

項目反応理論によるクリティカルシンキング測定のための尺度開発

若山 昇^{1,2} 宮澤 芳光³ 梶谷 真司⁴ 植野 真臣⁵

¹帝京大学 ²CRET ³東京学芸大学 ⁴東京大学 ⁵電気通信大学

< 概要 >

クリティカルシンキング(CT)能力は現代社会を生きるうえで重要な能力になっている。現在、CT能力を評価する試験は世界的には数多く提案されている。CTの試験は、受験者に長い思考時間を要求するので、多くの項目を出題することは困難である。このためCTの能力全体を網羅し、かつ、できるだけ少ない項目数でCTの能力を測定できる試験が必要となる。本稿は、項目反応理論に基づきCTの能力の尺度開発とその評価を試みた。大学生736人を分析した結果、CTの3尺度(分析、推論、読解)は、相関、散布図、情報量、固有値から統計的に概ね独立で、尺度の有用性が示唆された。

キーワード：クリティカルシンキング、尺度開発、試験、項目反応理論、評価

1. はじめに

さまざまな情報にあふれた現代社会を生きぬくには、クリティカルシンキング(以下「CT」という)は必須の能力になっており、21世紀型スキル(Griffin *et al.* 2011)においてもその重要性が謳われている。CTとは、先入観に囚われず、論理的に考え、合理的な決定を導き出す能力と意思である(若山 2009)。CTの試験では、思考する時間が長くなるので、多くの項目を出題することは困難となる。このためCTの能力全体を網羅しかつできるだけ少ない項目数でCTの能力を測定できる試験が必要となる。本研究の目的はCTの能力を測定する尺度を開発することである。なお、尺度開発には、異なる項目で構成された試験を、同一の基準で評価が可能である項目反応理論(Linn 1992)を用いた。

2. 研究方法

2.1 尺度開発

CTの能力を評価するには、行動観察や口頭試問などがあるが、本研究では多肢選択型試験で測定できるものと仮定した。CTの多肢選択型尺度はこれまでに世界中で数多く開発されてきた。本研究で、先行するCT尺度を質的に分析したところ、以下の3尺度(①分析的思考力、②論理・推論能力、③読解・理解能力)と対応していることが導きだされた。

① 分析的思考力

情報や問題を的確に把握し理解するための分解・解析する能力である。いわゆる公務員試験や法科大学院の適性試験では分析、数的処理といわれる分野がこれに含まれる。具体的には、順列組合せ、確率、順序・手順、位置・対応、嘘つき・暗号問題などが含まれが、数学的素養の有無による影響を極力低減するよう配慮した。例えば、2元連立方程式を立なければ解けない問題は排除している。

② 論理・推論能力

論理展開の妥当性・整合性の有無及びその背景を的確に把握する能力である。いわゆる公務員試験や法科大学院の適性試験では論理、推論といわれる分野がこれに含まれる。具体的には、集合、命題、逆・裏・対偶、演繹・帰納、論理の前提や飛躍・省略、論理構造、因果・相関、第3変数などがこれに含まれる。

③ 読解・理解能力

情報を有機的に結合し活用して、内容の本質を的確に理解する能力である。PISA の応用力はこれに含まれる。難解すぎる文章や長い文章は、国語力が左右するので避けた。さらに受検者の慣れ・不慣れや既存の教科の知識量による影響を避け、論理展開・構造が複雑すぎないものとした。十分な思考時間を確保するため設問の文章は概ね 500~1,000 文字程度とした。

2.2 調査方法・対象

問題項目に関するデータを収集するためにペーパー試験を実施した。試験は 2014 年 7 月から 2015 年 12 月に実施し、東京及び近郊の大学生 736 人のデータを分析した。年齢は概ね 18~22 歳であり留学生及び重複・欠損値のある回答は対象から除外している。倫理的配慮として、教育・研究目的以外には使用しないこと、優秀者はクラス内で表彰するものの成績に無関係であること、さらに結果については個人名は特定されないことを説明して、学生の了解を得た。

2.3 分析方法

本研究では、①分析、②推論、③読解の各尺度ごとに項目パラメータを推定する。各尺度のテストでは、各尺度の共通項目の 5 項目を含め、表 1 の通り実施した。これにより各尺度毎に同一基準で困難度と識別力のパラメータの推定が可能となる。時間は 1 項目 4~5 分とし、1 組 15 問で 60~75 分に設定した。各尺度で 5~6 組の試験を作成し 1 組に概ね 50 人程度が解答した。クリティカルシンキングに関連する大学の授業で設問を 1 組ずつ実施し、632 人からデータを得た。別途、土曜日に 55 人ずつ、2 日で 107 人(3 人が重複)が集まり、①分析と③読解の両方の設問と②推論の共通項目に解答してもらった。尺度の有効性を検討するために IRTPRO3, SPSS20 を用い相関、散布図、情報量、固有价值などの分析を行った。

3. 結果と考察

個人の能力における各尺度間の相関(表 2)、尺度間の散布図(図 1a~1c)を以下に示す。各アイテムバンクのテスト情報量(I)と標準誤差(S.E.)は以下のとおり(図 2a~2c)。なお、各尺度の項目数は①54、②71、③54 であり、2b のテスト情報量が高いのは、②の項目数の多いことによる。さらに各尺度の固有价值プロットを(図 3a~3c)示す。

また、固有价值プロットでは、どの尺度も第 1 因子と第 2 因子間で大きく減少しており 1 因子の可能性が示唆された。各尺度の相関は、.423~.570 となり、比較的小さくなった。散布図からは 2 次元以上の関係性は確認されず、偏りもそれほど大きくはなかったことで、独立的であることが示唆された。

4. おわりに

3 尺度の相関はそれほど大きくはなく、固有价值プロット、困難度・識別力、情報量から、各尺度は 1 因子で独立的であることが示唆された。これらの結果から、CT の能力は①分析的思考力、②論理・推論能力、③読解・理解能力の 3 尺度で測定可能であることが示唆された。今後は規模をさらに拡大して 3 次元尺度の最適性を探究することが望まれよう。

なお、本稿は若山ほか(2016)、若山ほか(2017)をもとに、速報としてまとめたものである。本研究は科研費(A)15H01772、(C)15K01088 及び CRET の助成を受けている。

表1 尺度毎の問題と受検者の概要

問題番号	①分析的思考力					②論理・推論能力					③読解・理解能力							
	1A	1B	1C	1D	1E	1共通	2A	2B	2C	2D	2E	2共通	3A	3B	3C	3D	3E	3共通
問題数	10	10	10	10	10	5	10	10	10	10	10	5	10	10	10	10	10	5
クリティカルシンキングに関連する授業で実施 632人	○					○					○							
週末に集合して実施 55人	○					○					○							
週末に集合して実施 55人	○					○					○							

注：授業と2つの週末において3人のダブルがある

表2 個人能力の尺度間相関

	Θ1	Θ2	Θ3
Θ1		.568**	.570**
Θ2			.423**
Θ3			

** : p < .01

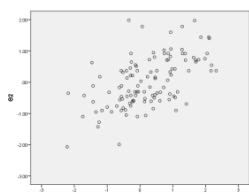


図 1a ①分析的思考力と②論理・推論能力

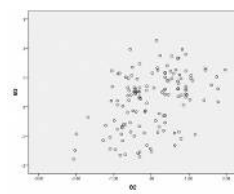


図 1b ②論理・推論能力と③読解・理解能力

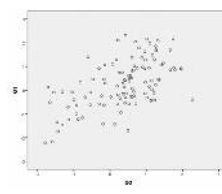


図 1c ③読解・理解能力と①分析的思考力

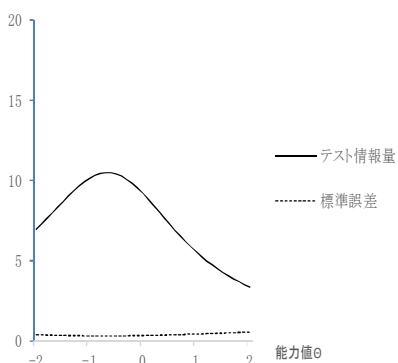


図 2a ①分析的思考力の情報量と標準誤差

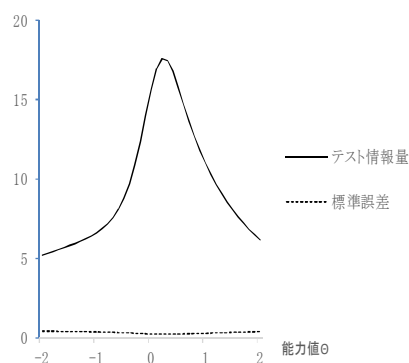


図 2b ②論理・推論能力の情報量と標準誤差

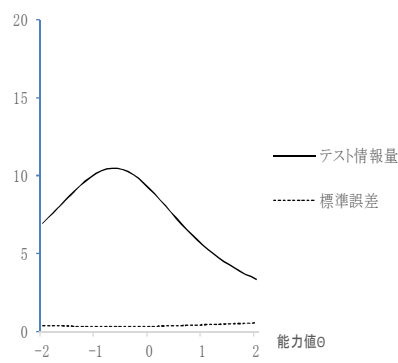


図 2c ③読解・理解能力の情報量と標準誤差

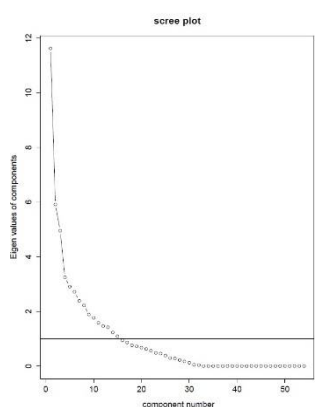


図 3a ①分析的思考力の固有値プロット

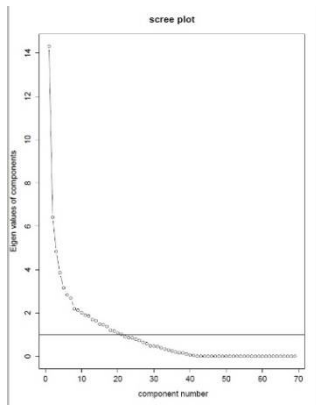


図 3b ②論理・推論能力の固有値プロット

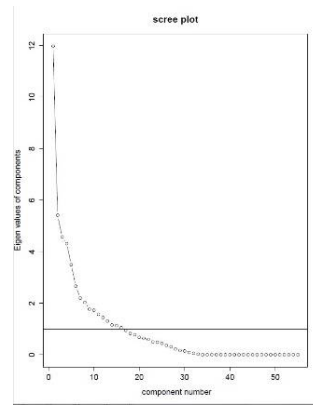


図 3c ③読解・理解能力の固有値プロット

<参考文献>

Griffin, P. et al. (2011) Assessment and Teaching of 21st Century Skills, Springer, New York, pp.17-66.

Linn, Robert L.池田央ほか(訳) (1992) 教育測定学 上下巻, 学習評価研究所

若山昇(2009) 大学におけるクリティカルシンキング演習授業の効果, 大学教育学会誌 31(1), pp.145-153

若山昇, 宮澤芳光, 梶谷真司, 植野真臣(2016) クリティカルシンキングの適応型テストの尺度開発, 日本テスト学会, 第14回大会発表論文抄録集, pp.58-61

若山昇, 宮澤芳光, 梶谷真司, 植野真臣(2017) クリティカルシンキング能力測定のための項目反応理論に基づいた尺度開発, 教育システム情報学会, Research Report, Vol.31, No7 pp.151-158