

## 資 料

### Likert 法における適応型選択肢の提案と基礎的検討

脇 田 貴 文 ・ 藤 岡 慧

### A proposal and basic examination of adaptive options when using the Likert method

Takafumi WAKITA and Satoshi FUJIOKA

#### Abstract

The Likert method presents certain problems, such as the distribution bias caused by phenomena related to social desirability, central tendency, etc. One conceivable solution to this is to correct the distribution by changing the expression of the option's anchor. In this study, we propose "adaptive options when using the Likert method" designed to change the option set according to the respondent's latent level. We compare two sets of options. In terms of the anchor expression, we used one expression involving a symmetric option (ver. A) and another that increases the number of options in the positive expression (ver. B). As a result, ver. B showed that the distribution of the scale score was corrected to a closed, normal distribution. Furthermore, we also obtained a reasonable result in terms of the psychological distance of options. These results suggest the possibility of manipulating the distribution of the scale score by manipulating the anchor expression, while the possibility of the practical use of "adaptive options" was also demonstrated.

Keywords: Likert method, anchor expression, item response theory

#### 抄 録

本研究では、Likert 法における社会的望ましさなどの影響による回答分布の偏りを解決する方法として、回答選択肢の評定尺度表現を変化させる適応型選択肢の考え方を提案した。また、その実現に向けて、同じ尺度に対して異なる 2 種類の評定尺度表現を用いることで回答分布等がどのように変化するかを検討した。その結果、評定尺度表現を変化させることにより、回答分布が変わることが確認された。

キーワード：Likert 法、評定尺度表現、項目反応理論

## 問題と目的

心理学では、扱う構成概念を測定するために、回答者に複数の項目と1セットの選択肢を提示して回答を求めるLikert法を用いた心理尺度を用いることが多い。Likert法は簡便な方法であるため、非常に多くの研究で用いられており、その結果をもとに多くの知見が得られている。しかし、Likert法にも様々な問題が指摘されている。例えば、両端の選択肢のみを使用して反応する極端反応、中間選択肢に回答が固まる中間反応（増田，2019）、社会的望ましさの影響、天井効果や床効果の問題などである。これらの現象は、研究において重要な「個人差の測定」という点で、ネガティブなものといえる。

本報告において提案する適応型選択肢は、その中でも天井効果や床効果、社会的望ましさの影響に対して改善できる可能性がある。はじめに、天井効果や床効果は、項目内容の難易度が高い（もしくは低い）ことにより、回答者の選択が一方に偏る現象である。これにより、測定値の分散が小さくなるため測定において望ましい現象とはいえない。具体的には、「私は日本語を話すことができる」という項目に対して、日本人であればほとんどの人が「あてはまる」という回答をする。一方で、「私はタイ語を話すことができる」という項目であれば、ほとんどの人が「あてはまらない」と答えるだろう。つまりこのような項目を日本人に提示し、回答を求めることはあまり意味がない。ここでは極端な例を提示したが、心理尺度においても「私は不安になることはない」（難易度は低い）や、「私は誰とでも仲良くなれる」（難易度は高い）という項目では、回答が偏ることがある。

社会的望ましさは、回答者の回答が「社会的に望ましい」方向に偏る傾向である。一般に、パーソナリティ検査に代表される典型値測定は、社会的望ましさの影響を受けやすいとされている（豊田・川端・松下，2005）。例えば、「信号無視はしてはいけない」という項目には、項目内容が刺激となって、実際の行動とは関係なく、多くの人が「してはいけない」という否定的な反応をするだろう。Edwards（1953）は、EPPSの項目内容の社会的望ましさと承認率（その項目に対して「はい」と答える人数の割合）には正の相関があることを示している。Likert法においても、肯定的な内容の項目に対して「ややあてはまる」や「あてはまる」といった回答が多くなると考えられる。

また学校現場で行う児童生徒が回答する調査では、教師が望んでいるだろうと思われる方向に回答が偏ることがある。例えば、「英語の授業は楽しい」という項目に対しては、実際がどうかはさておき、「あてはまる」という肯定的な反応が増加する。その他にも匿名性が保証されない場合や検査の得点が可否に関わる就職試験といった場合などのハイステー

クスの状況下では、社会的望ましさの影響を受けやすい（Tracey, 2016）。社会的望ましさに関しては、項目内容の社会的望ましさを考慮する方法（肥田野・柳井・塗師・繁樹・高根, 1971など）や社会的望ましさ尺度を用いる方法（Paulhus, 1984）など統制法に関する様々な研究がなされているが、その解決を直接的に解決できるだけの知見は得られていない。

天井効果や床効果と社会的望ましさは、発生要因は異なるが、回答分布が歪む（歪められる）という意味においては同様の現象を生じさせる。このような現象に対処する方策の1つとして、様々な難易度の項目を提示することが考えられる。具体的には、「英語の授業は楽しい」と「英語の授業は非常に楽しい」という項目は、いずれも「英語の授業が楽しい」かどうかを尋ねているが、その強度が異なり、前者では肯定反応をしやすいが、後者は比較的肯定反応をしにくい。

このように項目内容を調整することによってある程度回答分布を操作することは可能であるが、心理尺度の構成という意味では実用的でない。それは、通常、心理尺度は開発時に作成された項目によって測定概念が決定されていると考えるため、項目内容が異なれば、厳密には測定対象となる概念が異なることになる。英語に対する肯定的態度を測定するための尺度内の「英語の授業が楽しい」という項目を「英語の授業は非常に楽しい」という項目に変更した場合、前者の項目を使用して測定された肯定的態度と、後者を使用して測定された肯定的態度は異なる。

このように、項目内容を変更することによって、つまり項目の難易度を調整することによって回答分布を操作することは現実的とはいえない。もうひとつの解決方法は、尺度開発時から多様な難易度の項目を用意することであるが、これも必ずしも現実的とは言えない。

そこで、項目内容はそのままにして、回答選択肢を操作することで回答分布を操作することができないかを検討する。そして、それが可能であるならば、回答者の潜在特性値に応じた回答選択肢を提示し回答を求めることでより良い測定が可能となるのではないだろうか。本研究では、次に示す適応型選択肢のアイデアを提案し、実際のデータを用いて実用化に向けた基礎的な検討を行うことを目的とする。なお、このような選択肢に変更を加えることの是非に関しては、最後に述べる。

**適応型選択肢のアイデア** 適応型選択肢（Adaptive Options）は、Figure 1 に示す仕組みにより、回答者の潜在特性レベル（例：動機づけの高さ）に応じて、選択肢を変化させ、得点の天井効果や床効果を生じにくくするものである。具体的には、最初の項目（もしくは数項目）で、どのように回答したかによって、後に続く項目に対する選択肢を変化させ

る。例えば英語の授業に対する動機づけが高い生徒は、通常の選択肢セットを用いた場合、どの項目にも「5. あてはまる」と回答すると考えられる。

そこで、例えば、動機づけの高い生徒には、High Level Version の選択肢セットを提示すれば「3. かなりあてはまる」「4. よくあてはまる」「5. とてもよくあてはまる」の選択肢を選択することになるだろう。つまり、前者では「5. あてはまる」の選択肢を選ぶしかなかった回答者が、High Level Version では、3～5 番目の選択肢を選択する可能性が生まれたということである。これにより、動機づけが高い回答者の特性レベルを従来に比べてより精緻に測定できるようになると考える。最終的には、Computer Adaptive Testing のように、それまでに回答した項目により推定された潜在特性値を元に、次の項目の選択肢を変化させることが可能であろう。当然、この所作をした場合に、異なる選択肢セットによる回答をした回答者間の比較が可能かという問題は生じるが、その点に関しては Wakita et al. (2012) の知見を使うことで、異なる選択肢セットに回答していても、比較可能なスコアを算出可能である。

**本研究の目的** 本研究では上記で提案した適応型選択肢の実現に向け、基礎段階として選択肢を変更することにより、回答分布がどのように変化するかに関して検討を行った。具体例として中学校・高等学校において経年的に実施していた調査において、調査の見直し時期に、選択肢セットを変更したことによる影響を報告する。

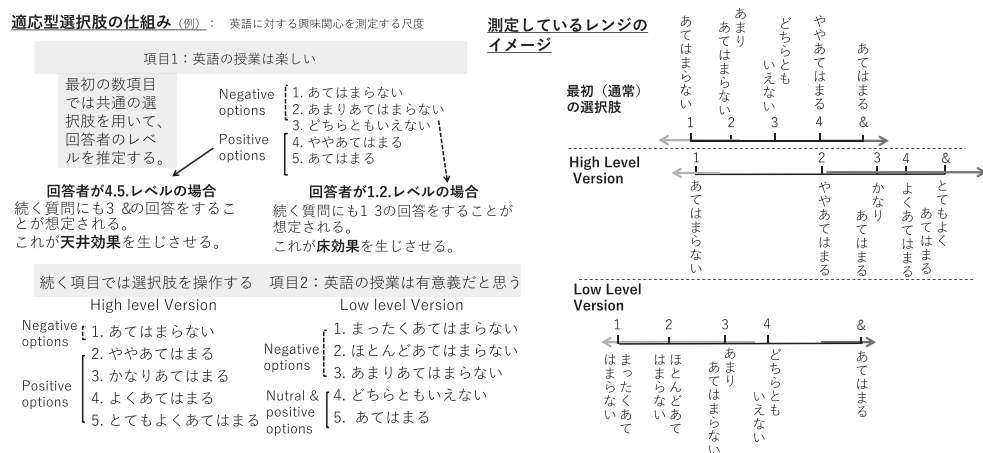


Figure 1 適応型選択肢のイメージ

## 方法

**質問紙** 中・高校生を対象とした、学びに関する項目を用いた。2回の調査のうち、項目内容が同一であった項目23項目に関して、因子分析を行い、最終的に14項目を抽出した。項目内容は「学習を進める中で、関連しそうな様々な情報を収集している」などの項目であった。使用した評定尺度表現を Figure 2 に示した。

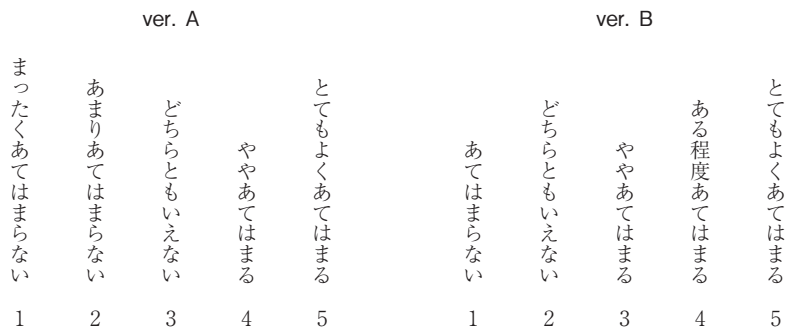


Figure 2 各パターンの評定尺度表現

**実施時期と対象者** ver. A: 2015年12月に中学生・高校生583名に回答を求めた。ver. B: 2016年12月に中学生・高校生590名に回答を求めた。

**分析方法** 各バージョンにおける回答者の分布、平均・標準偏差に関して検討を行った。IRT モデルにおけるパラメタ推定には、PARSCALE (Muraki & Bock, 2003) を用いた。本研究では、両パターンの category パラメタの比較に関心があるため、slope パラメタおよび location パラメタに関しては等値制約を置く DIF 分析の方法を用いた。

## 結果

**検討する尺度の構成** 尺度構成を行うために23項目に対して因子分析（主因子法）を行った。1因子の尺度を構成するため、因子負荷量が.50を下回る項目を削除し、最終的に14項目を選択した。信頼性係数の推定値として Cronbach's  $\alpha$  を求めたところ、いずれも.900であった。本研究では、この14項目を用いた検討を行う。

**項目ごとの回答分布とバージョン間の比較** 回答者が各選択肢を選択した割合を Table 1、

各項目の平均・標準偏差を Table 2 に示した。はじめに Table 1 に示した選択割合から検討すると、選択肢の位置の視点から、第 1 カテゴリー (ver. A まったくあてはまらない, ver. B あてはまらない) においては、すべての項目で ver. B、第 4 カテゴリー (ver. A ややあてはまる, ver. B ある程度あてはまる) では、ver. B よりも ver. A の選択割合が高かった。第 2 カテゴリー、第 3 カテゴリー、第 5 カテゴリーは大きな特徴は認められなかった。

続いて、選択肢の意味に着目すると、両者で同じ選択肢表現「どちらともいえない」(ver. A 第 3 カテゴリー、ver. B 第 2 カテゴリー) では、すべての項目において ver. A の選択割合が多い。「ややあてはまる」(ver. A 第 4 カテゴリー、ver. B 第 3 カテゴリー) では相対的に困難の高い 2 項目 (s29 と s52) を除き、ver. A の選択割合が高かった。

さらに選択肢の意味を考慮すると、否定形を選択肢 (ver. A まったくあてはまらない + あまりあてはまらない と ver. B あてはまらない) で比較すると、すべての項目で ver. A の方が選択割合が高く、平均では約 11% ほどの違いが認められた。肯定系を選択肢 (ver. A

Table 1 各選択肢の選択状況

	ver. A					ver. B				
	まったくあてはまらない	あまりあてはまらない	どちらともいえない	ややあてはまる	とてもよくあてはまる	あてはまらない	どちらともいえない	ややあてはまる	ある程度あてはまる	とてもよくあてはまる
s30	11 (.019)	67 (.116)	168 (.290)	238 (.411)	95 (.164)	35 (.060)	110 (.187)	194 (.330)	173 (.295)	75 (.128)
s40	12 (.021)	67 (.115)	162 (.279)	231 (.398)	109 (.188)	21 (.036)	70 (.119)	174 (.295)	180 (.306)	144 (.244)
s43	9 (.015)	60 (.103)	138 (.237)	259 (.444)	117 (.201)	36 (.061)	80 (.136)	172 (.293)	190 (.323)	110 (.187)
s36	22 (.038)	78 (.134)	175 (.301)	242 (.417)	64 (.110)	33 (.056)	103 (.175)	212 (.361)	156 (.265)	84 (.143)
s45	11 (.019)	58 (.100)	159 (.273)	253 (.435)	101 (.174)	23 (.039)	73 (.125)	188 (.322)	186 (.318)	114 (.195)
s50	10 (.017)	64 (.110)	173 (.297)	240 (.412)	95 (.163)	17 (.029)	64 (.109)	177 (.302)	197 (.336)	131 (.224)
s12	22 (.038)	84 (.144)	166 (.285)	219 (.376)	92 (.158)	22 (.037)	72 (.123)	165 (.281)	200 (.341)	128 (.218)
s51	26 (.045)	89 (.153)	191 (.328)	199 (.342)	77 (.132)	43 (.074)	114 (.195)	178 (.304)	158 (.270)	92 (.157)
s29	30 (.052)	134 (.230)	178 (.306)	167 (.287)	73 (.125)	59 (.100)	143 (.243)	180 (.306)	138 (.235)	68 (.116)
s52	29 (.050)	96 (.165)	207 (.356)	189 (.325)	61 (.105)	41 (.070)	116 (.198)	204 (.348)	147 (.251)	78 (.133)
s24	20 (.034)	90 (.154)	152 (.261)	228 (.391)	93 (.160)	41 (.070)	112 (.191)	214 (.365)	148 (.252)	72 (.123)
s01	16 (.027)	80 (.137)	160 (.274)	214 (.367)	113 (.194)	34 (.058)	86 (.146)	154 (.261)	202 (.343)	113 (.192)
s07	40 (.069)	112 (.192)	169 (.290)	178 (.305)	84 (.144)	64 (.109)	92 (.156)	167 (.284)	168 (.286)	97 (.165)
s21	25 (.043)	86 (.148)	173 (.297)	198 (.340)	100 (.172)	51 (.087)	111 (.189)	193 (.329)	146 (.249)	86 (.147)

( ) は割合

のややあてはまる＋ととてもよくあてはまると ver. B のややあてはまる＋ある程度あてはまる＋ととてもよくあてはまる）でみるとすべての項目で ver. B の選択割合が高く、約25%ほどの違いが認められた。

続いて、Table 2 に項目の平均および標準偏差を示した。項目平均に着目すると、s12は「自分がなぜそのように考えたかを相手に話すようにしている」（差は0.11）、s50は「考えた解決法を自分なりの言葉で説明できる」（差は0.03）の2項目を除き、ver. A の平均が高くなっていた。標準偏差に関しては、大きな違いは認められなかった。

Table 2 項目ごとの平均・標準偏差

	ver. A		ver. B		ver. Aとver. B の平均値の差
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
s30	3.59	0.96	3.24	1.08	0.35
s40	3.62	0.98	3.60	1.09	0.02
s43	3.71	0.95	3.44	1.12	0.27
s36	3.43	0.98	3.26	1.08	0.17
s45	3.64	0.94	3.51	1.06	0.13
s50	3.59	0.94	3.62	1.04	-0.03
s12	3.47	1.04	3.58	1.07	-0.11
s51	3.36	1.04	3.24	1.15	0.12
s29	3.20	1.09	3.02	1.16	0.18
s52	3.27	1.02	3.18	1.11	0.09
s24	3.49	1.04	3.17	1.09	0.32
s01	3.56	1.04	3.47	1.13	0.09
s07	3.26	1.13	3.24	1.22	0.02
s21	3.45	1.07	3.18	1.16	0.27

**尺度得点の回答分布** 通常、Likert タイプの心理尺度では、複数の項目の項目得点を加算する（もしくは平均）することにより尺度得点として扱う。そこで、14項目による尺度得点を算出したところ、ver. A: 3.48 ( $SD=0.67$ )、ver. B: 3.35 ( $SD=0.82$ ) であり、評定尺度表現を厳しめにしたパターン B で尺度得点が低い一方で、標準偏差に関しては若干大きかった。また尺度得点の分布を Figure 3 に示した。

**回答選択肢間の等間隔性の観点** IRT による項目パラメタを Table 3 に示した。なお、本研究では選択肢間の心理的距離に関心があるため、slope パラメタ、location パラメタは ver. A、ver. B で固定し、category パラメタを比較する DIF モデルを用いた。選択肢間の心理的距離として、IRT の項目パラメタである category パラメタの値から計算される変換後尺度値を Figure 4 に示した。その結果、等間隔性という意味では相対的にみて ver. B の方が等間隔性を満たしているという結果が示された。

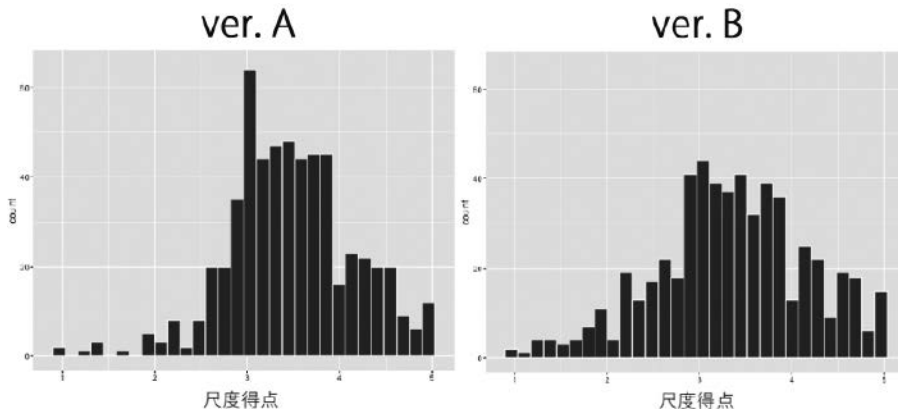


Figure 3 各パターンの尺度得点の分布

Table 3 IRT 項目パラメタ ( ) は s.e.

	slope		location	
s30	0.970	(0.044)	-0.462	(0.026)
s40	0.821	(0.038)	-0.740	(0.029)
s43	0.823	(0.038)	-0.692	(0.029)
s36	0.880	(0.040)	-0.387	(0.027)
s45	0.829	(0.039)	-0.689	(0.029)
s50	0.833	(0.039)	-0.731	(0.029)
s12	0.735	(0.035)	-0.639	(0.031)
s51	0.661	(0.032)	-0.356	(0.032)
s29	0.634	(0.031)	-0.116	(0.032)
s52	0.721	(0.034)	-0.248	(0.030)
s24	0.669	(0.032)	-0.388	(0.032)
s01	0.623	(0.031)	-0.649	(0.034)
s07	0.548	(0.028)	-0.307	(0.036)
s21	0.579	(0.029)	-0.384	(0.035)
category	ver. A		ver. B	
2	1.667	(0.046)	1.532	(0.046)
3	0.537	(0.028)	0.610	(0.028)
4	-0.313	(0.024)	-0.538	(0.024)
5	-1.890	(0.029)	-1.604	(0.029)

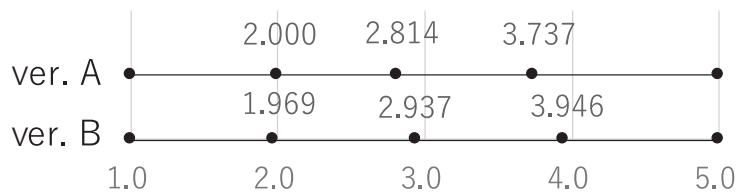


Figure 4 各パターンの変換後尺度値



**尺度得点の求め方の工夫** 先述した尺度得点は、従来の Likert 法の得点化方法である、各選択肢に整数値を割り当て、その和（もしくは平均）を求めたものである。適応型選択肢の実用化に向けては、異なる選択肢セットで得られたデータであっても比較可能にすることが望まれる。そこで、得点化の工夫として、Figure 4 に示した変換後尺度値を用いて尺度得点を算出する方法が考えられる。厳密な比較は、被験者内計画で収集されたデータを用いる必要があるが、本研究では資料的意義を鑑み Figure 5 にその分布を示した。ver. A: 3.33 ( $SD=0.67$ )、ver. B: 3.31 ( $SD=0.82$ ) であった。

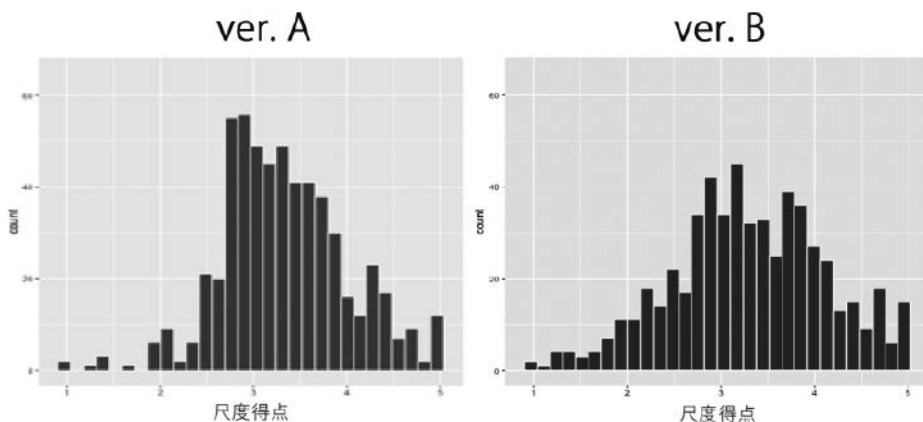


Figure 5 各パターンの尺度得点（変換後尺度値を使用）の分布

## 考察

**項目ごとの回答分布とバージョン間の比較** はじめに、回答選択肢ごとの選択割合を検討すると、ver. A では第 1 カテゴリーの選択割合が 0.019 から 0.069、平均は 0.035 ( $SD=0.015$ ) と高くなく、選択肢が十分な機能を果たしていないと考えられる。対して、ver. B では平均が 0.063 ( $SD=0.023$ ) であり十分機能していると考えられる。

第 4 カテゴリーに関しては、ver. A では 0.287 から 0.435 の範囲であり、その平均も 0.375 ( $SD=0.047$ ) と高く、多くの回答者の回答が第 4 カテゴリーに集まっている。個人差の測定という意味において、必ずしも望ましい現象であるとはいえない。一方で、ver. B の平均は 2.91 ( $SD=0.036$ ) であり ver. A に比べて望ましい結果と考えて良いだろう。

選択肢の意味に焦点をあてると、「どちらともいえない」という中間選択肢（選択肢の意味において中間）はすべての項目で ver. A の選択割合が高くなっている。本調査が学校と

いう場で行われているものであり、学習に対する項目であることを考えると、これらの項目に「どちらともいえない」と回答することは、厳密に中間の意味をもつというより、否定的な意味をもつと考えられる。ver. Bでは選択肢の位置からみると、中央より左寄り（否定）側にあるため、より明確にそのメッセージが回答者に伝わった可能性がある。

否定形を選択肢（ver. A まったくあてはまらない+あまりあてはまらないと ver. B あてはまらない）で比較した結果、すべての項目で ver. A の選択割合が高くなっているが、これは選択肢数が異なるため当然の結果と考えられる。また、肯定系を選択肢に関する結果も、選択肢数が異なるため単純な比較はできない。

先述したとおり、ver. A の肯定的選択肢（ややあてはまる+とてもよくあてはまる）と ver. B の肯定的選択肢（ややあてはまる+ある程度あてはまる+とてもよくあてはまる）では、これらに該当する回答が ver. B の方が25%ほど増加していた。肯定的なカテゴリをまとめて、肯定的な回答割合として何かしらの報告をする場合には、ver. B の方が肯定的な回答が増えるため解釈には注意が必要である。しかし、心理測定において、そのような所作をすることは少なく、項目得点で議論することが多いため、大きな問題とはならないだろう。

続いて、項目ごとの平均・標準偏差を検討したところ、2項目を除き ver. A の方が高い結果となった。これは ver. B のほうが選択肢の困難度が高い（上記の例では High Level version）であったためである。ver. B の方が平均が高かった s12は「自分がなぜそのように考えたかを相手に話すようにしている」（差は0.11）、s50は「考えた解決法を自分なりの言葉で説明できる」（差は0.03）の2項目であり、これらは他の項目 s43「解き方がわからない問題でも、いろいろな知識を用いて考えようとしている」よりも項目自体の困難度も高い。

これらの項目は、項目の困難度が高いことにより、困難度が低い ver. A であっても肯定的な反応をすることがそもそも抑制されており、その結果として ver. B と逆転していると考えられる。しかし、その差は相対的に小さく誤差の範囲である可能性がある。

**尺度得点の回答分布** 尺度得点レベルの検討を行ったところ、上述した回答選択肢割合の違いから想定されるように、ver. B の方が尺度得点が低くなった。また標準偏差に関しては、項目レベルで見るとそれほど差がなかったが、尺度得点にすると若干大きくなっていった。あらためて確認すると、すべての項目で ver. B の方が分散が大きくなっている。ver. B では第1カテゴリから第5カテゴリまで、選択者がいたことがこの結果をもたらしている。先述したとおり個人差の測定という意味においては ver. B の方が望ましい結果と

言えるだろう。

Figure 1 の分布を見ても ver. B の方が尺度得点のレンジ全体に得点分布が散らばっており、特に ver. A で最も頻度の大きな 3 点部分が ver. B では消えている。ver. A の 3 点は「どちらともいえない」の中間選択肢であり、回答の中心化傾向の結果であると考えられる。一方で、ver. B では明確な中心化傾向は認められず、意味的に中間である第 2 カテゴリー「どちらともいえない」の 2 点付近に分布のピークが来ているわけではない。この結果だけから判断することはできないが、ver. B のような選択肢を用いることで、中心化傾向が改善される可能性が示唆された。

**回答選択肢間の等間隔性の観点** Likert 法において、現在のような常識ともいえるべき一般的な形（ver. A）が定着した理由の 1 つとして、選択肢間の心理的距離の観点がある。ver. B のような選択肢は、従来の Likert 法の使用法の常識からは外れるものといえるかもしれない。例えば、鈴木（2011）では、奇数個の選択肢がある場合、すべての例において中央に「どちらともいえない」という中間選択肢を提示している。Stevens (1946) のいう、尺度水準でいえば、Likert 法で得られる数値は順序尺度水準である。それを間隔尺度水準として扱うためには、選択肢間の等間隔性が必要である。しかし、選択肢間の等間隔性を数値として示すことが困難であったため、最低限選択肢に与える選択肢表現は等間隔になるように設定するべきであるという理由が大きいと考えられる。

しかし、Wakita et al. (2012) で提案した方法は、IRT を用いて選択肢間の等間隔性を数値として示すことができるため、ver. B のような、意味的には等間隔とはならない選択肢セットを使用することも許容されるのではないだろうか。この方法を用いて、選択肢間の心理的距離を示したところ両者共に、特に ver. B に関しても等間隔性が満たされていることが確認された。したがって、ver. B のような選択肢であっても、項目得点の加算や平均を使用しても大過はないことが示唆された。

**尺度得点の求め方の工夫** Figure 5 に示したとおり変換後尺度値を用いて求めた尺度得点の分布は、Figure 3 とは印象が異なり、若干左に寄っている。これは ver. A では、従来の方法では 3 点が 2.814 点、4 点が 3.737 点になっていることに起因する。この結果に対する積極的な解釈は困難であるが、今後、被験者内計画を用いて選択肢セットが異なった場合の比較を行うことが必要であると考えられる。

## まとめ

本研究では、提案した適応型選択肢の実用化に向けて、異なる選択肢セットを用いて、選択肢表現が回答に与える影響を検討した。その結果、選択肢の難易度を調整することで、回答分布を操作することができること、および、選択肢の難易度を変えるために、選択肢表現を変えることが選択肢間の等間隔性を損なわないことを示した。

本研究では、通常の形である ver. A とより困難度が高い選択肢セットである ver. B を比較した。選択肢ごとの回答割合、尺度得点に与える影響いずれも解釈可能な結果であった。これらの結果は、適応選択肢の実用化に向けて有益なものであると考えられる。また、適応型選択肢を考えない場合も、社会的望ましさの影響を受けやすい尺度、中心化傾向が生じやすい尺度に関しては、回答選択肢の表現を工夫することで、これらの影響を低減させることができる可能性が示唆された。

いくつか残された課題としては以下の事柄が考えられる。始めに、適応型選択肢を用いた場合に問題になってくるのは、異なる選択肢セットに回答した場合に、その得点をどのようにするのか、回答者間で比較可能な得点を算出することができるのかという問題ある。その点に関しては、Wakita et al. (2012) の方法を用いることで解決可能ではないだろうか。ただし、本研究で示したものであれば、両端の選択肢には1点、5点を与えることになる。そこを固定して算出することの是非に関しては詳細な検討が必要である。

第2に、どちらの選択肢セットで得られた得点が、より真の得点を表しているのかという問題がある。本研究では、回答者の回答割合でみたときに選択肢が機能していること、求められた尺度得点の分布、得点の分散が個人差の測定という観点でより望ましいこと、選択肢間の等間隔性が満たされていることという3つの視点で検討した。選択肢表現を変えた場合の測定が妥当であるかは、例えば外的指標との関連などとの検討が必要であろう。例えば、本研究のように学校現場で用いられる学習の尺度であれば、実際の成績などとの関連を検討することも1つの方法である。この点に関しては今後の課題である。

また、当然のことながら、一般化の問題に関しては留意することが必要である。本研究の結果はあくまで検討に用いた項目、学校現場という状況に即したものである。この結果があらゆる尺度、状況に適用できるかはデータの蓄積は不可欠である。例えば、回答分布でみた場合、s12やs50など、必ずしも想定通り、解釈可能な結果となっていない項目も存在した。これは先述したとおり、項目の難易度が影響していると考えられる。この点に関しては、別の尺度においても項目困難度が高い場合に同様の現象が生じるのかなど、詳細

な検討が望まれる。

最後に、選択肢そのものの表現によって、回答を誘導させているのではないかという議論がある。例えば、鈴木（2011）では避けるべき、Peterson（2000）は、バイアスを避ける為に回答選択肢は連続体上でバランス良く配置するべき、であると述べている。一方で、Sarlis & Gallhofer（2014）は、回答者が一方の側に偏ることが明らかな場合には、通常の形式を用いることはあまり効果的ではなく、非対称にする方が適切（appropriate）かつ精緻（precise）であると述べている。また、山田（2010）では場合によっては有効だが注意が必要と述べている。

このように様々な主張がなされているのが実情である。肯定的な回答をした人が〇%であるというような報告をする場合には問題になることは否定できないが、心理尺度のように得点化する場合、個人差の測定という観点からすれば大きな問題とはならないだろう。

なお、「とてもよくあてはまる」という表現と「ある程度よくあてはまる」という表現で比較した場合も、言葉上は肯定しているが、後者では「ある程度」といっている以上、完全に「あてはまる」と回答しているわけではない。「満足していますか？」という問いに対して、「ある程度あてはまる」と回答することは、すくなく不満な部分も存在するということであろう。言い換えれば、自信を持って「満足している」と言えないということでもある。調整した選択肢を用いて調査した場合に、この回答が何%であるという報告の仕方をする場合には、その解釈に十分注意する必要がある。

このような課題はあるものの、選択肢を変化させることの心理測定上の効果、分布の偏りを是正できることの可能性を検討することは意義のあることであると考えられる。上記の議論も同様であるが、尺度内容もしくは項目内容によっても適しているもの、適していないものがある可能性が高い。この点に関しては今後の検討が望まれる。

#### 引用文献

- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37, 90.
- 増田 真也 (2019). 心理尺度の回答カテゴリに関する検討 哲学, 142, 245-267.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muraki, E., & Bock, R. D. (2003). PARSCALE: Parameter Scaling of Rating Data [Computer program]. Chicago, IL: Scientific Software.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality*

*and social psychology*, 46, 598.

Peterson, R. A. (2000). *Constructing effective questionnaires (Vol.1)*. Thousand Oaks, CA: Sage Publications.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, New Series*, 103, No. 2684, 677-680.

鈴木 淳子 (2011). 質問紙デザインの技法 ナカニシヤ出版

田崎 勝也・申 知元 (2017). 日本人の回答バイアス——レスポンス・スタイルの種別間・文化間比較——, 心理学研究, 88, 32-42.

豊田 秀樹・川端 一光・松下 信武 (2005). 採用場面におけるEQ検査の改善 教育心理学研究, 53, 456-466.

Tracey, T. J. (2016). A note on socially desirable responding. *Journal of Counseling Psychology*, 63, 224.

Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement*, 72, 533-546.

山田 一成 (2010). 聞き方の技術 リサーチのための調票作成ガイド 日本経済新聞出版社

#### 謝辞

本研究はJSPS科研費 17K13926の助成を受けたものである。

本研究の一部は、日本心理学会第83回大会においてポスター発表を行ったものである。

—2019.12.5 受稿—