

中学校理科のテスト場面における確信度判断の正確さ
—パーソナリティ特性と個人差変数が感度と特異度に与える影響—
Accuracy of confidence judgments for science test in junior high school students

○今野浩幸*¹, 久坂哲也*¹

Hiroyuki KONNO*¹, Tetsuya HISASAKA*¹

*¹ 岩手大学

*¹Iwate University

【要約】 本研究は、中学校理科のテスト場面における確信度判断の正確さに影響を与えるパーソナリティ特性と個人差変数を検討することを目的とした。中学校第2学年を対象に、学業的自己概念、自己愛傾向、Big Fiveを測るための質問紙およびローカル判断、グローバル判断の正確さを測るためのテスト作成し、調査を実施した。得られた126名のデータを分析した結果、成績は γ 係数と感度に正の影響を与えることが示された。また、外向性は感度に正の影響を、評価過敏性は特異度に正の影響を与えることが示された。

【キーワード】 確信度判断, バイアス, キャリブレーション, γ 係数, 感度, 特異度

I. 問題と目的

自己の学習を遂行する上で、正確に確信度判断 (confidence judgment) を行うことは極めて重要である。正確な判断によって得られる情報は、メタレベルで使用することができるため、その後のパフォーマンスを調整するのに役立つと考えられる (Schraw et al., 2013)。

テストの文脈における確信度判断とは、自身のパフォーマンスに対して行う判断である。確信度判断は、ローカル判断 (local judgment) とグローバル判断 (global judgment) に大別される。ローカル判断は、個々の項目ごとに正解か不正解か判断を行う一方、グローバル判断は、テスト全体に対して判断を行う。しかし、誤った判断をしてしまう理由は、まだ十分に理解されていない (Serra & DeMarree, 2016)。

先行研究において、確信度判断の正確さは、パーソナリティ特性の影響を受けることが報告されている。例えば、Händel et al. (2020) の大学生を対象とした研究において、領域固有の学業的自己概念は、感度 (sensitivity) に正の影響を与え、特異度 (specificity) に負の影響を与えることを示している。

Händel et al. (2020) は、確信度判断と個人差の関連を検討した先行研究において、主にバイアスなどの1つの指標のみで評価することを問題としている。また、Schraw et al. (2013) は、確信度判断の正確さを感度と特異度によって測定すべきであると指摘している。さらに、多くの国で子どもが大規模な評定付きテストに

直面するのは中学校であることから (Nederhand et al., 2020)、中学生段階で確信度判断の正確さに着目した研究を行うことには意義があると考えられる。しかし、中学生の理科学習における確信度判断の正確さを感度と特異度によって測定し、個人差との関連を検討した研究は、管見の限り見受けられない。そこで、本研究では正確さの指標としてバイアス、キャリブレーションに加えて、 γ 係数、感度、特異度を用いて個人差との関連について検討することを目的とした。

II. 方法

1. 調査協力者

2020年10月上旬に、岩手県内の中学校1校の第2学年4学級の生徒140名に調査協力を依頼した。

2. 手続き

調査は、2020年10月に各学級の理科授業内において担当教師の監督のもと実施された。質問紙への回答後、テストを行った。調査の冒頭に、調査結果は理科の成績と一切関係のないこと、個人を特定して分析したり、個人の結果を公表したりすることがないことを事前に説明し、了承を得た。

3. 材料

テスト テストは、中学校第1学年の理科の内容から作成した。物理、化学、生物、地学の領域から各5問、計20問で構成され、配点は各1点とした。テスト形式は、各領域から多肢選択

型 2 問, 穴埋め型 2 問, 論述型 1 問で作成した。テストの各解答欄の右欄に解答が正解だと思う場合は○を, 不正解だと思う場合は×を記入するよう求めた (i.e., ローカル判断)。また, テスト終了後に 20 問中何問正解したと思うか 0~20 の整数で記入するよう求めた (i.e., グローバル判断)。

ローカル判断の正確さの指標は γ 係数, 感度, 特異度によって評価される。それぞれの指標は 2×2 の分割表を基に算出される (表 1)。

表 1 ローカル判断における 2×2 分割表

判断	実際のパフォーマンス	
	正解	不正解
正解	a	b
不正解	c	d

γ 係数は, 2×2 の分割表において, 以下の式で表される。 γ 係数は, -1 から 1 までの値をとり, 1 に近いほど正確さが高い。

$$\gamma \text{ 係数} = \frac{ad - bc}{ad + bc}$$

感度は, 正確に検出された正解の割合を表し,

特異度は, 正確に検出された不正解の割合を表す。 2×2 の分割表において, それぞれ以下の式で表される。感度と特異度は, 0 から 1 までの値をとり, 1 に近いほど正確さが高い。

$$\text{感 度} = \frac{a}{(a + c)}$$

$$\text{特異度} = \frac{d}{(b + d)}$$

グローバル判断の正確さの指標は, バイアスとキャリブレーションによって評価される。バイアスは, 判断値とテスト得点の差である。値が正になれば過大評価, 負になれば過小評価を表す。一方, キャリブレーションは, 判断値とテスト得点の差の絶対値である。テスト得点に対して, 判断がどれだけ正確かを表す。判断値を c_i , テスト得点を p_i としたときバイアスとキャリブレーションは, それぞれ以下の式で表される。

$$\text{バイアス} = c_i - p_i$$

$$\text{キャリブレーション} = |c_i - p_i|$$

理科の成績 成績の指標として, 調査協力校で実施された理科の定期試験の成績を用いた。

表 2 各変数の記述統計量と信頼性係数および相関係数

	M	SD	α	1	2	3	4	5	6	7	8
テスト得点	12.71	3.12		.52***	-.05	.19*	-.03	-.04	-.01	.13	-.10
グローバル判断											
バイアス	-1.59	3.07		.30**	-.07	.26**	.18	.24**	-.08	.22*	.05
キャリブレーション	2.63	2.23		-.21*	-.02	-.18*	-.21*	-.06	-.01	-.16	.00
ローカル判断											
γ 係数	0.78	0.26		.27**	.10	.13	.26**	.03	-.04	.14	.09
感度	0.73	0.21		.50***	-.02	.31**	.28**	.09	-.10	.34***	.00
特異度	0.76	0.20		-.25**	.24**	-.20*	.05	-.09	.18*	-.22*	.02
パーソナリティ特性											
1. 学業的自己概念	3.45	1.09	.90		-.10	.38***	.23**	.12	-.04	.49***	.12
自己愛傾向											
2. 評価過敏性	2.94	1.05	.88			.17*	-.14	-.01	.60***	-.14	-.16
3. 誇大性	2.61	0.86	.81				.15	.19*	-.08	.52***	.21*
Big Five											
4. 外向性	3.99	1.09	.83					-.14	-.12	.50***	.14
5. 誠実性	3.09	0.81	.74						.02	-.04	.43***
6. 情緒不安定性	3.84	1.02	.80							-.18*	-.19*
7. 開放性	3.46	0.88	.78								.18*
8. 調和性	3.59	0.82	.73								

* $p < .05$ ** $p < .01$ *** $p < .001$

4. 質問紙

質問紙は、以下の構成とし、全回答を6件法(1:全く当てはまらない-6:とても当てはまる)で回答を求めた。また、選択番号を得点として下位尺度ごとに加算平均を算出した。よって、理論的中間値は3.50である。項目総数は46項目であった。

学業的自己概念尺度 富岡(2011)のSDQ-1の日本語版を参考に、理科固有の学業的自己概念尺度を作成した。尺度は、5項目(例“私は理科のみ込みが早い方である。”)であった。

自己愛傾向尺度 中山・中谷(2006)の評価過敏性-誇大性自己愛尺度を援用した。尺度は、12項目であった。下位尺度は、評価過過敏性(例“自分の欠点や失敗を少しでも悪く言われると、とても落ち込む。”など6項目)と誇大性(例“私にはもって生まれた素晴らしい才能がある。”など6項目)で構成される。

Big Five 尺度 並岡ほか(2012)のBig Five尺度短縮版を援用した。尺度は、29項目であった。下位尺度は、外向性(例“社交的”など5項目)、誠実性(例“几帳面だ”など7項目)、情緒不安定性(例“不安になりやすい”など5項目)、開放性(例“興味が広い”など6項目)、調和性(例“心が広い”など6項目)で構成される。

Ⅲ. 結果

欠席者や回答に不備のあった者を除いた結果、有効回答者数は質問紙136名、テスト126

名であった。全てのデータが揃ったものは126名であった。

理科の成績を除いた各変数の記述統計量とパーソナリティ特性の信頼性係数および相関係数を表2に示す。パーソナリティ特性における信頼性係数は $\alpha = .73-.90$ と概ね良好な信頼性が得られたといえる。テスト得点の平均値は、12.71点(正答率64%)であった。このことから、テストの難易度は、中程度だったことが示された。次に、バイアスは $M = -1.59, SD = 3.07$ であることから、全体としてテスト得点に対し過小評価していることが示された。また、感度よりも特異度が高いことから、正解よりも不正解を正確に検出していることが示された。

相関係数を見ると、学業的自己概念とバイアス、 γ 係数、感度の間に正の相関が、キャリブレーション、特異度の間に負の相関が見られた。また、誇大性と感度の間に正の相関が、特異度の間に負の相関が見られた。

正確さの各指標に影響を与える要因について調べるために、正確さの各指標を従属変数として階層的重回帰分析を行った。Step 1のモデルでは、個人差変数として性別と理科の成績を投入した。Step 2のモデルでは、各パーソナリティ特性を投入した。キャリブレーションを従属変数とした階層的重回帰分析では、分散分析の結果が有意にならなかったため使用を断念した($F(2,119) = 2.01, p = .14$)。分析の結果を表3に示す。バイアスに対し成績と誠実性が正の

表3 階層的重回帰分析の結果

	グローバル判断		ローカル判断					
	バイアス		γ 係数		感度		特異度	
	ΔR^2	β						
Step 1	.06*		.08**		.18***		.05*	
性別		-.15		-.03		-.11		.14
成績		.19*		.28**		.39***		-.17†
Step 2	.15*		.13*		.15**		.14*	
学業的自己概念		.10		.06		.21†		-.10
評価過過敏性		-.06		.25*		.11		.26*
誇大性		.15		-.01		.08		-.13
外向性		.16		.31**		.21*		.18†
誠実性		.32**		-.01		.14		-.12
情緒不安定性		-.02		-.12		-.14		.02
開放性		.04		-.06		.08		-.14
調和性		-.12		.11		-.05		.12
Total R^2	.21*		.21*		.33**		.20*	

† $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

影響を与えていた。また、 γ 係数に対し成績、評価過敏性、外向性が正の影響を与えていた。さらに、感度に対し成績と外向性が正の影響を、特異度に対し評価過敏性が正の影響を与えていた。

IV. 考察

初めに、グローバル判断の正確さについて考察する。バイアスの平均値は $M = -1.59$, $SD = 3.07$ と過小評価傾向であった。Serra & DeMarree (2016) では、大学生が自身の学習に対し、過大評価することを指摘している。しかし、中学生を対象とした先行研究では、パフォーマンスを過小評価する結果も見られる (e.g., 久坂・及川, 2019; Nederhand et al., 2020)。これらから、中学生の確信度判断の正確さに目を向けると、大学生と比べて過小評価する傾向の可能性が示唆された。

次に、ローカル判断の正確さについて考察する。 γ 係数と感度に対し、成績が正の影響を与えていた。成績の高い学習者は、学習内容に関しての知識を有しており、評価、熟考、予測の調整を行うことができたため (Serra & DeMarree, 2016)、正確な判断を行ったと考えられる。また、感度に対し、外向性が正の影響を与えていた。Händel et al. (2020) において、外向性のようにポジティブな期待感をもつ学習者は、自分の成果をより好意的に判断すると指摘しており、本研究においても同様の結果が得られたと言える。さらに、特異度に対し、評価過敏性が正の影響を与えていた。テスト実施後に、そのテストの成績が自己価値や自己価値を低められる証拠になり得るため (久坂・及川, 2019)、評価過敏性の高い学習者は、不正解を選択することで自己評価や自己価値を低く見積もっていると考えられる。

本研究において、理科固有の学業的自己概念は、感度と特異度に影響を与えていなかった。これは、大学生を対象として数学固有の学業的自己概念と感度および特異度の関連について検討した Händel et al. (2020) と対照的な結果である。中学生を対象としたことや、領域の違いによる影響が考えられる。

V. まとめと今後の課題

本研究では、中学校理科のテスト場面における確信度判断の正確さについて、性別、成績と

いった個人差や学業的自己概念、自己愛傾向、Big Five といったパーソナリティ特性による影響について検討することを目的として調査を行った。その結果、成績が γ 係数と感度に正の影響を与えることや、外向性が感度に正の影響を与えること、評価過敏性が特異度に正の影響を与えることが示された。

しかし、本研究は中学校 1 校を対象としたため、集団特性の影響を受けたことを否定できない。今後は、他の中学校においても調査を行うことで、より一般化可能な知見を得ることができると考える。

文献

- Händel, M., de Bruin, A. B. H., & Dresel, M. (2020) : Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, 15, 51-75.
- 久坂哲也・及川宏輝 (2019) : 中学生の理科学習に対するメタ認知的判断の正確さ (II) — 自己愛傾向と社会的望ましさはメタ認知を歪めるか —, 日本科学教育学会研究会研究報告, 34, 1, 67-70.
- 中山留美子・中谷素之 (2006) : 青年期における自己愛の構造と発達的变化の検討, 教育心理学研究, 54, 188-198.
- 並岡努・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.
- Nederhand, K. L., Tabbers, H. K., Jongerling, J., & Rikers, R. M. J. P. (2020) : Reflection on exam grades to improve calibration of secondary school students : a longitudinal study. *Metacognition and Learning*, 15, 291-317.
- Schraw, G., Kuch, F., & Gutierrez, A. P. (2013) : Measure for measure : Calibrating ten commonly used calibration scores. *Learning and Instruction*, 24, 48-57.
- Serra, M. J., & DeMarree, K. G. (2016) : Unskilled and unaware in the classroom : College students' desired grades predict their biased grade predictions. *Memory and Cognition*, 44, 1127-1137.
- 富岡比呂子 (2011) : 日米の小学生の自己概念 — 自己記述質問紙 (SDQ-1) の心理測定的検討 —, パーソナリティ研究, 19, 3, 191-205.