OUKA
The University of Osaka
Institutional Knowledge Archive

Title	日本語版科学的推論尺度の心理測定的特徴の検討
Author(s)	後藤, 崇志; 塩瀬, 隆之; 加納, 圭
Citation	大阪大学教育学年報. 2023, 28, p. 1-10
Version Type	VoR
URL	https://doi.org/10.18910/90190
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

日本語版科学的推論尺度の心理測定的特徴の検討

後藤　崇志・塩瀬　隆之・加納　圭

要旨

本研究の目的は、Drummond & Fischhoff (2017) の作成した Scientific Reasoning Scale を元にした日本語版の科学的推論尺度を作成し、その心理測定的特徴について検討することであった。成人（調査 1: N = 343）、大学生（調査 2: N = 93）、高校生（調査 3: N = 250）と異なる年代の人々を対象とした調査を行い、科学的推論尺度の心理測定的特徴を検討した。項目反応理論を用いた分析の結果から、科学的推論尺度は一次元構造であり、科学的推論能力の高い人から低い人まで幅広い能力を持つ人々の個人差を把握できる尺度であることが示唆された。また、科学的推論尺度の得点は科学の知識テストや批判的思考態度、パーソナリティ特性の開放性と正に関連しており、科学的推論能力を測定する妥当な尺度であることを示唆する結果も得られた。科学的推論尺度と三段論法形式のテストで測定される論理的推論能力との関連については一貫した結果が見られなかったが、高校生では両者の間には正の相関がみられており、部分的には尺度の妥当性を支持する結果が得られた。項目反応理論で得られた項目の特性推定値をもとに、測定上で留意するべき点や、項目の改善・拡充に必要な視点などを議論する。

1. 問題

科学リテラシーとは自然界の現象や科学・技術を理解し、証拠に基づいて結論を導く能力とされる。科学リテラシーは (1) 科学の知識と (2) 科学についての知識に大別できる (Miller, 2004)。前者は物理法則や生物の仕組み、化合物の性質といった、科学的な探究により蓄積されてきた知識を指す。後者は科学的な探究の性質や、科学の社会的影響に関する知識を指す。2000 年代初めまでは、科学リテラシーを測定する場合に用いられてきた調査材料には科学の知識に関するものが多く、科学についての知識を測定し、かつ、心理測定的な視点から調査材料の性質について十分に検討されたものは少ないことが指摘されてきた (e.g., Pardo & Calvo, 2004)。

科学についての知識を測定する尺度としては、科学の本質や科学の社会的影響についての認識を自己報告式で尋ねるものが多く作成されている。例えば、Views on Science-Technology-Society (Aikenhead & Ryan, 1992) は科学の本質や科学の社会的影響の認識について多肢選択式で回答を求める尺度である。また、科学の本質について多次元的に捉えようとする尺度としては、リッカート形式を採用した Views on Science and Education Questionnaire (Chen, 2006) や Student understanding of science and scientific inquiry (Liang et al., 2008) などがある。また、児童が持つ科学の本質の認識を捉える尺度としては Nature of science instrument for elementary school students (Hacıeminoğlu et al., 2012) なども作成されている。これらの尺度は科学の暫定性や科学的な知見が受けうる社会的・文化的影響、科学的手法の理解についての主観的な認識を尋ねるものである。

主観的な認識ではなく、実際の推論・判断の正誤から科学についての知識の一側面を測定しようとする尺度もある。Drummond & Fischhoff (2017) の作成した Scientific Reasoning Scale はその一つであり、科学的探究を行ったり、科学的な知識が得られた背景を理解したりする上で必要な科学

的推論能力を測定する尺度である。Scientific Reasoning Scale では科学の研究に関する架空のシナリオを読んだ上で、そのシナリオの内容に基づく推論やシナリオ内の手続きに関する質問について「正しい」か「誤っている」かを判断して回答するように求める。二重盲検法や因果推論、交絡変数など、科学的な手続きに基づいて推論を行うために必要な基礎的な視点を問う項目で構成されている。文章を読み進めて回答を選択する形で調査は進むため、調査の実施は比較的容易で、採点も客観性が担保されており、大人数に同時に実施しやすい。図表の読み取りなど、PISA で問われるような科学リテラシーの問題を包摂できていない面もあるが、簡便に個人の科学的推論能力を把握するためには有用な尺度である。

本研究では Drummond & Fischhoff (2017) の作成した Scientific Reasoning Scale を元にした日本語版の科学的推論尺度を作成し、その測定的特徴について検討する。測定的特徴を検討する上で、作成された尺度が幅広い年代に利用可能であるかは重要な視点であるため、成人（調査 1）、大学生（調査 2）、高校生（調査 3）と異なる年代の人々を対象とした調査を行う。調査 1 では、項目反応理論を用いて、作成された日本語版科学的推論尺度が一次元的に推論能力を捉えているものであるか、また、能力特性が高い人から低い人まで幅広い層の能力を把握可能な尺度であるかを検討する。また、調査 1 ～ 3 を通じて、科学リテラシーの構成要素である科学の知識や、科学的推論と関連すると考えられる概念である、証拠に基づいて論理的で偏りのない思考をしようとする傾向である批判的思考態度や、知的好奇心や創造的思考と関わるビッグファイブのパーソナリティ特性のひとつである開放性、三段論法形式のテストで測定される一般的な論理的推論能力との関連を検討することで、科学的推論尺度の測定上の妥当性の検討も行う。

2. 調査 1

2.1. 調査回答者

web 調査会社の登録モニタから公募された 400 名が調査に回答した。回答者は 20 代、30 代、40 代、50 代の男女がそれぞれ 50 名ずつになるように集められた。注意確認設問（後述）に適切に回答できた 343 名を分析対象とした（男性 176 名、女性 167 名／平均年齢 39.9 歳、SD = 11.1）。

2.2. 調査に用いた尺度

科学的推論尺度　科学的推論尺度は、Scientific Reasoning Scale (Drummond & Fichhoff, 2017) を参考に作成した。原著者 (Dr. Caitlin Drummond Otten) の許可を得たうえで、11 種類の架空の研究シナリオを日本語訳しつつ、日本人に馴染みの薄いと考えられた事物・表現は他の事物・表現に修正した。回答者はそれぞれの架空の研究シナリオ（例「ある企業の研究員が、白いカップにコーヒー A を注ぎ、黒いカップにコーヒー B を注いで、参加者に飲み比べてもらう品質テストを行うことにした。品質テストでは、実験助手がコーヒーを参加者に渡し、表情からどちらが美味しいと感じているかを判断する。」）を読んだ後、下線を付して与えられた記述（例「このテストを行う時、実験助手はカップにコーヒーが注がれているところを見てはいけない。」）が科学的な推論の結果として、あるいは科学的な推論を行うための手続きとして正しいか誤っているかを判断するように求められた（上述の例の場合、正答は「正しい」）。翻訳時の誤りにより項目 7（歴史的視点）の正誤が原版と逆転してしまったが、他の 10 項目については正誤が原版と同じものになるようにした。また、注意確認設問として、科学的推論能力を必要とせず、シナリオを精読すれば正誤を判断できる設問を新たに 1 問作成した。具体的な項目は付録 1 に報告している。

科学の知識テスト　科学の知識テストは、各国の社会調査などでもよくつかわれるテストである Trend Factual Knowledge of Science Scale を用いた。調査によって項目セットが異なるが、本研究では栗山他 (2011) で使われたものから選出した 10 項目を使用した。回答者はそれぞれの記述を読んだ後、記述されている内容が科学的な知見として正しいか誤っているかを判断するように求められた（例「すべての放射性物質は人工的に作られたものである」：正答は「誤っている」）。具体的な項目は付録 2 に、項目ごとの正答率は付録 3 に報告している。

批判的思考態度尺度　批判的思考態度尺度は、楠見・平山 (2013) で用いられた 12 項目短縮版の尺度を用いた。この尺度は 4 つの下位尺度から構成されており、論理的思考への自覚（例「議論の前提や用語の定義を正確にとらえて考えようとする」）、探究心（例「いろいろな考え方の人と接して多くのことを学びたい」）、客観性（例「いつも偏りのない判断をしようとする」）、証拠の重視（例「判断をくだす際は、できるだけ多くの事実や証拠を調べる」）のそれぞれの下位尺度につき 3 項目の計 12 項目で構成されている。回答者はそれぞれの項目が自身の考え方にあてはまるかどうかを「1. まったくあてはまらない」～「5. とてもよくあてはまる」の 5 段階で回答するように求められた。

ビッグファイブ尺度　ビッグファイブの測定には、小塩他 (2012) で作成された 10 項目版尺度である TIPI-J を用いた。この尺度はビッグファイブの 5 つの側面である開放性（例「新しいことが好きで、変わった考えをもつと思う」）、勤勉性（例「しっかりしていて、自分に厳しいと思う」）、外向性（例「活発で、外向的だと思う」）、協調性（例「人に気をつかう、やさしい人間だと思う」）、神経症傾向（例「心配性で、うろたえやすいと思う」）をそれぞれ順項目・逆転項目の 2 項目ずつで測定するものである。回答者はそれぞれの項目が自身にあてはまるかどうかを「1. 全く違うと思う」～「7. 強くそう思う」の 7 段階で回答するように求められた。

2.3. 結果と考察

項目反応理論による科学的推論尺度の構造の検討　科学的推論尺度のそれぞれの項目への正答率は表 1 に報告した。まず、科学的推論尺度の一因子性を確認するため、各項目への回答データから得られたテトラコリック相関行列の固有値をもとめた。固有値は 2.828、1.357、1.175、1.070…と、第 1 要素から第 2 要素にかけて大きく減少していたことから、一因子構造を仮定できると判断した。

表 1 科学的推論尺度の調査ごとの正答率

項目		調査 1	調査 2	調査 3
1	二重盲検法	50.7%	37.6%	−
2	因果推論	68.8%	66.7%	65.5%
3	交絡変数	60.3%	74.2%	52.2%
4	構成概念	88.9%	97.8%	83.1%
5	統制条件	83.4%	78.5%	−
6	生態学的妥当性	68.6%	71.0%	−
7	歴史的視点	63.3%	78.5%	−
8	実験パラメータの調整	87.8%	92.3%	78.8%
9	無作為割当	70.0%	78.3%	60.0%
10	測定の信頼性	32.4%	31.5%	43.0%
11	反応バイアス	64.7%	74.2%	−

注: 調査 3 では短縮版を用いたため、調査に用いた項目の正答率のみを報告している

続いて、それぞれの項目の測定的特徴を検討すべく、項目反応理論の3パラメータロジスティックモデルを用いて推定を行った。3パラメータロジスティックモデルは回答者 i が項目 j に正答する確率 $P_j(\theta_i)$ が式 (1) のモデル式に従うと仮定する

$$P_j(\theta_i) = c_j + (1 - c_j) \times \frac{1}{1 + \exp\{-D \times a_j(\theta_i - b_j)\}} \quad \text{式(1)}$$

θ_i は回答者 i がもつ能力特性値である。a_j, b_j, c_j はそれぞれ項目 j がもつ項目特性値であり、a_j は識別力、b_j は困難度、c_j は当て推量である。科学的推論尺度は正誤の二値選択を求めるものであり、当て推量により正答に至る可能性があると考え、本研究では3パラメータロジスティックモデルを用いることとした。D は固定値であり、本研究では慣例に従って 1.7 に指定した。モデルの推定には stan を用いた。3パラメータロジスティックモデルはサンプルサイズが小さいと推定が不安定になることから、豊田 (2012) に従って事前分布を設定した。

上記モデルについて stan を用いて推定を行った。長さ 8000 のチェインを 4 つ発生させ、バーンイン期間を 4000 とし、得られた 16000 の乱数を利用した。項目特性値を推定するために用いた stan コードは付録 4 に報告している。モデルの適合度の指標である WAIC は 4147.7 であった。なお、比較のために c_j=0 とした2パラメータロジスティックモデルに基づく推定も行ったところ、WAIC は 4170.1 であった。WAIC は3パラメータロジスティックモデルの方が小さかったことから、本研究では3パラメータロジスティックモデルの結果を報告することとした。得られた推定値は表2に示した通りであり、項目1（二重盲検法）は困難度が他の項目に比べて極めて高くなっていた。また、項目3（交絡変数）、項目5（統制条件）、項目6（生態学的妥当性）は他の項目に比べてやや識別力が低い傾向にあった。しかし、そのほかの項目はおおむね識別力が同程度であり、困難度は正の値をとるものがやや多いが、低いものから高いものまで幅広く見られていた。従って、科学的推論尺度は能力特性の高いものから低いものまで広く回答者の能力特性を把握するために用いることができる尺度であると解釈できる。

以降の分析では2つの方法で科学的推論尺度の得点を求め、分析に用いる。一つは、項目1から項目11の11項目の正答率を用いる。もう一つは、得られた項目特性の MED 推定値をもとに個人の回答パターンから推定された能力特性値（式 (1) の θi）を用いる。回答者の能力特性値を推定するために用いた stan コードは付録5に報告している。

表2. 3パラメータモデルで推定された科学的推論尺度の項目特性の MED 推定値と 95% 確信区間

項目		識別力	困難度	当て推量
1	二重盲検法	1.117	5.043	0.504
		[0.447, 2.838]	[2.645, 10.272]	[0.452, 0.556]
2	因果推論	1.137	0.456	0.504
		[0.518, 2.775]	[-0.163, 1.081]	[0.343, 0.617]
3	交絡変数	0.71	1.901	0.534
		[0.3, 2.119]	[0.638, 7.365]	[0.358, 0.628]
4	構成概念	1.254	-1.028	0.491
		[0.654, 2.769]	[-1.691, -0.478]	[0.299, 0.669]
5	統制条件	0.598	-0.754	0.519
		[0.317, 1.175]	[-1.906, 0.389]	[0.311, 0.708]
6	生態学的妥当性	0.87	0.564	0.51
		[0.412, 2.155]	[-0.193, 1.44]	[0.336, 0.628]
7	歴史的視点	1.113	0.408	0.413
		[0.581, 2.411]	[-0.057, 0.953]	[0.276, 0.528]
8	実験パラメータの調整	1.304	-0.934	0.483
		[0.674, 2.89]	[-1.568, -0.405]	[0.296, 0.661]
9	無作為割当	1.391	-0.035	0.399
		[0.749, 2.948]	[-0.427, 0.406]	[0.253, 0.529]
10	測定の信頼性	1.137	1.947	0.279
		[0.495, 2.689]	[1.295, 5.779]	[0.203, 0.353]
11	反応バイアス	1.049	0.52	0.456
		[0.511, 2.509]	[-0.038, 1.172]	[0.303, 0.572]

科学的推論尺度と他の変数との関連　科学的推論尺度と年齢、科学の知識テスト、批判的思考態度尺度、ビッグファイブとの間の相関係数を、それぞれの変数の平均値、標準偏差とともに表 4 に示した。なお、科学の知識テストは 10 項目の正答率を得点として用いた。また、批判的思考態度尺度は下位尺度ごとの項目評定値の平均値を下位尺度得点として用いた。ビッグファイブは順項目の評定値と逆転項目の評定値を逆転させたものの平均値を下位尺度得点として用いた。科学的推論尺度の得点は正答率・能力特性値ともに一貫して科学の知識テストの得点と弱いながらも有意な正の相関を示していた。科学的推論と科学の知識は共に広義の科学リテラシーを構成する要素であることから、両者に正の相関関係が見られることは尺度の基準関連妥当性を支持する結果と言える。一方で、得られた相関係数は同一の概念を測定していると考えられるほど大きなものでもなく、互いに科学リテラシーの異なる側面を捉える尺度だと解釈できる。

パーソナリティ特性に注目すると、科学的推論尺度の得点は正答率・能力特性値ともに一貫して批判的思考態度の論理的思考への自覚、ビッグファイブの開放性と弱いながらも有意な正の相関を示していた。批判的思考態度は自分の推論過程を意識的に吟味しようとする態度傾向である。開放性は好奇心や創造的思考と関わるパーソナリティ特性である。いずれも科学的探究への志向性との関連が想定される特性であることから、科学的推論との間に正の関連が見られたことは構成概念妥当性を支持する結果といえる。

表 3. 調査 1 における科学的推論尺度と他変数との相関係数・要約統計量

変数	正答率との相関係数	能力特性値との相関係数	平均値	標準偏差
科学的推論尺度				
正答率	—	—	0.67	0.18
能力特性値	—	—	0.04	0.75
年齢	.01 [-.10, .11]	-.01 [-.11, .10]	39.88	11.09
科学の知識テスト	.14* [.04, .25]	.20* [.10, .30]	0.71	0.17
批判的思考態度				
論理的思考への自覚	.12* [.01, .22]	.13* [.02, .23]	3.41	0.73
探究心	.04 [-.07, .14]	.02 [-.08, .13]	3.65	0.88
客観性	.01 [-.09, .12]	.03 [-.08, .13]	3.60	0.69
根拠の重視	.08 [-.02, .19]	.07 [-.03, .18]	3.55	0.71
ビッグファイブ				
開放性	.13* [.02, .23]	.11* [.01, .21]	3.98	1.23
勤勉性	.00 [-.10, .11]	-.02 [-.12, .09]	3.90	1.24
外向性	.01 [-.10, .12]	-.02 [-.13, .09]	3.80	1.36
協調性	-.07 [-.18, .03]	-.04 [-.15, .07]	4.76	1.06
神経症傾向	-.10 [-.20, .01]	-.10 [-.20, .01]	4.26	1.26

注　* $p < .05$, 相関係数の括弧内の数値は 95%信頼区間の下限・上限

3. 調査 2

3.1. 調査回答者

関西地方の大学生 100 名が授業中に集合形式にて調査に回答した。科学的推論尺度の項目 12（注意確認）に適切に回答できた 93 名を分析対象とした（男性 57 名、女性 36 名／平均年齢 19.2 歳、SD = 1.1、年齢幅 = 18 〜 22 歳）。

調査に用いた尺度

科学的推論尺度・科学の知識テスト・批判的思考態度尺度　調査１と同じ尺度を用いた。得点化の方法も調査１と同様であった。

論理的推論能力　論理的推論能力の測定には、BAROCO-short(Shikishima et al., 2011) を用いた。BAROCO-short は知能の一側面である三段論法の論理的推論能力を少数の項目で測定できるように作成された尺度である。２つの前提が与えられ（例: あるオウムはババロである。どのババロも鳥でない）、４つの命題（例：あるオウムは鳥である、すべてのオウムは鳥である、あるオウムは鳥でない、どのオウムも鳥でない）から論理的に正しい結論を選択することが求められる。いずれの命題も結論として正しくないときには「どれでもない」と回答することもできる。以上の形式の５つの問題について回答を求めた。以降の分析では正答数を論理的推論能力として用いる。

3.3. 結果と考察

科学的推論尺度と年齢、科学の知識テスト、批判的思考態度尺度、論理的推論能力の相関係数を、それぞれの変数の平均値、標準偏差とともに表４に示した。科学的推論と科学の知識テストは、正答率を用いた場合には非有意であったものの、能力特性値を用いた場合には有意な正の相関が見られた。調査１と同様に、科学的推論尺度が科学リテラシーの一側面として科学の知識とは異なる要素を測定していることを示唆する結果である。一方で、BAROCO-short で測定された論理的推論能力との間には、正答率・能力特性値のいずれも有意な関連は見られなかった。このことから、科学的推論尺度は論理的推論能力には必ずしも還元されない、科学領域に特有な推論能力を測定していると解釈できる。

また、批判的思考態度尺度の下位尺度と科学的推論尺度の得点の間には有意な関連は見られなかった。能力特性値を得点と用いた場合、論理的思考の自覚、客観性との間に正の相関関係は見られていた（それぞれ、$p = .06$、$p = .07$）。調査１と合致する傾向だが、サンプルサイズが小さかったために統計的に有意な結果は得られなかった可能性も考えられる。批判的思考態度との間に正の相関関係が見られたことは、相関係数が有意でなかったことには留意する必要があるものの、構成概念妥当性を支持する結果といえる

表 4. 調査２における科学的推論尺度と他変数との相関係数・要約統計量

変数	正答率との相関係数	能力特性値との相関係数	平均値	標準偏差
科学的推論尺度				
正答率	—	—	0.71	0.15
能力特性値	—	—	0.22	0.65
年齢	.07 [-.14, .27]	.10 [-.11, .30]	19.19	1.14
科学の知識テスト	.17 [-.10, .41]	.25* [.00, .47]	0.74	0.15
批判的思考態度				
論理的思考の自覚	.09 [-.12, .30]	.20 [-.01, .38]	3.59	0.61
探究心	-.08 [-.29, .13]	-.01 [-.22, .19]	3.89	0.76
客観性	.18 [-.03, .38]	.19 [-.02, .38]	3.71	0.78
根拠の重視	.03 [-.18, .23]	.05 [-.16, .25]	3.54	0.76
論理的推論	.08 [-.13, .29]	.14 [-.07, .33]	3.72	1.17

注　* $p < .05$, 相関係数の括弧内の数値は 95%信頼区間の下限・上限

4. 調査３

4.1 調査回答者

関西地方の高校生 364 名が集合形式にて調査に回答した。科学的推論尺度の項目 12（注意確認）に適切に回答できた 250 名を分析対象とした（1 年生 79 名、2 年生 74 名、3 年生 97 名）。

4.2. 調査に用いた尺度

科学的推論尺度・科学の知識テスト　調査の時間・労力面での負担を考慮し、それぞれ 6 項目ずつ（注意確認設問は除く）からなる短縮版を作成した。具体的には、調査 1 のデータから 3 パラメータロジスティックモデルで推定された項目特性値をもとに、テスト情報量の分布が大きく異ならないようにしつつ、「正しい」を正答とする問題と「誤っている」を正答とする問題の両者が含まれるように項目を選択した。科学的推論尺度の全項目版と短縮版（項目 2、3、4、8、9、10）のテスト情報曲線は図 1 に示している。科学的推論尺度の尺度得点は全項目版と同様に、設問への正答率と 3 パラメータロジスティックモデルによる能力特性値の両者を用いた。科学の知識テストの尺度得点は設問への正答率を用いた。

論理的推論能力　調査 2 と同様に、BAROCO-short(Shikishima et al., 2011) を用いた。尺度得点は調査 2 と同様に正答数を用いた。

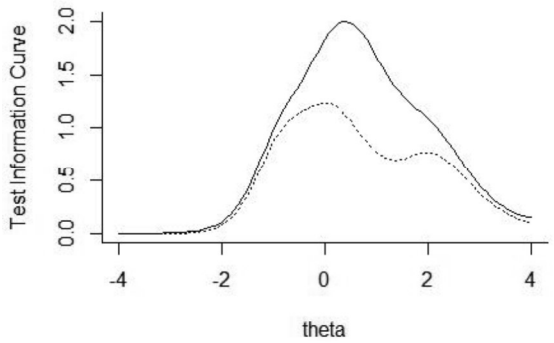


図 1. 調査 1 の項目特性推定値にもとづく科学的推論尺度の
全項目版（実線）・短縮版（破線）のテスト情報曲線

4.3. 結果と考察

科学的推論尺度と科学の知識テスト、論理的推論能力の相関係数を、それぞれの変数の平均値、標準偏差とともに表 5 に示した。尺度得点に正答率・能力特性値のどちらを用いた場合も結果は一貫しており、科学的推論尺度は科学の知識テスト、論理的推論能力のどちらとも有意な正の相関関係が見られていた。科学の知識テストに関しては調査 1・調査 2 から一貫した傾向が見られており、科学的推論尺度が科学リテラシーの一側面として科学の知識とは異なる要素を測定していることを示唆する結果である。論理的推論能力に関しては調査 2 とは異なる傾向が見られた。どちらも推論という共通の思考プロセスに関わる能力であることを考慮すると、正の相関関係が見られたとはいえ、同一の概念を測定していると判断されるほど大きな相関係数ではなかったという結果は、構成概念妥当性を支持する結果であると考えられる。

5. 総合考察

本研究の目的は Drummond & Fischhoff (2017) が作成した Scientific Reasoning Scale をもとにした日本語で利用可能な科学的推論尺度を作成し、その測定的特徴を検討することであった。インターネット調査会社の回答者モニタに登録している成人（調査 1）、大学生（調査 2）、高校生（調査 3）と異なる年代の人々を対象とした調査を行い、作成された尺度について幅広い世代での測定的特徴を検討した。

項目反応理論を用いた分析の結果からは、科学的推論尺度が一次元構造を持っており、能力の低いものから高いものまで幅広い層の能力特性を把握するために用いることができる尺度であることが示唆された。項目反応理論による推定には、能力特性値に基づいて正答する確率を規定する識別力と困難度に加えて、当て推量により正答する確率を考慮した 3 パラメータロジスティックモデルを適用した。他の尺度との間の相関関係を検討する際に正答率よりも項目反応理論で推定された能力特性値の方が調査間で一貫した傾向を示しやすかったことから、科学的推論尺度を用いる際には簡便に計算可能な正答率だけでなく、当て推量で正答する可能性も考慮した能力推定値を用いた検討も同時に行うことが望ましいと考えられる。

項目反応理論を用いて項目特性値の推定を行ったことには、能力特性値の推定が可能なこと以外にも、科学的推論尺度の項目の柔軟な活用を可能にするという利点がある。例えば、本研究においても調査 3 では調査を実施する時間と労力の負担を押さえるために、全 11 項目のうち 6 項目のみからなる短縮版を用いた。科学的推論尺度はそれぞれの項目が研究シナリオと正誤判断を問う記述から構成されるため、調査回答者への負担が大きい。広い視点から科学的推論能力の測定を目指すことを考えると多様な手法・論点が項目群に含まれることが望ましいが、簡便な能力推定を目的とするのであれば項目特性値をもとに測定的特徴が損なわれにくいように作成された短縮版を用いることも一つの手段である。

また、項目特性値が既知の項目と新規に作成された項目特性値が未知の項目を混合させた調査を行うことで、既知の項目特性値を基準として未知項目の測定的特徴を議論できるように特性値を変換する尺度項目の等化も可能である。本研究で作成された科学的推論尺度は科学の知識テストや批判的思考態度との間に相関が見られたことなどから、一定の妥当性を示唆する根拠を持った尺度であるといえる。一方で、尺度の具体的な項目の記述に目を向けると心理学や人の行動に関わる研究シナリオに関する記述が多かったり、「誤っている」を正答とする項目が多かったりと、項目には偏りがあることが窺え、改訂・拡充の余地がある。等化手続きを行いながら項目の改訂・拡充を行うことで、本研究で得られた尺度構成や妥当性の知見を活かしながら尺度の改訂を進めることが可能であろう。

本研究では、科学リテラシーを構成する知識の一つである科学についての知識の一側面として、科学的推論能力を測定する尺度を作成した。科学についての知識は科学的推論に限らず、科学の社会的影響や科学的探究の本質・限界点についての知識といった他の側面も含まれる。科学的推論尺度はこうした他の科学についての知識や、科学の知識、あるいは一般的な論理的推論能力や批判的思考を支える能力・態度と科学的推論能力との関連を検討する上で有用なツールである。科学リテラシーの発達の様相を捉え、教育実践への活用へとつなげる研究へと活用されることが期待される。

付記

本研究の付録は https://osf.io/9k5jx/ に公開している。本研究の調査 1、調査 2 は、日本科学教育学会第 42 回年会にて報告された「科学的推論尺度の心理測定的特徴」に追加分析と加筆をおこなったものである。各調査では他の研究の目的で尋ねている質問項目があったが、本稿では分析に用いた尺度のみを報告している。本研究は科学研究費補助金 18H01070 の助成を受けて行われた。

引用文献

Aikenhead, G.S., & Ryan, A.G. 1992 “The development of a new instrument: “Views on science-technology-society” (VOSTS)” Science Education, 76, pp.477-491.

Chen, S. 2006 “Development of an instrument to assess views on nature of science and attitude toward teaching science” Science Education, 90, pp.803-819.

Drummond, C. & Fischhoff, B. 2017 “Development and validation of the scientific reasoning scale” Journal of Behavioral Decision Making, 30, pp.26-38.

Hacıeminoğlu, E., Yılmaz-Tüzün, Ö., & Ertepinar, H. (2012) “Development and validation of nature of science instrument for elementary school students” International Journal of Primary, Elementary and Early Years Education, 42, pp.258-283.

楠見 孝・平山るみ 2013 「食品リスク認知を支えるリスクリテラシーの構造―批判的思考と科学リテラシーに基づく検討」『日本リスク研究学会誌』23, 165-172 頁 .

Liang, L.L., Chen, S., Chen, X., Jaya, O. N., Adams, A. D., Macklin, M. & Ebenezer, J. (2008) “Student understanding of science and scientific inquiry (SUSSI): Revision and further validation of an assessment instrument”. Paper presented at the Annual Conference of the National Association for Research in Science Teaching, San Francisco, CA.

小塩真司・阿部晋吾・カトローニピノ 2012 「日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み」『パーソナリティ研究』21, 40-52 頁 .

Pardo, R. & Calvo, F. (2004) “The cognitive dimension of public perceptions of science: Methodological issues” Public Understanding of Science, 13, pp. 203-227.

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J. (2011) “A simple syllogism-solving test: Empirical findings and implications for g research” Intelligence, 39, pp. 89-99.

Psychometric properties of Scientific Reasoning Scale in Japanese

GOTO Takayuki , SHIOSE Takayuki, KANO Kei

Abstract: This study develops and examines the psychometric properties of the Scientific Reasoning Scale (SRS) in Japanese, based on the SRS developed by Drummond and Fischhoff (2017). People of different ages were asked to complete the survey: adults (N = 343; Study 1), university students (N = 93; Study 2), and high school students (N = 250; Study 3). Analysis using item response theory suggests that the SRS has a unidimensional structure and can capture individual differences among people with a wide range of abilities, from high to low scientific reasoning ability. The results also suggest that scores on the SRS were positively correlated to factual knowledge of science, critical thinking dispositions, and openness of experiences, suggesting that it is a valid measure of scientific reasoning ability. Although the results were not consistent for the relationship between the SRS and reasoning ability assessed by a syllogism-solving test, the results revealed a positive correlation between them in high school students, partially supporting the validity of the scale. Based on the psychometric properties of the SRS in Japanese, we discuss considerations in using it to assess scientific reasoning ability and in improving and expanding the item set of the SRS.