05$, Cramer's $V = .04$)。以上の結果から、SP群でDKNA回答の割合が高くなってはいないと判断された。

非識別化回答 非識別化回答の測定には映像・動画視聴に関する質問 (18項目) を用いた。この質問には背反関係になりうる項目のペアが3 組含まれており (注12)、全項目選択や全項目非選択は矛盾する回答パターンとなるため、非識別化回答の指標となりうると考えられた。なお、この質問はスプリット法により回答形式が2 種類用意されており、一方はあてはまる項目をいくつでも回答するマルチアンサー (MA) 形式、もう一方は項目ごとにあてはまるかどうかを回答する個別強制選択 (FC) 方式であった。そこで、それぞれの回答形式ごとに、回答デバイス各群間で全項目選択 (および、1 項目非選択) と、全項目非選択 (および、1 項目選択) の割合を比較した (1 項目非選択や1 項目選択を準指標としたのは、回答者が全項目同一回答へのアラートを回避する目的で、1 箇所だけ回答を変える可能性があったためである)。

結果を見ると、まず、MA形式については全項目選択と1 項目非選択は皆無であったが、全項目非選択はPC群22.4%、SP群12.4%、TB群23.1%とSP群で低かった ($N=2,070$; $\chi^2(2)=33.0, p<.001$, Cramer's $V = .13$)。また、1 項目選択はPC群18.4%、SP群20.6%、TB群17.9%と同程度であった。次に、FC形式については、全項目選択はPC群0.6%、SP群0.6%、TB群0.0%と各デバイスとも希少であった ($N=2,070$)。また、1 項目非選択もPC群1.1%、SP群1.0%、TB群0.0%と同様であった。他方、全項目非選択はPC群3.0%、SP群2.9%、TB群1.2%で、回答デバイスとの有意な関連は認められなかった。また、1 項目選択はPC群7.2%、SP群4.1%、TB群3.5%で、SP群で高くなる傾向は見られなかった。

なお、上記の傾向がテレビとスマートフォンの所有状況の影響を受けている可能性も考えられたため、テレビとスマートフォンの両方を所有する層 (MA形式: $N=1,553$ 、FC形式: $N=1,580$) について同様の分析を行ったが、結果は上記と同様であった。以上の結果から、SP群で非識別化回答の割合が高くなってはいないと判断された。

ストレートライニング ストレートライニング (SL) の測定には、改訂版テレビ親近感尺度 (以下TAS、江利川・山田, 2012) の項目 (7 項目・5 件法)、および、視聴している番組・動画のジャンルに関する質問 (15項目・4 件法) を用いた。

まず、TAS項目については、7 項目中に逆転項目が1 項目含まれていたため、選択肢両端におけるSLは不正回答の疑いが濃厚となるが、回答デバイス各群間には選択肢両端におけるSLの割合にまったく差がみられなかった (注13)。また、5 件法の項目では中央選択肢でSLが増えることが知られているが (山田・江利川, 2021)、そうした回答の割合はPC群2.6%、SP群1.0%、TB群2.4%とSP群で低かった ($\chi^2(2)=12.6, p<.01$, Cramer's $V = .06$)。なお、こうした結果は、テレビ所有層を対象と

Table 7 心理尺度質問におけるストレートライニング

尺度	Device	選択肢別					合計	N
		1	2	3	4	5		
TAS	PC	0.5	0.7	2.6	0.2	0.4	4.3	2,260
	SP	0.4	0.7	1.0	0.1	0.4	2.6	1,602
	TB	1.2	0.0	2.5	0.0	0.0	3.7	163

注：数字は%（回答者はテレビ所有者）。TASは改訂版テレビ親近感尺度。Deviceは回答デバイス。SPはスマートフォン、TBはタブレット端末。選択肢の言語ラベルは1から5へ「あてはまる、まああてはまる、どちらともいえない、あまりあてはまらない、あてはまらない」。

Table 8 視聴行動質問におけるストレートライニング

Device	SL・NSL	選択肢別				合計	N
		1	2	3	4		
PC	SL	0.0	0.2	0.1	1.2	1.5	2,333
	NSL	0.2	0.9	0.2	1.2	2.4	
SP	SL	0.1	0.2	0.1	0.9	1.3	1,643
	NSL	0.3	0.3	0.6	1.1	2.3	
TB	SL	0.0	0.0	0.6	0.6	1.2	164
	NSL	0.6	0.6	0.6	1.2	3.0	

注：数字は%。Deviceは回答デバイス。SPはスマートフォン、TBはタブレット端末。SLはstraight-lining、NSLはnear straight-lining。選択肢の言語ラベルは1から4へ「よく見る、ときどき見る、あまり見ない、まったく見ない」。

した場合もまったく同様であった（Table 7）。

ただし、この質問は調査冒頭に配置されており、1画面に表示される項目数も7項目と少なかったため、SLが生じにくかった可能性もある。そこで、より調査後方に位置し（13問中5問目）、1画面の項目数（判断回数）が15項目と多く、回答者の疲労感や飽きの影響を受けやすかった質問として、視聴している番組・動画のジャンルに関する質問におけるSLとNSL（near straight-lining）について検討した（NSLは1項目のみ別回答の準SL）。結果はTable 8に示すとおりで、SL・NSLは各選択肢とも希少で、そうした割合には回答デバイス各群間に差が認められなかった。以上の結果から、SP群ではSL・NSLの割合は高くなっていないと判断された。

尺度項目の内的一貫性 上述のTASについて、回答デバイス各群間で信頼性係数（ α 係数）の値に差があるかどうか検討した。その結果、まず、TAS 7項目の α 係数は、PC群0.94（ $N=2,333$ ）、SP群0.93（ $N=1,643$ ）、TB群0.93（ $N=164$ ）で回答デバイス間に大きな違いはなく、3群とも十分に高い値であった。なお、テレビ所有層を対象とした場合でも、TAS 7項目の α 係数は、PC群0.94（ $N=2,260$ ）、SP群0.93（ $N=1,602$ ）、TB群0.93（ $N=163$ ）と上述とまったく同様の結果であった。以上の結果から、SP群における尺度項目の内的一貫性は他の2群と同程度であると判断された。

総合考察と結論

本研究では日本の公募型 Web 調査について、回答デバイス種別の割合を調べるとともに、回答デバイス各群の回答者属性と回答行動の違い、および、回答の質の違いについて検討した。結果を見ると、まず、スマートフォン (SP) からの回答は約 4 割であった。また、SP 回答者は女性で多く、SP 回答者の 6 割弱が若い年代 (20代と30代) によって占められていた。次に、回答行動に関しては、SP 群では PC 群に比べ回答ストレス感が非常に高く、回答中断行動が若干多くなっていた。ただし、回答時間については、年代を統制したところ、PC 群よりも SP 群で長時間化する傾向がうかがわれたが、対数化回答時間の平均値については回答デバイス間に有意差が認められなかった。なお、回答の質については、DKNA 回答、非識別化回答、ストレートライニング、尺度項目の内的一貫性という 4 種類のパラデータについて検討した結果、SP 群で PC 群より回答の質が低下することを示す結果は得られず、この点は海外の先行研究と整合的な結果となった。以下では、こうした結果の含意について検討する。

まず、注目すべきなのは、日本の公募型 Web 調査においてもスマホ・パラドックスが認められたことである。本研究では、SP は画面が小さく狭いため回答負荷が大きく、ディストラクションが生じやすい可能性があることが示された。にも拘わらず、SP 回答は明らかな長時間化を示さず、回答の質は PC 回答に劣るとは言えないことも示された。こうした逆説的な結果は、公募型 Web 調査を行う立場から見ると、それ自体としては好ましいものだと言ってよいだろう。

なお、こうした結果が生じる原因としては、まず、SP に慣れ親しみ SP 操作に長けた回答者 (登録モニター) が、SP を Web 調査の回答デバイスとして選択・利用した可能性が考えられる。そして、それが主たる原因であれば、今後の SP の普及や登録モニターの世代交代によって、公募型 Web 調査における回答デバイスの違いは問題ではなくなっていくことも十分予想される。ただし注意が必要なのは、SP による回答に困難を感じた回答者や、SP によって低質の回答を行いがちな回答者は、脱落によってサンプルに含まれなかった可能性がある、という点である。また、そうだとすれば、SP 回答者には PC 回答者よりも、真面目で粘り強いパーソナリティ特性の持ち主が多く含まれているかもしれない。この点は今後の重要な検討課題であると考えられる。

しかし、その他の原因として、そもそも本研究で行った調査が、全体として質問数が少なく、不良回答や不正回答が発生しにくいものであった可能性も考えられる。こうした点への積極的な言及は、先行研究においても意外なほど少ないが、質の低い回答は回答デバイスや回答者の特性によって生じるだけでなく、調査の特性によっても生じるはずである。長大な調査票やリッカートグリッドの連続、目的や意義が理解しにくい調査内容など、回答者の回答への動機付けを低下させる要素がないかどうか、調査それ自体についても検討されるべきである。

また、そのような意味では、本研究の結果の一般化には一定の制約があることになる。本研究で実

施した調査は2017年に一都三県の20歳～69歳を対象としたものであるが、制約はそれだけに留まらない。本研究の本調査の質問数は13問と比較的少なく、回答時間の中央値も5分程度に過ぎなかった。また、1質問の項目数やジャッジ（判断回数）が20を超える質問はほとんどなく、リッカートグリッドも3問であった。さらには、調査内容が誰にでも容易に回答可能な日常的情報行動であったことも、回答の質が低下しなかった原因となっていたかもしれない。そのため今後は、より質問数が多い調査や目的を異にする調査においても、回答デバイスの影響が生じないかどうか、さらなる検証が必要であると言わねばならない（注14）。

次に、公募型 Web 調査の設計において、回答デバイスをPCだけに制限すべきかどうかという問題（スマホ・ジレンマ）について考えておきたい。Web 調査については、不正回答や不良回答を除外するための方法として、回答時間を計測し超短時間回答者を除外するという方法が知られている（Smyth et al., 2006）。また、従来は、SP回答の質がPC回答とは異なる可能性があることをもって、回答デバイスをPCだけに制限することが可能であり、それによって、回答時間によるデータクリーニングも容易となっていた。

しかし、本研究の結果が示すように、SP回答とPC回答の間に質の差がないことが一般化可能だった場合には、SP回答を排除する積極的な根拠が減ることになる（「スマホ・パラドックスによるスマホ・ジレンマの解消」と言えるかもしれない）。ただし、回答デバイスを制限しない場合には、回答デバイスごとに超短時間回答者を識別する必要も生じうるため、データクリーニングの手間が増加することも懸念される。そのような意味では、本研究の結果は分析のコスト増につながりうるものであり、手放して好ましい結果だと言えるわけではない。

また、最後に、これまであまり議論されていない論点として、公募型 Web 調査の回答者がSP回答を望んでいるかどうか、という問題があることにも触れておきたい。ここまでの議論のなかで、SP回答者の割合が増加していることや、SP群とPC群の回答の質に差がなかったことが明らかとなったため、SP回答者は自ら進んでSPを利用したとのイメージが生まれているかもしれない。しかし、本当にそうだろうか。本研究の結果では、SP群で回答ストレス感が顕著に高まっていたが、そうしたSP回答者は、その後も喜んでSPで回答し続けているのだろうか。また、公募型 Web 調査の場合、回答者の脱落率についての情報は入手困難であるが、回答デバイスごとに脱落率に大きな差があるのではないだろうか。そうした考察から導かれるのは、登録モニターの回答頻度や継続年数によって、回答デバイスの選好に違いがあるかどうかというリサーチ・クエスチョンである。回答デバイスの制限が調査のカバレッジにどのような影響を与えるかを考えるうえで、こうした検討が不可欠であることは言うまでもないだろう。

今後も公募型 Web 調査の利用は増加し、SPの普及もさらに進んでいくものと予想されるが、そうした動向のなかで明らかになりつつあるのは、公募型 Web 調査は「お手軽でコストパフォーマンスの良い手法」であるとは限らない、ということである。公募型 Web 調査には適切な使い方があり、そうした使い方を具体的に知るためには、十分な基礎研究と情報の公開・共有が必要とされる。本研

究の結果が指し示しているのも、そうした当たり前の事実である。

注

- 1) 「インターネット調査品質ガイドライン 第2版」(一般社団法人日本マーケティング・リサーチ協会インターネット調査品質委員会、2020年5月刊行)。
- 2) 例えば、低所得層、低教育水準層、マイノリティなど (Keusch & Yan, 2017)。
- 3) SP回答における時間増の原因としては、通信回線速度の遅さ、画面サイズの小ささ(スクロールや読み取り)、携帯に起因するディストラクションなどが考えられるが、Couper & Peterson (2017) は大学生データを分析し、時間増の主な要因はスクロールであり、特に質問形式がグリッドであるときにそうである、と述べている。
- 4) 同時期のレビューである Wells (2015) や Couper et al. (2017) においても同様の見解が示されている。ただし、Couper et al. (2017) には一貫した結果が得られていない論点への言及もある。
- 5) LISS (オンライン確率パネル) はオランダの研究機関 CentERdata のデータ。GESIS Panel (確率ミックスモードパネル) はドイツのライプニッツ社会科学研究所 GESIS のデータ。
- 6) 順位法 (Revilla, 2017; Revilla et al., 2017; Revilla & Couper, 2018b)、グリッド (Revilla & Couper, 2018a)、選択肢順序 (昇順・降順) (Krebs & Höhne, 2021)、マルチタスキング (Höhne et al., 2020)、などが挙げられる。
- 7) そうしたなか、二瓶 (2015)、齊藤・二瓶 (2016)、工藤ほか (2018)、水野ほか (2018)、高橋・成田 (2019)、眞嶋・中村 (2021) は貴重な研究報告となっている。
- 8) 本論文の一部は2017年の日本社会心理学会第58回大会で発表された内容を再分析・再構成したものである。
- 9) 調査は株式会社 TBS テレビ・マーケティング部 (現: 総合マーケティング室) と本論文の著者を構成員とするアクティブメディア研究会が共同で実施した。データ利用を許可していただいた株式会社 TBS テレビ・総合マーケティング室に感謝する。なお、調査は調査会社に委託されたが、回答者の個人情報保護は調査会社の個人情報保護ポリシーに従って管理されており、プライバシーに関する問題はないと判断された。また、本研究で行うのはバラデータの分析であり、倫理的な問題はないと判断された。
- 10) スクリーニング後の本調査において、標準的な回答者が回答する質問は全13問であった (回答内容によっては「非該当」による質問のスキップが生じうる)。そのため、以下では特に断りのない限り、本調査の質問数を便宜的に「13問」と見なす。
- 11) 一元配置分散分析では、各調査票種別ごとに各回答デバイスの調査回答時間の分布を調べ、長時間方向の外れ値を除外した (除外数は Table 5 の「除外」欄に表示)。
- 12) 視聴画面志向 (大画面/スマホ画面)、違法動画意識 (良くない/気にならない)、視聴形態志向 (少しずつ/一気に) の3ペア。
- 13) これらの項目では NSL は十分ありうる回答である。また、調査会社のなかには、項目数が少ない調査画面では全項目同一回答へのアラートを表示しない仕様を採用している会社もあるため、回答者がそれを学習している可能性も考えられた。そのため、ここでは SL のみを回答の質の指標とした。
- 14) ただし、注意が必要なのは、今後 SP 群において回答の質の低下が認められても、それによって直ちに SP という回答デバイスが全否定されるわけではないという点である。というのも、そうした質の低下が示唆しているのは回答デバイスの適否であるとともに、「質問量の多さ」に代表されるような「調査自体の問題点」でもあるからである (質問量を著しく増やすなど、回答の質を低下させることは容易である)。また、そのような意味では、真に必要とされているのは「質問量の上限」に関する情報や経験則だと言うこともできるだろう。

引用文献

- 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.
- Couper, M.P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., & West, B.T. (Eds.) *Total survey error in practice: Improving quality in the era of big data*. New York: Wiley, pp.133-154.
- Couper, M.P. & Peterson, G.J. (2017). Why do web surveys take longer on smartphones?. *Social Science Computer Review*, 35, 357-377.

- 江利川滋・山田一成 (2012). 改訂版テレビ親近感尺度の信頼性と妥当性 心理学研究, **82**, 547-553.
- Höhne,J.K., Schlosser,S., Couper,M.P., & Blom,A.G. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, **111**, 106417.
- Keusch,F. & Yan,T. (2017). Web versus mobile web: An experimental study of device effects and self-selection effects. *Social Science Computer Review*, **35**, 751-769.
- Krebs,D. & Höhne,J.K. (2021). Exploring scale direction effects and response behavior across PC and smartphone surveys. *Journal of Survey Statistics and Methodology*, **9**, 477-495.
- 工藤公久・村上智章・岸田典子・二瓶哲也・出口敬子 (2018). インターネット調査の課題と品質向上に向けた取り組み：持続可能な調査環境を目指して 政策と調査, **15**, 5-12.
- 眞嶋良全・中村絨子 (2021). 携帯端末とマトリックス提示はオンライン調査の質を低下させる？ 日本認知心理学会発表論文集 日本認知心理学会第18回大会 (OS5-2).
- 水野一成・鈴木孝幸樹・吉良文夫 (2018). モバイル動向調査でみる回答の傾向の差異：スマートフォン回答者とパソコン回答者の特性 政策と調査, **15**, 13-19.
- 二瓶哲也 (2015). インターネット調査の新潮流：スマホユーザーの増加とその活用可能性 政策と調査, **9**, 59-66.
- Revilla,M. (2017). Are there differences depending on the device used to complete a web survey (PC or smartphone) for order-by-click questions?. *Field Methods*, **29**, 266-280.
- Revilla,M. & Couper,M.P. (2018a). Comparing grids with vertical and horizontal item-by-item formats for PCs and smartphones. *Social Science Computer Review*, **36**, 349-368.
- Revilla,M. & Couper,M.P. (2018b). Testing different rank order question layouts for PC and smartphone respondents. *International Journal of Social Research Methodology*, **21**, 695-712.
- Revilla,M., Toninelli,D., & Ochoa,C. (2017). An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones. *Telematics and Informatics*, **34**, 30-42.
- 齊藤ひとみ・二瓶哲也 (2016). PC・スマホの調査画面における回答方法の考察：アイトラッキングによる検証 政策と調査, **11**, 59-66.
- Smyth,J.D., Dillman,D.A., Christian,L.M., & Stern,M.J. (2006) . Comparing check-all and forced-choice question formats in Web surveys. *Public Opinion Quarterly*, **70**, 66-77.
- Sommer,J., Diedenhofen,B., & Musch,J. (2017). Not to be considered harmful: Mobile-device users do not spoil data quality in web surveys. *Social Science Computer Review*, **35**, 378-387.
- 高橋伸彰・成田健一 (2019). Web 調査における画面レイアウトの違いは回答姿勢に影響を与えるか？：同一の調査項目を調査会社 1 社のモニタに施行して 日本心理学会大会発表論文集 日本心理学会第83回大会 (1A-002).
- Tourangeau,R., Maitland,A., Rivero,G., Sun,H., Williams,D., & Yan,T. (2017). Web surveys by smartphone and tablets: Effects on survey responses. *Public Opinion Quarterly*, **81**, 896-929.
- Tourangeau,R., Sun,H., Yan,T., Maitland,A., Rivero,G., & Williams,D. (2018). Web surveys by smartphones and tablets: Effects on data quality. *Social Science Computer Review*, **36**, 542-556.
- Wells,T. (2015). What market researchers should know about mobile surveys. *International Journal of Market Research*, **57**, 521-532.
- 山田一成・江利川滋 (2021). 公募型 Web 調査におけるストレートライニングについて：SD法を用いた企業イメージ測定に関する事例研究 東洋大学社会学部紀要, **59**, 5-17.
- Zou,S., Tan,K.P.S., Liu,H., Li,X., & Chen,Y. (2021). Mobile vs. PC: the device mode effects on tourism online survey response quality. *Current Issues in Tourism*, **24**, 1345-1357.

【Abstract】

Effects of Response Devices on Response Quality in a Volunteer Panel Web Survey

Kazunari YAMADA
Shigeru ERIKAWA

The purpose of this study is to examine the effects of response devices on response quality in volunteer panel Web surveys. For this purpose, a volunteer panel Web survey was conducted in the Tokyo metropolitan area (4,140 Japanese adults in March 2017). The results revealed the following. Firstly, smartphone respondents (SPs) accounted for about 40% of the total respondents, and SPs were more likely to be female and younger than PC respondents (PCs). Secondly, SPs exhibited significantly higher subjective response-stress scores, and slightly higher distraction scores than PCs. However, survey completion times did not differ significantly between the above two groups. Furthermore, similarly to previous studies, response quality did not differ significantly between SPs and PCs. Finally, based on the above results, the practical implications of smartphone as a response device are discussed.