

Web 調査における Visual Analogue Scale の有効性評価

Evaluating the Effectiveness of Visual Analogue Scales in a Web Survey

山田一成

Kazunari YAMADA

江利川滋*

Shigeru ERIKAWA

1. 問題

本研究の目的は社会科学領域の Web 調査における Visual Analogue Scale (VAS) の有効性について検討することである (注1)。

VAS は回答者が自己の評定結果を連続帯である線分上の長さで自由に回答できる尺度であり、両極尺度では線分の両端または両端と中点に (図1)、単極尺度では線分の両端に (図4)、それぞれカテゴリーラベルが配置される (Reips & Funke, 2008; Couper, Tourangeau, Conrad, & Singer, 2006)。また、回答者はそうした線分上で、評定結果に対応するポイントを指定して回答することになる。

VAS の初出は Hayes & Patterson (1921) とされ (Ahearn, 1997)、その後、主に医療看護系の領域を中心に、回答者の主観的な痛み、疲労、気分などの測定に広く使用されるようになり、今日では VAS を使用した研究論文も膨大な数に達している (注2)。

こうした VAS の利点としては、間隔尺度水準のデータが得られる、回答がカテゴリーラベルの解釈に影響されない、回答への動機づけや満足度が高まる、微細な差や変化が測定可能である、などの点が挙げられてきたが (注3)、欠点としては、回答に練習を要する、回答に時間がかかる、脱落や無回答が増える、データのコーディング (質問紙における線分の長さの計測) にコストがかかる、電話調査では利用できない、などの点が挙げられてきた (Couper et al., 2006; Flynn, van Schaik, & van Wersch, 2004)。

ただし、そうした欠点はあるものの、それらが実際の調査場面で回答に大きな悪影響を及ぼすことがなければ、社会科学領域の調査研究においても、VAS の使用には以下のようなメリットがあると考えられる。まず、VAS によって間隔尺度水準のデータが得られれば、大規模調査であっても精度

*株式会社 TBS テレビ

の高い統計分析が可能となる。また、SD法のように、回答者が複数の項目間の関係を考慮して回答したい場合、リッカート尺度(LS)では同じ評定段階での回答で妥協せざるを得ないが、VASであれば微細な差を反映させた回答が可能となる。さらに、因子分析などの多変量解析を前提とした調査研究では、内容・形式が類似した質問項目が多数並ぶため、回答者の動機づけの低下や疲労による測定への悪影響が懸念されるが、VASであれば、使い方によっては、回答者の動機づけを高め、調査への興味を持続させることで、測定の質を向上させることも可能となる。

なお、世論調査や消費者調査に代表されるような大規模調査では、紙の調査票でVASを使用した場合のコーディングコストが膨大なものとなるため、これまでVASは積極的に用いられる方法とは言い難かった。しかし、今日では、回答者がPCに回答を直接入力するCASI(Computer-Assisted Self-Interviewing)方式が様々な形で実用化されており、数百段階のVASの回答も自動的に数値化可能であるため、少なくともWeb調査においては、VASのコーディングコストは欠点ではなくなったと言ってよい(注4)。

そうした状況のなか、近年では社会科学領域でも、Web調査におけるVASの有効利用への関心が高まり、その方法論的基礎に関する実証研究が盛んに行われている。例えば、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、ラジオボタン(21段階)、数値入力(1~21)の3形式によるWeb調査の回答を比較している。そして、VASの平均や分散が他の形式とほとんど変わらないこと、および、VASでは脱落や無回答が増え、回答時間が長くなることなどを報告している。

また、Reips & Funke (2008)は、Web調査の回答者があらかじめ指定された値をVASでどの程度正確に回答できるか検討している。そして、値の形式(%と分数)とVAS線分の長短(50pixel, 200pixel, 800pixel)を組み合わせた6条件で、いずれもほぼ指定通りの値が回答されたことや、VASにより間隔尺度水準のデータが得られたことなどを報告している。

さらに、Funke, Reips, & Thomas (2011)は、Web調査においてVAS(slider scale)とLS(ラジオボタン)を比較し(注5)、VASで脱落が多く、回答時間が長くなること、および、そうした傾向が低学歴層で顕著であることなどを報告している。

なお、Funke & Reips (2012)は、Web調査でのSD法(1画面13項目)について、VAS(250段階)とLS(5件法)とで、平均値、脱落、無回答、回答時間に差が見られないことを報告するとともに(注6)、VASではLSよりも回答者が回答を調整することが頻繁で、それにより良質のデータが得られるため、Web調査のSD法にはVASの採用が推奨される、と述べている。

このように、既に英語圏では、Web調査によって可能となった調査技法の有効利用のひとつとしてVASに注目が集まり、参考情報が様々な形で蓄積・公開されている段階である(注7)。しかし、日本においては、こうした実証研究はほとんど行われていないようである(注8)。

そこで本研究では、こうした研究動向を踏まえ、社会科学領域のWeb調査におけるVASの有効利用のために、首都圏において成人男女を対象とするWeb調査を実施し、スプリット法によってVAS

と LS の比較を行い、両者の互換性、および、心理測定法としての相対的優劣について検討する。

なお、意識や態度などの測定においては、質問内容と回答形式（回答様式、選択肢配列形式、選択肢ワーディング）の双方が回答に影響すると考えたほうがよい場合が多いが、これらのうち、本研究では特に回答形式に注目した検討を行う。また、本研究で分析の対象となる調査データは公募型 Web 調査のデータであるが、分析にあたっては統計的検定を実施する。

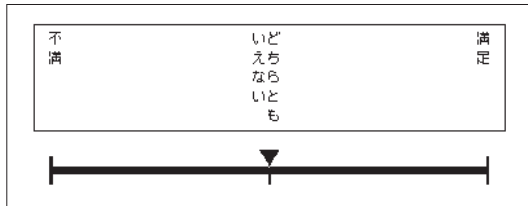


図1 両極型 VAS (▼: スライダー・初期位置中央)

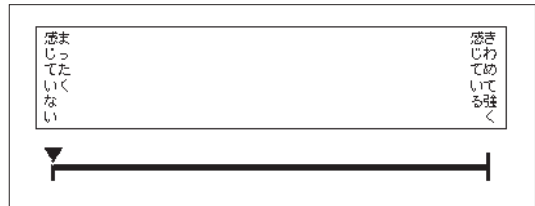


図4 単極型 VAS (▼: スライダー・初期位置左端)



図2 両極型11件法ラジオボタン

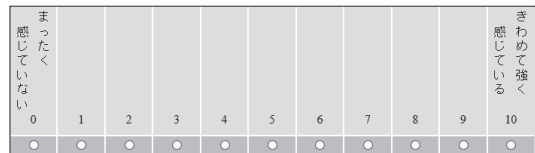


図5 単極型11件法ラジオボタン

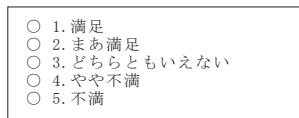


図3 両極型5件法ラジオボタン

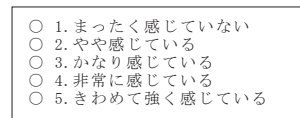


図6 単極型5件法ラジオボタン

2. VAS の種類と有効性評価

2.1. VAS の種類

CASI で利用される VAS は現状で 2 種類ある。一つは、回答となるポイントをマウスのクリックやタッチペンの操作で直接指定する方式であり、タブレット PC を利用した調査などで用いられる (例: Stubbs, Hughes, Johnstone, Rowley, Reid, Elia, Stratton, Delargy, King, & Blundell, 2000)。もう一つは、スライドバーを利用した方式で、Web 調査で多用されている (例: Funke et al., 2011)。スライドバーとは、マウスのドラッグ操作で Web 画面の線分上を直線移動させることが可能なアイコンであり、回答者が尺度上の回答ポイントを指定するために用いるものである。本研究では、これらのうち、後者のスライドバーを利用した VAS について検討する。

2.2. VAS の有効性評価

本研究では、回答時間、回答形式満足度、回答分布の3指標によってVASの有効性を評価する(注9)。まず、VASの第1の利点は間隔尺度水準の測定が可能になる点であるが、その代償として回答時間が増加するという欠点も予想される。そこで、LSと比較した回答時間の増減について検証する必要がある。なお、回答時間の比較は個人内でも行われるべきであるが、今回の調査データで可能だったのは群間の比較のみであった。

また、VASの第2の利点は、回答形式の特性に起因する回答への満足度(回答形式満足度)の向上が期待できる点である。しかし、VASに不慣れな回答者が微細な主観的程度を回答できたとしても、回答に手間取ったことに影響され、回答形式満足度が低くなる可能性もある。そこで、VASの回答形式満足度がLSより増加するかどうかについても検証する必要がある。なお、回答形式満足度の比較は個人内でも行われるべきであるが、今回の調査データで可能だったのは群間の比較のみであった。

さらに本研究では、VASとLSの回答分布の一致度についても検討する。もしも二つの尺度の回答分布に大きな差が見られないのであれば、両者を互換的に用いることが可能であると考えられる。また、こうしたVASとLSの比較に関しては、LSの評定段階数の影響についても検討されるべきである。そこで、本研究では5件法と11件法のLS(ラジオボタン)を用いて、この点に関する検証を行う。

なお、無回答の増減もVASの有効性評価指標となりうるが、本研究では調査票が無回答を許容しない設計であったため、検討対象外とした。

3. 方法

3.1. 調査概要

【調査方法】 Web調査(著者が所属するWeb調査研究会と株式会社TBSテレビ・マーケティング部が共同で実施。実査は調査会社に委託)。

【調査対象】 東京・埼玉・千葉・神奈川在住の20~69歳男女。調査会社の登録モニターを利用し、人口比例に基づいて本調査での各性年代の回答者数を割当。

【事前調査】 期間:2010年3月25日(木)~26日(金)。配信数:26,000名。回収数:5,891名。対象者数:5,465名。追加抽出条件:回答環境が著しく異なる少数のダイヤルアップ回線利用者を除外するため、通信回線がFTTH・ADSL・CATVのいずれかである者を抽出。

【本調査】 期間:2010年3月26日(金)~28日(日)。配信数:1,986名。回収数:1,563名。有効数:1,559名。

【調査票】 3種類の調査票によるスプリット法で、回答者をプログラム制御によりランダムに各票に割当。回答者数はA票519名(ラジオボタン5件法)、B票518名(ラジオボタン11件法)、C票522

名 (VAS)。

[その他] 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。また、本調査の回答者全員に調査会社による報酬ポイントが付与された。調査名は、事前調査が「生活に関するアンケート」、本調査が「日常生活に関するアンケート」と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、VASを含む回答形式比較のための質問項目に関して、文脈効果が懸念されるような項目群は含まれていなかった。

3.2. 回答形式比較のための質問項目

本研究では、社会科学領域の変数について、下記のような両極尺度と単極尺度をそれぞれ設定した。調査票における質問項目の位置は、どの調査票でも全25問中の10問目が両極尺度、12問目が単極尺度である。なお、VASはJavaScriptで動作する仕様であった。

[両極尺度] QOL研究でのVAS使用実績を踏まえ、経済生活満足度を測定した。質問の共通主文は「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、VASと5件法では「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」と続く。また、11件法では「『不満』を0、『満足』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答えください」と続く。

回答形式は図1～図3のとおりで、VASが300pixelの線分、11件法と5件法がラジオボタン。VASと11件法は水平方向に、5件法は垂直方向に、回答選択肢ないし線分が提示されている。いずれも左端(5件法は最下段)に「不満」、右端に「満足」(5件法は最上段)というラベルがある。その間は、VASと11件法では中点だけに「どちらともいえない」というラベルがあり、5件法では「やや不満」「どちらともいえない」「まあ満足」というラベルがある。VASのスライドバーは中央を初期位置とした(図1)。

[単極尺度] 一方向に程度が増大する変数として対人ストレス度を測定した(注10)。質問の共通主文は、「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、VASでは「あなたのお気持ちに最も近いところにスライドバーをマウスで動かしてお答えください」と続く。また、11件法では「『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答えください」と続き、5件法では単極尺度と同一の文章が続く。

回答形式は図4～図6のとおりで、VASが300pixelの線分、11件法と5件法がラジオボタン。VASと11件法は水平方向に、5件法は垂直方向に、回答選択肢ないし線分が提示されている。いずれも左端(5件法は最上段)に「まったく感じていない」、右端(5件法は最下段)に「きわめて強く感じている」というラベルがあり、5件法はさらに「やや感じている」「かなり感じている」「非常に感じている」という中間ラベルがある。VASのスライドバーは左端を初期位置とした(図4)。

なお、両極尺度、単極尺度とも、VASを提示された回答者がスライドバ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ドバーの位置をご回答とされる方も、一度、スライドバ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。また、技術的には、スライドバーの移動に対応する回答数値とその変化を、同一画面上に逐次表示することも可能であった。しかし、それによって回答が5の倍数に偏ることも予想されたため (Couper et al., 2006)、本研究では画面表示しない方式を採用した。

3.3. VASの有効性評価指標

【回答時間】 本研究では各質問ごとに、設問画面の表示時刻（開始時間）と同画面での「次へ」ボタン押下時刻（終了時間）をミリ秒単位で測定し、その差を回答時間とした。以下の分析では、ミリ秒単位を秒単位に変換し、小数点以下を四捨五入して整数化した回答時間を使用した。

【回答形式満足度】 3種類の回答形式のそれぞれについて、両極および単極尺度の回答直後に回答形式への満足度をたずねた。すなわち、「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、「十分表せた」「まあ表せた」「どちらともいえない」「あまり表せなかった」「まったく表せなかった」の5件法で回答させて、値が大きいほど満足度が高くなるよう1～5点を与えた。

【回答分布】 3形式による回答分布の相互比較を可能にするため、両極および単極尺度について、11件法と5件法の回答値の範囲がVASと同一の0～300となるよう変換した。これにより、LS回答の階級値は、11件法では「0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300」、5件法では「0, 75, 150, 225, 300」となった。なお、VASとLSの得点分布を比較する方法としては、①VASの得点を比較対象となるLSの段階数に合わせて離散変量化する方法 (Funke et al., 2011; Flynn et al., 2004) や、②VASの得点を0～100のレンジに線形変換するとともに、LSの各評定カテゴリに「0～100をカテゴリ数に応じて等間隔に分割した値」を与える方法 (Couper et al., 2006; Funke & Reips, 2012) などが知られているが、本研究では②の方法を採用する（ただしレンジは0～300）。なお、この方法によれば、VASと11件法と5件法を同時に比較することが可能となる。

3.4. その他の質問項目

【回答者属性項目】 回答者の基本属性として、性別（男性49.9%、女性50.1%）、年齢 ($M=44.65$, $SD=13.52$)、未既婚（未婚35.6%、既婚64.4%）、子供の有無（子供あり54.0%、子供なし46.0%）、最終学歴（高校卒以下25.9%、短大・専門学校卒24.6%、4年制大学卒以上46.7%、在学中1.7%、NA1.0%）について回答を得た。

【PCストレス項目】 回答環境に対する心理的ストレスについて回答形式条件間で差があるかどうか検討するために、回答者が回答に用いたと考えられる「ふだん使っているパソコン」について、下記3項目への該当の有無を尋ねた。すなわち、①「ふだん使っているパソコンは処理速度が遅いの

で、いつもストレスを感じている」(該当30.0%)、②「ふだん使っているパソコンはディスプレイ画面が狭いので、いつもストレスを感じている」(該当5.6%)、③「インターネットを使うとき、ふだん使っているブラウザの表示が遅いので、いつもストレスを感じている」(該当25.1%)である。

4. 結果

4.1. 回答形式条件群間の属性比較

分析に先立ち、調査全体の回答時間が極度に短い回答者と極度に長い回答者を分析の対象外とした。前者は報酬ポイント獲得のみを目的とする不正回答者である可能性があると考えられたためであり、後者は回答の中断など他の回答者と回答状況が異なる可能性があると考えられたためである。具体的手続きとして、全質問の回答時間合計の分布を確認し、非連続性が認められた150秒未満と1,800秒超の回答者を除外した。除外された回答者数はA票(5件法)12名、B票(11件法)13名、C票(VAS)4名であり、残り1,530名の全質問回答時間の平均は487.3秒であった。

次に、残りの1,530名について、VAS、11件法、5件法、の三つの回答形式条件間で、基本属性分布を比較した。その結果、性別($\chi^2(2, N=1530) = 0.04$)、年齢($F(2, 1527) = 0.02$)、既未婚($\chi^2(2, N=1530) = 0.71$)、子供の有無($\chi^2(2, N=1530) = 0.94$)、最終学歴($\chi^2(8, N=1530) = 6.75$)のいずれも有意差は見られなかった。

また、PCストレス項目も同様に、①「処理速度が遅い」($\chi^2(2, N=1530) = 1.31$)、②「ディスプレイ画面が狭い」($\chi^2(2, N=1530) = 0.28$)、③「ブラウザの表示が遅い」($\chi^2(2, N=1530) = 4.10$)と、回答形式条件間に有意差は見られなかった。

4.2. 回答時間

両極尺度(経済生活満足度)と単極尺度(対人ストレス度)で個別に回答時間の分布を確認すると、両極尺度では極度に回答時間の長い回答者がいたが(最長は1,289秒)、単極尺度で180秒を超える回答者はいなかった。そのため、以下の分析では、経済生活満足度の回答時間が180秒を超える回答者4名を除外した。また、両極尺度でも単極尺度でも、回答時間の分布に正の歪みが確認されたため(表1の歪度参照)、以下では常用対数変換した値を分析に用いることとした。

まず、両極尺度について回答時間(対数変換値)の分散分析を行うと、群間に有意な差が見られた($F(2, 1523) = 137.70, p < .001$)。また、各群の等分散性が棄却されなかったため(Levene 統計量 $L(2, 1523) = 1.16, ns$)、Tukey の HSD 法(5%水準)による多重比較を行うと、3群間に有意な差が見られた(VAS > 11件法 > 5件法)。

次に、単極尺度について回答時間(対数変換値)の分散分析を行うと、群間に有意な差が見られた($F(2, 1523) = 27.48, p < .001$)。また、各群の等分散性が棄却されなかったため($L(2, 1523) = 0.76$,

ns)、Tukey の HSD 法 (5% 水準) による多重比較を行うと、3 群間それぞれに有意な差が見られた (VAS>11 件法>5 件法)。

以上の結果から、両極尺度・単極尺度ともに、評定段階数が増加するにつれて回答時間も増加する傾向が認められた。表 1 に示すように、回答時間の中央値は、両極尺度の VAS では 11 件法より 3.1 秒増加し、5 件法より 5.6 秒増加していた。同様に、単極尺度の VAS では 11 件法より 1.2 秒増加し、5 件法より 2.2 秒増加していた。

4.3. 回答形式満足度

回答形式満足度の尺度得点分布は表 2 に示す通りである。まず、両極尺度の分散分析では 3 群間に有意な得点差が見られた ($F(2, 1523) = 36.71, p < .001$)。また、各群の等分散性が棄却されたため ($L(2, 1523) = 26.22, p < .001$)、Games-Howell 法 (5% 水準) による多重比較を行うと、3 群間に有意な得点差が見られた (VAS>11 件法>5 件法)。

次に、単極尺度について分散分析を行った結果、3 群間に有意な得点差が見られた ($F(2, 1523) = 29.62, p < .001$)。また、各群の等分散性が棄却されたため ($L(2, 1523) = 36.98, p < .001$)、Games-Howell 法 (5% 水準) による多重比較を行ったところ、3 群間に有意な得点差が見られた (VAS>11 件法>5 件法)。

以上の結果から、両極尺度・単極尺度ともに、評定段階数が増加するにつれて回答形式満足度も増加する傾向が認められた。なお、回答形式満足度の肯定的回答 (「十分表せた」と「まあ表せた」の合計) の割合を比較すると (表 2)、両極尺度では、VAS は 11 件法よりも 19 ポイント、5 件法よりも 25 ポイントほど高かった。また、単極尺度では、VAS は 11 件法よりも 10 ポイント、5 件法よりも 20 ポイントほど高かった。

4.4. 回答分布

尺度得点について、得点範囲が 0~300 となるよう変換した後の記述統計量は表 3 に、相対度数累積分布は図 7 (両極尺度) と図 8 (単極尺度) に示す通りである。まず、両極尺度について分散分析を行った結果、3 群間に有意な得点差は見られなかった ($F(2, 1523) = 1.43, ns$)。一方、単極尺度について分散分析を行った結果、3 群間に有意な得点差が見られた ($F(2, 1523) = 103.63, p < .001$)。また、各群の等分散性が棄却されたため ($L(2, 1523) = 44.07, p < .001$)、Games-Howell 法 (5% 水準) による多重比較を行ったところ、3 群間に有意な得点差が見られた (11 件法>VAS>5 件法)。

表 1 回答時間の記述統計量 (秒)

	経済生活満足度 [両極]			対人ストレス度 [単極]		
	VAS	11件法	5件法	VAS	11件法	5件法
平均値	16.3	12.9	9.4	12.3	11.4	9.5
SD	12.2	12.5	10.0	10.4	12.1	6.7
中央値	13.3	10.2	7.7	10.0	8.8	7.9
最小値	2.6	2.0	1.7	2.3	2.0	2.2
最大値	125.9	162.5	149.5	143.6	160.3	79.7
歪度	3.9	6.5	9.2	7.1	7.2	4.5
幾何平均	13.8	10.6	7.8	10.5	9.3	8.2
<i>n</i>	518	502	506	518	502	506

表 2 回答形式満足度 (列%)

	経済生活満足度 [両極]			対人ストレス度 [単極]		
	VAS	11件法	5件法	VAS	11件法	5件法
十分表せた	18.0	11.2	7.5	20.8	15.9	11.3
まあ表せた	61.4	49.8	46.6	61.0	56.4	51.0
どちらともいえない	13.7	28.7	31.4	13.5	19.7	24.3
あまり表せなかった	6.2	9.2	12.8	4.4	7.2	11.9
まったく表せなかった	0.8	1.2	1.6	0.2	0.8	1.6
<i>n</i>	518	502	506	518	502	506

表 3 回答分布の記述統計量

	経済生活満足度 [両極]			対人ストレス度 [単極]		
	VAS	11件法	5件法	VAS	11件法	5件法
平均値	116.5	115.8	123.6	139.5	155.1	89.2
SD	81.6	75.4	86.2	84.6	74.9	67.4
第一四分位数	52.0	60.0	75.0	64.0	90.0	75.0
中央値	112.0	120.0	150.0	146.0	150.0	75.0
第三四分位数	174.0	150.0	225.0	208.3	210.0	75.0
歪度	0.3	0.2	0.0	0.1	-0.1	1.4
<i>n</i>	518	502	506	518	502	506

注：数値の範囲は0～300にそろえてある。

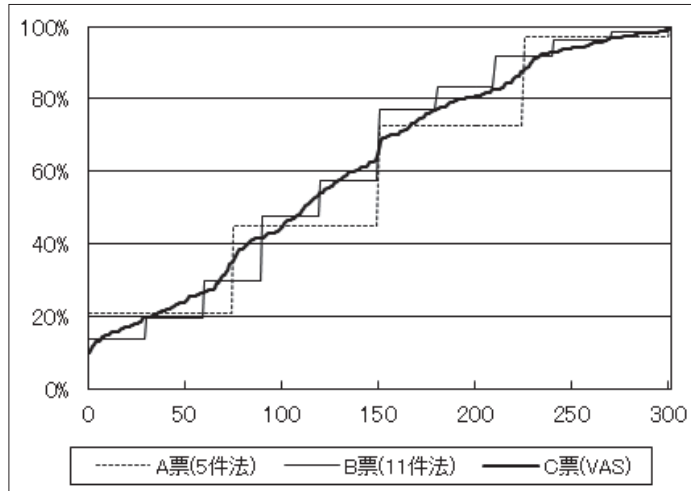


図7 経済生活満足度（両極尺度）回答の累積相対度数分布
注：「満足」方向にスコア大。

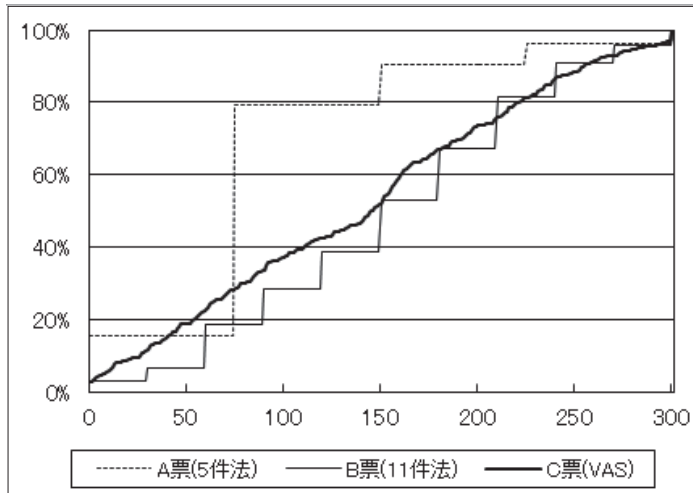


図8 対人ストレス度（単極尺度）回答の累積相対度数分布
注：「ストレス感有」方向にスコア大。

5. 考察

以上の分析結果から、まず、両極尺度については、評定段階数が多いほど回答時間と回答形式満足度が増加することが示された（VAS>11件法>5件法）。ただし、回答時間の増加量は1回答あたり中央値で数秒程度であり、調査全体に悪影響が及ぶほどではないと考えられる。また、両極尺度については、評定段階数によって回答分布に差が認められなかった。そのため、両極尺度のVASはLSと

互換可能であり、間隔尺度化、および、回答形式満足度の上昇というさらなる効用を得ることが可能であると考えられる。ただし、こうした点が経済生活以外の満足度や両極尺度一般にあてはまるかどうかを確認するためには、さらなる調査研究が必要である。

次に、単極尺度については、評定段階数が多いほど回答時間と回答形式満足度が増加することが示されたが（VAS>11件法>5件法）、回答分布についてはVASでLSとは著しく異なる結果が得られた（図8）。例えば、75という尺度得点値に対応する累積相対度数は、VASでは28.4%であるが11件法では18.3%とVASより10ポイント低く、5件法では79.2%とVASより50ポイントほど高かった。こうした差異が生じる要因としては、以下の可能性が考えられる。

まず、VASの特性が特殊な回答傾向を生み出しているという可能性である。すなわち、VASのスライドバー操作行為が、回答者の回答内容への関与度を高めて情報処理を精緻化し、その結果としてLSとは異なる回答が得られた可能性である。しかし、そうならば、他のVASでも同様の回答傾向がうかがえるはずであるが、それが見られない以上、その可能性だけをもって回答分布の差異が単極尺度だけに認められた原因とすることはできない（注11）。

そこで、次に考えるべきなのは、尺度に中央の目安がある両極尺度に比べて、両端にしかラベルがない単極尺度のVASでは中間回答が分散しやすい、という可能性である。今回用いた単極尺度では、VASと11件法を比較すると、中央値や第三四分位数はほぼ同一であるが、第一四分位数はVASのほうが小さく（表3）、尺度の前半部でVASの回答が分散していることがうかがえる。ただし、こうした説明は11件法については可能であっても、5件法の回答分布が他の形式とは大きく異なることが、それだけで説明できるようには思われない。

そこで、5件法にのみ影響を与えうる要因について検討すると、ラベルによる程度差の表現が不十分なために、回答が「やや感じている」という選択肢に集中した可能性が考えられる。また、本研究では、調査会社の仕様により5件法の選択肢のみ縦方向に配置されたが（11件法やVASは横方向）、このことと、上述のラベル表現の不十分さとの交互作用により、5件法のみ特異な分布となった可能性が考えられる。現段階では可能性の指摘に留まらざるを得ないが、この問題は、VASの有効性と別には、単極型LSのレイアウトに関わる問題として極めて重要な論点であり、今後の研究課題であると考えられる（注12）。

Web 調査の急速な普及にともない、今後、VAS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。また、上述のとおり、VASに測定値の精度や回答形式満足度の向上という大きな利点があることも事実であり、変数によってはVASの利用が実証研究に大変有効であると考えられる。ただし、単極型VASの回答分布がLSとは大きく異なる可能性がある以上、単極型VAS（および単極型LS）の設計と使用には十分な注意と事前検討が必要であると言わざるを得ない（注13）。今後のWeb 調査の有効活用のためにも、VASの更なる基礎研究が必要である。

注

- 1) 本研究の一部は日本社会心理学会第51回大会 (2010年) で発表された。
- 2) 国立情報学研究所学術情報ナビゲータ (CiNii) で文字列「visual analogue scale」を引用符で囲んで検索するとヒット数は512件に達する (2014年6月20日現在)。また、同様の検索を Google Scholar で行うと、ヒット数は日本語のページのみで約1,120件に、ウェブ全体からで約120,000件に達する (2014年6月20日現在)。なお、こうした状況は「visual analog scale」で検索しても同様である。
- 3) VAS と同様の尺度として古くから知られているのがグラフ (評定) 尺度 (GRS : graphic rating scale) である。GRS は尺度の両端だけでなく中間にも複数の程度表現のラベルが加わるとされるが、回答者が線分の長さで回答するという点においては VAS と同一である。なお、GRS は古くから心理測定法の研究者によって高く評価されており、例えば Guilford (1954=1959:p.334) は次のように述べている。「グラフ評定尺度の長所は多いが、短所は比較的少ない。それを支持するためにしばしばあげられる利点は、それが簡単で、実施が容易であるということである。この尺度は、興味のもてるものであって、別に動機づけを加える必要はほとんどない。記入も早くすむ。また、評定者を数で悩ますこともない。これらの特徴は、評定者には魅力となるものである。また、研究者の観点からは、グラフ尺度は、評定者にできるかぎりの精細な弁別を求める機会を与えるものであり、採点の精細さを意のままに大きくすることができるものである。グラフ尺度の不利な点として考えられるものは、たいてい、どの型の尺度にも当てはまるもので、当てはまらないものは1つもない。ただし例外的に、ある形式のグラフ尺度は、他の型の尺度より、採点上いくぶん大きな労力を要するということがある」。なお、Kerlinger (1964: p.516) もこうした Guilford の高い評価を引用しながら、「Guilford は少し言い過ぎているが、しかし、彼の主張はよくわかる」と賛同の意を表明している。
- 4) タッチパネル方式の PC ・スマートフォン ・ゲーム機器などが広く普及した今日、マウスによるスライドバー操作やタッチペンによる回答に特別な練習や教示が必要であるとは思われない。
- 5) スライドバー (またはスライダ) で回答する形式の尺度を、VAS と区別して slider scale と呼ぶこともあるが (Reips, 2012; Funke et al., 2011)、本研究ではそうした区別を行わない。なお、VAS と slider scale を区別する根拠のひとつとなるのは、slider scale では回答時にスライドバーを動かしながら納得のいくまで値の調整を行える点であると考えられる。
- 6) 回答時間に差が見られなかった点については、両形式とも1画面中の13項目を相互に比較しながらの回答が可能であったことが関係すると考えられる。
- 7) 調査様式の違い (例 : 質問紙と Web) が結果に及ぼす影響についても実証研究が行われているが、本研究では質問紙調査を並行して実施していないため、この問題は取り扱わない。
- 8) 国立情報学研究所学術情報ナビゲータ (CiNii) で文字列「visual analogue scale」と「web survey」をそれぞれ引用符で囲んで AND 検索すると、ヒット数は0件である (2014年6月20日現在)。また、同様の検索を Google Scholar で行うと、ヒット数は日本語のページのみで0件、ウェブ全体からでも97件である (2014年6月20日現在)。なお、こうした状況は「visual analog scale」で検索しても同様である。
- 9) この他にも、他変数との相関関係の比較や、再テスト信頼性の検証が必要であると考えられるが、後者については本研究のデータでは検証不能であり、別途パネル調査が必要となる。また、前者については紙幅の都合もあるため、他日を期したい。
- 10) 対人ストレス度項目は臨床場面での使用を想定したものではなく、世論調査におけるストレス関連質問を想定したものである。なお、そうした質問項目の具体例としては、時事通信社の「ストレスに関する世論調査」や、東京都生活文化スポーツ局の「健康に関する世論調査」が挙げられる。
- 11) ここでは、対人ストレス度が経済生活満足度とはまったく異なる傾向を示すという積極的な根拠がないと仮定して議論を進める。
- 12) 調査実施時には調査会社の仕様により5件法の横レイアウトが不可能であったが、現在では当該の調査会社でも5件法の横レイアウトが可能である。
- 13) 本研究で得られた結果は、現在の日本における調査会社委託のポイント制 Web 調査についてのものであり、無前提に一般化できるわけではない。この点にも注意が必要である。

引用文献

Ahearn, E.P. (1997). The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatric Re-*

- search*, 31, 569–579.
- Couper, M.P., Tourangeau, R., Conrad, F.G., & Singer, E. (2006). Evaluating the effectiveness of visual analogue scales: A web experiment. *Social Science Computer Review*, 24, 227–245.
- Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, 20, 49–58.
- Funke, F., & Reips, U.-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, 24, 310–327.
- 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.
- Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill. (ギルフォード, J.P. 秋重義治 (監訳) (1959). 精神測定法 培風館)
- Hayes, M.H., & Patterson, D.G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, 18, 98–99.
- Kerlinger, F.N. (1964). *Foundations of behavioral research: Educational and psychological inquiry*. New York: Holt Rinehart & Winston.
- Reips, U.-D. (2012). Using the Internet to collect data. In H.Cooper, P.M. Camic, R. Gonzalez, D.L. Long, A. Panter, D. Rindskopf, & K.J. Sher (Eds.), *APA Handbook of Research Methods in Psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. Washington. DC: American Psychological Association. pp.291–310.
- Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods*, 40, 699–704.
- Stubbs, R.J., Hughes, D.A., Johnstone, A.M., Rowley, E., Reid, C., Elia, M., Stratton, R., Delargy, H., King, N., & Blundell, J.E. (2000). The use of visual analogue scales to assess motivation to eat in human subjects: a review of their reliability and validity with an evaluation of new hand-held computerized systems for temporal tracking of appetite ratings. *British Journal of Nutrition*, 84, 405–415.

【Abstract】

Evaluating the Effectiveness of Visual Analogue Scales in a Web Survey

Kazunari YAMADA
Shigeru ERIKAWA

The purpose of this study was to examine the effectiveness of visual analogue scales (VASs) in a social science Web survey. For this purpose, an experimental Web survey was conducted (split method: $N=519$, $N=518$, $N=522$; Tokyo metropolitan area; March, 2010), and VASs were compared with Likert scales (LSs) in response distribution, response time, and respondent satisfaction concerning form. The results showed that (1) VASs needed slightly longer response time than LSs; however, such slight differences did not affect the Web survey adversely. In addition, (2) respondents were more satisfied with form of VASs than that of LSs. Meanwhile, (3) there could be a great difference in response distribution between the unipolar VAS and LS. Based on the above results, the possibility and limitation of VAS in Web surveys were discussed.