

Title	心理尺度の回答カテゴリに関する検討
Sub Title	Classification of psychological scales based on the response category
Author	増田, 真也(Masuda, Shinya)
Publisher	三田哲學會
Publication year	2019
Jtitle	哲學 (Philosophy). No.142 (2019. 3) ,p.245- 267
JaLC DOI	
Abstract	The response formats in self-rated psychological scales are known to influence how people respond to the question items of these scales. However, response formats are not given careful consideration when designing questionnaires. In this article, the effects of different methods of displaying options on response behaviors were reviewed; and the type of response formats in scales developed in Japan and Western countries was classified based on the number of options, order, and labels. The main results of classification types indicated the following: (a) items of scales developed in Japan are usually measured on a five-point scale, whereas those developed in Western countries use a seven-point scale; (b) the response options of nearly all the scales has an ascending order; and (c) response categories of many scales developed in Japan has the anchors "strongly disagree–strongly agree". Moreover, types of response options used in recent years tend to be homogenized. The problems of response formats in psychological assessment and issues remaining for clarification in future studies are discussed.
Notes	特集：坂上貴之教授 退職記念号#寄稿論文
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00150430-00000142-0245

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

心理尺度の回答カテゴリに関する検討

— 増 田 真 也* —

Classification of Psychological Scales Based on the Response Category

Shinya Masuda

The response formats in self-rated psychological scales are known to influence how people respond to the question items of these scales. However, response formats are not given careful consideration when designing questionnaires. In this article, the effects of different methods of displaying options on response behaviors were reviewed; and the type of response formats in scales developed in Japan and Western countries was classified based on the number of options, order, and labels. The main results of classification types indicated the following: (a) items of scales developed in Japan are usually measured on a five-point scale, whereas those developed in Western countries use a seven-point scale; (b) the response options of nearly all the scales has an ascending order; and (c) response categories of many scales developed in Japan has the anchors “strongly disagree—strongly agree”. Moreover, types of response options used in recent years tend to be homogenized. The problems of response formats in psychological assessment and issues remaining for clarification in future studies are discussed.

1. はじめに

心理学はもちろん、社会、経済、医療、福祉などの様々な研究領域で、

* 慶應義塾大学看護医療学部

自記式の心理尺度が利用されている。例えば山口（2011）は、2009年7月から2010年6月までに公刊された「パーソナリティ研究」の96.4%、「教育心理学研究」の57.9%、「心理学研究」の49.0%の論文で、心理尺度を作成したり、利用していることを報告している。また筆者が、看護学分野での国内における代表的な学術誌である「日本看護科学会誌」を調べたところ、2011年度中に公刊された実証系論文の中で、心理尺度の開発に関する論文が20.8%あり、他に何らかの心理尺度を利用した論文が23.4%を占めていた。山口（2011）は、「パーソナリティ研究とは心理尺度づくりなのだろうか」と述べているが、この嫌疑は他の研究領域にもあてはまりそうである。

心理尺度は、回答者にとっても研究者にとっても、簡便で使用しやすい。しかし、自記式であることの問題や、類似の尺度が乱立している現状から、安易な作成や使用に関する懸念が示されたり、注意が呼びかけられたりしている（村上、2007; 山口、2011）。また質問紙法での測定には、様々な反応バイアス（response bias）が含まれることが知られており、回答を過信することが危険であることも古くから指摘されてきた（小嶋、1975）。精確な研究結果を得ようとするならば、「目に見えない構成概念を測るだけに、質問紙法ほど難しい検査はない」（山崎・内田、2005）のである。しかしながら、心理尺度を含む質問紙法は、他者が直接観察できない主観的体験にアプローチする数少ない方法であり、今後も重用されるものと思われる。

質問紙法等での設問への回答は、1) 理解・解釈、2) 関連情報の想起、3) 判断の算出、4) 回答形式に合わせた判断の編集、の4段階からなるとされる（Borgers, Hox & Sikkels, 2004; De Leeuw, Hox & Huisman, 2003; Schwarz, Knäuper, Oyserman & Stich, 2008）。このとき回答者は、調査場面に含まれる様々な情報から、何が尋ねられているのかを理解したり、どのような回答が求められているのかを判断したりする（Tourangeau,

Rips & Rasinski, 2000). また回答者は回答へのモチベーションが常に高いわけではなく、上記の4段階を適切に辿らなかつたり、十分な認知的努力を伴わなかつたりする(最小限化 (satisficing); Krosnick, 1991). したがって、回答形式は最終の選択に強い影響を及ぼすものと思われる。

それにも関わらず、調査研究においては、項目文の文言の良し悪しだけに注意が向けられがちで(Rammstedt & Krebs, 2007)、尺度作成に関するテキスト(鎌原・宮下・大野木・中澤, 1998; 村上, 2007)でも、回答形式が回答に与える影響についてはほとんど触れられていない。回答形式に関するガイドラインはなく、研究者の勘や経験などの主観、紙幅の制限などから便宜的に決められているのが現状であり(Weijters, Cabooter & Schillewaert, 2010)、調査法を用いた学術論文にも、回答形式については、何件法で回答を求めたのかくらいしか記載されていない。

反応バイアスは、設問が曖昧であったり、回答が難しかったりするとき特に現れやすいと考えられており(Johnson, 1981; Podsakoff, MacKenzie, Lee & Podsakoff, 2003; Shulman, 1973), Weng & Cheng (2000) は、曖昧な項目を除くと回答形式の影響を最小にできると述べている。したがって、良い項目文が作られれば、回答形式について深慮する必要は無いのかもしれない。しかしながら、項目文の良し悪しの判定が難しいことから、質の高い回答を得るために、回答形式にもっと注意が向けられるべきであろう。

本稿では、まず心理尺度で用いられている回答形式に関する知見をまとめ、その上で、国内外で開発された心理尺度の回答形式の特徴を整理把握し、心理尺度の作成や利用に関する課題について考察する。

2. 回答形式の構成要素とその影響

まず、回答形式の構成要素について整理しておこう。ここでは、口頭での調査ではなく、紙やコンピュータ画面などに視覚的に設問が提示され、

複数の回答カテゴリの中の1つを選ぶという、多肢型の場合を想定している。したがって、回答カテゴリ数、カテゴリの並べ方、カテゴリに付与されているラベルなどから、回答形式を分類することができる。

2-1. 回答カテゴリ数

回答者にとってちょうど良い回答カテゴリがないと、回答できなかつたり、反応バイアスが生じたりする。回答カテゴリ数は「はい／いいえ」、「賛成／反対」等の2択が最少であるが、この場合は中間的な意見や態度があっても反映されない。また3択では、程度の強弱（「やや賛成」か「強く賛成か」など）が示せない。実際に、カテゴリ数が2や3のときよりも、7以上のときの方が信頼性や妥当性が高いことを示した先行研究が多い（Churchill & Peter, 1984; Preston & Colman, 2000; Weng, 2004）。しかし、6件法より4件法の方が信頼性が高い（Chang, 1994）とか、4件法と5件法に差がない（Spagna, 1984）という報告もある。また7より多くなると、各カテゴリの意味が不明確になり（Krosnick & Presser, 2010）、10以上だと信頼性が低下する（Preston & Colman, 2000）という研究結果もある。

Weathers, Sharma & Niedrich (2005) は、回答カテゴリ数が多いとカテゴリ間の差が小さくなって選択が難しくなるため、現状維持 (Status quo) ヒューリスティックによって、複数の設問で同じ回答カテゴリを選ぶようになると考えた。そして一部の回答カテゴリばかり選ばれる結果、信頼性係数の値は高まるものの、回答の正確さ（構成概念妥当性）は低くなることを見出した。このように、回答カテゴリ数は一概に多ければ良いというわけではない。

なおこの場合の回答カテゴリは順序尺度を構成するが、間隔尺度水準以上を想定している統計分析を適用することも多い。しかし、5件法以上であればピアソンの積率相関を算出しても、大きな問題はないという（萩生

田・繁樹, 1996). 以上のような議論から, 一般にカテゴリ数は5以上が良いとされ (Matell & Jacoby, 1971, 1972; 小島, 1975; Preston & Colman, 2000; Weng, 2004), Krosnick & Presser (2010) は, 多くの場合で7件法が最適であるという結論を下している.

しかしながら, Weathers, Sharma & Niedrich (2005) は, 適切なカテゴリ数は回答者の特性によって異なると述べている. また, 測定しようとする概念や尺度の使用目的によっても, 何が適切であるかの基準は異なるであろう. 脇田 (2004) は, 項目反応理論を利用して, 自尊感情尺度のポジティブな表現からなる項目と逆転項目であるネガティブな項目とでは, カテゴリ間の心理的距離が異なることを示した. 具体的には, ネガティブな項目では回答者はカテゴリを比較的適切に弁別しているが, ポジティブな項目では様々な潜在特性を有する人が中間カテゴリを選択し, カテゴリ間の心理的距離が広がったり狭かったりしたのである. 心理的距離が著しく小さい場合, 両方の回答カテゴリを呈示する意味が無いと考えられることから, カテゴリ数を減らす方が適切であると思われる.

なお, 「どちらともいえない」など, 複数の回答カテゴリの中で意味的に中立となる, もしくは視覚上で中央に位置する中間カテゴリが, 過度に選ばれることが知られている (Bishop, 1987; Schuman & Presser, 1981). そのため, カテゴリ数が奇数かどうかの違いも重要である. 中間カテゴリは, 必ずしも中間の意見や態度を有しているからではなく, 「わからない」, 「答えたくない」といった理由から選ばれることがあり (Kulas, Stachowski & Haynes, 2008 他), 解釈や分析がしにくくなることから, これを設けないほうが良いと考える研究者もいる (小松, 2005). さらに中間カテゴリは選ばれやすいので, 回答分布や平均点の高低, 散らばりの大きさに影響を及ぼすことがある (Baumgartner & Steenkamp, 2001; 増田・坂上・北岡・佐々木, 2016; Masuda, Sakagami, Kawabata, Kijima & Hoshino, 2017; Paulhus, 1991).

しかし、中間カテゴリがないと無理に意見を作り出してしまうことがある（De Vaus, 2002）、ある方が妥当性・信頼性が高いという（Buchanan & Henderson, 1992; Saris & Gallhofer, 2007）。例えば Herbert（1971）は、実在しない架空の集団に対しても偏見が見られることを示した研究は、中間カテゴリを含まない態度尺度を用いたことによる artifact であると述べている。このように、中間カテゴリが回答結果に与える影響は必ずしもはっきりせず、中間カテゴリがある方が良いかどうかに関して、一般的な合意は得られていない（Choi & Pak, 2005）。中間カテゴリに関する研究については、増田・坂上（2014）を参照されたい。

2-2. 回答カテゴリのラベル

回答カテゴリには、その意味や順序関係を示す「あてはまる-あてはまらない」、「思う-思わない」などのラベルが付与される。各回答カテゴリの意味を言語で示すラベル（言語ラベル）は、すべてのカテゴリに付されていることもあれば、両端のカテゴリだけとか、両端と中間だけという場合もある。言語ラベルの有無に関しては、全部のカテゴリに言語ラベルがある方が信頼性が高い（Alwin & Krosnick, 1991; Krosnick & Berent, 1993; Weng, 2004）という結果が多いが、無関係という報告もある（Churchill & Peter, 1984）。

また Klockars & Yamagishi（1988）では、両端にのみ言語ラベルがあつて、他は数字だけでカテゴリの違いが示されているときに、回答者はカテゴリ間を最も等間隔と見なすことを示した。しかし、脇田（2012）では、5件法で両端と中間にだけ言語ラベルがあるときに、最も等間隔と見なされていた。Weijters et al.（2010）は、平均やパーセンテージを求めるような場合には、尺度の意味がわかりやすくなることから、全カテゴリに言語ラベルをつけることを勧めている。一方、変数間の関係や線形モデルの推定をするときには、両端にだけ言語ラベルがある方が望ましいと述べて

いる。

回答カテゴリの数値ラベルは、本来、カテゴリ間の順序関係を反映しているだけであるが、その数値の印象で回答が変わることがある。例えば、自分の人生が成功か否かを11段階で尋ねる設問で、「0-10」の数値が示されている場合と、「-5+5」が示されている場合とが比べられたところ、負の数値ラベルが用いられた回答カテゴリの選択率が低下した。これは、負の数値が強い失敗を示すものと解釈され、選択されにくくなったためだとされる (Schwarz, Knauper, Hippler, Noelle-Neumann & Clark, 1991)。また数値ラベルの最小値として、「1」を用いるときよりも、「0」を用いるときの方が、「まれに」の意味を低く捉え、評定値平均が高くなった (Schwarz, Grayson & Knauper, 1998)。

2-3. 回答カテゴリの並べ方：配置の視覚的な問題

各回答カテゴリは、縦 (vertical) に並べられる場合と、横 (horizontal) に並べられる場合とがある。Toepoel, Das & Soest (2009) では、縦よりも横に並べたときの方が得点が高くなったが、住本 (2008) では差は見られなかった。Toepoel et al. (2009) は、横に並べたときの方が、回答の際に他の要因の影響を受けにくいと述べている。

さらに、回答カテゴリは横に1行で配置される (linear) ことが多いが、カテゴリ数が多いと、複数行にわたって並べられる (nonlinear) ことがある。Christian & Dillman (2004) は、選択肢が複数行にわたるとき、最初の行だけ読んで回答することがあるため、同じ回答カテゴリでも最初の行に配置したときの方が選択されやすいと予想し、2問中1問でそれが支持された。

Tourangeau, Couper & Conrad (2004) では、「女性の就業機会を男性と同等にするために、連邦政府がどれくらい取り組んでいると思うか」といった項目で、「非常に少ない」から「非常に多い」までの5件法に、「わ

からない」と「意見がない」という選択肢を加えて回答を求めた。ただし、ある回答フォーマットでは計7つの選択肢が等間隔で並べられていたが、別のフォーマットでは5つの回答カテゴリと、「わからない (Don't know)」「意見がない (no opinion)」という選択肢の間に線が引かれていたり、広いスペースがあったりして、見かけ上の中央に位置する選択肢が異なっていた。結果は、同じ選択肢であっても、中央にあるように見えるときの方が、そうでないときよりも多く選ばれていた。Tourangeau et al., (2004) によると、こうした差が生じる理由の1つは、項目への回答の際に「中間が典型」(Middle means typical.) というヒューリスティックが働くためであるという。

2-4. 回答カテゴリの並べ方：昇順か降順か

ある特性に関して、そのあてはまりの程度を一方向で尋ねる尺度を単極尺度 (unipolar scale) と呼ぶ。そして正反対の特性を回答カテゴリの両端に配置して (例えば、賛成-反対)、どちらにより近いかを尋ねる尺度を両極尺度 (bipolar scale) と呼ぶ。Schaeffer & Presser (2003) によると、単極尺度では選択カテゴリの一方の端に0か1の数値が割り当てられていることが多く、回答者は右端か左端を原点として項目内容に当てはまる程度を検討する。一方、両極尺度では中間カテゴリが原点となり、回答者はまず中央からどちらかの方向を定め、次にその程度を見積もるという。

後述するように、心理尺度においては単極尺度が用いられることが多いが、問われている内容に対して、その程度が最も小さかったり、頻度が低かったりするようなラベルが最初の回答カテゴリに付与されていて、次第に大きくなったり、多くなったりする方向で並べられているような配置をここでは昇順と呼ぶ。そして、逆の順に並べられているときを降順と呼ぶ。この昇降順は、言語ラベルと数値ラベルの両方で考えることができる。すなわち、言語ラベルに関しては、設問に対して否定から肯定 (例、「あてはまらない-あてはまる」) の方向で並べられるときが昇順で、逆の

場合（「あてはまる-あてはまらない」）が降順である。数値ラベルについては、最小値が最初にあって最大値が最後にある（例えば1-5）場合が昇順、逆の場合（例えば5-1）が降順である。

自記式調査のように視覚的に設問が示されるときには、初頭効果（primacy effect）が生じることがある。そのため、回答カテゴリが横並びで呈示されている場合は、肯定的な意味を持つカテゴリ（strongly agree など）が左にあるとき（降順）の方が、ポジティブな評価を得やすい（Albanese, Prucha, Barnet & Gjerde, 1997; Chan, 1991; Hartley & Betts, 2010; Krebs & Hoffmeyer-Zlotnik, 2010; Nicholls, Orr, Okubo & Loftus, 2006; 住本, 2008）。ただし、必ずしも全項目で初頭効果が生じているというわけではないし、Johnson (1981), Power, Morrow, Goudy & Keith (1977), Weng & Cheng (2000) では昇順と降順の差はなかった。また、Toepoel et al. (2009) では、縦並びのときにむしろ下方に位置する回答カテゴリが選ばれやすかった。選択率以外では、昇順か降順かで項目間相関係数が異なっていたが、潜在因子の次元には影響が無かった（Krosnick & Alwin, 1987）。

初頭効果は、回答者が必ずしも最適なカテゴリを選ぶとは限らず、最初の受け入れ可能な選択肢を選ぶという、最小限化原理に従うために生じると考えられている（Krosnick, 1991）。したがって、回答カテゴリ数が多く、カテゴリ間の意味の差が小さいと、初頭効果が生じやすいものと思われる。

2-5. 言語ラベルと数値ラベルの昇降順の一致性

回答カテゴリに、言語ラベルと数値ラベルの両方が用いられることも多い。Rammstedt & Krebs (2007) は、数値ラベルの最小値がネガティブな極（例えば「1. 全く賛成でない」）に、最大値がポジティブな極（例えば「5. 非常に賛成」）に割り当てられる方が直感的に理解しやすいので、一貫した判断がなされると考えた。そして、同一対象者に Big five 尺度へ

の回答を2時点で求めた。

時点1では、全回答者が言語と数値の昇降の一致したフォーマットで回答した。すなわち、否定的な回答カテゴリに1、肯定的な回答カテゴリに8が付されていて、かつ1から8の昇順で並べられていた(1.Strongly disagree-8. Strongly agree)。そして約3週間後に、条件1では同じく昇降が一致しているが、8から1の降順で並べられたフォーマットで回答した(8. Strongly agree-1.Strongly disagree)。一方、条件2では数値ラベルは昇順であるが、言語ラベルが降順という不一致なフォーマットであった(1. Strongly agree-8.Strongly disagree)。その結果、条件1では時点間の平均、標準偏差、相関係数のパターンに違いはなかったが、条件2では2つの下位尺度で標準偏差が大きく、平均値が低くなった。また時点1と時点2での同じ変数間の相関が、条件1より低く、信頼性が低かった。

回答形式が異なると、項目内容が同じであっても回答が変わってしまうことがある。例えば、「あなたにとって音楽はどれくらい重要か」といった項目で「1.とても重要」から「5.全く重要ではない」の選択肢が提示されていて、その中の1つを選ぶ場合(Polar Point Scale)より、空欄に1-5の数値のいずれかの記入を求める場合(Number Box)の方が、平均得点が高くなることが多くの研究で報告されている。しかしながら、「重要でない」の方向で、数値ラベルが大きくなるような選択肢は直観的に理解しづらく、ラベルを目にしながらかける Polar Point Scale よりも、特に Number Box のときに誤った回答をしやすと思われる。逆に、数値の大小と重要性の評価の方向が一致しているときには、どちらの回答形式でもミスが少ないだろう。実際、Heerwegh (2011) が選択肢の言語ラベルと数値ラベルを入れ替えて、「1.全く重要ではない -5.とても重要」として回答を求めたところ、Polar Point Scale と Number Box 間の平均得点の差は見られなくなった。このように、言語ラベルと数値ラベルの昇降順を一致させることは、適切な回答を得る上で重要であると思われる。

3. 国内外で開発された心理尺度の回答形式の特徴

次に、国内外で開発された心理尺度において、どのような回答形式が用いられているのかを調べた。本稿では、「心理測定尺度集」(サイエンス社、以下尺度集と呼ぶ)に収められている255尺度の回答形式について集計した。これは全6巻からなっており、1990-1999年に公刊された159尺度が、I巻「自己・個人内過程」(堀監修、山本編、2000)、II巻「対人関係・価値観」(堀監修、吉田編、2001)、III巻「適応・臨床」(堀監修、松井、2001)に分けて収められている。さらに2000年以降に公刊された96尺度に関して、V巻「自己・対人関係・価値観」(堀監修、吉田・宮本編、2011)、VI巻「集団・組織・適応」(堀監修、松井・宮本、2011)が追加されている。なおIV巻は児童用であることから、今回は分析対象から除いた。尺度集に含まれていないものもあると思われるが、国内において、最も網羅的に心理尺度が収集されており、また項目文だけでなく、回答形式も記載されていることから、本研究において最適な資料であると判断した。

尺度集に含まれている尺度の回答形式の多くは、各設問に対して回答カテゴリを示す数値や線分等が横並びになっていて、いずれか1つをチェックするものであった(237, 92.9%)。その他に、数値・記号を空欄に直接記入するものや、Visual Analogue Scale、縦並び、項目ごとに異なった回答形式となっている、といった尺度が18あったが、これらについては、検討対象としなかった。

欧米文献で発表された尺度については、Handbook of Marketing Scales (Bearden, Netemeyer & Kelly, 2011, 以下 Handbook と呼ぶ)に収録されている187尺度を検討した。この本には、マーケティング、購買、消費行動、組織環境健康娯楽行動、性格特性、認知傾向、感情など、幅広い尺度が掲載されている。年代別に見ると、1980年代以前に発表された63尺度、1990年代の64尺度、2000年以降の52尺度が収録されている。

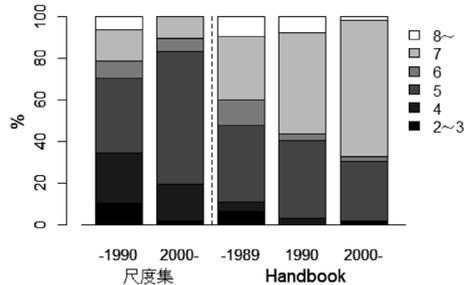


図1 発表年別に見た心理尺度の回答カテゴリ数

尺度集と Handbook では、尺度の発表年代や対象とする研究領域が完全に同じというわけではない。また Handbook には、尺度集ほど詳細に回答形式が記されていないなど、留意すべき点もある。そのため、あくまでも参考として比較を試みる。

3-1. 回答カテゴリ数

図1に発表年別に見た、回答カテゴリ数の割合を示す。尺度集、Handbook のいずれも、回答カテゴリ数は最少2、最多11であった。また全体的に5件法と7件法が多かった。特に尺度集では5件法が多く、2000年以降、さらに増大していた。一方 Handbook は7件法の増加傾向が見られる。また尺度集の方がカテゴリ数4以下が多く、8以上が少なかった。すなわち、欧米で作成された尺度の方が、回答カテゴリ数が多い傾向があった。カテゴリ数が奇数であったのは、尺度集で63.1%、Handbook で88.8%であり、多くの尺度で中間カテゴリが含まれていた。

3-2. 回答カテゴリのラベルと昇降順

尺度集の237尺度の全てで、何らかの言語ラベルが付されていた。ただし、2尺度が両極尺度であったため、これらを除いた235尺度で言語ラベルに関する割合を算出した。また回答カテゴリに数値ラベルを用いているのは210で、その他は空欄等にチェックを求めるといったものであった。Handbook に関しては、27尺度でラベルに関する詳しい記載が無かった。

そこで、これらを除いた160尺度を分母として割合を算出した。

表1に結果を示す。まず、言語ラベル、数値ラベルともに昇順が多かった。また年代の経過に従って増加していた。尺度集では1999年までは降順の言語ラベルが多かったが、2000年以降では昇順が多かった。すなわち、近年作成された尺度のほとんどが、言語ラベル、数値ラベル共に昇順で並べられていた。言語ラベルと数値ラベルの昇降順は、ほとんどで一致していた。特に尺度集では、2000年以降に公刊されたすべての尺度で、昇降順が一致していた。

なお数値ラベルの最小値は、尺度集の94.3%、Handbookの97.3%が1で、残りが0であった。すなわち、ほとんどの数値ラベルが1-5, 1-7となっており、同じ5件法であっても0-4が用いられることはまれであった。

次に尺度集では、両端だけでなく、すべての回答カテゴリに言語ラベルがあることがほとんど(93.7%)で、両端だけ(4.6%)や、両端と中間の3箇所(1.7%)という尺度はわずかであった。Handbookにおいては、この点について記載されていないことが多く、検討することができなかった。

表1 回答カテゴリのラベルの昇降順 (%)

		～1989	1990～1999	2000～
言語ラベルが昇順	Handbook	55.6	73.7	91.7
	尺度集		45.4	76.0
数値ラベルが昇順	Handbook	58.8	89.5	100.0
	尺度集		62.5	77.8
言語ラベルと数値ラベルの昇降順が一致している割合	Handbook	87.5	78.9	96.3
	尺度集		90.0	100.0
数値ラベルの最小値が1	Handbook	95.5	98.5	98.1
	尺度集		90.0	100.0

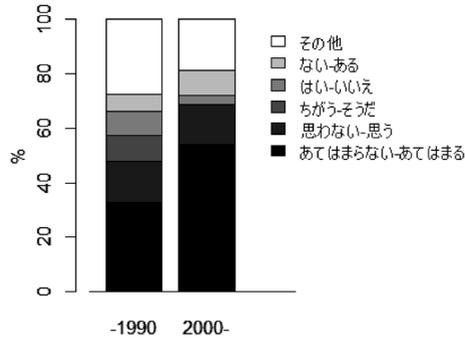


図2 尺度集における回答カテゴリの言語ラベル

尺度集に関して、用いられている言語ラベルの文言の割合を算出したところ、「あてはまらない-あてはまる」が最も多く（43.8%）、次に「思わない-思う」（16.0%）であった（図2）。さらに年代別に見ると、「あてはまらない-あてはまる」が2000年以降に増えていた（～1999年：32.7%、2000年～：54.2%）。中間カテゴリに関しては、95.4%で何らかの言語ラベルが記されており、その中で「どちらともいえない」が66.3%、「どちらでもない」が13.1%を占めていた。

4. おわりに

欧米で開発された尺度は7件法が多く、日本では5件法が多かった。これらのほとんどで、言語ラベルと数値ラベルの昇降順は一致していた。さらに、2000年以降に作られた尺度の形式はバリエーションが小さかったが、特に国内においては、単極の5件法、昇順、数字ラベルの最小値が1、言語ラベルが「あてはまらない-あてはまる」で昇降順は一致しているという回答形式が典型となっており、均一化していく傾向が見られた。

回答形式に関する先行研究からは、5件法以上で言語ラベルと数値ラベルの昇降順が一致していることが良いという。であるならば、近年作成されている尺度の多くが、望ましい特徴を備えていると考えて良いのかもし

れない。一方で、回答形式が均一化しているのは、尺度作成の手順が形式的になっていて、十分に検討されていないことを反映している可能性もある。

さらに、回答形式が同じだと、複数の回答カテゴリの中から同じカテゴリばかりを選ぶという同一回答 (straight line response もしくは non-differential response) が生じ、項目間の関連が強くなることが懸念されている (Krosnick, 1991)。実際、Rammstedt & Krebs (2007) の縦断調査では、言語と数値の昇降順が一致しているかどうかで、同一対象者による2時点で測定された同じ尺度間の相関が変わっていた。したがって横断調査でも、例えば2つの尺度の昇降順が同じときと違うときとで、相関の強さが異なる可能性もある。回答形式の均一化が、尺度間相関などを不当に高めているおそれがあることには留意すべきであろう。

先述したように、心理尺度を用いた論文においては、回答形式に関する情報は何件法なのか程度しか記されていないことが多い。しかし、少なくともラベルの昇降順については明記すべきと思われる。また例えば、ラベルについて触れず、評定段階数だけが記されていて、尺度得点の平均値が算出されている論文を散見するが、同じ5件法でも数値ラベルが「0-4」なのか「1-5」なのかでは、当然、平均値の意味が異なる。さらに数字ラベルの最小値が「0」のときよりも「1」のときの方が、多く選択されてしまう可能性がある (Schwarz et al., 1998)。数値ラベルの最小値が「0」であるような尺度はほとんど見られないが、基本的な情報として、論文には数値ラベルについても記載すべきであろう。

最後に、本稿で十分に取り上げることができなかった課題について述べる。

(1) 単極尺度と両極尺度

黙従傾向を抑え、内容的妥当性を高めるために、尺度に逆転項目を含めることが望ましいとされる (Paulhus, 1991)。しかし Friborg,

Martinussen & Rosenvinge (2006) によると、ポジティブな特性の否定は直感的に不自然なので、逆転項目を作成するのが難しいという。そして、両極尺度で尋ねると、構造方程式モデリングによる適合度が高くなったことから、ポジティブな概念を測定するときには、Likert 尺度のような単極尺度よりも、SD 法のような両極尺度の方が適切であると述べた。尺度集や Handbook を見ると、両極尺度は少ないが、こうした回答形式の適用を検討すべき場合があるものと思われる。

(2) 言語ラベルの違い

両端の回答カテゴリの言語ラベルに関して、「非常にあてはまる-全くあてはまらない」のように、「非常に」、「常に」、「全く」、「よく」などの修飾語が加わることがある。小野寺 (2002) によると、こうした修飾語の有無や違いによっても回答結果が変わる。

また中間カテゴリは最小限化から選ばれたり (Krosnick, 1991)、判断の目安になったりする (Tourangeau et al., 2004) が、言語ラベルの有無や違いによる影響があることが示されている (Klopfer & Madden, 1980)。増田 (2009) では、言語ラベルがあるときの方が、無いときよりも中間カテゴリがよく選ばれていた。

さらに尺度集については、回答カテゴリの言語ラベルとして「あてはまらない-あてはまる」が最も多く、言語ラベルがある中で 43.8% を占めていたが、同じ内容を尋ねる場合でも、異なる言語ラベルを用いることができる場合もあろう。したがって、言語ラベルの有無や文言の違い、修飾語の有無やその違いによる影響も、さらに検討すべきである。

(3) 項目数や継時的に回答することの問題

心理尺度には複数の項目が含まれていることから、項目数や項目の並べられ方も回答形式の問題と関連するだろう。例えば、一般的に 5 件法や 7 件法が望ましいとしても、項目数が非常に多いときには、もっとカテゴリ数が少ない方が回答者の負担にならず、適切な回答が得られるかもしれな

い。

また増田・坂上・北岡（2017）では、5因子性格検査の50項目を、回答者ごとにランダムに呈示することで項目ごとの差異を相殺しても、各回答カテゴリの順番ごとの選択率は同じにならず、後半では中間選択が増大し、「2.ちがう」や「4.そうだ」といった弱い選好を示す回答カテゴリが減少した。項目の位置に依存して回答が変わるのであれば、研究結果の再現は項目が同じ順番で並んでいたことを反映したにすぎないかもしれないし、逆に結果の違いが生じたとしても、それは単に順番が異なっていたためかもしれない。本稿で見てきたように、多くの心理尺度の項目で中間カテゴリがあるが、中間カテゴリが無いときに回答の変化があるかどうかは確認されていない。そのため、継時的な回答という観点から見たときに、何が適切な回答形式なのかが変わってくる可能性がある。

（4）視覚的要因に関して

Web調査が盛況になるのにつれて、調査票のデザインにおける視覚的側面（色、大きさ、矢印などの記号、スペースの大小、イラストの使用の有無など）の重要性も注目されるようになった（Christian & Dillman, 2004; Schaeffer & Dykema, 2011; Tourangeau, Couper & Conrad, 2004, 2007）。特に、回答者が何を尋ねられているのか、どう答えれば良いのかわからないとき、非言語的なコンテキストから情報を得ようとするという（Christian & Dillman, 2004）。Web調査は、他の調査法よりも最小限化が促進されやすいという指摘がある（Heerwegh & Loosveldt, 2008）が、そうであればこそ、Web調査の特徴を活かすことで、回答の質を上げるための努力をしていかなければならない。回答形式に関しても、様々な工夫の余地があることから、今後こうしたテーマの研究が増大していくものと思われる。

引用文献

- Albanese, M., Prucha, C., Barnet, J. H., & Gjerde, C. L.(1997). The effect of right or left placement of the positive response on Likert-type scales used by medical students for rating instruction. *Academic Medicine*, 72, 627-630.
- Alwin, D. F., & Krosnick, J. A.(1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods & Research*, 20, 139-181.
- Baumgartner, H., & Steenkamp, J.(2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Bearden, W. O., Netemeyer, R. G., & Kelly, L. H.(2011). *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. SAGE.
- Bishop, G. F.(1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 210-232.
- Borgers, N., Hox, J., & Sikkkel, D.(2004). Response effects in surveys on children and adolescents: The Effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity*, 38, 17-33.
- Buchanan, B., & Henderson, P.(1992). Assessing the bias of preference, detection, and identification measures of discrimination ability in product design. *Marketing Science*, 11, 64-75.
- Chan, J.(1991). Response-order effect in Likert-type scales. *Educational and Psychological Measurement*, 51, 531-540.
- Chang, L.(1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18, 205-215.
- Choi, B. C. K., & Pak, A. W. P.(2005). A catalog of biases in questionnaires. *Preventing Chronic Disease*, 2, 1-13.
- Christian, L.M. & Dillman, D.A.(2004). The influence of graphical and symbolic language manipulations to self-administered questions. *Public Opinion Quarterly*, 68, 57-80.
- Churchill, G. A., & Peter, J. P.(1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21, 360-375.
- De Leeuw, E., Hox, J. & Huisman, M.(2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19, 153-176.

- De Vaus, D.(2002). *Surveys in social research*. Routledge.
- Friborg, O., Martinussen, M., & Rosenvinge, J. H.(2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*, 873-884.
- 萩生田伸子・繁榊算男 (1996). 順序付きカテゴリカルデータへの因子分析の適用に関するいくつかの注意点, *心理学研究, 67*, 1-8.
- Hartley, J., & Betts, L. R.(2010). Four layouts and a finding: The effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology, 13*, 17-27.
- Heerwegh, D.(2011). Internet survey paradata. In Das, M., Ester, P., & Kaczmarek, L.(Eds.)*Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*(pp. 325-348), Routledge.
- 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.
- Herbert, C. F.(1971). Fictitious groups and the generality of prejudice: An artifact of scales without neutral categories. *Psychological Reports, 29*, 359-365.
- 堀洋道監修, 松井豊編 (2001). 心理測定尺度集 III—心の健康をはかる〈適応とライフイベント〉, サイエンス社.
- 堀洋道監修, 松井豊・宮本聡介編 (2011). 心理測定尺度集 VI—現実社会とかわる〈集団・組織・適応〉, サイエンス社.
- 堀洋道監修, 山本真理子編 (2001). 心理測定尺度集 I—人間の内面を探る〈自己・個人内過程〉, サイエンス社.
- 堀洋道監修, 吉田富二雄編 (2001). 心理測定尺度集 II—人間と社会のつながりをとらえる〈対人関係・価値観〉, サイエンス社.
- 堀洋道監修, 吉田富二雄・宮本聡介編 (2011). 心理測定尺度集 V—個人から社会へ〈自己・対人関係・価値観〉, サイエンス社.
- Johnson, J.(1981). Effects of the order of presentation of evaluative dimensions for bipolar scales in four societies. *The Journal of Social Psychology, 113*, 21-27.
- 鎌原雅彦・宮下一博・大野木裕明・宮澤潤 (1998). 心理学マニュアル質問紙法. 北大路書房.
- Klockars A. J., & Yamagishi M.(1988). The influence of labels and positions in

- rating scales. *Journal of Educational Measurement*, 25, 85-96.
- Klopfert, F. J., & Madden, T. M.(1980). The middlemost choice on attitude items ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin*, 6, 97-101.
- 小嶋外弘 (1975). 質問紙調査法の技法に関する検討. 村上英治編 心理学研究法 9 質問紙調査法 (pp. 224-270). 東京大学出版会.
- 小松洋 (2005). 中間選択肢の扱い. 大谷信介・木下栄二・後藤範章・小松洋 (編) 社会調査へのアプローチ—論理と方法第2版 (pp. 111). ミネルヴァ書房.
- Krebs, D., & Hoffmeyer-Zlotnik, J. H. P.(2010). Positive first or negative first? Effects of the order of answering categories on response behavior. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 118-127.
- Krosnick, J.(1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A., & Alwin, D. F.(1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201-219.
- Krosnick, J., & Berent, M.(1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, 37, 941-964.
- Krosnick, J., & Presser, S.(2010). Question and questionnaire design. In P. V. Marsden, J. D. Wright(Eds.), *Handbook of Survey Research*. 2nd(pp. 263-313). Bingley, UK: Emerald Publishing Group.
- Kulas, J. T., Stachowski, A. A., & Haynes, B. A.(2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, 22, 251-259.
- 増田真也 (2009). 選択肢に付与された数値が回答に与える影響—中間選択肢との関係. 日本社会心理学会第50回大会・日本グループダイナミクス学会第56回大会合同大会発表論文集, 396-397.
- 増田真也・坂上貴之 (2014). 調査の回答における中間選択—原因, 影響とその対策一. 心理学評論, 57, 472-494.
- Masuda, S. Sakagami, T., Kawabata, H., Kijima, N. & Hoshino, T.(2017). Respondents with low motivation tend to choose middle category: Survey questions on happiness in Japan. *Behaviormetrica*, 44, 593-605.
- 増田真也・坂上貴之・北岡和代・佐々木恵 (2016). 回答指示の非遵守と反応バイ

- アスの関連. 心理学研究, 87, 354-363.
- 増田真也・坂上貴之・北岡和代 (2017). 多くの項目に回答することによる中間選択の増加. 行動計量学, 44, 117-128.
- Matell, M. S., & Jacoby, J.(1971). Is there an optimal number of alternatives for Likert scale items? Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674.
- Matell, M. S., & Jacoby, J.(1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56, 506-509.
- 村上宣寛 (2007). 心理尺度のつくり方. 北大路書房
- Nicholls, M., Orr, C., Okubo, M., & Loftus, A.(2006). Satisfaction guaranteed the effect of spatial biases on responses to Likert scales. *Psychological Science*, 17, 1027-1028.
- 小野寺典子 (2002). 「非常に」と「かなり」で異なる回答. 放送研究と調査, 52, 62-75.
- Paulhus, D. L.(1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman(Eds.), *Measures of Personality and Social Psychological Attitudes*(pp. 17-59). San Diego, CA: Academic Press.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P.(2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- Powers, E., Morrow, P., Goudy, W., & Keith, P.(1977). Serial order preference in survey research. *Public Opinion Quarterly*, 41, 80-85.
- Preston, C. C., & Colman, A. M.(2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.
- Rammstedt, B., & Krebs, D.(2007). Does response scale format affect the answering of personality scales? *European Journal of Psychological Assessment*, 23, 32-38.
- Saris, W., & Gallhofer, I.(2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1, 29-43.
- Schaeffer, N. C., & Dykema, J.(2011). Questions for surveys: Current trends and future directions. *Public Opinion Quarterly*, 75, 909-961.
- Schaeffer, N. C., & Presser, S.(2003). The Science of asking questions. *Annual*

- Review of Sociology*, 29, 65-88.
- Schuman, H., & Presser, S.(1981). *Questions and Answers in Attitude Surveys*. Academic Press New York.
- Schwarz, N., Grayson, C. E., & Knauper, B.(1998). Formal features of rating scales and the interpretation of question meaning. *International Journal of Public Opinion Research*, 10, 177-183.
- Schwarz, N., Knauper, 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.
- Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C.(2008). The psychology of asking questions. In E. D. de Leeuw, J. Hox, & D. Dillman(Eds.), *The International Handbook of Survey Methodology*(Vol. 22, pp. 18-34). New York: Lawrence Erlbaum Associates.
- Shulman, A.(1973). A Comparison of two scales on extremity response bias. *Public Opinion Quarterly*, 37, 407-412.
- Spagna, G.(1984). Questionnaires: Which approach do you use? *Journal of Advertising Research*, 24, 67-70.
- 住本隆 (2008). 調査票の質問項目選択肢の表示方式が回答行動に及ぼす影響—インターネットによる実験調査からみた事例紹介—. *行動計量学*, 35, 161-176.
- Toepoel, V., Das, M., & Soest, A. V.(2009). Design of web questionnaires: The effect of layout in rating scales. *Journal of Official Statistics*, 25, 1-35.
- Tourangeau, R., Couper, M. P., & Conrad, F.(2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68, 368-393.
- Tourangeau, R., Couper, M. P., & Conrad, F.(2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, 71, 91-112.
- Tourangeau, R., Rips, L. J., & Rasinski, K.(2000). *The Psychology of Survey Response*. Cambridge University Press.
- 脇田貴文 (2004). 評定尺度法におけるカテゴリ間の間隔について—項目反応モデルを用いた評価方法—. *心理学研究*, 75, 331-338.
- 脇田貴文 (2012). Likert法における回答選択肢のレイアウトが選択肢間の心理的距離に与える影響. *関西大学社会学部紀要*, 43, 135-144.
- Weathers, D., Sharma, S., & Niedrich, R. W.(2005). The impact of the number of scale points, dispositional factors, and the status quo decision heuristic on

- scale reliability and response accuracy. *Journal of Business Research*, 58, 1516-1524.
- Weijters, B., Cabooter, E., & Schillewaert, N.(2010). The effect of rating scale format on response styles: The number of response categories and response. *International Journal of Research in Marketing*, 27, 236-247.
- Weng, L. J.(2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972.
- Weng, L. J., & Cheng, C. P.(2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement*, 60, 908-924.
- 山口陽弘 (2011). 心理尺度に着目したこの一年の概観—パーソナリティ研究とは「心理尺度づくり」なのだろうか?—, 教育心理学年報, 50, 97-107.
- 山崎勝之・内田香奈子 (2005). 調査研究における質問紙の作成過程と適用上の諸問題. 鳴門教育大学研究紀要 (教育科学編), 20, 1-10.