

クリティカルシンキング検定試験の妥当性

— 実験参加者のデータによる検討 —

若山 昇¹² 宮澤芳光³ 宇都雅輝⁴ 植野真臣⁴

¹帝京大学 ²CRET ³大学入試センター ⁴電気通信大学

<概要> さまざまな情報が溢れる現代を生き抜くためには、的確な意思決定が必要であり、クリティカルシンキングが不可欠となる。本研究の目的は、項目反応理論(IRT)により開発したクリティカルシンキング検定試験の妥当性を受験データで検討することである。このため大学の研究者・教員、学部学生、高校生に実際に受験してもらいデータを分析したところ、能力値に有意差がみられたことから、本尺度の妥当性が確認された。

キーワード：試験，妥当性，項目反応理論，クリティカルシンキング，尺度

1. はじめに

現代はさまざまな情報が満ち溢れているので、これを生き抜くためには、情報を十分に吟味して体系的に理解し、的確に意思決定を行う必要がある。このためクリティカルシンキングは不可欠である。クリティカルシンキングとは、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)。21世紀型スキル(Griffin *et.al.* 2015)においても、その重要性は謳われている。さらに、我が国では直面する課題や将来の状況を鑑み、求められる人材像・目指すべき新しい大学像を念頭に「クリティカルシンキング等を問う新たな共通テストの開発」(文部科学省 2012)が求められている。また、クリティカルシンキングは就活時でも重視されている。日本語での試験は限られており、教育効果を測定できる試験開発も望まれている(平山 2011)。本研究は、開発しているクリティカルシンキング検定試験について実際の受験データにより妥当性を検討することを目的とする。

2. 方法

2.1 項目反応理論

クリティカルシンキングの試験では、思考力を測るため時間がかかるので、多くの項目を出題するのは難しい。そこで、クリティカルシンキングの全体を網羅し、少ない項目で正確に測定できる試験が必要となる。このため項目反応理論(IRT)(Lord 1980)を利用する。

IRTでは、項目の特性と実験参加者(以下、参加者)の能力から、ある項目に参加者がどの程度の確率で正答するかをモデル化する。これにより、既存のテスト問題(以下、項目)では、同じテストを受けた者同士でしか能力比較ができないが、IRTでは異なる項目によって構成されたテストでも同一の尺度上で評価できる。さらに、IRTではその項目によって測定したい能力および参加者の能力を、どの程度確かに測定できるかを算出できる。

2.2 アイテムバンク

これまでに開発された既存のクリティカルシンキング試験で測定されている能力を分析したところ、下位の領域として①分析的思考力、②論理・推論能力、③読解・理解能力が導かれた。

なお, 論理・推論能力の問題(イメージ)を図 1 に示す。

問: 「人は死ぬ」「ソクラテスは人である」この 2 つの命題を合わせることから「ソクラテスは死ぬ」という 100% 確実な結論になる。以下の設問において, 結論「?」に最適なものを(1)~(5)の中から選べ。

「飛行機に乗ると遅れないで着く」「あの人は, 遅れないで着いた」したがって「?」

- (1) あの人は, 飛行機に乗った
- (2) 遅れるのは飛行機に乗らないからだ
- (3) 遅れないためには飛行機に乗る
- (4) 飛行機は, 時速 500km を超えている
- (5) 何も入りえない

図 1 クリティカルシンキングの論理・推論能力の問題(イメージ)

図 2 に示すように各尺度では 15 問の試験を 5 組~6 組作成し, 等化の推定のために共通項目を各尺度内それぞれ 5 問とした。試験時間は 1 問 4~5 分として, 60~75 分に設定し, 1 組に合計で 50 人程度が解答した。大学の授業などで試験を 1 組ずつ行い合計 736 人のデータを得た。IRTPRO3, SPSS24 で分析したところ, 散布図, 相関, 情報量及び因子分析によるスクリープロットにより, 各尺度は概ね一因子であり独立的であることが示唆された。

| 受検者グループ | 項目番号 | | | | | | | | | | | | | | | | | |
|---------|------|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 6 | 16 | 26 | 36 | 46 | 101 | 106 | 116 | 126 | 136 | 146 | 201 | 206 | 216 | 226 | 236 | 246 |
| | 5 | 15 | 25 | 35 | 45 | 55 | 105 | 115 | 125 | 135 | 145 | 155 | 205 | 215 | 225 | 235 | 245 | 255 |
| 1A | ✓ | ✓ | | | | | | | | | | | | | | | | |
| 1B | ✓ | | ✓ | | | | | | | | | | | | | | | |
| 1C | | | | | ✓ | | | | | | | | | | | | | |
| 1D | ✓ | | | | | ✓ | | | | | | | | | | | | |
| 1E | ✓ | | | | | ✓ | | | | | | | | | | | | |
| 2A | | | | | | | ✓ | ✓ | | | | | | | | | | |
| 2B | | | | | | | ✓ | | ✓ | | | | | | | | | |
| 2C | | | | | | | ✓ | | | ✓ | | | | | | | | |
| 2D | | | | | | | ✓ | | | | ✓ | | | | | | | |
| 2E | | | | | | | ✓ | | | | | ✓ | | | | | | |
| 3A | | | | | | | | | | | | | ✓ | ✓ | | | | |
| 3B | | | | | | | | | | | | | ✓ | | ✓ | | | |
| 3C | | | | | | | | | | | | | ✓ | | | ✓ | | |
| 3D | | | | | | | | | | | | | ✓ | | | | ✓ | |
| 3E | | | | | | | | | | | | | ✓ | | | | | ✓ |

図2 各尺度の項目と受検者の概念図

3. 妥当性の確認

試験の妥当性を評価する場合は, 類似した試験間との相関を調べることで評価するケースが多い(Campbell 1959, Zahner 2013)。しかるに, 既存の試験は英語のものが多く, 日本に即したものにすることが必要があり, そのまま和訳して評価することは困難とされている(川島 2007, 井上 1974)。そこで, 本研究では「クリティカルシンキング力は, 平均的な高校生より平均的な大学生の方が高く, 平均的な大学生より大学の研究者の方が高い」という仮説を立て, これを受験データで確認することで, 試験の妥当性を検討することとした。

中堅の公立の高校生のべ 182 人, いわゆる偏差値 50 程度の大学の学生のべ 319 人, さらに必ずしも十分な人数とはいえないが大学の研究者のべ 44 人に参加してもらった。なお, 倫理的配慮として, 教育・研究目的以外には使用せず個人名は特定されないことを実験参加者に説明し, 承諾を得ている。

4. 結果と考察

実験参加者の尺度ごとの平均能力値を図 3 に示す。尺度 1 と尺度 2 の能力値 Θ を分散分析したところ, 群間の能力値の差が有意で ($F(2,128)=25.2, p<.001, F(2,283)=30.9, p<.001$) あったことから, 多重比較(Tukey の HSD)を行ったところ, 尺度 1 では高校生と大学生, 大学生と研究者にともに $p < .001$ の有意差があった。尺度 2 では大学生と研究者に $p < .001$ の有意差があったものの, 高校生と大学生では差がなかった。この理由は, 尺度 2 に参加した高校がいわゆる中高一貫校の 6 年制であり, 特別講義の実施などクリティカルシンキングに注力していることから, 一般の平均的高校レベルよりクリティカルシンキングの能力が高かったことが考えられた。

また、尺度 3 では高校生データは得られてないので、研究者と大学生での t 検定を行ったところ、($t(126)=2.76, p<.001$)有意差が現れ、仮説どおりとなった。なお、現状では、高校生から多くのデータを取得するのは難しく、今後の課題となる。

5. むすびに

これらの分析結果から開発したクリティカルシンキングの 3 尺度には、妥当性が有ることが示唆された。妥当性を完璧に検証することは難しい。特に衆目の一致するような外的基準が存在しない場合には、他の方法で試験の妥当性を検証するべく、さらなる検討が望まれよう。

なお、本稿は若山ら(2017)の一部をもとに発展させ考察等を加えてまとめたものである。

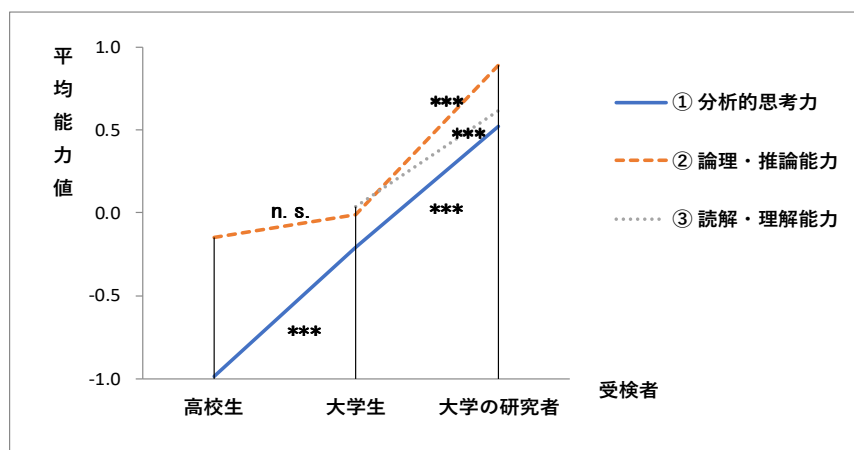


図3 尺度別の受験者の平均能力値

6. 参考文献

Campbell, D. T. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81- 105.

Griffin, P., Care, E. (2015) *Assessment and teaching of 21st century skills*, Springer Science+Business Media, Dordrecht

平山るみ (2011) 批判的思考の測定, 楠見孝ほか (編) 批判的思考力を育む, 有斐閣

井上尚美ほか (1974) 批判的思考力テスト日本語版作成の試み, 日本教育心理学会総会発表論文集 16(0) : 418-419

川島範章 (2007) 批判的思考態度の形成と深化に関する研究, 雄松堂出版

Lord, F.M. (1980) *Applications of item response theory to practical testing problems*, Lawrence Erlbaum Associates, Hillsdale, NJ

文 部 科 学 省 (2012) 大 学 改 革 実 行 プ ラ ン
http://www.mext.go.jp/b_menu/houdou/24/06/___icsFiles/afieldfile/2012/06/05/1312798_01_3.pdf
 (参照日2019.4.29)

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 (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 (2017-3) 151-158

Zahner, D. (2013) *Reliability and validity CLA+*, Council for Aid to Education (CAE)
http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf (参照日2018.1.30)

若山 昇, 宮澤 芳光, 宇都 雅輝, 植野 真臣