

公募型Web調査におけるスライダー尺度の有効性評価（2）： 単極型スライダー尺度による主観的ストレスの測定

Evaluating the Effectiveness of Slider Scales in Volunteer Panel Web Surveys (2) : Measuring Subjective Stress Using Unipolar Slider Scales

山田 一成
Kazunari YAMADA

問 題

本研究では前研究（山田, 2020）に引き続き、社会科学領域の公募型Web調査における連続型スライダー尺度の有効性について検討する（注1）。なお、前研究では両極型スライダー尺度を取り上げたが、本研究では単極型スライダー尺度について検討する。

単極型スライダー尺度とは、文字通りスライダーを使って回答する単極尺度（unipolar scale）のことを指すが、その単極尺度としての特性については、両極尺度（bipolar scale）との対比において以下のように説明される。

まず、単極尺度は対象となる特性の程度について回答する尺度形式であり、尺度の一方の端には程度がゼロとなる点が配置され、そこからもう一方の端に向かって程度が強くなるように構成されている。例えば、対象への興味の程度を測定するリッカート尺度の場合であれば、両極尺度では選択肢の言語ラベルが「非常に退屈である」から中央の「どちらともいえない」を経て「非常に興味深い」まで配置されるのに対し、単極尺度では「まったく興味がない」から「非常に興味深い」まで配置され、「退屈」などの言語ラベルは配置されないことになる。

また、こうした単極尺度と両極尺度のどちらを使用するかは、測定対象となる構成概念の特性に依存すると考えられている（Dillman et al., 2014）。例えば、客観的な行動頻度を測定する場合であれば、頻度について「負の頻度」を設定することは不可能でありナンセンスであるため、単極尺度が利用されることになる。しかし、測定内容によっては、単極尺度と両極尺度の両方を構成することが可能となる。例えば、対象に対する満足度を測定する場合、単極尺度であれば、選択肢の言語ラベルを「まったく満足していない」から「非常に満足している」まで配置することになるが、同時に、両極尺度として、言語ラベルを「非常に不満」から「どちらともいえない」を経て「非常に満足」まで配置することも可能である。そして、そのような場合、どちらの形式で質問を構成するかについては、事前に

複数の論点について検討することになる(注2)。

こうした単極尺度は医療看護領域の調査研究においては一般的に使用されており、その具体例としては、主観的ストレスを測定するリッカート尺度のような離散型の尺度に加え、主観的な痛みの測定に使用される単極型のVAS (visual analogue scale) や、怒りやストレスの測定に使用される感情温度計 (feeling thermometer) など、複数の連続型の尺度形式を挙げることができる。

本研究で注目するのもそうした連続型の単極尺度であるが、本研究ではそのなかから、尺度形式の汎用性と研究結果の一般化可能性が相対的に高いと考えられる単極型スライダー尺度を取り上げる(注3)。また、その測定内容としては、これまで医療看護領域においてVASによって測定されてきた心理状態のなかから、主観的ストレス(例:鈴木ほか, 1997; 小磯ほか, 2020)を取り上げる。

なお、本研究における主観的ストレス(対人ストレス)の測定は、臨床場面を想定したものではなく、世論調査を想定したものであり、その測定内容は主観的なものに過ぎないと言わざるを得ない面もある。しかし、世論調査において、そのような主観的ストレスが質問されることは決して希ではない(注4)。また、現時点では、そうした調査において利用される尺度のほとんどがリッカート尺度であるが、今後は、Web調査の利用が増加するにつれて、主観的な心理状態の測定を目的として、スライダー尺度の利用が検討される機会も増えてゆくものと予想される。というのも、前研究でも述べたとおり、スライダー尺度には連続量を測定できるという大きなメリットがあり、調査に対する回答者の倦怠感を低下させることも期待できるからである。

ただし、同時に、リッカート尺度の代わりにスライダー尺度を使うことで、回答形式の変化によって判断の方法と内容が変化することや、それに伴い他の項目との相関が変化することも予想されるため、使用に際してはリッカート尺度との互換性について十分な事前検討が必要となる。しかし、現状では、社会科学領域の公募型Web調査において、単極型スライダー尺度の有効性を検討した実証研究はほとんど行われていないようである。

そうしたなか、日本における先行研究としては、山田・江利川(2014)を挙げることができる。この研究では主観的ストレス(対人ストレス)の測定に関し、単極型スライダー尺度と単極型リッカート尺度(ラジオボタン)との間でパラデータが比較され、その結果、回答時間はスライダー尺度のほうが平均値で1~2秒長いこと、および、回答形式満足度はスライダー尺度のほうが顕著に高いことなどが明らかにされている。

ただし、有効性の参考情報となる回答分布については、両尺度形式間に著しい違いがあり、リッカート尺度では5件法で2番目に低度の「やや感じている」という選択肢に回答が著しく集中し、この選択肢までの累積相対度数には、スライダー尺度との間に50ポイントもの差があったことが報告されている。また、その原因については、言語ラベルによる程度差の表現の不十分さ、および、選択肢の配置方向(垂直/水平)、という2つの要因が考えられるとされているが、具体的な検証はなされておらず、推論に留まっている。

もしも、こうした結果が常に生じるようであれば、単極型スライダー尺度は、少なくとも回答分布

に関しては、単極型リッカート尺度とはまったく互換性がないことになる。また、そうであれば、その原因を特定する必要も生じる。公募型Web調査の利用が一般化するなか、この問題は再現性の確認も含め、Web調査の方法論的基礎研究における急務となっているように思われる（注5）。

そこで本研究では、単極型スライダー尺度と単極型リッカート尺度（ラジオボタン）の違いについて、まず、山田・江利川（2014）で報告された結果が再現されるかどうか確認し、回答分布に著しい違いがあれば、その原因について検討する。また、単極型スライダー尺度の有効性については、先行研究ではほとんど検討されていない再検査信頼性や、参考情報となる他の項目との相関についても検討を加える。

方 法

調査概要

本研究では前研究と同様に、「2017年調査」と「2018年調査」（パネル調査）という2つの公募型Web調査（調査会社委託・ポイント報酬制）のデータを分析対象とし、ラジオボタン5件法・垂直配置、同・水平配置、ラジオボタン11件法、スライダー尺度という回答形式4条件間（Figure 1～Figure 4）の異同について、スプリット法による分析を行った。これらの調査の概要とスプリット法の詳細については前研究の該当箇所を参照されたい。また、本研究でも、前研究と同様の理由により超短時間回答者の除外は行っていないが、この点については前研究の注3を参照されたい。

Figure 1 ラジオボタン5件法・垂直配置（RB5V）

Figure 3 ラジオボタン11件法（RB11）

Figure 2 ラジオボタン5件法・水平配置（RB5H）

Figure 4 単極型スライダー尺度（SLD）

なお、以下では、ラジオボタン5件法・垂直配置はRB5V、同・水平配置はRB5H、ラジオボタン11件法はRB11、スライダー尺度はSLDと略記する。また、本研究におけるスライダーの初期位置は左端（「まったく感じていない」）であった。

回答形式比較のための質問項目

測定内容は主観的ストレス（対人ストレス）であった。質問文は「あなたは毎日の生活の中で、人

間関係や人づきあいで、どれくらいストレスを感じていますか」とし、その後、SLDとRB5V・RB5Hでは「次の中から、あなたのお気持ちに最も近いものを選んでお知らせください」と依頼した。また、RB11では「『まったく感じていない』を0、『きわめて強く感じている』を10とした場合に、あなたのお気持ちに最も近いものを選んでお知らせください」と依頼した(注6)。調査票における質問項目の位置は、2017年調査ではどの条件でも18問中の9問目であり、2018年1回目調査ではどの条件でも17問中の7問目、2018年2回目調査ではどの条件でも11問中の6問目であった。

なお、ラジオボタンの言語ラベルについては、2017年調査のRB5VとRB5Hでは、山田・江利川(2014)で報告された結果の再現性を確認するために、言語ラベルを山田・江利川(2014)に合わせた(「まったく感じていない」「やや感じている」「かなり感じている」「非常に感じている」「きわめて強く感じている」の5段階)。また、2018年調査では両回とも、山田・江利川(2014)とは異なる言語ラベルを採用した(「まったく感じていない」「あまり感じていない」「どちらともいえない」「やや感じている」「きわめて強く感じている」の5段階)。

有効性評価の指標

前研究と同様、単極型スライダー尺度の有効性評価の指標として、回答時間、回答形式満足度、再検査信頼性を設定した(それぞれの詳細については前研究を参照)。

結 果

回答形式条件間の属性比較

前研究で述べたとおり、分析に先立ち、回答形式条件間で回答者の基本属性とPCストレス度の回答分布を比較した結果、全ての調査で、回答形式と性別、年代、子供の有無、最終学歴、世帯年収、自由裁量所得、PCストレス度との有意な関連は認められなかった。なお、2018年2回目調査でSLD条件の未婚比率がやや高かったが($\chi^2 = 6.5, df = 2, p < .05$)、それ以外の調査では回答形式と既未婚に有意な関連は認められなかった。

回答時間

回答形式の違いによって回答時間に差があるかどうか調べるために、自然対数化された回答時間について、以下のとおり一元配置分散分析を行った。

まず、2017年調査について、回答時間の外れ値(長時間方向:RB5Vで3名、RB5Hで1名、SLDで3名)を除外して分散分析を行ったところ、条件間に有意な差が認められた($F(3, 1390) = 11.48, p < .001, \eta^2 = .02$)。また、各条件の等分散性が棄却されなかったため、TukeyのHSD法(5%水準)による多重比較を行ったところ、条件間に有意な差が認められた(SLD > RB5V, RB5H, RB11)。次に、2018年1回目調査について、回答時間の外れ値(長時間方向:RB11で5名、SLDで2名)を除外して分散分析を行ったところ、条件間に有意な差が認められた($F(2, 584) = 15.09, p < .001, \eta^2 = .05$)。また、各条件の等分散性が棄却されなかったため、TukeyのHSD法(5%水準)

による多重比較を行ったところ、条件間に有意な差が認められた (SLD > RB11, RB5V)。なお、2018年2回目調査について、回答時間の外れ値 (RB5Vで1名、RB11で6名、SLDで1名) を除外して分散分析を行ったところ、条件間に有意な差が認められた ($F(2, 542) = 16.30, p < .001, \eta^2 = .06$)。また、各条件の等分散性が棄却されなかったため、TukeyのHSD法 (5%水準) による多重比較を行ったところ、条件間に有意な差が認められた (SLD > RB11, RB5V)。

以上のように、スライダー尺度ではラジオボタンよりも回答時間が長くなることが示された。また、回答時間の中央値 (第2四分位) は、RB5VよりもSLDのほうが0.8~1.6秒長くなっていた (Table 1)。なお、RB5VとRB5Hの間に回答時間の差は認められなかった。

Table 1 主観的ストレスの回答時間 (秒)

調査時点	2017年				2018年1回目			2018年2回目		
	RB5V	RB5H	RB11	SLD	RB5V	RB11	SLD	RB5V	RB11	SLD
最小値	1.8	1.5	1.7	2.5	1.5	1.4	2.7	1.2	1.4	2.4
最大値	42.3	43.2	48.0	47.3	27.4	34.0	29.3	25.5	19.2	26.9
第1四分位	5.2	5.2	4.9	6.2	4.3	4.7	5.6	3.8	4.5	5.3
第2四分位	7.4	6.9	6.8	8.2	5.8	6.4	7.4	5.2	6.2	6.8
第3四分位	9.6	10.0	9.4	10.8	7.6	9.1	10.1	7.3	7.7	9.3
平均値	8.4	8.1	8.1	9.7	6.5	7.4	8.7	6.0	6.7	7.9
SD	5.3	5.2	5.4	5.9	3.5	4.7	4.7	3.3	3.4	4.2
歪度	2.7	3.4	2.8	2.9	2.2	2.5	1.9	2.2	1.3	2.0
<i>n</i>	355	351	364	324	195	209	183	184	193	168

注: RB5Vはラジオボタン5件法・垂直配置、RB5Hはラジオボタン5件法・水平配置、RB11はラジオボタン11件法、SLDはスライダー尺度。

回答形式満足度

回答形式満足度は2017年調査と2018年1回目調査で測定され、それぞれの回答分布はTable 2に示す通りであった (上述の回答時間の分析で外れ値を示した回答者は除外された)。この回答形式満足度に回答形式間で差があるかどうか調べるために、以下のとおり一元配置分散分析を行った。

まず、2017年調査について分散分析を行ったところ、条件間に有意な得点差が認められた ($F(3, 1390) = 12.45, p < .001, \eta^2 = .03$)。また、各条件の等分散性が棄却されたため (Levene統計量 $L(3, 1390) = 14.13, p < .001$)、Games-Howell法 (5%水準) による多重比較を行ったところ、条件間に有意な得点差が認められた (SLD > RB11, RB5H, RB5V)。次に、2018年1回目調査について分散分析を行ったところ、条件間に有意な得点差が認められた ($F(2, 584) = 12.38, p < .001, \eta^2 = .04$)。また、各条件の等分散性が棄却されたため ($L(2, 584) = 9.98, p < .001$)、Games-Howell法 (5%水準) による多重比較を行ったところ、条件間に有意な得点差が認められた (SLD > RB11, RB5V)。

以上のように、両調査とも、RB5V・RB5HやRB11よりもSLDのほうが回答形式満足度が有意に

高かった。また、回答形式満足度の肯定的回答の割合は、SLDではRB5V・RB5Hより10～18ポイント、RB11より20ポイントほど高かった（注7）。

Table 2 主観的ストレスの回答形式満足度（列％）

調査時点 回答形式	2017年				2018年 1 回目		
	RB5V	RB5H	RB11	SLD	RB5V	RB11	SLD
十分表せた	13.2	12.5	12.9	21.9	11.3	13.4	23.5
まあ表せた	53.2	50.1	47.8	58.3	55.4	41.6	53.6
どちらともいえない	24.5	29.3	31.6	16.4	26.2	37.8	20.8
あまり表せなかった	7.0	5.1	6.0	2.5	5.6	4.8	2.2
まったく表せなかった	2.0	2.8	1.6	0.9	1.5	2.4	0.0
<i>n</i>	355	351	364	324	195	209	183

注：回答形式のラベルは Table 1 と同様。

再検査信頼性

2018年調査（パネル調査・4週間間隔）の1回目調査と2回目調査の主観的ストレスについて相関係数を算出し、再検査信頼性を回答形式ごとに調べた（Table 3）。まず、パネル調査の有効回答者（携帯・スマホから回答していない536人）から、回答時間が外れ値の回答者を除外した群（VLD群）については、全ての回答形式で再検査信頼性係数（*r*）が約.70で、各回答形式の*r*の95%信頼区間（95% CI）は相互に重なっており、有意な差は認められなかった。なお、係数の値の高低については明確な基準があるわけではないが、心理学領域では.70という数値が直ちに不十分と判断されることはないようである（小塩, 2016）。

次に、パネル調査の2回目調査で、1回目調査の後にライフイベントがあった者をVLD群から除き、除去後の群（NLE群）について*r*を算出したが、結果はVLD群の場合と同様であった。また、こうしたNLE群についての結果は、既未婚を考慮しても同様であった。以上の結果から、SLDの再検査信頼性はRB5V・RV11とほぼ同等であると判断された。

Table 3 主観的ストレスの再検査信頼性（*r*）

分析対象	回答形式	<i>r</i>	[95%CI]	<i>n</i>
VLD	RB5V	.68***	.59-.75	175
	RB11	.69***	.61-.76	186
	SLD	.69***	.59-.76	163
NLE	RB5V	.68***	.58-.76	137
	RB11	.72***	.63-.79	156
	SLD	.71***	.61-.78	137

注：回答形式のラベルは Table 1 と同様。*** $p < .001$

VLDは有効回答者（回答時間が外れ値の回答者を除外）。

NLEはVLDからライフイベントがあった者を除外。

回答分布

主観的ストレスの回答分布を回答形式間で比較するために、まず、RB5V、RB5H、RB11の回答値の範囲をSLDと同一の0～300となるように変換した。変換は、回答値の範囲（0～300）を各尺度の件数で均等分割し、各分割範囲の中央を階級値とする方法で行った（回答値が大きいほど主観的ストレスが高くなるように得点化した）。そのうえで、各調査ごとに主観的ストレスの記述統計量を算出し（Table 4）、一元配置分散分析を行った（上述の回答時間の分析で外れ値を示した回答者は除外された）。

Table 4 主観的ストレスの回答分布

調査時点	2017年				2018年1回目			2018年2回目		
	RB5V	RB5H	RB11	SLD	RB5V	RB11	SLD	RB5V	RB11	SLD
平均値	101.7	108.5	148.4	139.3	154.6	150.9	140.1	152.0	146.7	136.1
SD	58.6	60.1	65.5	86.1	62.7	64.3	89.5	63.0	66.6	90.4
第1四分位	90.0	90.0	95.5	65.3	90.0	95.5	64.0	90.0	95.5	58.5
第2四分位	90.0	90.0	150.0	140.0	150.0	150.0	148.0	150.0	150.0	135.0
第3四分位	90.0	90.0	204.5	211.0	210.0	204.5	217.0	210.0	204.5	205.0
最頻値	90.0	90.0	150.0	300.0	90.0	150.0	300.0	150.0	150.0	300.0
歪度	1.54	1.35	0.18	0.22	0.15	-0.05	0.09	0.11	-0.02	0.24
<i>n</i>	355	351	364	324	195	209	183	184	193	168

注：回答形式のラベルはTable 1と同様。主観的ストレスは得点が高いほどストレスが高い。

SLD以外の回答形式の得点はSLDと比較可能な得点に変換。

その結果、平均値については、まず、2017年調査について分散分析を行ったところ、条件間に有意な得点差が認められた（ $F(3, 1390) = 39.96, p < .001, \eta^2 = .08$ ）。また、各条件の等分散性が棄却されたため（ $L(3, 1390) = 41.22, p < .001$ ）、Games-Howell法（5%水準）による多重比較を行ったところ、条件間に有意な得点差が認められた（RB11, SLD > RB5H, RB5V）。

なお、こうした結果について、ラジオボタンによる回答の単純集計を示したのがTable 5とTable 6である。表中のSLD5は、スライダー尺度の回答を範囲均等分割により5段階化したものであるが、3種類の回答形式間で回答分布を比較すると、ラジオボタンとSLD5との間には著しい違いがあることがわかる。こうした結果は山田・江利川（2014）によって報告された結果とまったく同様であり、回答分布の著しい違いは再現性の高い頑健なものであると判断された。

また、RB5VとRB5Hの間に回答分布の差が認められなかったことにより、RB5VとSLD5の著しい回答分布の差は、ラジオボタンの配列方向に起因するものではないことも明らかとなった。そこで、2018年度調査では、RB5Vの言語ラベルを変更し、2017年調査のRB5Vと回答分布に違いが見られるかどうか検証を行った。その結果、まず、平均値については、一元配置分散分析を行ったところ、

2018年1回目調査でも2回目調査でも条件間に有意な得点差は認められなかった。なお、こうした結果について、ラジオボタンによる回答の単純集計を示したのがTable 6であるが (SLD5はSLDの範囲均等分割による5段階化)、回答形式間で回答分布を比較すると、1回目調査でも2回目調査でも、Table 5で見られた回答分布の著しい違いは生じていなかった。ただし、RB5VとSLD5を比較すると、両回とも、RB5Vで両端の選択肢の回答割合が少なくなる傾向がうかがわれ、その程度は「まったく感じていない」という程度ゼロの選択肢で顕著であった。

Table 5 主観的ストレスの回答分布 (2017年, 列%)

調査時点	2017年		
	RB5V	RB5H	SLD5
まったく感じていない	16.3	13.1	22.5
やや感じている	64.8	62.4	21.6
かなり感じている	8.5	12.0	24.1
非常に感じている	3.9	5.7	15.4
きわめて強く感じている	6.5	6.8	16.4
<i>n</i>	355	351	324

注：回答形式のラベルはTable 1と同様。

選択肢の言語ラベルはラジオボタンのもの。

SLD5はSLDの範囲均等分割による5段階化。

Table 6 主観的ストレスの回答分布 (2018年, 列%)

調査時点	2018年1回目		2018年2回目	
	RB5V	SLD5	RB5V	SLD5
まったく感じていない	3.6	23.0	5.4	26.2
あまり感じていない	30.8	19.1	28.3	19.0
どちらともいえない	29.2	26.2	32.6	24.4
やや感じている	27.2	14.2	25.0	13.1
きわめて強く感じている	9.2	17.5	8.7	17.3
<i>n</i>	195	183	184	168

注：回答形式のラベルはTable 1と同様。

選択肢の言語ラベルはラジオボタンのもの。

SLD5はSLDの範囲均等分割による5段階化。

なお、各尺度の最頻値 (Table 4) については、SLDは全ての調査で300 (尺度右端の「きわめて強く感じている」) であった。その割合は、2017年が5.6% (左端1.9%)、2018年1回目が6.0% (左端4.9%)、2018年2回目が6.0% (左端3.6%) と安定していた。これに対し、RB11の最頻値は全ての調査で中央の「5」であり (Figure 3も参照)、2017年が22.8%、2018年1回目が23.0%、2018年2回目が23.3% と安定して突出していた。

項目間相関

主観的ストレスと他の項目との相関を回答形式間で比較した結果がTable 7である(注8)。まず、2017年調査においては、RB11を除く3つの回答形式の間では、共通して有意な相関が示される項目は認められなかった。また、RB5VとSLDに限定してみても、両形式とも相関が有意な項目は年齢だけであり、それ以外の項目については、相関の有意性が示されたのはいずれか一方の形式においてのみであった。なお、こうした結果のなかで、友人数(注9)と主観的ストレス(対人ストレス)との相関の有意性がRB5VとSLDで異なる点は、社会関係資本に関する実証研究の結果に尺度形式の違いが影響しうることを示唆しており、そのような意味でも注意すべき結果であると考えられる。

Table 7 主観的ストレスと他の項目との相関

調査時点 回答形式	2017年				2018年1回目		
	RB5V	RB5H	RB11	SLD	RB5V	RB11	SLD
年齢(連続量)	-.16**	-.09	-.20***	-.22***	-.23**	-.18**	-.14
世帯年収(10)†	-.05	.00	-.05	.01	-.05	-.11	.01
自由裁量所得(10)†	-.07	-.12*	-.07	.03	-.08	-.25**	.00
平日自由時間(9)	.00	-.05	-.17**	-.13*	-.18*	-.20**	-.18*
休日自由時間(9)	-.02	-.01	-.09	-.08	-.07	-.16*	-.07
ネット利用時間(9)	.18**	.06	-.01	.03			
テレビ視聴時間(9)	.04	-.08	-.05	-.12*			
Web調査回答数(9)	-.03	.01	.00	-.01	.02	-.03	.06
友人数WS(5)	-.13**	-.08	-.05	-.01	-.13	-.08	-.04
友人数NWS(5)	-.12*	-.19**	-.15**	-.04	-.08	-.14	-.13
性別(2)	.16	.07	.05	.07	.12	.17	.08
最終学歴(3)†	.10	.08	.07	.15	.11	.24**	.11
<i>n</i>	355	351	364	324	195	209	183

注:回答形式のラベルはTable 1と同様。()内はカテゴリー数。†はDK・NAを除外。性別と最終学歴はCramer's V (RB11とSLDは5段階化)。* $p < .05$ 、** $p < .01$ 、*** $p < .001$

次に、2018年調査については、年齢を別にとすると、RB5VでもSLDでも有意な相関は平日自由時間と主観的ストレスの相関のみであった。また、年齢と主観的ストレスの相関は、両形式とも方向は負であり、SLDでは有意傾向が示された($p = .06$)。このように、2018年調査では2017年調査よりも、他の項目との相関が回答形式間で一致する傾向がうかがわれた。また、こうした結果には、2018年調査のRB5VとSLDの回答分布の違いが、2017年調査よりも著しく小さかったことが影響している可能性が高いと考えられる。ただし、友人数と主観的ストレスとの相関をみると、RB5Vでは友人数WSでのみ有意傾向($p = .07$)が示され、SLD5では友人数NWSでのみ有意傾向($p = .09$)が示されており、そうした点においては、必ずしも回答形式間での一致を積極的に主張できる結果ではなかった。

総合考察と結論

本研究では連続型の単極型スライダー尺度の有効性について検討した。その結果、まず、スライダー尺度の回答時間はラジオボタンと比べ若干増加することが示された。ただし、そうした時間増はわずかであり、調査に悪影響を及ぼすほどのものではなく、時間増はスライダー尺度のデメリットではないと判断された。なお、こうした結果は前研究（山田, 2020）における両極尺度についての結果と同様であるが、増加時間は両極尺度よりも単極尺度でやや短くなる傾向がうかがわれた。その原因としては測定内容、尺度形式、呈示順序などが考えられるが、本研究では全ての調査で収入満足度（両極尺度）の質問の後で主観的ストレス（単極尺度）の質問を行っており、呈示順序を変えた比較を行っていないため、この点については今後の検討課題と言わざるを得ない。

次に、スライダー尺度の回答形式満足度はラジオボタンよりも有意に高いことが示された。こうした結果も前研究における両極尺度についての結果と同様であり、スライダー尺度の大きなメリットを示す頑健性の高い結果であると判断された。

なお、スライダー尺度の再検査信頼性については、再検査信頼性係数の値は.70前後で、回答形式間に有意な差は認められなかった。また、こうした結果はパネル調査期間中のライフイベントの有無を考慮しても同様であった。ただし、両極尺度の係数の値が約.80であったことに比べると、単極尺度の係数の値は若干低くなっていた。その原因としては、測定内容、尺度形式、呈示順序などが考えられるが、この点についての検討も今後の課題である。

以上の結果から、単極型スライダー尺度は、上記の3指標に関しては、日本の公募型Web調査において十分な有効性を有すると判断された。ただし、以上の結果は主観的ストレスに関するものであり、結果の一般化については前研究（山田, 2020）の両極尺度と同様の制約のもとにあると言わねばならない。

なお、本研究では、単極型スライダー尺度の有効性検討の参考情報として、回答分布についても検討された。この検討の目的は、まず、山田・江利川（2014）によって報告された「単極型ラジオボタンと単極型スライダー尺度の間に見られた回答分布の著しい違い」の頑健性を確認することであったが、本研究においても先行研究と同様の著しい違いが確認された。また、本研究では、その原因についても検討されたが、ラジオボタンの配置方向が回答分布に影響しないことが示されたため、ラジオボタンの言語ラベルを変更したところ、特定選択肢への回答の集中は消失した。また、2017年調査と2018年調査の結果を比較すると、スライダー尺度でもラジオボタン11件法でも、回答分布に調査時点間の差は認められなかった（Table 4）。そのため、上述の「著しい違い」は単極型ラジオボタンの言語ラベルの程度表現に起因する可能性が高いと判断された。

なお、本研究では、有効性検討の参考情報として、単極型スライダー尺度と他の項目との相関についても検討されたが、相関係数の方向と有意水準を基準にすると、どの回答形式でも同じ傾向が示さ

れるとは言いがたい結果であった。また、そうした結果は、ラジオボタンとスライダー尺度の分布の違いが著しかった2017年調査でより顕著であった。以下では、このような結果に基づき、単極型ラジオボタンと単極型スライダー尺度の互換性について検討する。

まず、この問題について、スライダー尺度の回答分布を基準にすると、ラジオボタンの言語ラベルは2017年調査よりも2018年調査のほうが互換性が高い、ということになる。また、その原因は、中央に「どちらともいえない」が配置され、低度の選択肢が増えたことによる、と考えることになる。

ただし、2018年調査のラジオボタン（言語ラベル）は以下のような問題を抱えている。まず挙げられるのは、依然として、ラジオボタンとスライダー尺度の間には分布のズレが認められる点である。具体的に言えば、スライダー尺度では低度の回答が多くなる傾向があるのに対し、ラジオボタンでは両端の回答割合が少なく、特に程度がゼロの回答が少なくなる傾向が顕著である。そのため、スライダー尺度の分布を基準にした場合には、単極尺度では言語ラベルが回答分布に与える影響が非常に大きいため、中央に「どちらともいえない」を配置したほうがよいが、配置してもラジオボタンとスライダー尺度の間には回答分布に違いが生じる、と判断することになる。そして、それを根拠に、両回答形式の間には、回答分布については、必ずしも十分な互換性があるわけではない、という結論を得ることになる（注10）。

次に、ラジオボタンの回答分布を基準にすると、以下のような議論も可能となる。まず、「ストレス」という言葉が日常的な常套句（cliché）であり、低度の主観的ストレスの表明自体は非常に多くの人が行うと仮定すると、低度の選択肢に回答が集中するはずであり、そうだとすると、2017年調査の言語ラベルに大きな問題はない、ということになる。また、そうした2017年調査のラジオボタンの回答分布を基準にすると、2018年調査のラジオボタンの言語ラベルは、「どちらともいえない」という中央選択肢によって、回答分布を両極尺度の分布に近くなるように歪める効果を持っていた、とも考えられる。

ただし、ここで問題になるのは、中央に「どちらともいえない」が配置された尺度を単極尺度と言えるか、という点である。2018年調査の言語ラベルは、尺度の一端を程度がゼロの点とし、もう一方の端に向かって一方向に程度が強くなるように選択肢が配置されており、その点では単極尺度を構成していると言える。しかし、中央の「どちらともいえない」という選択肢に注目すると、右方向に「感じている」という肯定的な言語ラベル、左方向に「感じていない」という否定的な言語ラベルが配置されており、2つの正反対の方向性を有するという点では、2018年調査の言語ラベルは両極尺度の特性を併せ持つ「準単極尺度」だと考えることもできる（注11）。そして、そうだとすると、単極尺度よりも準単極尺度の回答分布のほうがスライダー尺度の回答分布に近いという本研究の結果は、単極型スライダー尺度を両極尺度と見なして回答する回答者が多いことを示唆しているとも考えられる。このように、ラジオボタンの回答分布を基準にすると、尺度の互換性に関する議論は、回答ヒューリスティックによる「単極尺度の両極化現象」という新しい問題にたどり着くことになる。

もっとも、単極型ラジオボタンに「どちらともいえない」を配置せず、中間の3つの言語ラベルを

「わずかに感じている、わりと感じている、かなり感じている」などとすれば、ラジオボタンでもスライダー尺度でも、低度の回答がやや多い回答分布が共通して示され、両尺度の間には十分な互換性があるという結論が得られる可能性も残る(注12)。そのため、スライダー尺度の回答分布の互換性については、さらなる検討が必要であると言わざるを得ないことになる。

なお、以上のような議論に対し、「測定方法によって構成概念を定義する」という操作主義的な立場から、「測定方法が異なる構成概念の回答分布が相互に異なることは、何ら問題ではない」との批判的な指摘があるかもしれない。しかし、本論では、そうした議論の意義を認めつつも、過度に科学哲学的な議論には立ち入らず、調査研究や調査実務における現実的な要請を重視し、「測定形式の違いによる回答分布の差が、実際に、どの方向に、どの程度生じうるかを経験的に把握しておくことが必要である」という立場を取る(注13)。なお、このような立場は、具体的な尺度の特性を経験的に明らかにし、使用する尺度形式の適不適を使用目的に応じて判断するための情報を提供するという点で、単なる「操作主義的な相対主義」とは異なると考えられる。また、そのような意味では、本研究の知見のなかで最も重視すべきことは、スライダー尺度との対比において、単極型リッカート尺度(ラジオボタン)では、言語ラベルが回答分布に大きな影響を及ぼしうることが経験的に示されたことであるとも考えられる。

Web調査の急速な普及に伴い、今後スライダー尺度の利用が検討される機会も増えてゆくと予想される。また、新型コロナウイルスへの感染拡大に伴い、各種調査において主観的ストレスの測定が増加する可能性もあるが、そうした調査がWeb調査として実施されれば、その測定にスライダー尺度が利用される機会も増えてゆくと予想される。

そのような視点から見ると、本研究で取り上げた単極型スライダー尺度は、連続量の測定が可能であり、十分な再検査信頼性を有し、回答者の調査への倦怠感を緩和することも期待でき、主観的ストレスの測定に大変有益であると考えられる。ただし、単極型ラジオボタンと単極型スライダー尺度の間には、回答分布や他の項目との相関という点では、十分な互換性があるとは言えないようである。単極型リッカート尺度(ラジオボタン)は、言語ラベルが回答分布に大きな影響を及ぼす尺度形式であり、言語ラベルの程度表現や段階数にも言語的な制約がある(織田, 1970)と言わざるを得ない。そのような意味では、単極型スライダー尺度の利用に際しては、まず単極尺度という形式自体の是非が議論されるべきであり、また、単極尺度を利用するとしても、単極型ラジオボタンとの回答分布の一致のみに注目するのではなく、スライダー尺度のメリットとデメリットを十分把握したうえで、目的に応じた利用が可能かどうか、事前に十分な検討を行うことが必要となる。

注

- 1) 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。また、本研究の調査は東洋大学大学院社会学研究科研究倫理委員会の承認を受けた。
- 2) そうした論点として、両極尺度の両端が互いに反対の意味かどうか(Green & Goldfried, 1965)、両極尺度に

- における中央選択肢の意味(続, 1974; Schaeffer & Presser, 2003; 増田・坂上, 2014)、単極尺度の「まったく～ない」というカテゴリーに両極尺度の「非常に～である」から「どちらともいえない」までが含まれるかどうか(Schaeffer & Presser, 2003)、選択肢が数値の場合のポジティブ・バイアス(Schwarz et al., 1991)、尺度極性による信頼性の違い(Alwin, 2007; Menold, 2021)などが挙げられる。
- 3) 感情温度計はグラフ評定尺度の一種であり、広く知られてはいるが、測定内容が感情に特化されがちであり、使用に際しては説明文や温度計イラストなどが必要とされることも多く、特殊な測定方法であると言わざるを得ない。
 - 4) 内閣府の「体力・スポーツに関する世論調査」(2004年2月)では「精神的な疲労、ストレス」の程度を尋ねる4件法の質問が設けられている。また、NHK放送文化研究所の「健康に関する世論調査」(2009年2月)でも「イライラやストレス」を感じる頻度を尋ねる4件法の質問が設けられている(山田・酒井, 2009)。なお、近年では、厚生労働省の「国民生活基礎調査」で大規模調査時に「日常生活での悩みやストレス」の有無を尋ねる質問が設けられている(例: 2016年6月)。また、NHK放送文化研究所が加盟する国際比較調査グループISSPの調査「仕事と生活(職業意識)」(2015年)でも「仕事にストレスを感じる頻度」を尋ねる質問が設けられている(村田, 2018)。
 - 5) ただし、前研究において議論したように、リッカート尺度とスライダー尺度との間で、回答分布や他の項目との相関が異なることは、直ちに、いずれかの回答形式が不適切であることを意味しない。そのため、本研究でも、前研究と同様、回答分布や他の項目との相関を有効性評価の指標とはせず、尺度特性に関する参考情報として報告・検討する。
 - 6) 調査会社の仕様により、RB11では個々のラジオボタンの上に0～10の数値が付記されている。そのため本研究では主にRB5VとSLDの差異に注目する(RB11についての結果は厳密に言えば参考情報である)。
 - 7) 自分の気持ちを「表せた」という肯定的回答をしなかった層は、ラジオボタンでは3割強～4割強にも及んでおり(Table 2)、単極尺度についても、回答者の否定的な意識が懸念される結果となった。
 - 8) 相関の検討に使用した項目については山田(2020)の注7を参照。
 - 9) 友人数WSは、今の仕事・学校との関係で「一緒に飲食や買い物などに行くような間柄」の人の数であり、友人数NWSは、今の仕事・学校との関係以外で同様の間柄の人の数である(どちらも家族・親戚を除く人数)。
 - 10) スライダー尺度の分割範囲を均等にせず、両端の範囲を中間の範囲の2分1にすることも可能である。そのようにSLDの範囲を不均等に5分割した場合の回答分布(SLD5NE:%)は、2018年1回目では、主観的ストレスの低度から高度に向かって[18.6, 20.2, 30.1, 21.3, 9.8]、2018年2回目では、同方向に[17.9, 24.4, 28.0, 17.3, 12.5]であった。こうした分布と比較しても、RB5Vには程度ゼロの回答割合が少なくなる傾向がうかがわれる。
 - 11) こうした2方向性は「どちらともいえない」を「中間」という表現にしても、また、「どちらともいえない」を配置せずに選択肢数を偶数にしても解消されない。なお、こうした概念と尺度の極性の組み合わせについては、Saris & Gallhofer(2007)が、共に両極、共に単極、単極尺度を伴う両極概念の3形態があることを指摘しているが、DeCastellarnau(2018)は両極尺度を伴う単極概念もありうることを、および、極性の組み合わせが尺度の解釈に及ぼす影響についての研究が行われていないことを指摘している。
 - 12) ただし、Höhne & Krebs(2020)は、2019年にドイツで行われた公募型スマホ調査に基づき、単極尺度における尺度の極性(polarity)と中央選択肢の言語ラベルのマッチ/ミスマッチは回答分布にほとんど影響を与えないことを報告している。もっとも、そこで検討の対象とされたのは、主観の程度を問う質問ではなく、賛否を問う質問である点には注意が必要である。
 - 13) 継続調査の調査主体が、回答の質の向上を目的として、回答形式の変更について検討する場合には、こうした要請が生じると予想される。

引用文献

- Alwin, D.F. (2007). *Margins of error: A study of reliability in survey measurement* (Vol. 547). Hoboken, NJ: John Wiley & Sons.
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, 52, 1523–1559.
- Dillman, D.A., Smyth, J.D., & Christian, L.M. (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.

- Green,R.E. & Goldfried,M.R. (1965). On the bipolarity of semantic space. *Psychological Monographs: General and Applied*, 79, 1-31.
- Höhne,J.K. & Krebs,D. (2020). Mismatching middle options: Consequences for attitude measurement in smartphone surveys. *International Journal of Social Research Methodology*, 04 Aug 2020, doi.org/10.1080/13645579.2020.1803621.
- 小磯京子・木下直彦・本間美知子・度會裕子・淡島正浩・瀧口徹 (2020).福島原発事故による県外避難者の重大ライフイベントが主観的ストレス度に及ぼした相対的影響 日本放射線看護学会誌, 8, 11-21.
- 増田真也・坂上貴之 (2014). 調査の回答における中間選択：原因、影響とその対策 心理学評論, 57, 472-494.
- Menold,N. (2021). Response bias and reliability in verbal agreement rating scales: Does polarity and verbalization of the middle category matter?. *Social Science Computer Review*. 39, 130-147.
- 村田ひろ子 (2018). 何が仕事のストレスをもたらすのか：ISSP国際比較調査「仕事と生活（職業意識）」から 放送研究と調査, 68, 38-50.
- 織田揮準 (1970). 日本語の程度量表現用語に関する研究 教育心理学研究, 18, 166-176.
- 小塩真司 (2016). 心理尺度構成における再検査信頼性係数の評価：「心理学研究」に掲載された文献のメタ分析から 心理学評論, 59, 68-83.
- Saris,W.E. & Gallhofer,I.N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, NJ: John Wiley & Sons.
- Schaeffer, N.C. & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65-88.
- Schwarz,N., Knäuper,B., Hippler,H.-J., Noelle-Neumann,E., & Clark,L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570-582.
- 鈴木伸一・嶋田洋徳・三浦正江・片柳弘司・右馬埜力也・坂野雄二 (1997). 新しい心理的ストレス反応尺度 (SRS-18) の開発と信頼性・妥当性の検討 行動医学研究, 4, 22-29.
- 続有恒 (1974). 調査：心理学実験演習Ⅱ 金子書房
- 山田亜樹・酒井芳文 (2009). 現代日本人の健康意識：「健康に関する世論調査」から 放送研究と調査, 59, 2-21.
- 山田一成 (2020). 公募型Web調査におけるスライダー尺度の有効性評価 (1)：両極型スライダー尺度による記入満足度の測定 東洋大学社会学部紀要, 58, 35-50.
- 山田一成・江利川滋 (2014). Web調査におけるVisual Analogue Scaleの有効性評価 東洋大学社会学部紀要, 52, 57-70.

【Abstract】

Evaluating the Effectiveness of Slider Scales in Volunteer Panel Web Surveys (2) : Measuring Subjective Stress Using Unipolar Slider Scales

Kazunari YAMADA

The purpose of this study is to examine the effectiveness of unipolar slider scales (SLDs) in volunteer panel Web surveys concerning social sciences. For this purpose, two experimental Web surveys were conducted in Japan (split method, Tokyo metropolitan area). One was a one-shot survey (January 2017; $n = 1,444$), and the other was a panel survey (January to February 2018; $n = 622$). Using these data sets, SLDs were compared with radio buttons (RBs: Likert scales) on response time, respondent satisfaction concerning the format, and test-retest reliability. The results of subjective stress (interpersonal stress) showed that (1) SLDs required a slightly longer response time than RBs, (2) respondents were more satisfied with the format of SLDs than that of RBs, and (3) there were no differences in test-retest reliability between scale formats. However, (4) the response distribution forms for SLDs differed greatly from those using RBs, suggesting that verbal labels on unipolar Likert scales have a significant effect on response distribution forms. The usefulness and limitations of unipolar SLDs in volunteer panel Web surveys are discussed based on the above results.