

令和 元年 9 月 11 日現在

機関番号：32643

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K01088

研究課題名(和文)クリティカルシンキング検定試験用の多次元項目反応理論に基づく適応型テストの開発

研究課題名(英文)Development of Adaptive Testing Based on Item Response Theory for Critical Thinking Test

研究代表者

若山 昇(Noboru, WAKAYAMA,)

帝京大学・法学部・准教授

研究者番号：90439589

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：1. 研究の概要：クリティカルシンキングが重要であるとされている。しかし、この領域は広く、能力の測定には時間がかかる。一方、テスト理論では、効率的に能力を測定するため、項目反応理論(以下IRT)に基づく適応型テスト(以下CAT)が利用される。そこで本研究では、クリティカルシンキングの検定試験のため、IRTに基づくCATの開発を研究目的とし、以下の課題に取り組んだ。

2. 研究実績：(1)尺度開発：IRTにより更なる尺度開発を推進した。分析的思考力 論理・推論能力 読解・理解能力においてパラメータの最適な推定をした。(2)実験参加者からデータを収集し、信頼性・妥当性の検証を行った。

研究成果の学術的意義や社会的意義

クリティカルシンキングの能力測定においても、項目反応理論に基づく適応型テストを用いて効率的に測定することが可能になり、測定精度を向上することができるため、ハイスタークスのテストに利用することが可能となった。

研究成果の概要(英文)：Summary of research: Critical thinking is considered important. However, this area is wide and measuring the ability takes time. On the other hand, test theory measures ability efficiently using computer adaptive test (CAT) based on item response theory (IRT). Therefore, in this research, we aimed at the development of CAT based on IRT for the examination of the critical thinking test, and addressed the following issues.

Research Results: Scale development: IRT promoted further scale development by optimal estimation of parameters. (1) Analytical thinking ability scale (2) Logic and reasoning ability scale (3) Reading and understanding ability scale. In addition, data were collected from the participants in the experiment to verify their reliability and validity.

研究分野：クリティカルシンキング

キーワード：クリティカルシンキング 項目反応理論 試験

## 1. 研究開始当初の背景

現代は様々な情報があふれており、十分に吟味した情報を体系的に理解し、的確な意思決定を行うためには、クリティカルシンキングは不可欠である。社会人基礎力(経済産業省 2006)、学士力(文部科学省 2008)、21 世紀型スキル(Griffin 他 2011)においても、その重要性は謳われている。なお、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)。

クリティカルシンキングが重要であるとされている。しかし、この領域は広く、能力の測定には時間がかかる。一方、テスト理論では、効率的に能力を測定するため、項目反応理論(以下 IRT)に基づく適応型テスト(以下 CAT)が利用される。

## 2. 研究の目的

近年、資質・能力の育成の重要性が指摘されて、その中でもクリティカルシンキング能力は現代社会を生きるうえで必須の能力になってきている。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<sup>(1)</sup>。現在、クリティカルシンキングの能力を測定するためのテストは、世界中で数多く提案され、それぞれのテストでは複数の領域が存在し、それらの領域の項目を全て解答することによってクリティカルシンキングの能力を測定している。

研究の目的は、クリティカルシンキングの検定試験の開発であり、このため IRT に基づく CAT の開発を研究目的とし、以下の課題に取り組む。

- 1) クリティカルシンキングの領域で IRT のためのアイテムバンクを作成する。
- 2) アイテムバンクを有効活用するべく、IRT のアルゴリズムを開発し、実装する。
- 3) 検定試験のためのシステムを開発し、その成果を、印刷物・Web や学会を通して公開する。

## 3. 研究の方法

クリティカルシンキングの領域において、3 尺度の項目反応理論(IRT)用にスコア化を行い、アルゴリズム作成し、CAT に実装して、検定試験のシステムを開発した。

### (1) 項目反応理論

本研究では、項目の分析に項目反応理論を用いる。項目反応理論の利点として以下の点が知られている。

- 1) 受検者が異なる項目で構成されたテストを受けた場合においても、同一尺度上で評価することができる
- 2) データの中に欠損値があっても容易にパラメータを推定することができる

項目反応理論では、2パラメータロジスティックモデルが最も一般的によく使われている。

受検者  $j$  の項目  $i$  への正答率  $P(u_{ij} = 1 | \theta_j)$  を以下のように定義する。

受検者  $j$  の項目  $i$  への正答率  $P(u_{ij} = 1 | \theta_j)$  を以下のように定義する。

$$P(u_{ij} = 1 | \theta_j) = \frac{1}{1 + \exp\{-1.7(\alpha_i(\theta_j - \beta_i))\}}$$

ただし、

$$u_{ij} = \begin{cases} 1: \text{受検者 } j \text{ が項目 } i \text{ に正答したとき} \\ 0 \text{ 上記以外} \end{cases}$$

識別力パラメータ  $\alpha_i$  (discrimination parameter) と難易度パラメータ  $\beta_i$  (difficulty parameter) の二つのパラメータを項目ごとに推定する。

式  $(1)$  の  $\alpha_i$  は、識別力パラメータいわれ項目の識別力を示す。図 1 は、項目反応関数 (item response function: IRF) の例であり、x 軸の受検者の能力パラメータ  $\theta_j$  であり、y 軸はその能力パラメータの受検者の正答確率を示している。図 1a には識別力パラメータ  $\alpha_i$  の傾きの小さい方から ( $\alpha_i = 0.5$ ,  $\alpha_i = 1$ ,  $\alpha_i = 2$ ) 異なる三つの項目反応関数を示す。図のように識別力パラメータ  $\alpha_i$  の値が大きいと、図 1 の IRF の傾きが大きくなる。

同様に式  $(1)$  の  $\beta_i$  は、難易度パラメータであり項目の困難度を示している。図 1b の横軸は受検者の能力パラメータであり、縦軸はその能力パラメータ  $\theta_j$  の受検者の正答確率を示している。図 1b は難易度パラメータ  $\beta_i$  が左から ( $\beta_i = -1$ ,  $\beta_i = 0$ ,  $\beta_i = 1$ ) の項目反応関数を示す。難易度パラメ

一タ  $b_i$  の値が大きいと、図 2 の IRF は右にシフトする。また IRF の勾配が最大になるのは、 $\beta_i = \theta_j$  となるところであり、傾きが 0.5 となる。

項目反応理論では、個々の項目が受験者の能力をどの程度の精度で測定できるかを以下のフィッシャー情報量により評価できる<sup>(13)</sup>。

$$I(\theta_j | \alpha_i, \beta_i) = 1.7^2 \sum_{i=1}^n \alpha_i^2 P(u_{ij} = 1 | \theta_j) (1 - P(u_{ij} = 1 | \theta_j)) \quad (2)$$

また、図 2 は項目情報関数の例であるが、例えば能力パラメータ  $\theta_0$  を推定する場合、項目 1 が項目 2, 3 より項目情報量が大きくなるため、能力パラメータ推定の精度が高くなる。

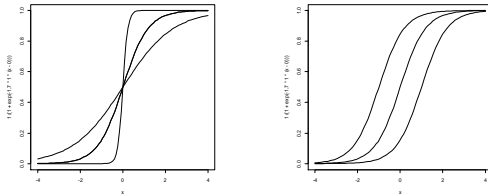


図 1a：能力値による正解率の変化 a 図 1b：能力値による正解率の変化 b

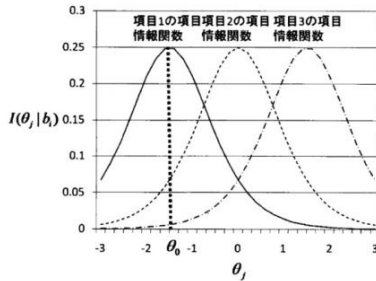


図 2 項目情報関数例

## (2) 尺度の開発

### ① 既存尺度の分類と整理

既存の尺度とその内容を分類・整理することを試みた。まず、既存の尺度及びその下位の領域の共通点・相違点について比較した。そこで本研究において、いかなる下位の尺度を設定すべきかを検討するべく、探索的に下位の尺度数とその内容を変化させるところ、3 尺度(i)分析的思考力、(ii)論理・推論能力、(iii)読解・理解能力が導きだされた。

### ② 開発する尺度の下位の尺度

下位の尺度は以下の 3 尺度となる

(i)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。

(ii)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。

(iii)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。

クリティカルシンキングでは、知識を覚えていることより、考えることの方が重要であるので、設問を解答する際には、極力知識量に左右されないように細心の注意を払って項目を設定した。

## (3) 分析方法

尺度作成のために策定した項目群を分析するために、東京及び近郊の大学生 700 人超を対象として、試験を実施した。倫理的配慮として、成績に無関係で教育・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説明し、学生の了解を得た。

また、等化の推定のための共通項目を、各尺度内それぞれ5問として、各尺度では15問の試験を5組～6組作成した。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。大学の授業では試験を1組ずつ行いデータを得た。これらで得られたデータを、IRTPRO、SPSSを用いて相関、散布図、信頼性係数、情報量を分析し、さらに、因子分析により固有値プロットにより、尺度の一因子性を検討した。

#### 4. 研究成果

(1) 尺度開発：IRTにより尺度開発を推進した。

(i)分析的思考力 (ii)論理・推論能力 (iii)読解・理解能力においてパラメータの最適な推定をした。

(2) 実験参加者からデータを収集し、妥当性の検証を行った。

クリティカルシンキングの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独立で,信頼性が確認された。さらに分析の尺度を例に,大学研究者,大学生,高校生の結果の差異から尺度の妥当性が示された。

#### 5. 主な発表論文等

〔雑誌論文〕(計10件)

若山昇,大浦宏邦,長谷川成海,植野真臣,クリティカルシンキングに対する志向性について,教育テスト研究センター年報,査読有,第1号,40-42,2016,

若山昇,大浦宏邦,長谷川成海,植野真臣,クリティカルシンキングに対する志向性に関する検討,情報文化学会誌,査読有,23(2),43-50,2016

若山昇,宮澤芳光,梶谷真司,植野真臣,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2016年度特集論文研究会研究会報告(JSiSE Research Report),Vol.31, No7,151-158,2017

若山昇,宮澤芳光,梶谷真司,植野真臣,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,教育テスト研究センター年報,査読有,第2号,61-63,2017

若山昇,宮澤芳光,梶谷真司,宇都雅輝,植野真臣,クリティカルシンキングの設問における識別力・困難度-ベイズ,4枚カード問題を項目反応理論で分析-,教育テスト研究センター年報,査読有,第3号,28-30,2018

宮澤芳光,宇都雅輝,石井隆稔,植野真臣:測定精度の偏り軽減のための等質適応型テストの提案,電子情報通信学会論文誌.D,査読有,Vol.101, No.6,909-920,2018

Maomi Ueno, Yoshimitsu Miyazawa IRT-Based Adaptive Hints to Scaffold Learning in Programming, IEEE Transactions on Learning Technologies, IEEE computer Society, 査読有, Vol.11, Issue 4, 415-428, 2018

Masaki Uto, Maomi Ueno: Empirical Comparison of Item Response Theory Models with Rater's Parameters. Heliyon, Elsevier, 査読有, Vol.4, No 5, 2018, 2018

宇都雅輝,植野真臣:ピアアセスメントにおける異質評価者に頑健な項目反応理論. 電子情報通信学会論文誌D, 査読有, Vol. 101, No.1, 211-224, 2018,

若山昇,立野貴之,河村一樹,飯箸泰宏,大学生の非認知的能力の向上-「日常のなぜ」をクリティカルに探究する試み-,帝京大学高等教育開発センターフォーラム,Vol.6,2019,43-63

〔学会発表〕(計20件)

若山昇,立野貴之,クリティカルに考えることを育成するための教材開発~「なぜなぜカルタ」による質問力・仮説構築力の向上~,日本情報科教育学会,第8回全国大会,2015

立野貴之,加藤尚吾,若山昇,加藤由樹,ビジネスゲームを利用した授業におけるフィードバックの方略の違いが学習に与える影響,日本情報科教育学会,第8回全国大会,2015

若山昇,立野貴之,学生にクリティカルシンキングを促すための授業実践(2)-ディベートに頼らず,疑問を投げかける効果について-,教育システム情報学会,第40回全国大会,2015

若山昇,大浦宏邦,長谷川成海,クリティカルシンキングに対する志向性とその能力,日本教育工学,第31回全国大会,2015

宮澤芳光,住政二郎,植野真臣:知識活用力を測定するための適応型テスト,教育システム情報学会,第40回全国大会,2015

若山昇,立野貴之,館秀典,学生の聴く態度が変化するプレゼンテーション実践,情報コミュニケーション学会第13回全国大会発表,2016

若山昇, 小澤伸二, 立野貴之, 「情報の科学」における問題解決能力を目的としたeラーニング教材の開発 - クリティカルシンキングの教材活用 - (学会賞受賞), 日本情報科教育学会 第9回全国大会, 2016

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, クリティカルシンキングの適応型テストのアイテムバンク構築, 教育システム情報学会, 第41回全国大会, 2016

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, 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, 2016

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, クリティカルシンキングの能力測定のための適応型テスト, 日本教育工学会, 第32回全国大会, 2016

Noboru Wakayama, Takashi Tachino, Shinji Ozawa, Hidenori Tachi, Development of Critical Thinking Support Program, Asia-Pacific Conference on Social Sciences and Management, 2017

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, 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 2016年度 特集論文研究会, 2017

若山昇, 宮澤芳光, 立野貴之, 植野真臣, 高校生のクリティカルシンキング能力をIRTにより測定する試験開発, 日本情報科教育学会, 第10回全国大会, 2017

Noboru Wakayama, Yoshimitsu Miyazawa, Shinji Kajitani, Maomi Ueno, Critical Thinking Ability Scale Development on Item Response Theory, Conference of The International Federation of Classification Societies, IFCS-2017

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, クリティカルシンキングの難問を項目反応理論で分析する試み-4枚カード, ベイズ問題を例に -日本テスト学会, 第15回大会, 2017

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, クリティカルシンキングの適応型テスト開発のシミュレーション, 教育システム情報学会, 第42回全国大会, 2017

若山昇, 宮澤芳光, 梶谷真司, 植野真臣, 項目反応理論によるクリティカルシンキング能力試験の妥当性, 日本教育工学会, 第33回全国大会, 2017

若山昇, 立野貴之, 木原裕紀, 納庄聡, Webの記事情報に対してクリティカルな思考を促す授業実践, 日本情報科教育学会 第11回全国大会, 2018

若山昇, 宮澤芳光, 梶谷真司, 宇都雅輝, 植野真臣, クリティカルシンキングの試験開発における妥当性, 日本行動計量学会 第46回大会, 2018

若山昇, 宮澤芳光, 梶谷真司, 宇都雅輝, 植野真臣, 高校生のクリティカルシンキングの伸びを測定, 日本教育工学会 第34回全国大会, 2018

## 6. 研究組織

### (1)研究分担者

植野真臣 (UENO, Maomi)

電気通信大学・情報システム学研究科・教授  
研究者番号: 50262316

宮澤芳光 (MIYAZAWA, Yoshimitsu)

独立行政法人大学入試センター・研究開発部・助教  
研究者番号: 70726166

立野貴之 (TACHINO, Takashi)

松蔭大学・観光メディア文化学部・准教授  
研究者番号: 50564001  
(2016→2018)

### (2)研究協力者

梶谷真司 (KAJITANI, Shinji)

東京大学・総合文化研究科・教授  
研究者番号: 50365920