川崎医療福祉学会誌　Vol. 25　No. 1　2015　23－35

原　著

項目反応理論に基づく順序尺度の等間隔性
－質問紙調査の回答選択肢（3～5件法）の等間隔性と回答のしやすさ－

井　上　信　次*1

要　　約

本稿の目的は，質問紙調査における，回答のしやすさと統計学的な妥当性が成立する尺度のカテゴリ数について明らかにすることである．筆者は，医療福祉の利用者を調査対象とする場合，最小限の負担であることが好ましく，また尺度のカテゴリ数は少ない方がよいと考える．回答のしやすさと統計学的な妥当性が成立する尺度のカテゴリ数，さらには質問項目数の分析が必要である．医療福祉分野の研究の中で，以上の関心をもった研究をみることはできない．調査は2015年1月下旬に筆者が行う講義において，質問紙を配布し，回収した（集合調査法）．対象はＡ大学のＺ学部の学生である．調査票は，130部配布し，103部を回収した．有効回答は89票 (86.4%) であった．調査項目は主に性別，3件法，4件法，5件法質問票（同内容でカテゴリ数が異なる3種類の調査票），回答のしやすさについてである．項目反応理論(IRT)により3～5件法質問紙における変数の等間隔性について分析を行った．その結果，3つの結論を得た．第1に3件法尺度について等間隔性が認められた．第2に性別によって回答のしやすさが異なっていた．第3に，対象者の多くが女性であるとして回答のしやすい5件法調査票のみを用いることは，尺度の等間隔性という観点から分析時に分析手法が限定される可能性がある．にもかかわらず，パラメトリック手法が多く用いられる現状には課題が残る．

1．緒言

質問紙調査において，尺度水準の設定は大きな課題の一つである．特にリッカート尺度は，順序尺度と間隔尺度のどちらに分類するかについて研究上の長い歴史がある[1)]．両者の違いは，尺度の正規性とカテゴリ（選択肢）間の等間隔性に要約される．尺度の正規性については，ピアソン等によって古くから議論され[2)]，Kolmogorov-Smirnov 検定を始め様々な検定が用いられる．一方，カテゴリ間の等間隔性については検定までに至っていない．

5件法のリッカート尺度の場合，統計学的には，例えば「4」と「5」との間の等間隔性を確認しなければならない．尺度の等間隔性が保証され，さらに正規性を持つのであれば，間隔尺度として扱うことができる．例えばＦ検定は順序尺度に対して用いてもよいという議論もあるが[1)]，それらの根拠の大半は調査者の経験則である．統計学の教科書は，例えば算術平均値は間隔尺度以上でしか計算できないと指摘するが，多くの統計学的分析では順序尺度を間隔尺度とみなしている．

西里はリッカート尺度を比例尺度として扱うことは誤りであるが，それを検証する方法が身近にないとした[3)]．狩野は3～4件法は「グレーゾーン」，5件法以上は連続尺度とみなしてもよいと考えた[4)]．脇田は間隔尺度では各カテゴリ間の間隔が等間隔であるという仮定を満たす必要があるとした[5)]．その他，様々な議論があるが，等間隔性に関する数十年にわたる議論は必ずしも決着をみていない．

ここで項目反応理論（Item Response Theory, 以下 IRT）に着目する．IRT は特に心理学領域で研究が進んでおり[6-9)]，近年になり社会福祉分野でも適用され始めた[10)]．これらの研究の中では，IRT は間隔尺度，順序尺度の尺度水準を下げずに，調査で得られた情報を最大限にいかすために用いられてい

*1 川崎医療福祉大学　医療福祉学部　医療福祉学科
（連絡先）井上信次　〒701-0193　倉敷市松島288　川崎医療福祉大学
E-mail : inoshin@mw.kawasaki-m.ac.jp

る．職業性ストレス簡易調査票は4件法によって構成された質問紙調査であるが，IRT によって間隔尺度を担保している[11]．尺度開発上，このような研究は重要であり IRT が有効である．しかしながら筆者はこのような研究は分析者視点によるものであり，回答のしやすさという回答者視点を欠いていると考える．

どのような尺度が回答しやすいかは，調査対象者や調査内容によるが，筆者は，医療福祉の利用者を調査対象とする場合，回答が利用者にとって最小限の負担であることが好ましく，また尺度のカテゴリ数は少ない方がよいと考える．回答のしやすさと統計学的な妥当性が成立する尺度のカテゴリ数，さらには質問数の分析が必要ではないか．にもかかわらず医療福祉分野の研究の中で，以上の関心を持った研究をみることはできない．また調査対象者が「回答してもいい」という理由の一つが質問の量と難易度だが[12]，具体的にどの程度の量と簡便さが必要かまでは，教科書的な記述を除き，統計学的に言及されることはない．高齢者施設の待機者本人を対象にした5件法自記式質問紙調査があるが，回収率が低く，欠損値も多くなっている[13]．

どの程度の尺度のカテゴリ数が回答しやすいかを明らかにするためには，調査対象者を同じにし，同内容の質問を，複数の尺度のカテゴリで分析する必要がある．しかしながら同内容のものを複数回，短時間の間隔で調査をすれば，調査対象者が1回目の調査内容を記憶している可能性が高いため（学習効果），正確な結果が得られない可能性がある．一方，時間を空けすぎると対象者に大きな変化が起こりうる．さらに日を空けて複数回の調査を行う場合，個人を固定する必要があるが，個人情報保護の観点からも困難であり，また2回目以降の調査でケースの脱落（摩耗）が起こる可能性がある．以上のような調査の困難さもあり，とかく医療福祉分野の研究において，どの程度の尺度のカテゴリ数が回答しやすいかに関する研究は十分でなかったと考える．

２．研究目的

質問紙調査における，回答のしやすさと統計学的な妥当性が成立する尺度のカテゴリ数について明らかにすることである．

３．研究方法

３．１　対象者及び配布・回収方法

2015年1月下旬に筆者が行う講義において，質問紙を配布し，回収した（集合調査法）．対象はＡ大学Ｚ学部の学生で，大半が3年生である．対象者の標本抽出は行っていない．

3.2に示す3件法，4件法，5件法のそれぞれの質問紙は無作為に並んでいることから，回答者によって回答する質問紙の順番は異なる．それぞれの質問紙における質問の順番は乱数によって無作為に並べた．3種類の質問紙への回答は，約20分おきに行い，一度回答した質問紙の訂正は行わない旨を回答者に伝えた．以上を通して，同じ質問を繰り返し行うことで生じる回答者の学習効果を最小限にした．

３．２　調査項目

質問は「性別」「3件法，4件法，5件法質問票」（表1），「3件法，4件法，5件法質問票に対する回答のしやすさの順」（以下，回答のしやすさ），「SOC 日本語版13項目」から構成した．3件法，4件法，5件法質問票は国民生活基礎調査の項目を参考に，学生が回答しやすい人生等への「不安」に関する14項目を筆者が独自に作成した．3つの質問票のそれぞれは同じ質問項目であり，選択肢が3件法，4件法，5件法になっている．回答のしやすさについては，「最も回答しやすかった質問紙」，「次に回答がしやすかった質問紙」，「最も回答がしにくかった質問紙」について3択で順に回答を求めた．「SOC 日本語版13項目」[14]は，例えば目の前の課題に対する処理可能感，有意味感といった下位概念によって構成されている．SOC（Sense of coherence）の程度によって回答方法に違いがあるかを確認するために挿入した．なお，本稿では紙幅の都合でSOCの分析を割愛した．

３．３　分析対象

調査票は，130部配布し，103部を回収した．3件法，4件法，5件法質問票の項目について，一つでも無回答がある場合は無効回答票とした．その上で，「性別」，「回答のしやすさ」については無回答がある場合でも有効票とした．その結果，89票を有効回答票（68.4%）とした．また，「性別」，「回答のしやすさ」に関する分析ではペアワイズによる除去を行ったため，それぞれの有効回答票は，88票（67.7%），78票（60.0%）となった．

３．４　分析方法

3件法，4件法，5件法質問票は IRT の段階反応モデルによる分析を行った．性別と回答のしやすさの関係は，クロス集計表を作成し，χ^2検定を行った．IRT は Muthen らが開発した MPlus7.1を，それ以外は IBM SPSS Statistics21を用いた．

３．５　分析モデル

３．５．１　項目反応理論における尺度の等間隔性

IRT は，テスト項目に対する受検者（受験者，回答者）の反応（解答や回答）に基づいて心理的特性値の大きさを推定し，項目やテストの特徴を統計

表 1　各質問項目の度数分布及び基礎統計

	C1	C 2	C 3	C 4	C 5	平均値	中央値	標準偏差
1 自分の将来について不安になった R	52 (58.4)	24 (27.0)	13 (14.6)			1.6	1.0	0.7
	28 (31.5)	42 (47.2)	18 (20.2)	1 (1.1)		1.9	2.0	0.7
	21 (23.6)	43 (48.3)	6 (6.7)	13 (14.6)	6 (6.7)	2.3	2.0	1.2
2 たくさんの友人に相談することが多かった R	36 (40.4)	31 (34.8)	22 (24.7)			1.9	2.0	0.8
	18 (20.2)	35 (39.3)	23 (25.8)	13 (14.6)		2.4	2.0	1.0
	14 (15.7)	27 (30.3)	22 (24.7)	16 (18)	10 (11.2)	2.8	3.0	1.2
3 今の自分は幸せであると感じることが多かった R	12 (13.5)	32 (36)	45 (50.6)			2.4	3.0	0.7
	8 (9.0)	25 (28.1)	38 (42.7)	18 (20.2)		2.7	3.0	0.9
	7 (7.9)	12 (13.5)	25 (28.1)	26 (29.2)	19 (21.3)	3.4	4.0	1.2
4 自分の就職について不安になった	59 (66.3)	19 (21.3)	11 (12.4)			1.5	1.0	0.7
	29 (32.6)	43 (48.3)	14 (15.7)	3 (3.4)		1.9	2.0	0.7
	27 (30.3)	42 (47.2)	5 (5.6)	10 (11.2)	5 (5.6)	2.2	2.0	1.1
5 何をするにも骨折りだと感じることが多かった	18 (20.2)	42 (47.2)	29 (32.6)			2.1	2.0	0.7
	7 (7.9)	23 (25.8)	47 (52.8)	12 (13.5)		2.7	3.0	0.8
	6 (6.7)	20 (22.5)	20 (22.5)	31 (34.8)	12 (13.5)	3.2	3.0	1.2
6 自分は価値のない人間だと感じることが多かった	20 (22.5)	30 (33.7)	39 (43.8)			2.2	2.0	0.8
	6 (6.7)	23 (25.8)	41 (46.1)	19 (21.3)		2.8	3.0	0.8
	4 (4.5)	11 (12.4)	28 (31.5)	29 (32.6)	17 (19.1)	3.5	4.0	1.1
7 目の前の課題に対して，どうしていいか分からなくなることが多かった R	35 (39.3)	26 (29.2)	28 (31.5)			1.9	2.0	0.8
	14 (15.7)	38 (42.7)	28 (31.5)	9 (10.1)		2.4	2.0	0.9
	11 (12.4)	30 (33.7)	19 (21.3)	21 (23.6)	8 (9.0)	2.8	3.0	1.2
8 うれしいと思うことが多かった	11 (12.4)	30 (33.7)	48 (53.9)			2.4	3.0	0.7
	5 (5.6)	23 (25.8)	43 (48.3)	18 (20.2)		2.8	3.0	0.8
	3 (3.4)	9 (10.1)	26 (29.2)	31 (34.8)	20 (22.5)	3.6	4.0	1.0
9 落ち着かないことが多かった	28 (31.5)	30 (33.7)	31 (34.8)			2.0	2.0	0.8
	11 (12.4)	34 (38.2)	30 (33.7)	14 (15.7)		2.5	2.0	0.9
	7 (7.9)	25 (28.1)	26 (29.2)	22 (24.7)	9 (10.1)	3.0	3.0	1.1
10 神経過敏に感じることが多かった R	25 (28.1)	28 (31.5)	36 (40.4)			2.1	2.0	0.8
	11 (12.4)	28 (31.5)	35 (39.3)	15 (16.9)		2.6	3.0	0.8
	9 (10.1)	23 (25.8)	13 (14.6)	27 (30.3)	17 (19.1)	3.2	3.0	1.3
11 勉強が苦痛だと感じることが多かった	43 (48.3)	26 (29.2)	20 (22.5)			2.1	2.0	0.8
	22 (24.7)	40 (44.9)	22 (24.7)	5 (5.6)		2.1	2.0	0.8
	20 (22.5)	32 (36.0)	13 (14.6)	19 (21.3)	5 (5.6)	2.5	2.0	1.2
12 気分が沈み込んで，何が起こっても気が晴れないように感じることが多かった	28 (31.5)	26 (29.2)	35 (39.3)			2.1	2.0	0.8
	9 (10.1)	32 (36.0)	35 (39.3)	13 (14.6)		2.6	3.0	0.9
	9 (10.1)	20 (22.5)	22 (24.7)	27 (30.3)	11 (12.4)	3.1	3.0	1.2
13 楽しいと思うことが多かった	11 (12.4)	30 (33.7)	48 (53.9)			2.4	3.0	0.7
	4 (4.5)	19 (21.3)	42 (47.2)	24 (27.0)		3.0	3.0	0.8
	3 (3.4)	10 (11.2)	22 (24.7)	34 (38.2)	20 (22.5)	3.7	4.0	1.1
14 自分の将来は明るいと感じることが多かった	32 (36.0)	45 (50.6)	12 (13.5)			1.8	2.0	0.7
	14 (15.7)	44 (49.4)	25 (28.1)	6 (6.7)		2.3	2.0	0.8
	10 (11.2)	25 (28.1)	31 (34.8)	16 (18.0)	7 (7.9)	2.8	3.0	1.1

C1～C5はカテゴリ．3件法：「C1　当てはまらない」「C2　どちらでもない」「C3　当てはまる」．4件法：「C1　当てはまらない」「C2　あまり当てはまらない」「C3　少し当てはまる」「C4　当てはまる」．5件法：「C1　当てはまらない」「C2　あまり当てはまらない」「C3　どちらでもない」「C4　少し当てはまる」「C5　当てはまる」．各変数の1段目は3件法, 2段目は4件法, 3段目は5件法のデータである．数字は人数, 単位は人である．括弧内の数字は % である．合計はすべて89人 (100.0) である．R は逆転項目．表中のデータはすべて逆転させた後のものである．

的に評価するために開発された[15]．IRT は，受検者の能力を１因子の構成概念でとらえる．その意味では，IRT は因子分析において1因子構造を設定することと，ほぼ同義である．IRT により，各変数の尺度を得点として単純加算するような古典的テスト理論ではなく，尺度や項目に関する精緻な情報を得ることができる[16]．テストを受けた受検者の能力だけでなく，テスト問題の作成や評価を同時に行うことができる[17]．さらに IRT の項目統計量は受検者の母集団の特徴，項目数や難易度とは独立に定義される[18]．

TOEFL では，正誤の合計から合計点を算出するのではなく，IRT から得点を算出している．それにより例えば TOEFL の得点が700点である場合，どの回の受験者であってもその能力は700点の能力があることを担保している．つまり，異なる問題を異なる受験者集団が解いたとしても，同一の評価を行うことが可能となっている[19]．

項目反応の出現率を定義するものを項目反応モデルという．特に本稿は，3件法以上の尺度を分析するが，この場合，段階反応モデルと呼ぶ．段階反応モデルのアルゴリズムは研究者によって表記が異なる．本稿では，比較的簡便である住のモデルを援用する[20]．

段階反応モデルでは，（潜在）能力値 θ $(-\infty < \theta < +\infty)$ の人が，項目 j で c 番目の選択肢（カテゴリ）を選択する確率 $p_{jc}(\theta)$ を考える．式（1）に示す．

$$p_{jc}(\theta) = p^*_{jc}(\theta) - p^*_{jc+1}(\theta) \qquad (1)$$

$p^*_{jc}(\theta)$ は，能力が θ である回答者が，項目 j において選択肢の一つを選択し（$c=i$），$c \geq i$ となる確率を意味する．例えば，項目 j が「1. いいえ」(c=1)「2. どちらでもない」(c=2)「3. はい」(c=3) の3件法を考える．このモデルでは，例えば，能力 θ を持つ者が，項目 j で「2. どちらでもない」を選択する確率 $p_{j2}(\theta)$ は，「2. どちらでもない」以上を選択する確率 $p^*_{j2}(\theta)$ から，項目 j で「3. はい」以上を選択する確率 $p^*_{j3}(\theta)$ を引いた値となる．式(2)に示す．

$$p_{j2}(\theta) = p^*_{j2}(\theta) - p^*_{j3}(\theta) \qquad (2)$$

回答者はどの選択肢を選んだ場合でも，$c \geq 1$ となるため，$p^*_{j1}(\theta)=1$ となる．3件法の場合，$c \geq 4$ は存在しないため，例えば，$c=4$ の場合は，$p^*_{j4}(\theta)=0$ となる．

$p^*_{jc}(\theta) - p^*_{jc+1}(\theta)$ は，識別力（Discrimination parameter または Slope parameter）と困難度（Difficulty parameter）から得られる．式（3）に示す．

$$p^*_{jc}(\theta) - p^*_{jc+1}(\theta) = \frac{1}{1+exp[-a_j(\theta - b_{jc})]} - \frac{1}{1+exp[-a_j(\theta - b_{jc+1})]} \qquad (3)$$

a_j は項目 j における識別力パラメタであり，曲線の傾きである．b_{jc} は項目 j で c 番目のカテゴリを選択する際の困難度パラメタである．識別力は，テスト全体で測っている特性を適切に反映しているかどうかの指標である[21]．識別力が高くなれば，特性を適切に反映しており，出現率を明確に示しているといえる．b_{jc} は，能力が θ である回答者が項目 j で c と $c \pm 1$ との選択を分別する閾値 τ_{jc} から計算される．3件法の場合，以下の通りである．

b_{j1}：「1」に対して出現率が50%になる能力の位置

b_{j2}：「1」と「2」に対する出現率が同じになる能力の位置

b_{j3}：「3」に対して出現率が50% になる能力の位置

b_{jc} について式（4）（5）（6）に示す．

$$\tau_{j1}=b_{j1} \qquad (4)$$

$$\frac{\tau_{j1}+\tau_{j2}}{2} = b_{j2} \qquad (5)$$

$$\tau_{j3}=b_{j3} \qquad (6)$$

3件法の場合，それぞれのカテゴリを選択する確率は式（7）（8）（9）から求められる．

項目 j で「1. いいえ」(c=1) を選択する確率：$p_{j1}(\theta)$

$$p_{j1}(\theta)=p^*_{j1}(\theta) - p^*_{j2}(\theta)= 1-\frac{1}{1+exp[-a_j(\theta - b_{j2})]} \qquad (7)$$

項目 j で「2. どちらでもない」(c=2) を選択する確率：$p_{j2}(\theta)$

$$p_{j2}(\theta)=p^*_{j2}(\theta) - p^*_{j3}(\theta)= \frac{1}{1+exp[-a_j(\theta - b_{j2})]} - \frac{1}{1+exp[-a_j(\theta - b_{j3})]} \qquad (8)$$

項目 j で「3. どちらでもない」(c=3) を選択する確率：$p_{j3}(\theta)$

$$p_{j3}(\theta)=p^*_{j3}(\theta) - p^*_{j4}(\theta)= \frac{1}{1+exp[-a_j(\theta - b_{j3})]} \qquad (9)$$

式（7）（8）（9）を3件法でシミュレーションしたのが図1である．図1は IRCCC（Item Response Category Characteristic Curve：項目カテゴリ反応曲線）である．横軸が能力（θ）であり，縦軸が出現する確率（probability）である．各図の一番左の曲線は「1 思わない」(C1)，2番目が「2 どちらでもない」(C2)，一番右が「3 思わない」(C3) である．各カテゴリがテストの得点であると考えると，能力が低い（θ が低い）者は，1点を得る確率が高くなるが，能力が高い（θ が高い）者は，3点を得

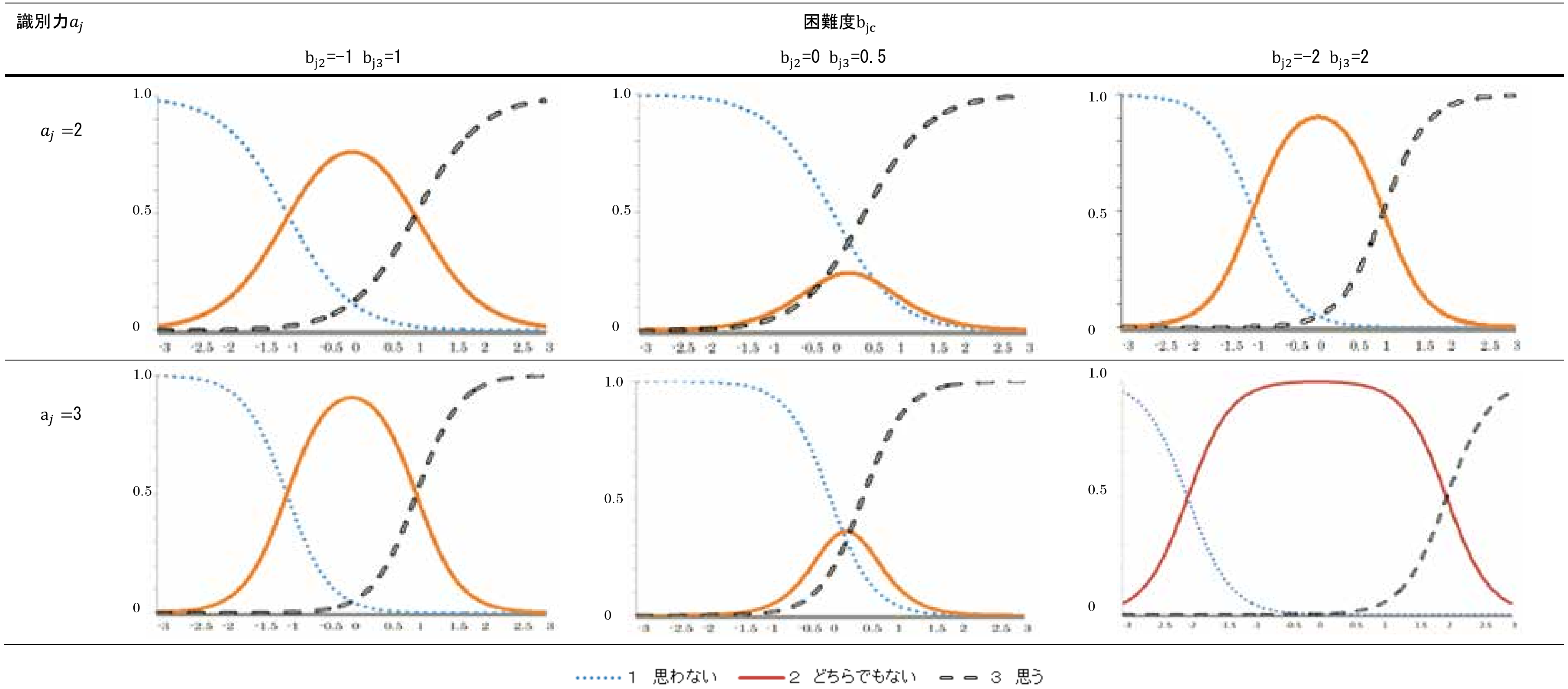


図１　3件法の IRCCC（Item Response Category Characteristic Curve: 項目カテゴリ反応曲線）シミュレーション

$p_{jc}(\theta)=\frac{1}{1+exp[-a_j(\theta-b_{jc})]}-\frac{1}{1+exp[-a_j(\theta-b_{jc+1})]}$ のシミュレーションである．各図の横軸は能力値(θ)を，縦軸は各カテゴリの出現率（Probability）を意味する．各図の曲線はカテゴリの出現率である．項目反応理論では能力値が高くなるにつれて，「1思わない」（C1）「2　どちらでもない」（C2）「3　思う」（C3）の順に，それぞれの出現率が高くなると考える．θが低い者は，C1を選択する確率が高くなるが，θが高い者は，C3を選択する確率が高くなる．

る確率が高くなる．C1は，θ が高くなるにつれて，出現率が低くなっていく．反対に，C3は，θ が高くなるにつれて，出現率が高くなっていく．

カテゴリごとの困難度が均等になればなるほど，各カテゴリが等間隔に出現する．つまり，尺度の等間隔性を保証できる．IRCCC では各曲線が左から均等にカテゴリ順に並んでいるとき，尺度が等間隔であると判断される．例えば，3件法では，C1とC2，C2と C3の交点が左から順に，また均等に並んでいる場合である．

３．５．２　カテゴリカル因子分析

IRT を用いる際，カテゴリカル因子分析により因子構造が1因子構造であることを確認した上で分析が行われることになる．Mplus でカテゴリカル因子分析を行う際，因子軸の回転は，ジュオミン（Geomin）回転が初期設定になっており，順序尺度の分析に適している．パラメタの推定は順序尺度の場合，多分相関係数（Polychoric Correlation）に基づく WLSMV（Weighted Least Squares estimation with Mean and Variance：重み付けのある最小二乗法）が一般に用いられるとされている．

因子構造について，IRT は1因子構造が分析条件であることから，すべての尺度において，第1固有値から第2固有値への減退率が大きいこと，かつ CFI（Comparative Fit Index）>0.95，TLI（Tucker-Levis index）>0.95，RMSEA（Root Mean Square Error of Approximation）<0.08という因子構造を満たす必要がある．CFI はモデルの適合度を示す指標であり，0～1.0の範囲を取る．TLI はモデルの適合度を示す指標であり，0～1.0の範囲を取る．RMSEA はモデル分布と真の分布との乖離を示す指標であり，0～∞の範囲を取る．

３．６　倫理的配慮

回答は無記名であり，本調査への回答が成績等に全く影響をしないこと，調査結果は学術目的にのみ使用し，調査に協力できない場合は，白紙にて返却する旨を伝えた．

４．調査結果

４．１　回答者の属性

回答者の約6割が女性であった（表2）．学年は不明であるが，大半が21～22歳である．

４．２　3件法，4件法，5件法質問票の基礎集計

3件法，4件法，5件法質問票のそれぞれの度数分布，及び基礎統計量として（算術）平均値，中央値，標準偏差を表1に示した．順序尺度の場合，基礎統計量を計算できないが，比較のために計算した．すべての質問について，5件法が最も標準偏差が大きく，

表２　対象者の性別

	人数	(%)
男性	32	(36.8)
女性	55	(63.2)
合計	87	(100.0)

無回答：2人

表３　固有値（λ）の減退状況及び因子負荷量（カテゴリカル因子分析）

	3件法	4件法	5件法
λ_1	2.970	2.506	2.772
λ_2	0.485	0.745	0.638
λ_1/λ_2	6.124	3.363	4.344
χ^2	331.142	232.649	430.567
CFI	0.966	1.000	1.000
TLI	0.999	1.016	1.009
RMSEA	0.035	0.000	0.000
ROTATED LOADINGS			
7	0.856	0.936	0.866
10	0.793	0.734	0.755
11	0.740	0.481	0.549
12	0.874	0.681	0.897

7, 10, 11, 12は質問項目番号．表1参照．λ_1/λ_2：固有値の減退状況．CFI：Comparative Fit Index．モデルの適合度を示す指標．0～1.0の範囲を取り，0.95以上がモデルの基準となる．TLI：Tucker-Levis index．モデルの適合度を示す指標．0～1.0の範囲を取り，0.95以上がモデルの基準となる．1.0を超える場合もある．RMSEA: Root Mean Square Error of Approximation. モデル分布と真の分布との乖離を示す指標．0～∞の範囲を取り，0.05以下がモデルの基準となる．ROTATED LOADINGS：回転後の因子負荷量．

3件法が最も小さかった．

４．３　IRT による分析結果

3件法，4件法，5件法質問票について，先述の方法・基準によるカテゴリカル因子分析を行った．その結果，質問項目7，10，11，12の4つの変数の組合せが，1因子構造として最適であることが認められた（表3）．それ以外の質問項目は，IRT の分析では除外された．表4はIRT に基づく各項目の識別力，閾値（τ），困難度（b）である．図2～4はそれらの IRCCC である．3件法は3本，4件法は4本，5件法は5本の曲線が描かれる．図及び表中の質問項目はカテゴリカル因子分析において IRT の分析モデルとして妥当であると認められた質問項目である7，10，11，12のみを記載した．

識別力について，Roznowski は0.5未満の項目[22]，豊田は0.2以下の項目は除去すべきであると指摘している[23]．本稿の場合，いずれの場合でも基準を満たした．困難度について豊田は絶対値が4.0を超え

表４　各項目の識別力・閾値（τ）・困難度（b）（項目反応理論）

質問項目	尺度	識別力	SE	τ_1	τ_2	τ_3	τ_4	b_1	b_2	b_3	b_4	b_5
7	3	2.825	0.858	-1.735	0.995			-1.735	-0.370	0.995		
	4	4.320	2.473	-4.649	1.082	5.876		-4.649	-1.784	3.479	5.876	
	5	3.010	0.651	-4.004	-0.341	1.488	4.609	-4.004	-2.173	0.574	3.049	4.609
10	3	2.300	0.577	-1.576	0.775			-1.576	-0.401	0.775		
	4	1.926	0.464	-3.003	-0.296	2.485		-3.003	-1.650	1.095	2.485	
	5	2.189	0.497	-3.525	-0.870	0.200	2.491	-3.525	-2.198	-0.335	1.346	2.491
11	3	0.993	-1.303	1.003	3.220			1.003	2.112	3.220		
	4	0.915	0.314	-1.292	1.008	3.225		-1.292	-0.142	2.117	3.225	
	5	1.280	0.381	-1.557	0.433	1.245	3.475	-1.557	-0.562	0.839	2.360	3.475
12	3	1.848	0.480	-1.101	1.899			-1.101	0.399	1.899		
	4	1.749	0.590	-3.247	-0.254	2.623		-3.247	-1.751	1.185	2.623	
	5	4.039	1.247	-5.548	-1.870	0.782	4.993	-5.548	-3.709	-0.544	2.888	4.993

カテゴリカル因子分析において，項目反応理論の因子構造として妥当であった質問項目7,10,11,12についてのみ記載した．SE: 標準誤差　3件法：τ_1= b_1, (τ_1＋ τ_2)/2=b_2, τ_3= b_3. 4件法：τ_1= b_1, (τ_1＋ τ_2)/2=b_2, (τ_2＋τ_3)/2=b_3, τ_4=b_4. 5件法：τ_1= b_1, (τ_1＋ τ_2)/2=b_2, (τ_2＋τ_3)/2=b_3, (τ_3＋τ_4)/2=b_4, τ_5=b_5. 識別力：Discrimination parameter もしくは Slope parameter. 各項目が，どの程度対象者の能力 (θ) を反映しているかを示す指標．困難度：Difficulty parameter. あるカテゴリと次のカテゴリとの出現率が50% になる位置を示す.

る項目は除去すべきであるとした．本稿の場合，質問7の4件法，5件法，質問12の5件法がこの基準に該当する．以上から以下の知見が導き出された.

1）3件法（図２）

質問項目7・10･11･12：θ が上がるにつれ，C1の出現率が下がり，C3の出現率が上がっている．すべての変数において各カテゴリの曲線が C1，C2，C3の順に，かつ均等に存在していた．ここからすべての変数に等間隔性が認められた.

2）4件法（図３）

・質問項目7:先述の基準により分析から除外された.

・質問項目10：各カテゴリの曲線が C1，C2，C3，C4の順に，かつ均等に存在したことから等間隔性が認められた.

・質問項目11：C3と C4の交点が大きく右にはみ出しているため，θ に C1から C4が等間隔に反応していないと認められた．ただし θ =3.5では C3と C4の交点がみられないが，θ >3.5のいずれかの段階で交点が生じる可能性がある．ここから等間隔性が認められるかどうかは判断できなかった.

・質問項目12：各カテゴリの曲線が C1，C2，C3，C4の順にかつ均等に存在したことから等間隔性は認められた.

3）5件法（図５）

・質問項目7:先述の基準により分析から除外された.

・質問項目10：C3と C4の交点が C2と C4の交点より左に認められた．各カテゴリ曲線が順に出現しなかったため，等間隔性は認められなかった.

・質問項目11：C3と C4の交点が，C2と C3の交点よりも左に認められた．各カテゴリ曲線が順に出現しなかったため，等間隔性は認められなかった.

・質問項目12:先述の基準により分析から除外された.

３．４　性別と最も回答がしやすかった尺度とのクロス集計表

回答のしやすさを表5に示した．4件法が最も回答しやすかったと回答した者は約11% であった．そこで4件法質問紙を除き，性別によって3件法，5件法について最も回答しやすかった質問紙に違いがあるかを明らかにするためにクロス集計表を作成し，χ^2検定を行った（表6)．その結果，性別と回答のしやすさに統計学的な差が認められた.

さらに，男性，女性それぞれの中で，最も回答がしやすかった尺度に違いがあるかどうかを明らかにするために，χ^2検定を行った．その結果，男性についてはその違いは認められなかったが（χ^2 = 0.040,p=N.S.)，女性には違いが認められた（χ^2 = 8.395,p<0.01).

以上から，特に男性については，最も回答がしやすかった質問紙に違いはないが，女性については3件法質問紙よりも5件法質問紙の方が最も回答しやすかったと回答する傾向があった.

５．考察

3件法はすべての質問項目，4件法は2変数に等間隔性が認められたが，5件法はすべての変数に等間隔性が認められなかった.

3件法については，どの質問項目についても尺度の等間隔性が認められたことから，順序尺度ではな

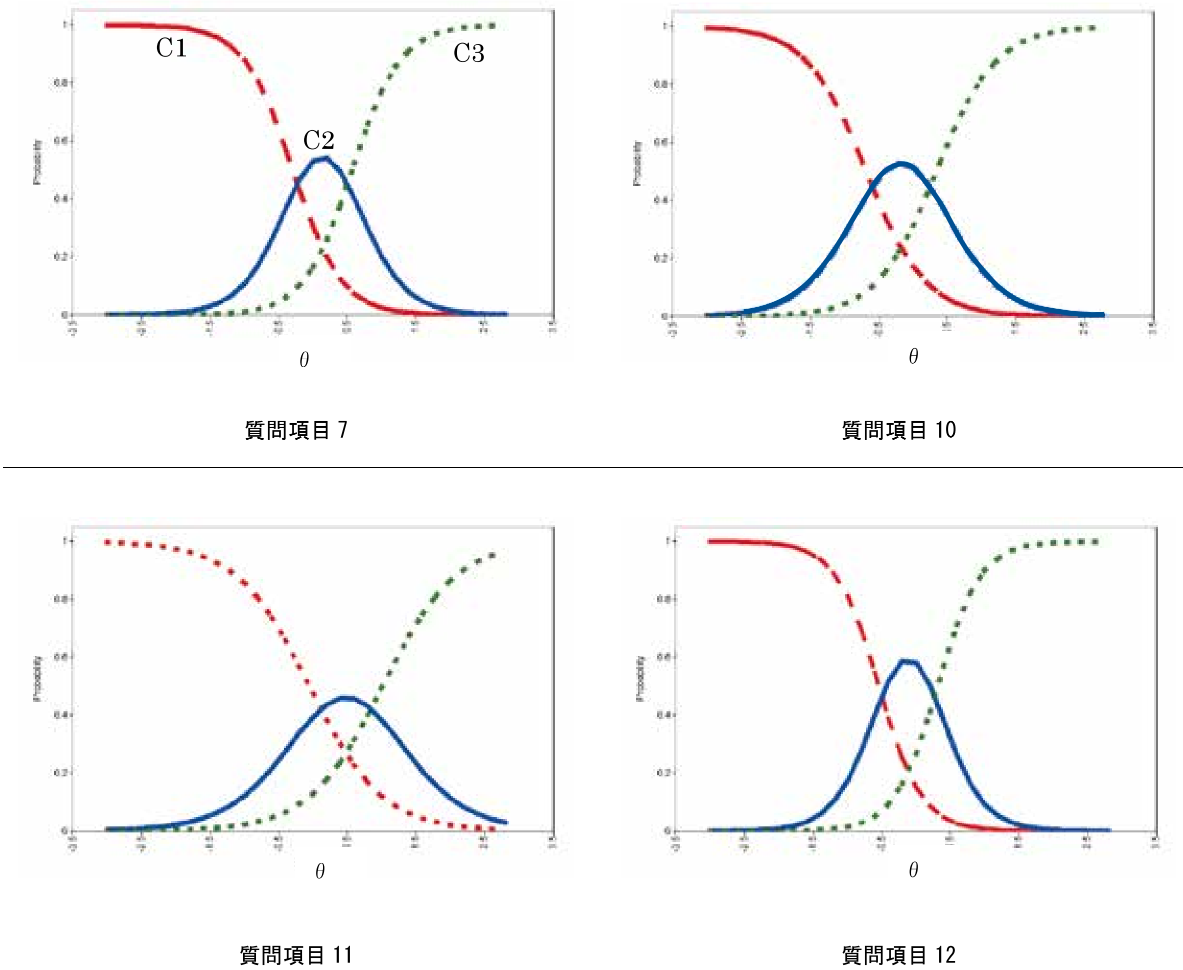


図２　質問項目ごとのIRCCC（項目カテゴリ反応曲線）－3件法

カテゴリカル因子分析において，項目反応理論の因子構造として妥当であった質問項目7,10,11,12についてのIRCCCのみを記載した．それ以外の質問項目は分析対象外である．各図の縦軸，横軸及び曲線の解釈は図1に準ずる．「C1　当てはまらない」「C2　どちらでもない」「C3　当てはまる」．C1からC3の各カテゴリの交点が左から順に，また均等である場合，等間隔であると認める．

表５　最も回答がしやすかった尺度（3～5件法）

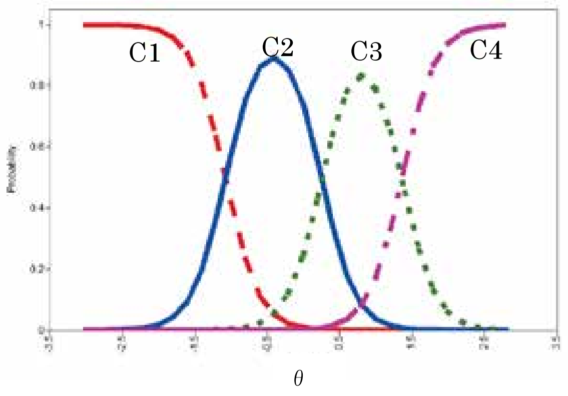
	人数	(%)
5-4-3	23	(29.9)
5-3-4	20	(26.0)
3-4-5	13	(16.9)
3-5-4	12	(15.6)
4-5-3	5	(6.5)
4-3-5	4	(5.2)
合計	77	(100.0)

3桁の数字は回答のしやすさを順に並べたものである．たとえば「5-4-3」は5件法が最も回答しやすく，3が最も回答しにくかったことを意味する．無回答 :12人

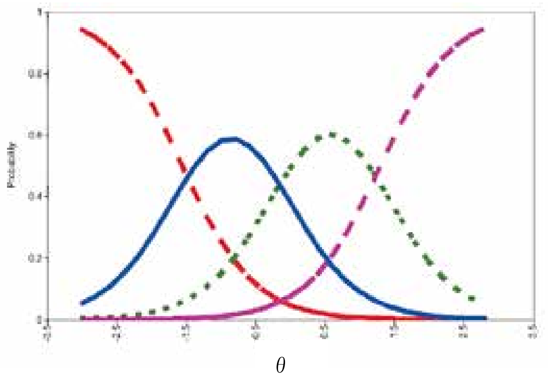
表６　性別と最も回答がしやすかった尺度（3件法・5件法）とのクロス表

	最も回答がしやすかった尺度		
	3件法	5件法	合計
男性	13 (52.0)	12 (48.0)	25 (100.0)
女性	12 (27.9)	31 (72.1)	43 (100.0)
合計	25 (36.8)	43 (63.2)	68 (100.0)

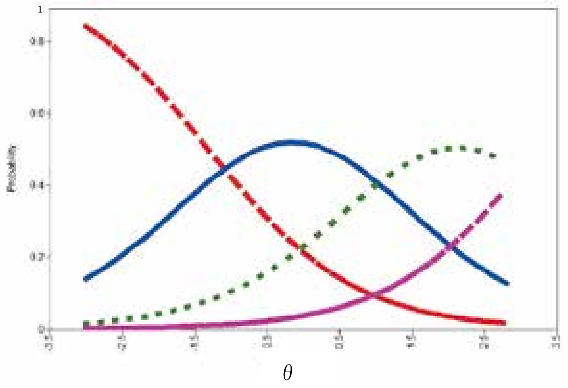
4件法は少数のため除去した．数字は人数，単位は人である．括弧内の数字は％である．χ^2=3.947, p<0.05



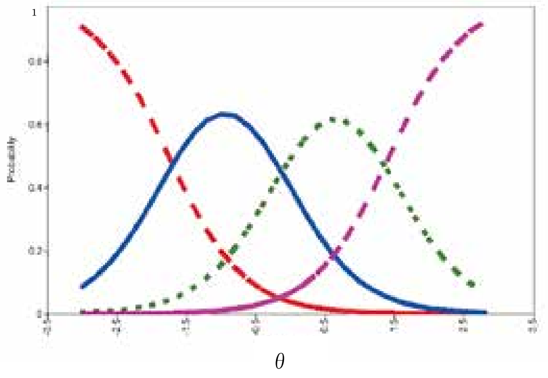
質問項目 7



質問項目 10



質問項目 11



質問項目 12

図３　質問項目ごとのIRCCC（項目カテゴリ反応曲線）－4件法

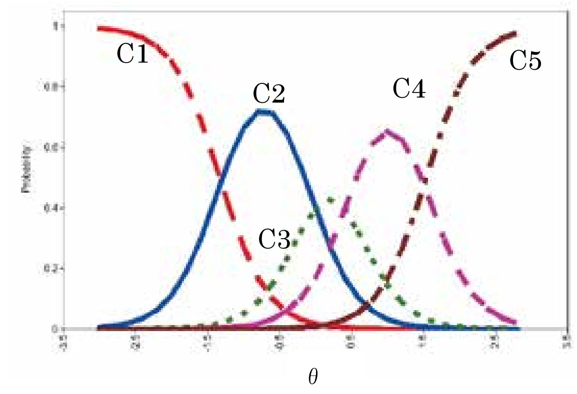
各図の縦軸，横軸及び曲線の解釈は図1に準ずる．質問項目の選定は図2に準ずる．「C1　当てはまらない」「C2　あまり当てはまらない」「C3　少し当てはまる」「C4　当てはまる」．C1からC4の各カテゴリの交点が左から順に，また均等である場合，等間隔であると認める．

く間隔尺度とみなせる可能性が示唆された．一方，尺度が4件法,5件法になるにつれて,質問項目によって尺度の等間隔性が担保できないという結果になった．ここから，特に5件法の分析時には，間隔尺度ではなく，順序尺度として扱う方が妥当であることが示唆された．

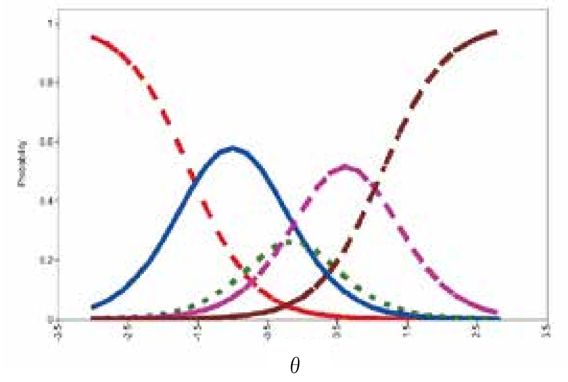
ただし，困難度の絶対値が4未満であるという基準は満たしていないが，識別力の視点から分析すると，質問項目7の4件法，5件法，質問項目12の5件法の識別力が高く，各カテゴリの交点が順に左から均等に存在していた．この点から，これらの質問項目が安易に順序尺度とはいえない．間隔尺度の場合，その分析方法はノンパラメトリック手法からｔ検定や分散分析といったパラメトリック手法まで多様な手法を用いることができる．一方, 順序尺度の場合，ノンパラメトリック手法に限定され，パラメトリック手法を用いることはできない．

性別と最も回答がしやすかった尺度とについて，ジェンダー（社会的性差）の影響があると推測できる．しかし本質問紙から詳細な分析をすることは不可能である．本調査が20歳前半の学生を対象にしていることから，すべてを普遍化することはできないが，調査対象者に女性が多い場合と，男性が多い場合とでは回答のしやすさに少なからず影響を及ぼす可能性が示唆できよう．

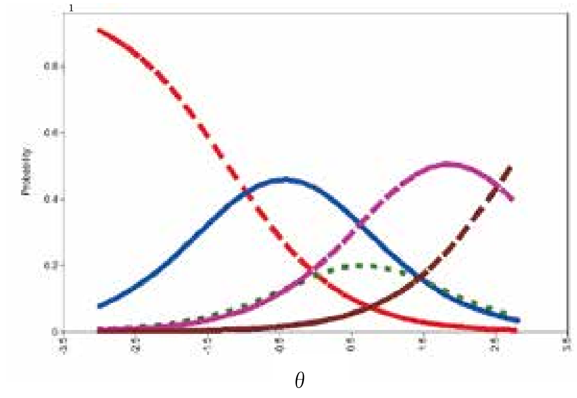
ただし，対象者の多くが女性であるとして回答がしやすい5件法調査票のみを用いることは，尺度の等間隔性という観点から分析時に分析手法が限定される可能性がある．なぜならば，先述の通り，5件法は必ずしも等間隔性を担保できないからである．



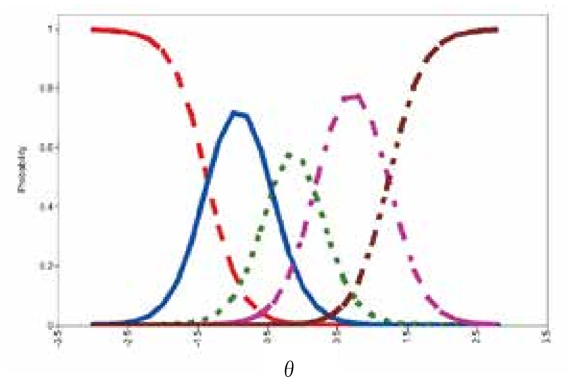
質問項目 7



質問項目 10



質問項目 11



質問項目 12

図４　質問項目ごとの IRCCC（項目カテゴリ反応曲線）－5件法

各図の縦軸，横軸及び曲線の解釈は図1に準ずる．質問項目の選定は図2に準ずる．「C1　当てはまらない」「C2　あまり当てはまらない」「C3　どちらでもない」「C4　少し当てはまる」「C5　当てはまる」．C1から C5の各カテゴリの交点が左から順に，また均等である場合，等間隔であると認める．

等間隔性が認められない場合，通常，例えば2群間の比較を行う際は，t 検定ではなく Mann-Whitney 検定を用いる必要がある．にもかかわらず，5件法調査票を用いた調査では，t 検定がよく用いられる．回答のしやすさと統計学的な妥当性との間に乖離が生じる可能性があり，パラメトリック手法が多く用いられる現状には課題が残るといえる．

両性別ともに4件法が最も回答しやすいと回答した者は少なかった．これは日本人という社会的性格が影響しているともいえる．国際比較調査において，欧米では5件法が多く用いられるが，日本では中心点がない4件法が用いられる．それは，日本人は中心点の「どちらでもない」に回答が集中しやすいという指摘があるからである[24]．これは逆にとれば，3件法や5件法が回答しやすいことを示唆する．この点から，本調査の結果は妥当であるといえるが，女性について，なぜ3件法より5件法の方が回答しやすいかは推測できない．

6．結語

IRT の観点から3～5件法質問紙における変数の等間隔性について分析を行った．その結果，第1に3件法尺度について等間隔性が認められた．ここから，より少ない選択肢（カテゴリ）数で質問紙調査を行い，間隔尺度とみなすことが統計学的に妥当であること，さらに「2. どちらでもない」という尺度の中心点が存在することで，日本人にとってより回答がしやすくなることが確認された．第2に性別によっ

て回答のしやすさが異なっていた.

第3に対象者の多くが女性であるとして回答のしやすい5件法調査票のみを用いることは，尺度の等間隔性という観点から分析時に分析手法が限定される可能性がある．にもかかわらず，パラメトリック手法が多く用いられる現状は否定しえない.

7．研究の限界と課題

第1に，本調査は調査対象が20歳前半の学生であるため，そのバイアスを排除できない．第2に3件法，4件法，5件法質問紙の回答の違いを明確にできる最適な質問であるとは科学的にいえず，普遍化しきれない．さらに回答のしやすさに対して，なぜその質問紙が回答しやすかったかについて直接的に尋ねていない．第3にカテゴリカル因子分析，IRT の分析について，ある程度大きい標本数が必要であるが，本調査では100以下の標本数における分析である．第4に尺度の正規性と等間隔性との関連について議論する必要がある.

以上の課題は，適切な層化抽出を行った大規模な調査を行うことで解決されるものであると考える.

文　献

1) Carifio J and Perla R：Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*，**42**(12)，1150－1152，2008.

2) Pearson ES，D'Agostino RB and Bowman KO：Test for normality：Comparison of powers. *Biometrika*，**64**(2)，231－246，1977.

3) 西里静彦：データ解析への洞察―数量化の存在理由．初版，関西学院出版会，兵庫，**25**，2007.

4) 狩野裕：AMOS，EQS，LISREL によるグラフィカル多変量解析．初版，現代数学社，京都，**156**，1997.

5) 脇田貴文：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**(4)，331－338，2004.

6) 鈴木綾子，豊田秀樹，小杉正太郎：項目反応モデルによるストレス反応尺度の構成とテスト特性曲線によるその深化の課程．心理学研究，**75**，389－396，2004.

7) 田中健吾：勤労者を対象として心理的ストレス反応尺度の項目反応理論による検討．大阪経大論集，**63**(3)，137－150，2012.

8) 笹川智子，金井嘉宏，村中泰子，鈴木伸一，嶋田洋徳，坂野雄二：他者からの否定的評価に対する社会的不安測定尺度（FNE）短縮版作成の試み：項目反応理論による検討．行動療法研究，**30**(2)，87－98，2004.

9) 狩野京子，李志嬉，中島望，實金栄，山口三重子，中嶋和夫：看護職者の「職業キャリア成熟測定尺度」に関する構成概念妥当性の検討．岡山県立大学保健福祉学部紀要，**19**(1)，19－29，2012.

10) 八重樫牧子：児童館の子育ち子育て支援．初版，相川書房，東京，2012.

11) 加藤正明（班長）：労働の場におけるストレス及びその健康影響に関する研究報告書 労働省「作業関連疾患の予防に関する研究」平成11年度．労働省，2000.

12) 安藤昌代：面接調査における回収率向上のための一考察－対象者アンケート 調査員アンケートの結果から－．社会と調査，**9**，76－89，2012.

13) 井上信次，岡本宣雄：特別養護老人ホームの待機者調査にみる待機者及びその介護者へのアウトリーチの必要性，川崎医療福祉学会誌，**20**(2)，331－345，2011.

14) アントノフスキー，A．山崎喜比古，吉井清子監訳：健康の謎を解く－ストレス対処と健康保持のメカニズム．初版，有信堂高文社，東京，2012.

15) 松原望，美添泰人，岩崎学，金明哲，竹村和久，林文，山岡和枝編：統計応用の百科事典，初版，丸善出版，東京，436，2011.

16) 並川努，谷伊織，脇田貴文，熊谷龍一，中根愛，野口裕之：Big Five 尺度短縮版の開発と信頼性と妥当性の検討．心理学研究，**83**，91－99，2012.

17) 時光順平，鳥越規央：項目反応理論を用いたプロ野球選手の評価について（統計的モデルの新たな展望とそれに関連する話題）．数理解析研究所講究録，**1804**，21－29，2012.

18) 矢冨直美，渡辺直登：心理的ストレス反応尺度（PSRS）の分析．経営行動科学，**10**(1)，23－34，1995.

19) 尾崎幸謙：テストのための数理モデル．統計数理研究所コラム110，http：//www.ism.ac.jp/ism_info_j/labo/column/110.html（2015.5.16）

20) 住政二郎：項目反応理論－1PLM，2PLM，3PLM，多段階反応モデル－．外国語教育メディア学会（LET）関西支部メソドロジー研究部会2013 年度報告論集，34－62，2013

21）豊田秀樹：項目反応理論［入門編］．初版，朝倉書店，東京，5，2002.
22）Roznowski，M：Examination of the Measurement properties of the job descriptive index with experimental items. *Journal of Applied Psychology*，74，805－814，1989.
23）豊田秀樹：項目反応理論［事例編］．初版，朝倉書店，東京，**29**，2002.
24）岩井紀子，宍戸邦章，佐々木尚之：East Asian Social Survey を通してみた国際比較調査の困難と課題．社会と調査，**7**，18－25，2011.

（平成27年６月15日受理）

Equal Distance Between Categories in Ordinal Scale on Item Response Theory – equal distance of 3, 4 and 5 category scale questions and easily answerable in the survey questionnaire

Shinji INOUE

(Accepted Jun. 15, 2015)

Key words : distance between categories, ordinal scale, Item Response Theory

Abstract

The purpose of this paper is to clarify the number of the scale category which satisfies the criteria of being easily answerable and producing statistical valid answers in survey questionnaires. The author recognizes that it is ideal to decrease the burden of answering and the number of the scale category when we survey users of medical welfare.

The questionnaire was distributed to 130 students (Z faculty, A university), of which 89 (86.4%) submitted valid responses. The questionnaire consisted of 3-category, 4-category and 5-category scale questions; and a question with respect to whether the questionnaire was easily answerable.

Item response theory (ITR) results revealed 3 conclusions. First, 3-category scales showed the same distinction between categories. Second, the ease of answerability was different for males and females. Third, the method of analysis will be limited if we use only 5-category scale questions because most of the respondents are female. Nevertheless, there remain many issues in this situation. Thus, parametric methods are commonly employed.

Correspondence to : Shinji INOUE

Department of Social Work
Faculty of Health and Welfare
Kawasaki University of Medical Welfare
Kurashiki, 701-0193, Japan
E-mail : inoshin@mw.kawasaki-m.ac.jp
(Kawasaki Medical Welfare Journal Vol.25, No.1, 2015 23－35)