

項目反応理論による安全風土調査の項目分析

An Item Analysis of Safety Climate Survey Made Utilizing Item Response Theory

西田 豊 (Yutaka Nishida) *1

要約 組織における安全風土や安全文化の状態を評価する方法として、質問紙によって測定する方法がある。安全風土のような実体のない構成概念を測定しようとするとき、使用する尺度が測ろうとしている対象を安定的に測定可能かどうかを検証しなくてはならない。また、尺度をより良い精度のものに改善しようとするとき、各質問項目がどのような特徴（たとえば項目困難度や項目識別力）を持っているかを把握しておかなければならない。本研究では項目反応理論を適用し、質問項目の特徴を明らかにするとともに、尺度の精度向上に資するテスト情報量、ならびに項目情報量を求めた。

キーワード 安全風土, 安全文化, 項目反応理論, 項目分析, 評価

Abstract There is a way to assess the safety climate or safety culture conditions in organizations by using a questionnaire survey. When we try to measure intangible constructs such as safety climate, we have to make sure that the scale used in the assessment measurement is reliable. Also, when we try to improve the questionnaire to get better accuracy, we need to understand what features (e.g., item difficulty and item discrimination) each question has. In this research, we used item response theory to clarify the features of the questions and to determine the test information and the item information that contribute to the improvement of the scale accuracy.

Keywords safety climate, safety culture, item response theory, item analysis, assessment

1. はじめに

1.1 安全風土とその測定

組織における安全確保のためには設備、機器など技術的要因に加え、人的及び組織的要因のマネジメントが重要であることが指摘されている（山本・関村, 2017）。近年の組織における安全意識の高まりにともなって事業者は様々な安全文化醸成活動を実施し、評価を行っている。

評価法の1つとして質問紙調査による安全文化、および安全風土の状態の測定があり（Zohar, 1980）、多くの研究が実施されている（西田, 2017）。原子力産業分野でも盛んに安全風土に関する研究がなされており（福井・吉田・山浦, 2000; 福井・吉田・吉山, 2001; 福井, 2012）、近年では回答者の属性に関する回答傾向の違いや（藤田, 2017;

福井, 2012; 西田, 2018）、属性による影響の程度が事業所間で異なること（藤田, 2018）が報告されている。

物理的な存在ではなく、安全風土のような構成概念を測定するためには、正確に安全風土の状態を測ることのできる尺度が必要となる。例えば、福井（2012）が開発した尺度は、信頼性と妥当性が確認されており、測りたいものを安定して測定できる尺度だと考えることができる。

既存の尺度をより精度の高い尺度にしていくためには、測定したい概念と異なるものを測定している項目を除いたり、十分な信頼性を持つ尺度にするため項目を増やしたりといった項目の編集が必要になる。そのためには、尺度を構成する質問項目が実施者が期待するような機能を果たしているかどうかを確認し、どのような特徴を持っているか把握しておかなければならない。このような目的のために行わ

*1 (株)原子力安全システム研究所 社会システム研究所

れる統計的な分析を項目分析という。

1.2 項目分析

項目分析は主に2つの観点から行われる。1つは項目困難度、もう1つは項目識別力である。従来は回答者の得点(測定値)を真値と誤差の和と考える古典的テスト理論(Lord & Novick, 1968; Novick, 1966)と呼ばれる枠組みで項目分析や信頼性の検討が実施されてきた。

項目困難度とは、ある項目に回答することの難しさを表す指標である。古典的テスト理論においては各項目における全回答者の平均得点で評価できる。平均得点が低ければ難しい項目であり、平均得点が高ければ易しい項目と解釈できる。

項目識別力とは、ある項目が、測定している特性の高い人と低い人を識別できるかどうかを表す指標である。古典的テスト理論においては項目得点と全項目の合計得点との相関係数で評価できる。これを項目-全体相関、もしくはI-T相関と呼ぶ。

信頼性とは、測定結果の安定性のことを意味する。同一尺度に対して同一人物が同一条件で回答した場合、同じ得点になるとその尺度の信頼性は高いと考えられる。古典的テスト理論において信頼性は測定値の分散に占める真値の分散の割合と定義される。真値は得られないため推定することになるが、信頼性係数の推定値としてクロンバックの α (Cronbach, 1951)が使われることが多い(岡田, 2015; 高本・服部, 2015)。

安全風土調査で用いられた各質問項目がどのような特徴を持っているかの検討を行うことが本研究の目的である。しかしながら、古典的テスト理論にはいくつかの問題点が指摘されている(村木, 2011)。1つは困難度や識別力といった項目に関する指標が、回答者集団に依存するという標本依存性の問題である。もう1つは尺度得点などの回答者の特性に関する情報が尺度項目に依存するという項目依存性の問題である。またクロンバックの α に対しても信頼性を過少推定してしまうことが指摘されており、その使用は推奨されない(岡田, 2011)。

2. 項目反応理論

近年では古典的テスト理論の問題点を解消した現代テスト理論と位置づけられる項目反応理論(Lord,

1952; Lord & Novick, 1968; 池田, 1994) が利用される。海外においてはTOEFLをはじめとする試験で広く用いられてきた。日本においてはIT パスポート試験、日本語能力試験など用いられることが増えている。項目反応理論の特徴は、回答者集団の性質には依存しない潜在特性を導入し、項目に関する指標と回答者に関する指標を切り離して推定することにある。

また、古典的テスト理論のもとでは測定の精度が信頼性係数であらわされるが、これは尺度全体の平均的な精度を評価する指標であった。しかし、ある尺度がすべての回答者に対して同じ精度を持つわけではない。一方、項目反応理論において、測定精度はテスト情報量や項目情報量として表現され、尺度全体もしくは項目ごとに示すことができる。これらの指標は潜在特性を表す連続尺度上の特性との関数として表現され、特性値がどのくらいの値をとるときに高い精度で測定が可能であるのかを示すこともできる。テスト情報量や項目情報量を用いれば、情報量の少ない項目を削ることにより、なるべく測定精度を落とさずに少ない項目数で運用できる短縮版の尺度を構成することが可能になる。

もともと項目反応理論は教育分野における試験などの2値データ(正答か誤答か)を扱う分析手法であったが、心理学研究で多用される性格特性や態度などの質問紙尺度から得られる多値データを扱うことのできるモデルも開発されている(Masters, 1982; Muraki, 1992; Samejima, 1969)。現在では組織心理測定にも利用され(渡辺・野口, 1999; 野口・渡辺, 2015)、尺度の分析や開発において広く用いられている。

3. 方法

3.1 データ

安全に関する組織の状態や従業員の意識について尋ねている、2018年度に実施された質問紙調査のデータを用いる。この調査で用いた質問項目は7つの下位尺度(本調査では原則と呼んでいる)を含んでおり、全78項目である。各原則は「安全最優先の価値観」、「リーダーシップ」、「安全確保の仕組み」、「円滑なコミュニケーション」、「個人・組織の姿勢」、「潜在的リスクの認識」、「活気のある職場環境」の概念を測定するものであり、それぞれが8~16項

目によって構成されている。各質問項目は「そう思わない」から「そう思う」の5件法のリッカートスケールによって測定され、1点から5点が付与される。

本研究では、全回答者のうち電力会社に所属し原子力発電所用の質問項目に回答した9026名分のデータを分析対象としている。

3.2 分析モデル

項目反応理論には多くの統計モデルが提案されているが、本研究では、順序尺度データに適用可能な段階反応モデル (Samejima, 1969) を用いた。段階反応モデルでは回答者 j が項目 i に対して、 K 個の反応カテゴリからカテゴリ k と反応する確率 p_{ijk} を k 以上のカテゴリが選択される確率 p_{ijk}^* と $k+1$ 以上のカテゴリが選択される確率 p_{ijk+1}^* の差分によって

$$p_{ijk} = p_{ijk}^* - p_{ijk+1}^*,$$

$$p_{ijk}^* = \begin{cases} 1 & (k = 1), \\ \frac{\exp(\alpha_i(\theta_j - b_{ik}))}{1 + \exp(\alpha_i(\theta_j - b_{ik}))} & (k = 2, \dots, K), \\ 0 & (k = K + 1). \end{cases}$$

と表す。

ここで θ_j は回答者 j の潜在特性を、 b_{ik} は項目 i において k 以上のカテゴリに回答する困難度を、 α_i は項目 i が潜在特性値 $\theta = b_{ik}$ の特性をどの程度識別できるかという識別力を意味する。また、1 以上のカテゴリが選択される確率は必ず選択されるため $p_{ij1}^* = 1$ 、 $K+1$ 以上のカテゴリが選択される確率は存在しないカテゴリであるため $p_{ijK+1}^* = 0$ としている。困難度パラメータ b_{ik} には順序制約 $b_{i2} < b_{i3} < \dots < b_{iK}$ が課されている。

本研究のように回答選択肢が5カテゴリである場合、それぞれの選択肢が選択される確率は

$$\begin{aligned} p_{ij1} &= p_{ij1}^* - p_{ij2}^* = 1 - p_{ij2}^* \\ p_{ij2} &= p_{ij2}^* - p_{ij3}^* \\ p_{ij3} &= p_{ij3}^* - p_{ij4}^* \\ p_{ij4} &= p_{ij4}^* - p_{ij5}^* \\ p_{ij5} &= p_{ij5}^* - p_{ij6}^* = p_{ij5}^* - 0 \end{aligned}$$

となる。

例として困難度が $b_1 = [-2.11, -0.63, 0.64, 2.20]$ 、

識別力が $\alpha_1 = 1.02$ である項目 ex1 と、困難度が $b_2 = [-2.00, -1.13, -0.15, 0.85]$ 、識別力が $\alpha_2 = 4.30$ である項目 ex2 について、各パラメータがどのように解釈されるかを示す。横軸に回答者の潜在特性 θ 、縦軸に選択肢 k に対する反応確率の曲線を描いたものを項目反応カテゴリ特性曲線 (IRCCC) という (図 1)。

困難度パラメータは図 1 (左) において曲線の横軸の位置を表すパラメータであり、 $b_{12} = -2.11$ は選択肢 1 と回答する確率が 0.5 になる特性の位置を、 $b_{13} = -0.63$ は選択肢 2 と回答する確率と選択肢 3 と回答する確率と同じになる特性の位置を、 $b_{14} = 0.64$ は選択肢 3 と回答する確率と選択肢 4 と回答する確率と同じになる特性の位置を、 $b_{15} = 2.20$ は選択肢 5 と回答する確率が 0.5 になる特性の位置を意味する。

識別力パラメータは曲線の変化率を表すパラメータであり、 α_1 の値が小さい項目 ex1 はなだらか山形を描いているのに対し (図 1 : 左)、 α_2 の値が大きい項目 ex2 はとがった山形を描いている (図 1 : 右)。

潜在特性値が高くなればなるほど高い選択肢に反応する確率が高くなるのがわかる。たとえば選択肢 3 への回答確率をしてみると項目 ex1 においては潜在特性値が 0 付近で最も反応確率が高くなっているのに対し、項目 ex2 では -0.6 付近で最も反応確率が高くなっている。これは、項目 ex1 よりも項目 ex2 のほうが低い潜在特性値でも回答可能であることを意味する。つまり、困難度は回答者の潜在特性値がいくらであれば、どの選択肢の回答するのを示す答えやすさ、答えにくさの指標と考えることができる。

また、項目 ex1 において選択肢 3 は潜在特性値が -2 から 2 の範囲で反応確率が高くなっているのに対し、項目 ex2 においては -1 から 0 の範囲で反応確率が高くなっている。これは識別力が高い項目 2 のほうが、回答者の潜在特性値を識別する能力が高いことを意味している。つまり、識別力は回答者の潜在

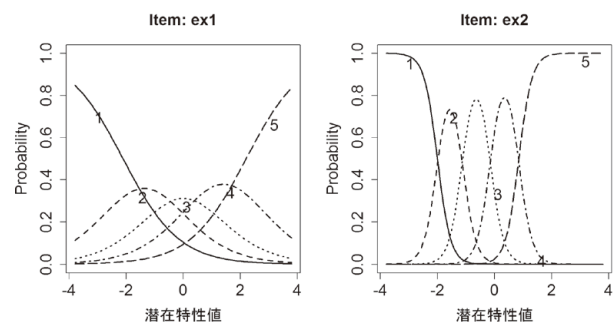


図 1 項目 ex1 と項目 ex2 の項目反応カテゴリ特性曲線

特性値の大きさの違いを識別できるかを示す敏感さの指標と考えることができる。

4. 結果と考察

4.1 潜在特性の一次元性の確認

項目反応理論による分析に先立ち、各尺度が測定しようとしている潜在特性が1次元であること確認した。表1に項目間のポリコリック相関行列の固有値分解により得られた固有値とその寄与率を大きい順に5つまで示した。第1固有値の寄与率は49%～67%となっており、第2固有値との差も十分に大きく、1次元性を有していると考えてよいだろう。

表1 各原則の上位5つまでの固有値とその寄与率

	固有値				
	第1	第2	第3	第4	第5
原則1	3.95	1.24	0.80	0.57	0.48
	49.39%	15.47%	10.03%	7.12%	5.98%
原則2	6.67	0.94	0.47	0.43	0.34
	66.65%	9.39%	4.70%	4.34%	3.36%
原則3	4.09	1.07	0.70	0.52	0.48
	51.13%	13.41%	8.70%	6.54%	6.06%
原則4	6.64	0.91	0.81	0.67	0.59
	55.29%	7.56%	6.76%	5.58%	4.89%
原則5	9.83	0.80	0.75	0.70	0.57
	61.41%	4.99%	4.70%	4.38%	3.58%
原則6	5.51	0.83	0.62	0.53	0.41
	61.20%	9.22%	6.88%	5.94%	4.60%
原則7	7.83	1.28	1.25	0.70	0.59
	52.19%	8.55%	8.36%	4.64%	3.91%

4.2 項目パラメータの推定値

各原則に含まれる項目を用いて、項目パラメータを周辺最尤法によって推定した。パラメータの推定値を表2に示す。各原則を見ると困難度の平均値が負の値になっており、平均的な潜在特性値よりも低い値を持つ回答者に対して有効な尺度になっており、高い潜在特性値を持つ回答者に対しては充分とは言えない。

以下、原則ごとに特徴的な項目を概観する。

4.2.1 原則1（安全最優先の価値観）

Q1, Q2は識別力が高く、困難度が低くなっており、よく似た特徴を持っている。平均以下の特性値を持つ回答者をよく識別できる項目であると考えられる。

Q1 あなたの職場では、担当業務が安全最優先の方針と矛盾することなく実践されている

Q2 あなたの職場では、安全最優先の方針に基づいた取り組みの必要な見直しを評価に基づいて継続的に実施している

Q7, Q8は識別力が低く、評定値が潜在特性値に依存しにくい項目となっている。

Q7 あなたは、自分が担当している業務・設備で、発電所の安全性に影響する弱み・課題について十分に説明できる

Q8 あなたは、発電所で起りうる事故の中で安全上の最悪のシナリオを他者に十分に説明できる

4.2.2 原則2（リーダーシップ）

すべての項目で高い識別力となっている。特にQ16, Q17が高くなっている。

Q16 安全に対する姿勢や取り組みを発電所の幹部は認めてくれる

Q17 安全に対する姿勢や取り組みを直属上司は認めてくれる

4.2.3 原則3（安全確保の仕組み）

Q22は識別力の高い項目となっている。

Q22 あなたの職場では、業務遂行に必要な規則やルールが、タイムリーに更新されている一方、Q21, Q25は低めの識別力となっている。特にQ25は特性値が低くても高い評定値をつけやすい基礎的な項目であると考えられる。

Q21 あなたは自分の職務の責任範囲がはっきりしていないと思う（逆転項目）

Q25 あなたの職場では、手間のかかる規則・ルールよりも容易な方法を選択する場合がある（逆転項目）

表2 各原則における項目パラメータ推定値と項目情報量

項目	困難度				識別力	情報量		項目	困難度				識別力	情報量			
	1	2	3	4		値	比率		1	2	3	4		値	比率		
原則1	Q1	-2.66	-2.25	-1.68	-0.13	2.82	7.80	20.2%	Q39	-2.63	-1.79	-0.87	0.91	2.00	5.79	5.2%	
	Q2	-2.91	-2.30	-1.52	0.10	3.05	9.63	25.0%	Q40	-3.64	-2.18	-0.64	1.58	1.22	3.46	3.1%	
	Q3	-2.26	-1.61	-0.57	0.95	1.92	5.27	13.7%	Q41	-2.89	-2.20	-1.22	0.57	1.99	5.70	5.1%	
	Q4	-1.63	-0.63	0.36	1.97	1.20	2.82	7.3%	Q42	-2.79	-2.11	-1.14	0.61	2.34	7.07	6.4%	
	Q5	-2.62	-1.92	-0.99	0.80	1.63	4.23	11.0%	Q43	-2.75	-1.84	-0.62	1.05	1.75	5.01	4.5%	
	Q6	-2.61	-1.75	-0.70	1.17	1.44	3.73	9.7%	Q44	-2.89	-2.05	-1.08	0.66	2.28	7.05	6.3%	
	Q7	-4.49	-2.96	-1.20	1.65	0.98	2.69	7.0%	Q45	-2.84	-2.07	-1.06	0.68	2.40	7.50	6.8%	
	Q8	-4.08	-2.63	-1.10	1.53	0.93	2.39	6.2%	Q46	-2.72	-2.00	-1.11	0.56	2.37	7.14	6.4%	
	平均		-1.21			1.75	4.82		原則5	Q47	-2.48	-1.57	-0.62	1.16	2.30	7.19	6.5%
SD		1.62			0.75	2.46		Q48	-2.34	-1.35	-0.31	1.42	1.95	5.82	5.2%		
原則2	Q9	-2.50	-1.87	-1.03	0.58	2.16	6.08	8.3%	Q49	-2.69	-2.05	-1.16	0.46	2.81	8.89	8.0%	
	Q10	-2.90	-2.42	-1.50	0.05	2.65	7.89	10.7%	Q50	-2.69	-1.94	-0.95	0.75	2.65	8.54	7.7%	
	Q11	-2.77	-1.98	-1.07	0.73	2.22	6.69	9.1%	Q51	-2.21	-1.29	-0.37	1.30	1.96	5.68	5.1%	
	Q12	-2.77	-2.00	-1.05	0.68	1.71	4.58	6.2%	Q52	-2.56	-1.84	-0.95	0.79	2.58	8.08	7.3%	
	Q13	-2.41	-1.78	-0.75	0.77	2.46	7.47	10.1%	Q53	-2.70	-1.99	-0.89	0.79	2.44	7.67	6.9%	
	Q14	-2.38	-1.74	-0.91	0.65	2.48	7.37	10.0%	Q54	-2.68	-1.97	-0.91	0.82	3.10	10.56	9.5%	
	Q15	-2.48	-1.88	-1.15	0.45	2.56	7.48	10.1%	平均		-1.15			2.26	6.95		
	Q16	-2.54	-1.97	-0.96	0.62	2.69	8.36	11.3%	SD		1.37			0.43	1.65		
	Q17	-2.66	-2.15	-1.25	0.28	3.29	10.61	14.4%	原則6	Q55	-2.78	-2.06	-1.10	0.76	1.86	5.19	8.0%
Q18	-2.64	-2.13	-1.12	0.53	2.42	7.14	9.7%	Q56		-2.79	-2.08	-1.17	0.69	2.40	7.37	11.3%	
平均		-1.28			2.46	7.37		Q57		-2.85	-1.81	-0.42	1.65	1.29	3.45	5.3%	
SD		1.20			0.39	1.47		Q58		-3.04	-2.36	-1.26	0.46	2.98	10.01	15.4%	
原則3	Q19	-2.44	-1.77	-0.87	0.75	2.04	5.69	14.1%		Q59	-2.95	-2.17	-1.28	0.33	3.04	10.11	15.6%
	Q20	-2.55	-1.68	-0.69	1.10	1.89	5.44	13.4%		Q60	-2.76	-2.08	-1.20	0.37	2.90	9.29	14.3%
	Q21	-2.76	-1.27	-0.43	0.99	1.07	2.41	5.9%		Q61	-2.68	-1.97	-1.13	0.50	2.88	9.23	14.2%
	Q22	-2.60	-1.76	-0.92	0.78	2.61	8.31	20.5%		Q62	-2.77	-1.89	-0.59	1.28	1.81	5.39	8.3%
	Q23	-2.47	-1.62	-0.65	1.07	2.07	6.15	15.2%		Q63	-2.68	-1.77	-0.72	1.12	1.73	4.90	7.5%
	Q24	-2.35	-0.85	0.33	1.77	1.24	3.14	7.8%	平均		-1.26			2.32	7.22		
	Q25	-4.05	-2.53	-1.06	0.48	1.07	2.64	6.5%	SD		1.38			0.62	2.39		
	Q26	-2.78	-1.93	-0.87	1.03	2.16	6.68	16.5%	原則7	Q64	-1.79	-1.22	-0.33	1.09	2.04	5.46	7.0%
	平均		-1.02			1.77	5.06			Q65	-1.85	-1.39	-0.48	0.75	2.30	6.21	7.9%
SD		1.45			0.54	1.98		Q66		-2.75	-1.96	-0.74	1.05	1.82	5.26	6.7%	
原則4	Q27	-2.52	-1.39	-0.34	0.90	1.43	3.58	5.2%		Q67	-2.83	-1.65	-0.53	0.74	1.14	2.60	3.3%
	Q28	-2.83	-2.23	-1.35	0.36	1.93	5.19	7.6%		Q68	-2.44	-1.83	-1.02	0.59	2.62	7.89	10.0%
	Q29	-3.00	-1.86	-0.60	0.70	1.45	3.80	5.5%		Q69	-2.03	-1.58	-0.72	0.46	3.19	9.70	12.4%
	Q30	-2.42	-1.63	-0.49	1.28	1.70	4.69	6.8%		Q70	-2.19	-1.62	-0.74	0.70	3.11	9.91	12.6%
	Q31	-2.94	-2.12	-1.07	1.01	2.07	6.32	9.2%		Q71	-2.90	-2.34	-1.51	0.13	2.30	6.54	8.3%
	Q32	-3.08	-2.23	-1.07	0.84	2.01	6.11	8.9%		Q72	-2.17	-1.49	-0.64	1.00	1.60	3.99	5.1%
	Q33	-3.44	-2.77	-1.76	0.04	2.04	5.88	8.6%	Q73	-1.47	0.13	1.38	2.62	0.92	2.02	2.6%	
	Q34	-2.65	-2.06	-1.26	0.36	2.39	6.88	10.0%	Q74	-1.93	-1.01	-0.03	0.71	1.45	3.21	4.1%	
	Q35	-2.43	-1.68	-0.80	0.93	2.42	7.41	10.8%	Q75	-3.07	-2.01	-0.98	0.06	1.40	3.30	4.2%	
Q36	-2.38	-1.65	-0.71	1.04	1.97	5.56	8.1%	Q76	-2.84	-1.95	-0.71	1.29	1.71	4.98	6.3%		
Q37	-3.10	-2.39	-1.32	0.50	2.21	6.69	9.8%	Q77	-3.00	-1.77	-0.70	0.83	1.41	3.69	4.7%		
Q38	-2.96	-2.23	-1.17	0.61	2.16	6.49	9.5%	Q78	-3.06	-1.89	-0.77	0.59	1.44	3.73	4.7%		
平均		-1.28			1.98	5.72		平均		-0.93			1.90	5.23			
SD		1.37			0.31	1.15		SD		1.34			0.66	2.35			

4.2.4 原則4 (円滑なコミュニケーション)

すべての項目で高い識別力となっている。特にQ34, Q35は識別力が高い。

- Q34 あなたの直属上司は職場間の連絡をうまくとっている
- Q35 あなたの職場では、他部署にまたがる問題の調整や解決が十分になされている

4.2.5 原則5 (個人・組織の姿勢)

すべての項目で十分な識別力となっている。特にQ49, Q54が高くなっている。

- Q49 あなたの職場では、事故や安全性の問題が率直に話し合われている
- Q54 あなたの職場では、安全確保のための意見やアイデアが活かされている

4.2.6 原則6 (潜在的リスクの認識)

すべての項目で十分な識別力となっている。特にQ58, Q59, Q60, Q61が高くなっている。

- Q58 あなたの職場の仲間は、能率の良い作業方法を思いついた場合でも実行する前に安全性を確認している
- Q59 あなたの職場の仲間は、作業エリアに危険が存在しないか、事前に確認している
- Q60 あなたの職場の仲間は、安全上の問題がないか現場に出向いて注意を払っている
- Q61 あなたの職場の仲間は、仕事量の多い時期でも安全性の確保を優先している

4.2.7 原則7 (活気のある職場環境)

識別力が高い項目はQ69, Q70である。

- Q69 あなたは今の職場の仲間の一員でいたい
- Q70 あなたの職場では、お互いがベストをつくすように励まし合う

一方、Q67, Q73の識別力が低い。Q73は識別力が低いが困難度は高くなっている。

- Q67 あなたの職場では、個人的な感情や利害を職場に持ち込んでいる人がいる(逆転項目)
- Q73 あなたは、これからの人生をあれこれと考えてみて「将来なんとなく不安だ」というような気がする(逆転項目)

4.3 測定の精度

4.3.1 テスト情報量

各原則のテスト情報関数を図2に示す。テスト情報関数は項目情報関数の和であるため、項目数が多い原則ほど情報量も多く、測定精度が高くなっていることがわかる。測定の標準誤差の観点から、古典的テスト理論における信頼性係数の.7, .8, .9はテスト情報量において3.33, 5.00, 10.00に相当する

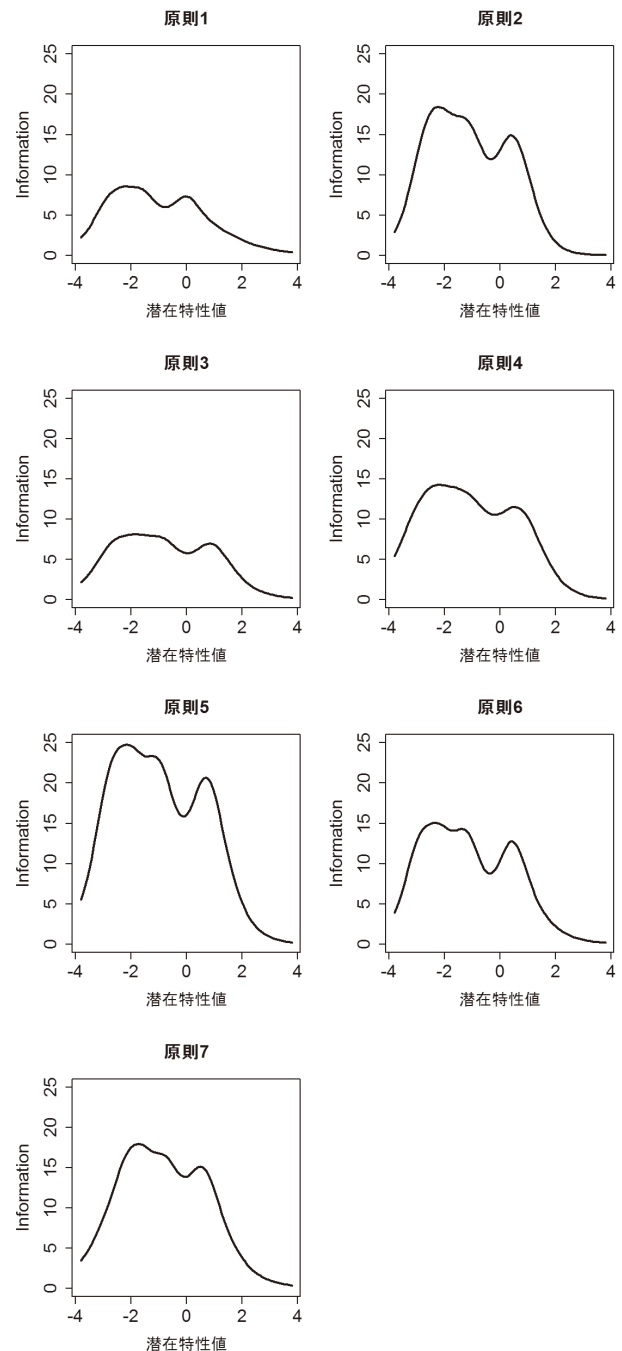


図2 各原則のテスト情報関数

(Thissen, 2000). この基準においてはすべての原則で特性値が-3~+1の回答者にとっては十分な精度を持っていると考えてよいだろう. いずれの原則においても, 潜在特性値が-2付近と+1付近で極値をとっており, このような値の潜在特性値を持つ回答者にとっては非常に有効な尺度になっていると考えられる. また多くの原則で潜在特性値が0付近でテスト情報量関数が落ち込んでいる. すなわち平均的特性値を持つ回答者への精度がやや低くなっていることも読み取れる. 潜在特性値が2以降では情報量が低下しており, 特性値が高い回答者に対してはあまり有効な尺度になっていないことがうかがえる.

4.3.2 項目情報量

項目情報量と項目情報量の合計に占めるその項目の情報量の比率を表2に示す. 項目に関する精度もテスト情報関数と同じように潜在特性値についての関数で得られるが, ここでは項目情報関数を全区間について積分した値を示している.

項目情報量が低い6項目の項目反応カテゴリ特性曲線を図3に示す. 原則1のQ7, Q8, 原則3のQ21, Q25, 原則7のQ67, Q73など識別力が低い項目が項目情報量も低くなっていることがうかがえる. 識別力が低い項目は各原則が測定しようとしている概念とのズレが生じている可能性が考えられる. しかしながら, 識別力が低い項目においても基礎事項の確認を行うような目的であれば必要な項目といえる.

5. まとめ

安全風土のような, 物理的に測定できない構成概念を測定対象としている質問紙調査において, 使用される尺度の妥当性と信頼性を確保することは何よりも重要である. 本研究では, 項目反応理論を用いて項目分析を行い, 今後の調査における尺度項目の改善の基礎となる情報を提供した.

尺度の項目数を増やすと, 尺度の信頼性が増したり, 測定誤差が減ったりする一方で, 回答者の負担が大きくなり回答率やデータの質の低下を招いてしまう (Franke, Rapp, & Andzulis, 2013). この意味で尺度の信頼性と回答者の負担はトレードオフの関係にあるが, 調査を実施するうえでは, なるべく高

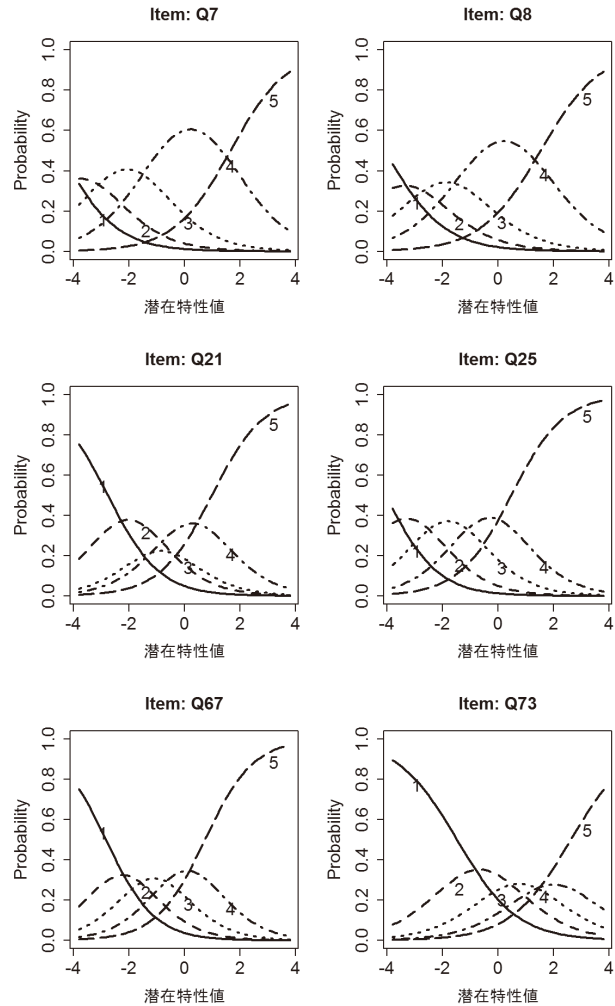


図3 項目情報量が低い6項目の項目反応カテゴリ特性曲線

い信頼性を保ったまま, 項目数を減らすことが求められる.

実施上の問題から多くの尺度において短縮版が作成されている. 近年では項目反応理論を利用した項目の選択により短縮版が作成されることが増えている. 短縮版作成の方法として因子分析で得られる因子負荷量を基に項目を選択することが多く行われているが (Goetz, Coste, Lemetayer, Rat, Montel, Recchia, Debouverie, Pouchot, Spitz, & Guillemin, 2013; 並川・谷・脇田・熊谷・中根・野口, 2012), 並川ら (2012) は因子負荷量に基づく項目選択では, 項目困難度に関する情報が反映されないことを指摘し, 項目反応理論の利用の有効性を議論している. 安全風土の関連ではHuang, Lee, Chen, Perry, Cheung, & Wang (2017) は項目反応理論を用いて, Zohar & Luria (2005) の安全風土尺度の短縮版を開発している.

本研究では項目困難度と項目識別力のパラメータ

から、潜在特性値の値によって回答しやすい評定値が項目ごとに異なること、また、潜在特性値の値を識別可能かどうか項目によって異なることが確認された。また項目情報量から、どの項目がどの程度尺度の信頼性に貢献しているかを把握することが可能になった。さらに、テスト情報量を参照することにより、潜在特性値の値と尺度の信頼性の関係を把握することができる。本研究の結果を参照することで項目選択が可能になり、少ない項目数で信頼性が高い尺度を運用できると考える。

謝 辞

本研究は一般社団法人原子力安全推進協会 (JANSI) と各事業所のご協力のもとに実施できたものである。厚く感謝申し上げます。

引用文献

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Franke, G. R., Rapp, A., & Andzulis, J. "Mick". (2013). Using shortened scales in sales research: Risks, benefits, and strategies. *Journal of Personal Selling and Sales Management*, 33, 319-328.
- 藤田智博 (2017). 安全確認を抑制するメカニズム - 知識・技能への自信に注目して -. *INSS JOURNAL*, 24, 48-57.
- 藤田智博 (2018). 原子力産業の安全風土調査へのマルチレベル分析の適用. *INSS JOURNAL* 25, 17-24.
- 福井宏和 (2012). 原子力発電所の安全風土に関する質問紙調査. *集団力学*, 29, 69-86.
- 福井宏和・吉田道雄・山浦一保 (2000). 原子力発電所職員の安全確認行動と組織風土の因果モデル. *INSS JOURNAL*, 7, 2-14.
- 福井宏和・吉田道雄・吉山尚裕 (2001). 原子力発電所における安全風土に関する研究. *INSS JOURNAL*, 8, 2-13.
- Goetz, C., Coste, J., Lemetayer, F., Rat, A.-C., Montel, S., Recchia, S., Debouverie, M., Pouchot, J., Spitz, E., & Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology*, 66, 710-718.
- Huang, Y.H., Lee, J., Chen, Z., Perry, M., Cheung, J.H., & Wang, M. (2017). An item-response theory approach to safety climate measurement: The Liberty Mutual Safety Climate Short Scales. *Accident Analysis & Prevention*, 103, 96-104.
- 池田央 (1994). 現代テスト理論. 朝倉書店.
- Lord, F.M. (1952). A theory of test scores (Psychometric Monograph No.7). Richmond, VA: Psychometric Corporation.
- Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.
- Masters, G. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- 村木英治 (2011). 項目反応理論. 朝倉書店.
- 並川努・谷伊織・脇田貴文・熊谷龍一・中根愛・野口裕之 (2012). 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 (2), 91-99.
- 西田豊 (2017). 安全風土と安全文化 - 概念, 測定と理論, 醸成について -. *INSS JOURNAL*, 24, 21-31.
- 西田豊 (2018). スパース判別分析による属性別安全風土の特徴抽出. *INSS JOURNAL*, 25, 25-30.
- 野口裕之・渡辺直登 (編著) (2015). 組織・心理テストの科学. 白桃書房.
- Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3 (1), 1-18.
- 岡田謙介 (2011). クロンバックの α に代わる信頼性の推定法について. *日本テスト学会誌*, 7 (1), 37-50.
- 岡田謙介 (2015). 心理学と心理測定における信頼性について - Cronbach の α 係数とは何なのか, 何でないのか -. *教育心理学年報*, 54, 71-

83.

- Samejima, F. (1969) . Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph, No17) . Richmond, VA: Psychometric Society.
- 高本真寛・服部環 (2015) . 国内の心理尺度作成論文における信頼性係数の利用動向. 心理学評論, 58 (2) , 220-235.
- Thissen, D. (2000) . Reliability and measurement precision. In H. Wainer (Ed) , Computerized adaptive testing: A primer (2nd ed., pp. 159-184) . Mahwah, NJ: Lawrence Erlbaum Associates.
- 渡辺直登・野口裕之 (編著) (1999) . 組織心理測定論. 白桃書房.
- 山本晃弘・関村直人 (2017) . 原子力発電所における安全文化醸成活動の実効性向上に関わる研究. 日本原子力学会和文論文誌. 16 (3) , 119-138.
- Zohar, D. (1980) . Safety climate in industrial organizations: Theoretical and applied implications. Journal of Applied Psychology, 65, 96-102.
- Zohar, D. & Luria, G. (2005) . A multilevel model of safety climate: cross-level relationships between organization and group-level climates. Journal of Applied Psychology, 90, 616-628.