

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収入満足度の測定

Evaluating the Effectiveness of Slider Scales in Volunteer Panel Web Surveys（1）： —Measuring Income Satisfaction Levels Using Bipolar Slider Scales—

山田 一成
Kazunari YAMADA

問 題

スライダー尺度（slider scale）はグラフ評定尺度（graphic rating scale: GRS）の一種であり、回答は、水平または垂直方向の線分に沿ってスライダー（スライドバー）を移動させる操作によって行われる。また、そうしたスライダー尺度により、連続量の測定が可能になるとともに、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。従来、サーベイではリッカート尺度が多用されてきたが、Web 調査の利用が増加するにつれてスライダー尺度への関心も高まり、その方法論的基礎に関する実証研究も盛んに行われるようになってきている（Chyung et al., 2018）。そうしたなか、本研究では社会科学領域の公募型 Web 調査におけるスライダー尺度の利用に焦点を合わせ、ラジオボタン（リッカート尺度）との比較を通して、その有効性評価を試みる（注1）。

なお、GRSを代表する尺度としてはVAS（visual analogue scale）が広く知られており、スライダー尺度のことをVASと呼称することもあるが（Couper et al., 2006）、両者は回答方法が異なっている。PCで回答する場合、VASの操作は、(1)マウスポインタを回答位置まで移動させる、(2)クリックして回答する、という2アクションとなる。これに対しスライダー尺度の操作は、(1)マウスポインタをスライダーまで移動させる、(2)クリックしてスライダーをつかむ、(3)スライダーを回答位置まで移動させる、(4)スライダーを放して回答する、という4アクションとなる。このように、スライダー尺度はVASよりも操作が複雑であり、回答行動の研究においては両者が明確に区別されることも多い（Funke, 2016）。

ただし、そうした研究においては、スライダー尺度とVASの差異よりも、GRSとリッカート尺度の違いが主たる関心事である場合が多い。というのも、スライダー尺度やVASには連続量を測定できるという大きなメリットがあるからである。また、特にスライダー尺度には、スライダー操作によって、回答作業への関与が深まる（Roster et al., 2015）、回答作業が楽しいものとなる（Sikkel et

al. 2014)、回答への満足度が高まる (山田・江利川, 2014)、といったメリットもあると考えられている。そして、最終的には、そうした事柄の帰結として、回答の質の向上が期待されていることは言うまでもない。

ただし、そうしたスライダー尺度の有効性については、以下のレビューで述べるように、必ずしも一貫した結果が得られているわけではない。

スライダー尺度の有効性

スライダー尺度の有効性に関する初期の研究としてはCook et al. (2001) を挙げることができる。Cookらは大学図書館サービスの評価に関するWeb調査データを用いて、心理尺度の信頼性係数を回答形式間で比較している。心理尺度にはリッカート尺度版 (9件法) とスライダー尺度版 (5件法、9件法、連続型) があったが、 α 係数を比較した結果、連続型スライダー尺度もリッカート尺度も十分な信頼性を示すこと、および、係数の値は前者で若干低いものの、両者の間に大きな差はないことなどが報告されている。Cookらは、こうした結果に基づき、いずれの尺度形式もデータ収集法として許容されると述べている。ただし、この研究は20年前のWeb環境下のものであり、スライダー尺度の回答者数が少ないことや、リッカート尺度が9件法であることなど、注意すべき点も少なくない。

これに対し、近年の研究としてはBosch et al. (2019) を挙げることができる。Boschらはノルウェーの大規模な確率オンラインパネルを対象とするWeb調査データを利用し、各種尺度の信頼性と妥当性について、多特性・多方法行列による検討を行っている。分析対象となった質問は移民資格と最高裁に関するもので、それぞれについて、リッカート尺度とスライダー尺度による測定結果が比較された。その結果、Boschらは、いずれの測定法も同じくらい推奨できるものであるが、リッカート尺度からスライダー尺度に移行する必要はないと述べている。

なお、スライダー尺度の回答の楽しさについてはSikkel et al. (2014) を挙げるすることができる。Sikkelらは、オランダの調査会社のモニターを対象とする調査において、ドラッグによる回答形式として、3種類のスライダー尺度 (連続型、離散型、絵文字付き離散型) を設定し、クリックによる回答との比較を行っている。その結果、ドラッグ回答群ではクリック回答群よりも、質問が興味深く、質問は自分に重要な事柄を含んでおり、質問への回答は楽しい、と回答する傾向が認められた。ただし、Sikkelらの研究では、上記3つのドラッグ形式が明確には区別されておらず、結果の解釈に議論の余地を残している。

スライダー尺度のデメリット

以上の例に代表されるように、これまでの研究の結果は、必ずしもスライダー尺度の利用を積極的に推奨するものとはなっていない。また、先行研究のなかには、以下のように、スライダー尺度の利用に否定的な見解を示す研究もある。

まず、Couper et al. (2006) は、人間の行動特性の原因について「100%遺伝」から「100%環境」までを範囲とする評価軸を用い、VAS (スライダー尺度)、ラジオボタン、数値入力 of 3形式によるWeb調査の回答を比較し、スライダー尺度では脱落や無回答が増え、回答時間が長くなることを指

摘している。

また、Funke et al. (2011) は、スライダー尺度のデータの質について、商品コンセプト評価に関する Web 調査のデータを用いて分析を行っている。回答形式はリッカート尺度とスライダー尺度の 2 種類で、脱落率、回答時間、測定値の分布が比較されたが、スライダー尺度では脱落率が高く回答時間が長かった。なお、特に注目すべきなのは、脱落率が教育水準の低い層で高かったことで、Funke らはこうした結果を踏まえ、スライダー尺度の使用は推奨しないと述べている。

さらに、Maloshonok & Terentev (2016) は、大学のオンライン授業後の Web 調査において、ラジオボタン、スライダー尺度、テキストボックスの回答を比較している。質問内容は「授業に費やす 1 週間の勉強時間」であったが、「わからない」の割合がラジオボタンでは 3.3% であったのに対し、スライダー尺度では 11.5%、テキストボックスでは 14.3% に及んだため、Maloshonok らは、よりよいのはラジオボタンの利用であると述べている。

有効性評価の問題点

ただし、以上のようなスライダー尺度のメリットとデメリットについては、一貫した結果が得られているとは言えないという指摘もある (Roster et al., 2015)。また、そうした総括とならざるを得ない理由としては、以下の 3 点が重要であると考えられる。

まず挙げられるのは、研究ごとに測定内容、回答者、回答デバイスなどが様々に異なっているという点である。もちろん、そこに、国や時期といった調査状況の違いが関わることは言うまでもない。そのため、スライダー尺度の評価に関する研究横断的な一般化は、そもそも不可能に近いと考えるべきなのかもしれない。

また、これまであまり言及されていない点として、スライダー尺度にも複数のバリエーションがあるという点が挙げられる。そこには、VAS との同一視という問題も含まれるが、スライダー尺度に限定しても、配置方向（水平／垂直）(Funke et al., 2011)、言語・数値ラベルの有無と数、スライダー位置の数値フィードバックの有無 (Couper et al., 2006)、スライダー初期位置の違い (Liu & Conrad, 2019)、といった様々な仕様の違いが存在する。さらに、スライダー尺度のなかには離散型の尺度も存在するため（注 2）、そうした点も結果の一貫性を論じることを非常に困難にしている。

しかも、さらに重要な点として、有効性評価の指標が共有されているわけではない、という点も挙げられる。なお、この点は実証研究における重要な論点であるため、以下、個々の指標に検討を加えながら、本研究における指標の選定を行うことにする。

有効性評価の指標

先行研究における有効性評価の指標としては、回答時間、DK・NA、脱落、回答分布、信頼性係数、回答心理などが挙げられるが、ここではそれらに、先行研究が見当たらない指標として項目間相関と再検査信頼性を加え、公募型 Web 調査を想定して、それぞれについて検討する。

回答時間 スライダー尺度はラジオボタンと比べ回答に必要なアクション数が多く、また、回答に熟慮を要する可能性もあるため、回答時間が長くなることが予想される。ただし、日本における先行

研究では、そうした時間増が中央値で数秒程度であり、調査に悪影響が及ぶほどではないという報告もある(山田・江利川, 2014)。そこで、本研究でも回答時間を有効性評価の指標とし、先行研究の結果が頑健かどうか検証する。

DK・NA 調査会社の公募型Web調査では、NA(無回答)が許容されない仕様となることも少なくない。本研究で実施した調査もそうした仕様であったため、本研究ではNAを有効性評価の指標とはしない。一方、「わからない」というDK回答については、トレードオフを含むような難しい質問(Couper et al., 2006)であれば、設定したほうがよいこともあるが、本研究では一般的な質問にスライダー尺度を利用するケースを想定している。そのため、本研究ではDK回答を有効性評価の指標とはしない。

脱落 調査会社の公募型Web調査では、原則として、回答が完了した調査票が先着順で回収票となっていき、所定の数だけ回収票が集まったと判断されたところで回収が打ち切られる。従って、回答が途中で放棄された調査票は回収票には含まれないことになる。本研究で実施した調査も脱落(attrition)が発生しない仕様であったため、本研究では脱落を有効性評価の指標とはしない。

回答分布 先行研究のなかには測定値の分散が大きいことを尺度のメリットと見なすものもある(Cooper et al., 2006)。回答が特定の選択肢や回答位置に集中すると、個人間の差が十分検出されないからである。ただし、そこで仮定されているのは、回答者が微妙な差異をリッカート尺度では回答できず、代替となる選択肢を選んでいくようなケースである。しかし、別のケースとして、回答形式の変化によって判断の方法と内容が変わり、それによって分散が増大することも予想される。そして、その場合には、分散の増大を尺度のメリットと見なすことには議論の余地があることになる。そのため、本研究では回答分布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・検討することにする。

信頼性係数 スライダー尺度によって内容の類似した項目を複数測定した場合には、 α 係数を回答形式間で比較することができる(Cook et al., 2001)。ただし、調査会社の公募型Web調査では、1回答画面に設定できるスライダー尺度の数に限りがあることも少なくない。また、回答画面の増加は料金増に直結し、調査実施上の大きな制約となる。そのため、スライダー尺度の大量使用には議論の余地があるが、本研究では大量使用を前提としていないため、信頼性係数を有効性評価の指標とはしない。

回答心理 スライダー尺度は回答への関与、回答の楽しさ、回答への満足度などを高める可能性があると考えられる。しかし、そうした回答心理への効果を実証した研究はほとんど見当たらない。日本では山田・江利川(2014)がスライダー尺度の回答形式満足度が高いことを報告しているが、それ以外には実証研究が見当たらない。そこで、本研究でも回答形式満足度を有効性評価の指標とし、先行研究の結果が頑健かどうか検証する。

項目間相関 スライダー尺度と他の項目との相関については、先行研究がほとんど見当たらない。理由として考えられるのは、以下のような問題点である。まず挙げられるのは、そうした項目間相関

の異同をどう評価すべきか、という問題である。構成概念妥当性という視点から見ると、項目間相関については、相関の有無や方向性の一致という緩やかな基準に照らし、回答形式間に違いがないことが望ましい。しかし、他方では、スライダー尺度の測定値が連続量であることが原因で相関係数が大きくなり、それがメリットとされる場合もあると考えられる。そのため、項目間相関の異同の評価については、研究ごとの個別具体的な判断に委ねるしかないようにも思われる。また、項目間相関といっても、単相関だけでなく、偏相関の検討も必要とするかどうか、という論点もある。さらに言えば、割当法による公募型Web調査において、スプリット各群の等質性をどの程度仮定できるか、という問題も残されている。このように、項目間相関については要検討点が多いため、本研究では項目間相関を有効性評価の指標とせず、尺度特性に関する参考情報として報告・検討する。

再検査信頼性 スライダー尺度では、連続量を回答できることにより再検査信頼性が低下することも予想される。そこで、本研究では再検査信頼性を有効性評価の指標とし、スライダー尺度の再検査信頼性が十分かどうか検証する。

本研究の目的

以上の検討を踏まえ、本研究では研究対象を社会科学領域の公募型Web調査における連続型の両極型スライダー尺度に限定し、その有効性について、回答時間、回答形式満足度、再検査信頼性を指標として、ラジオボタンとの比較検討を行う。また、尺度特性を把握するための参考情報として、スライダー尺度の回答分布と項目間相関についても報告・検討する。

方 法

調査概要

本研究では下記の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た（注3）。なお、これら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。

2017年調査：一都三県在住の男女20～69歳が対象。割当法（平成27年度国勢調査に基づく一都三県全体の性年代別人口構成比への割当、総数1,400人目処）。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（家族に従事者がいる者を含む）、および、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（携帯・スマホからの回答者を除いた分析対象者は1,401人）。なお、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。

2018年調査（パネル調査） 1回目調査：一都三県在住の男女20～69歳が対象。割当法（総数600人目処）。事前調査は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、2017年調査と同様の手続きで4,118人を抽出。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、622人の有効回答を得た(携帯・スマホからの回答者を除いた分析対象者は594人)。2回目調査：1回目調査の有効回答者622名に調査を依頼し、2018年2月16日(水)～19日(金)に568名(91.3%)から回答を得た(携帯・スマホからの回答者を除いた分析対象者は553人)。なお、両回とも、上記以外の主な仕様は2017年調査と同様。

スプリット法

回答形式間比較のために、上記のWeb調査ではスプリット法を採用し、回答者をプログラム制御によりランダムに各条件に割り当てた。2017年調査はラジオボタン5件法・垂直配置(RB5V)、ラジオボタン5件法・水平配置(RB5H)、ラジオボタン11件法(RB11)、スライダー尺度(SLD)の4条件(Figure 1～Figure 4)。2018年調査は両回ともRB5V、RB11、SLDの3条件(2回目は1回目と同一条件に回答)。各条件の回答者数はTable 1に示す通り。

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 両極型スライダー尺度 (SLD)

なお、SLDは300pixelの線分であり、スライダー初期値は中央とした。また、SLDを提示された回答者がスライダーを操作せずに次の画面に進もうとした場合は、警告を表示し、初期位置がそのまま得点化されないようにした。警告文は「最初に表示されたスライダーの位置をご回答とされる方も、一度、スライダーを動かしてからその位置にお戻しください」というものであった。

回答形式比較のための質問項目

測定内容は社会科学領域のサーベイにおける一般的な質問項目のなかから収入満足度を選定した。質問文は「あなたは現在のご自分の生活で、所得や収入の面では、どれくらい満足していますか」とし、その後に、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「『どちらともいえない』を0、『満足』や『不満』を5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答えください」と依頼した(注4)。

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7問目であり、2018年1回目調査ではどの条件でも17問中の5問目、2018年2回目調査ではどの条件でも11問中の5問目であった。

有効性評価の指標

回答時間 収入満足度の質問画面の表示時刻と、同画面での「次へ」ボタン押下時刻をミリ秒単位で記録し、その差を回答時間（画面提示時間）とした。

回答形式満足度 収入満足度の各回答形式について、回答直後に「その形式であなたのお気持ちを十分表すことができましたか、そうでもありませんか」とたずね、5件法で回答を得て、値が大きいほど満足度が高くなるよう1～5点を与えた。

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回目調査の収入満足度について相関係数を算出し、再検査信頼性の指標とした。分析の際には、有効票全体を対象とする分析に加え、1回目調査と2回目調査の間に、回答者に大きな影響を与えた可能性のある生活上の出来事（ライフイベント）があった者を除外した分析も行った。ライフイベントの質問項目は、就業状況、家計状況、健康状態、人間関係上のトラブル、転居、結婚・離婚、死別などについて、11項目設定し、これに「その他」を加え、合計12項目を調査の最後に配置した（個別強制選択形式であり、項目によっては配偶者の状況変化も回答可能）。なお、本パネル調査の期間中には、ボーナス（12月）や昇給（4月）の時期が含まれておらず、期間も4週間と短いため、収入満足度の安定を仮定した再検査信頼性の検討が可能であると判断した。

結 果

回答形式条件間の属性比較

分析に先立ち、各調査ごとに、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。その結果、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、自由裁量所得との有意な関連は認められなかった。また、全ての調査で、回答形式条件間にPCストレス度（注5）の有意差は認められなかった（一元配置分散分析）。なお、既未婚については、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条件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（ $F(3, 1395) = 74.93, p < .001, \eta^2 = .14$ ）。また、各条件の等分散性が棄却されたため（Levene統計量 $L(3, 1395) = 4.17, p < .01$ ）、Games-Howell法（5%水準）による多重比較を行ったところ、条件間に有意な差が認められた（SLD > RB11 >

RB5V, RB5H)。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：RB5Vで1名、RB11で1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 ($F(2, 587) = 38.45, p < .001, \eta^2 = .12$)。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（RB11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 ($F(2, 546) = 46.26, p < .001, \eta^2 = .14$)。また、各条件の等分散性が棄却されたため ($L(2, 546) = 3.36, p < .05$)、Games-Howell法（5%水準）による多重比較を行ったところ、条件間に有意な差が認められた (SLD > RB11 > RB5V)。

以上のように結果は一貫しており、回答時間の中央値（第2四分位）は、RB5VやRB5HよりもRB11のほうが1秒ほど長く、RB11よりもSLDのほうが2～3秒長くなっていた (Table 1)。また、RB5VとRB5Hの間に回答時間の差は認められなかった。

Table 1 収入満足度の回答時間（秒）

調査時点 回答形式	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	1.7	2.1	1.4	3.0	1.8	1.4	3.0	1.7	1.5	3.2
最大値	90.5	81.2	90.6	88.4	46.5	60.9	53.5	25.8	28.1	46.2
第1四分位	3.9	4.0	4.7	6.8	3.9	4.6	6.4	3.7	4.4	6.1
第2四分位	5.2	5.3	6.5	9.7	5.3	6.3	8.9	5.0	6.0	8.3
第3四分位	7.2	7.7	9.4	14.8	7.2	9.0	12.3	7.2	8.9	11.3
平均値	6.9	7.2	9.4	12.6	7.0	8.0	10.8	5.8	7.6	10.1
SD	7.7	7.5	10.6	10.1	6.7	6.8	7.3	3.0	4.9	6.6
歪度	6.8	6.2	4.4	3.3	4.1	4.7	2.7	2.4	1.9	2.5
<i>n</i>	356	352	364	327	194	213	183	185	198	166

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。

Table 2 収入満足度の回答形式満足度（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	10.4	11.9	10.2	20.5	8.2	9.9	19.1
まあ表せた	45.2	48.3	43.1	52.0	52.1	41.3	52.5
どちらともいえない	32.9	29.5	35.2	21.7	27.8	37.1	20.2
あまり表せなかった	9.0	8.2	9.1	4.6	8.2	8.9	8.2
まったく表せなかった	2.5	2.0	2.5	1.2	3.6	2.8	0.0
<i>n</i>	356	352	364	327	194	213	183

注：回答形式のラベルはTable 1と同様。

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 ($F(3, 1395) = 12.31, p < .001, \eta^2 = .03$)。また、各条件の等分散性が棄却されたため ($L(3, 1395) = 4.93, p < .01$)、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 > RB5V, RB5H, RB11)。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 ($F(2, 587) = 9.20, p < .001, \eta^2 = .03$)。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、条件間に有意な得点差が認められた (SLD > RB5V, RB11)。

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11～17ポイント、RB11より20ポイントほど高かった（注6）。

再検査信頼性

収入満足度の再検査信頼性を回答形式ごとに見たのがTable 3である。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）についてみると、全ての回答形式で再検査信頼性係数 (r) が.80前後であったため、どの回答形式も十分な再検査信頼性を有すると判断された。なお、 r の値自体はSLDで若干低めの値となっていたが、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。

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群の場合とほぼ同様であった。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ほぼ同様であった。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 収入満足度の再検査信頼性 (r)

分析対象	回答形式	r	95%CI	n
VLD	RB5V	.85***	.81-.89	176
	RB11	.84***	.79-.87	192
	SLD	.79***	.73-.84	161
NLE	RB5V	.83***	.77-.88	138
	RB11	.82***	.76-.86	161
	SLD	.80***	.73-.85	134

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

収入満足度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満足度が高くなるように得点化した）。そのうえで、各調査ごとに収入満足度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）。

その結果、平均値については条件間に有意差は認められなかった（2018年2回目調査のVLD群では平均値に有意差が認められたが、ライフイベントのなかったNLE群に限定すると有意差は認めら

Table 4 収入満足度の回答分布

調査時点 回答形式	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	136.2	127.5	134.7	127.5	137.0	130.3	127.2	141.6	129.2	119.3
SD	71.6	71.2	71.7	83.0	74.7	69.9	81.0	75.4	74.5	76.3
第1四分位	90.0	90.0	68.2	68.0	90.0	68.2	74.0	90.0	68.2	66.8
第2四分位	150.0	150.0	150.0	133.0	150.0	150.0	136.0	150.0	150.0	129.5
第3四分位	210.0	210.0	177.3	187.0	210.0	177.3	173.0	210.0	177.3	168.0
最頻値	150.0	210.0	150.0	0.0	210.0	150.0	0.0	210.0	150.0	0.0
歪度	-.028	-.033	-.026	.046	-.168	-.076	.053	-.147	.032	.079
n	356	352	364	327	194	213	183	185	198	166

注：回答形式のラベルはTable 1と同様。収入満足度は得点が高いほど満足度が高い。SLD以外の回答形式の得点はSLDと比較可能な得点に変換。

れなかった)。なお、分散については、値自体は常にSLDが最大であったが、SLDで有意に大きいのは2017年調査のみであった ($L(3, 1395) = 4.31, p < .01$)。

次に、分布形については、歪度を見ると、どの調査でもRB5Vが負、SLDが正であり、特に2018年調査では両回ともRB5Vが強い負の歪みを示していた。また、最頻値については、SLDの最頻値は全て0（尺度右端の「不満」）であった。その割合は、2017年が12.2%（左端1.8%、中央3.7%）、2018年1回目が11.5%（左端2.7%、中央7.1%）、2018年2回目（VLD群）が12.7%（左端0.0%、中央4.8%）と安定して突出していた。これに対し、RB11の最頻値は全て中央の「どちらともいえない」で、2017年が27.2%、2018年1回目が31.5%、2018年2回目が24.7%と安定して突出していた。

項目間相関

収入満足度と他の項目との相関を回答形式間で比較した結果がTable 5である（注7）。まず、2017年調査では、年齢についてRB11のみで有意な相関が認められなかった。また、カテゴリー数の多い7項目については、相関係数の方向と有意水準に、回答形式間にまったく違いが見られなかった（注8）。なお、カテゴリー数の少ない性別と最終学歴については、RB11とSLDを範囲均等分割に最も近くなるように5段階化し、Cramerの連関係数（ V ）を比較した。その結果、性別との関連については回答形式間に違いが見られ、RB11とSLDで連関係数が有意な値を示していた。次に、2018年1回目調査では、各回答形式とも収入満足度と年齢の間に有意な相関が認められなかったが、それ以外は2017年調査とほぼ同様の結果であった。

Table 5 収入満足度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)	.12*	.11*	.08	.16**	-.02	.07	.13
世帯年収(10) †	.36***	.41***	.43***	.37***	.46***	.42***	.48***
自由裁量所得(10)	.30***	.39***	.29***	.38***	.36***	.46***	.28***
平日自由時間(9)	.08	-.05	.00	.00	-.02	.04	.01
休日自由時間(9)	.08	-.04	-.06	.00	.00	.06	-.06
ネット利用時間(9)	-.11*	-.11*	-.14**	-.14*			
テレビ視聴時間(9)	.03	-.01	.04	.08			
Web調査回答数(9)	.00	-.07	-.08	-.02	.02	-.07	-.04
性別(2)	.15	.10	.20**	.18*	.14	.12	.30**
最終学歴(3) †	.13	.14	.10	.12	.16	.19	.08
<i>n</i>	356	352	364	327	194	213	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)。* $p < .05$, ** $p < .01$, *** $p < .001$

総合考察と結論

本研究では連続型の両極型スライダー尺度について検討した。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。こうした時間増は回答操作による時間増を考慮しても、スライダー尺度が回答者に熟慮を促すことを示唆しているが、そうした熟慮と時間増は調査に悪影響を及ぼすほどのものとは考えにくい結果であった。そのため、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。

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。こうした結果は山田・江利川 (2014) でも報告されており、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。一般にサーベイにおいては、リッカート尺度の連続等による回答者の倦怠感の高まりが懸念されるが、スライダー尺度の回答形式満足度の高さには、そうした状況を改善する可能性があると考えられる。

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.80前後と十分高く、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ほぼ同様であった。

以上の結果から、両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、以上の結果は収入満足度に関するものであり、質問内容に関わらず一般化できるとは限らない。また、本研究で行った調査は質問数があまり多くないため、質問数の多い調査や、そのなかの質問の位置によって結果が異なる可能性も考えられる。さらに、本研究の公募型Web調査では、脱落や無回答、および、携帯・スマホからの回答を許容していないため、そうした点については、今後さらに一般化に関する検討が必要とされる。

なお、本研究では、両極型スライダー尺度の回答分布や他の項目との相関についても検討された。その結果、平均値については、スライダー尺度で若干低めの値となる傾向がうかがわれたが、回答形式間に有意差は認められなかった。また、分散については、スライダー尺度でラジオボタンより有意に大きいのは3回の調査のうち1回だけであった。ただし、歪度と最頻値については回答形式間に顕著な違いが見られた。なお、他の項目との相関については、相関係数の方向と有意水準を見る限り、回答形式間の差はほとんど認められなかった。ただし、性別と収入満足度の関連については、スライダー尺度でラジオボタンよりも強く表れることも示された。以下では、こうした結果の含意について、主にラジオボタン5件法とスライダー尺度の互換性という視点から検討する (注9)。

まず、ラジオボタン5件法とスライダー尺度との間で、平均値、分散、他の項目との相関に大きな違いが見られなかったため、両回答形式には一定の互換性があるようにも思われた。しかし、回答の分布形については両者の間に顕著な違いがあった。まず、スライダー尺度では、最頻値が尺度右端の「不満」であった。こうした回答は、回答位置に言語ラベルがあることによっても考えられるが、「満

足」と「どちらともいえない」の割合は突出していなかった。また、スライダーが反応バイアス（極端回答傾向）を助長した可能性もあるが、「満足」の割合は突出していなかった。そのため、「不満」という回答内容にも原因があると考えざるを得ない。プロスペクト理論 (Tversky & Kahneman, 1981) に従えば、「スライダー尺度では否定的な極方向への極端回答が生じやすい」と考えることもできるため、それが頑健な尺度特性なのかどうか、今後検証が必要である。

次に、回答の分布形については、歪度の符号が示すように、ラジオボタン5件法を基準にすると、スライダー尺度では不満度を高める方向に反応バイアスが働き、スライダー尺度を基準にすると、ラジオボタン5件法では満足度を高める方向に反応バイアスが働くことが示唆された。原因として考えられるのは、ラジオボタン5件法の「やや満足」への回答の集中（注10）、および、スライダー尺度における上述の極端回答傾向である。こうした点は、回答の割合に注目する場合には、看過できない点になるものと思われる。

なお、本研究では、性別と収入満足度との関連がスライダー尺度で強く表れており、その原因として、上述の「不満」という極端回答の性差が疑われた。しかし、そうした回答に性別との有意な関連は認められなかった。従って、各条件の回答者が等質であると仮定できるなら、スライダー尺度は、「不満」という極端回答に限らず、男性に不満を強く回答させ、女性に満足を強く回答させる傾向があると考えざるを得ない。こうした点についても、それが頑健な尺度特性なのかどうか、今後さらなる検証が必要である。

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思われ、本研究で取り上げた両極型スライダー尺度は、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る。ただし、両極型スライダー尺度は常にラジオボタン5件法より分散が大きいわけではないようである。また、両極型スライダー尺度はラジオボタン5件法とは分布形が異なり、測定内容によっては特殊な反応バイアスの影響を受けることも懸念される。従って、ラジオボタン5件法と両極型スライダー尺度の間には、必ずしも十分な互換性があるわけではないと言わざるを得ない。また、そのため、両極型スライダー尺度の利用にあたっては、利用目的に応じ、十分な事前検討を行うことが推奨される。

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。
- 2) スライダー尺度がVASに劣ることを示した研究としてFunke（2016）が頻繁に引用されるが、この研究のVASとスライダー尺度が離散型である点には注意が必要である。
- 3) 回答者の公募は各調査ごとに独立に行われた。また、調査名は一般的なものであり、分析に用いる変数への影響はないと判断された。なお、公募型Web調査の分析においては、リッカート・グリッドのような回答負荷の高い回答画面で発生しやすい不正回答の除去を目的として、当該回答画面の超短時間回答者を除外することがある。しかし、本研究の主な従属変数である収入満足度や回答形式満足度は、1画面1質問で質問文も短く、1ジャッジ・1クリックで回答可能であった。また、実際に、回答時間の最小値は小さく、短時間方向の外れ

値も識別不能であった。そのため、回答画面ごとの超短時間回答者の除外は行わないこととした。

- 4) 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0～5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)。
- 5) PCストレス度7項目 (5件法) について因子分析 (最尤法、プロマックス回転) を行ったところ、全ての調査で、速度ストレス度と操作ストレス度による2因子構造が認められた。そこで、各因子の因子得点を各PCストレス度の尺度得点とした。
- 6) 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、ラジオボタンでは約4割にも及んでおり (Table 2)、サーベイに対する回答者の否定的な意識が懸念される。
- 7) 関連の検討に使用した項目は以下のとおり。「世帯年収」は昨年の税込み世帯年収で、分析時にはDK・NAを除外。「自由裁量所得」は1ヶ月間に自分の自由に使える金額 (生活費や預貯金を除く)。「平日自由時間」と「休日自由時間」は仕事・家事・睡眠などを除く1日の自由時間。「ネット利用時間」は平日1日のインターネット利用時間 (仕事や授業での利用を除く、全デバイスによる利用)。「テレビ視聴時間」は平日1日の視聴時間 (録画番組を含む番組視聴)。「Web調査回答数」は登録モニターとして回答した直近1ヶ月間の調査件数。「最終学歴」は「高卒以下」「専門・専修・短大・高専卒」「大卒以上」にカテゴリー併合し、分析時には「答えたくない」と「その他」を除外。
- 8) 世帯年収と自由裁量所得については、回答形式間で相関係数の値に多少の違いが見られたが、各相関係数の間に有意な差が認められたのは、2018年1回目調査の自由裁量所得のRB11とSLDの間のみ ($p < .05$) であった。
- 9) 本研究の結果では、RB5HはRB5Vとほぼ同様の傾向を示した。また、RB11はRB5Vと比べて、回答形式満足度が低く、分散も増えず、回答が中央に集中しがちであり、他の項目との関係も独自の傾向を示していた。そのため、RB5Vの代わりにRB5HやRB11を使用するメリットはないと判断された。
- 10) RB5VとSLDの回答分布 (%) は以下の通り (Figure 1の選択肢の上から順に表示。SLDは範囲均等分割による5段階)。2017年調査: RB5V [5.6, 26.1, 17.0, 22.2, 19.1]、SLD [9.2, 19.3, 25.4, 23.5, 22.6]。2018年1回目調査: RB5V [5.2, 29.4, 27.8, 13.9, 23.7]、SLD [9.3, 13.1, 32.2, 21.9, 23.5]。2018年2回目調査: RB5V [7.6, 28.6, 27.0, 15.7, 21.1]、SLD [9.0, 12.0, 32.5, 24.1, 22.3]。

引用文献

- Bosch, O.J., Revilla, M., DeCastellarnau, A., & Weber, W. (2019). Measurement reliability, validity, and quality of slider versus radio button scales in an online probability-based panel in Norway. *Social Science Computer Review*, 37, 119-132.
- Chyung, S.Y.Y., Swanson, I., Roberts, K., & Hankinson, A. (2018). Evidence-based survey design: The use of continuous rating scales in surveys. *Performance Improvement*, 57, 38-48.
- Cook, C., Heath, F., Thompson, R.L., & Thompson, B. (2001). Score reliability in Web- or internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61, 697-706.
- Couper, M.P., Tourangeau, R., & Conrad, F.G. (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, 24, 227-245.
- Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, 34, 244-254.
- Funke, F., Reips, U.-D., & Thomas, R.K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, 29, 221-231.
- Liu, M. & Conrad, F.G. (2019). Where should I start? On default values for slider questions in web surveys. *Social Science Computer Review*, 37, 248-269.
- Maloshonok, N. & Terentev, E. (2016). The impact of visual design and response formats on data quality in a web survey of MOOC students. *Computers in Human Behavior*, 62, 506-515.
- Roster, C.A., Lucianetti, L., & Albaum, G. (2015). Exploring slider vs. categorical response formats in web-based surveys. *Journal of Research Practice*, 11, 1-19.
- Sikkel, D., Steenbergen, R., & Gras, S. (2014). Clicking vs. dragging: Different uses of the mouse and their implications for online surveys. *Public Opinion Quarterly*, 78, 177-190.

Tversky,A.& Kahneman,D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.

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 (1) : —Measuring Income Satisfaction Levels Using Bipolar Slider Scales—

Kazunari YAMADA

The purpose of this study is to examine the effectiveness of bipolar 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 income satisfaction 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, (3) test-retest reliability was sufficiently high, and there was no difference in test-retest reliability between scale formats. However, (4) the response distribution forms for SLDs differed greatly from those using RBs, suggesting the occurrence of each response bias in each scale format. The usefulness and limitations of bipolar SLDs in volunteer panel Web surveys are discussed based on the above results.