

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

Critical Thinking Ability Scale Development

on Item Response Theory

若山 昇^{*1}

宮澤 芳光^{*2}

梶谷 真司^{*3}

植野 真臣^{*4}

Noboru WAKAYAMA^{*1}

Yoshimitsu MIYAZAWA^{*2}

Shinji KAJITANI^{*3}

Maomi UENO^{*4}

^{*1} 帝京大学/CRET

^{*2} 東京学芸大学

^{*3} 東京大学

^{*4} 電気通信大学

^{*1}Teikyo University/CRET ^{*2}Tokyo Gakugei University

^{*3}The University of Tokyo ^{*4}The University of Electro-Communications

Email: Wakayama.class@pobox.com

近年、資質・能力の育成の重要性が指摘され、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。現在、クリティカルシンキング能力を評価するテストは世界的には数多く提案されているが、それに伴い多くの尺度が存在する。被検者のクリティカルシンキング能力や学習効果の測定では、比較的短時間で正確に測る必要があるが、クリティカルシンキングには、複数の領域が関わるので容易ではない。本稿は、項目反応理論に基づきクリティカルシンキング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、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唆された。

キーワード：クリティカルシンキング、項目反応理論、尺度開発、試験、評価

1. はじめに

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⁽¹⁾。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。このため、クリティカルシンキングを測定するために現存する領域が多岐に渡り、クリティカルシンキングの能力が同程度であっても、それぞれのテストによって点数が

異なる。この問題を防ぐには、それらの領域の項目を全て解答する必要が、クリティカルシンキングの項目は、受験者に長い思考時間を要求し、多くの項目を解答させることは困難である。このため、受験者に出題する項目数にはトレードオフの関係が存在する。

本研究の目的は、新たな尺度開発を行うことにより、このトレードオフの問題の解決を試みるものである。このようなトレードオフの問題を解決するには、実現可能な少ない項目数において、十分に測定可能なテストが必要となる。そこで本研究では、具体的に(1)まずクリティカルシンキングの主な尺度間の関係を分かりやすく整理し、(2)これに基づき、能力全体を網羅し、かつ、実現可能な少ない領域数で漏れがなく、信頼性のある尺度を開

発する。

筆者らはこれまでに、クリティカルシンキングの能力を測定するための尺度開発を行ってきた^②。本研究では、これらを基に分析の尺度を実装して、実際に大学研究者、大学生、高校生に受検してもらい、その結果の差異などから尺度の妥当性を示す。

2. 尺度の開発

2.1 クリティカルシンキングの能力測定尺度

クリティカルシンキングにおける尺度については、面接・口頭試験や筆記試験があり、以下のように分類される(表 1)。筆記試験は、面接・口頭試験に比べて評価者の評価特性に左右されないのでも筆記試験を対象とする。

さらに、クリティカルシンキングの能力で測定は、質問紙の自己申告形式では個人間比較には適さない^④、対象から外す。試験形式では、選択式項目は採点の利便性があるが、クリティカルシンキングなど複数の領域にまたがる能力の測定には適してなく、記述式の方が優れている^⑤、指摘されることがあった。しかし、大学入試センター試験においては、理数系でない科目でも選択式問題が出題され、その能力測定に大きな問題は生じてない。最近の研究では同じ時間では選択式の方が信頼性が高いと報告されている^{⑥⑦}。そこで本研究では、クリティカルシンキングの能力は、2. 試験形式、(2)選択式(Multiple Choice)で十分に測定できるという前提下で、その尺度開発を行う。

2.2 既存尺度の分類と整理

既存の尺度とその内容を Liu et al の総説^⑧を参考に分類・整理することを試みた(表 2)。まず、既存の尺度及びその下位尺度の共通点・相違点について比較した。そこで本研究において、いかなる下位尺度を設定すべきかを検討するべく、探索的に下位尺度数とその内容を変化させるところ、下位尺度数が 3 尺度①分析的思考力、②論理・推論能力、③読解・理解能力であり、かつ、それらの尺度が以下の内容とする場合が最も適切であると判断された。分析については、大学教員でかつ教育に関する研究を行っている博士号を有する研究者 4 人で議論し、同意できるまで検討した。

なお、専門的な知識ではなく、クリティカルシンキングの力が重視されているといわれる我が国の法科大学院入学の適性試験の択一式では、論理的判断力、分析的判断力、長文読解力が出題されており、本研究における尺度と概ね合致している。

表 2 においては、強い関連のあるものは「2」、関連のあるものは「1」と表記している。既存の尺度を整理した表 2 の結果から、下位尺度を全て含めた場合には尺度の漏れは見あたらなく、②論理・推論能力が尺度の中心となり重要であることが浮かび上がってきた。

さらに、開発する尺度は、既存の主な尺度、下位尺度の内容との関係は明らかなので、開発する尺度で測定すれば既存の尺度の能力値を推定することができることになり、既存の尺度を複数回受験するという受検者の負担を減らすことが可能になる。

2.3 開発する尺度の下位尺度

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るが、数学の知識の有無の影響のないよう配慮した。例えば、変数 x 、 y の 2 つを使わないと解けない問題は避けた。

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・論理、命題、演繹・帰納、逆・裏・対偶、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。

表1 クリティカルシンキングにける尺度の分類

尺度	I. 面接・口頭試験		
	II. 筆記	1. 質問紙・自己申告形式(リッカート、7件法など)	
		2. 試験形式	(1) 記述式 (Essay)
			(2) 選択式 (Multiple Choice)

参考: 若山他 2014^③

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るべく、論理展開・構造が複雑すぎないものとし、概ね500～1000文字程度の文章とした。

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⁹⁾、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。クリティカルシンキングの定義に沿って能力を測るべく、項目の策定は、独自で作問したものに加えて、公務員試験など既存の項目の中から良問と考えられるもので構成した。なお、各尺度の設問例(イメージ)は以下のとおり。

各尺度の設問例

① 分析的思考力の設問例

コインを3枚同時に投げた。今、その3枚は表・表・表ではなく、裏・裏・裏でもないことが分かった。このとき3枚のうち2枚が裏である確率は、次のどれか。なお、コインの表裏の出る確率はそれぞれ50%である。

- (1) $\frac{1}{8}$ (2) $\frac{1}{4}$ (3) $\frac{1}{3}$ (4) $\frac{3}{8}$ (5) $\frac{1}{2}$ (6) $\frac{5}{8}$ (7) $\frac{2}{3}$ (8) $\frac{3}{4}$ (9) $\frac{7}{8}$ (10) これらのいずれでもない。

② 論理・推論能力の設問例

「人は死ぬ」「ソクラテスは人である」この2つの命題を用いたことから100%確実に導かれる結論は「ソクラテスは死ぬ」となります。2つの命題を必ず用いることから結論が100%確実に導かれるとき、「？」の内容で最適なものを選びなさい。

「新幹線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、新幹線に乗った
(2) 遅れるのは新幹線に乗らないからだ
(3) 遅れないためには新幹線に乗る
(4) 新幹線は、時速200kmを超えている
(5) 何も入りえない

③ 読解・理解能力の設問例

次の文章の要旨として最も適切なのはどれか。

通貨を統合すると何が起きるか。通貨を統合した国とは戦争が起こせないという事である。単純な話である。もし経済的に同一の国の中で戦争になれば、経済的には「国内紛争」となり、その「国」の国力を激減させる。経済的に見て敗者も勝者もなく、搾取側も被搾取側も存在しない。戦後処理が国内で負担されるのであるから、その「国」にとって基本的なメリットはない。自分で自分の首を締めることになるからだ。確かに、通貨統合の象徴である中央銀行の設立は、市民の意識を大きく変えることになる。明治維新後の日本銀行設立、アメリカ独立後のFRB設立、ユーロの中央銀行設立以降、国内の紛争は激減している。

そもそも面倒な通貨統合を成し遂げてまで目指そうとする基軸通貨とは、何だろうか。ユーロは通貨統合を成し遂げて世界の基軸通貨になることを目指している。現在、世界の基軸通貨はドルである。では、アメリカのドルが基軸通貨になると、アメリカにとって何が好都合なのであろうか。通貨であるドルが、アメリカ国外でも流通することでドル通貨圏が拡大する。端的に言えば、これによりアメリカの中央銀行(FRB)が世界経済の中央銀行になることに近づくことができる。つまりアメリカが世界経済を左右する事が可能となる。では、世界の経済が一つになり、完璧に融合されたときの状態はどうなるか。世界全体が「拡大アメリカ合衆国」となり、戦争が消え平和が訪れる、と考える有識者もいる。ある意味で真実だろう。しかし、世界の人が本当にそれを望んでいるだろうか。

大切な事は、欧米にできた事は、アジアでも十分にできると言うことである。それどころか現状のままでは、ドルとユーロの2大通貨がアジアを分断統治することになる。アジアが通貨覇権競争において草刈場となりつつある。西にあるトルコはユーロの勢力範囲に、東にあるニッポン、フィリピン、太平洋の島々はドルの勢力範囲になりつつあると見るのは、早合点であらうか。10年後にはインドを含めたアジアは人口、経済ともに世界1位になるだろう。今こそアジアの力の結集を検討するべきではないだろうか。アジアの英知が、いやその英知の有無までもが、今まさに試されようとしているのだから。

- (1) アジアの諸国は、ユーロで見られたような通貨統合を、検討するべきである。
(2) アメリカのドルを基軸通貨とする世界経済は安定しており、アメリカ中心の世界が戦争のない世界平和をもたらす。
(3) 通貨を統合すると、統合した国同士では戦争は起こせない。なぜなら、経済的には「国内紛争」となり、相手から搾取する意味がなくなるからである。
(4) 基軸通貨になると、その国の世界的な影響力が拡大する。なぜなら、その国の中央銀行が世界の中央銀行のような力を持つからである。
(5) ドルとユーロの2大通貨の勢力がアジア地域に及んでいる。西のトルコ、東のニッポンなどをみれば、明らかである。

表2 クリティカルシンキングの主な尺度とその下位尺度

	尺度1	尺度2	尺度3
California Critical Thinking Disposition Inventory (CCTDI)			
(a) 真実の探求	1	1	
(b) 視野の広さ		1	1
(c) 分析力	2	1	1
(d) 系統立てる力	1	1	1
(e) 推論の信頼性	1	2	1
(f) 好奇心、質問力	1	1	
(g) 判断の成熟度	1	1	1
California Critical Thinking Skills Test (CCTST)			
(a) 分析力	2	1	1
(b) 評価力	1	2	1
(c) 推論力	1	2	1
(d) 演繹法		1	
(e) 帰納法		1	
(f) 論理体系化力、理由づけ能力	1	2	1
Collegiate Assessment of Academic Proficiency (CAAP) Critical Thinking,			
(a) 議論の要点分析する力	1	1	1
(b) 議論を評価する力	1	2	1
(c) 発展させる力		1	1
Cornell Critical Thinking Test (CCTT) Level X: 71 items Grades 5-12+			
(a) 帰納法		1	
(b) 演繹法		1	
(c) 信頼性		1	1
(d) 前提条件を識別		1	
Cornell Critical Thinking Test (CCTT) Level Z is intended for students in Grades 11-12+			
(a) 帰納法		1	
(b) 演繹法		1	
(c) 信頼性		1	1
(d) 前提条件を識別		1	1
(e) 意味付け力		1	1
(f) 定義		1	1
(g) 計画予測力	1	1	1
ETS Proficiency Profile (EPP)			
(a) ノンフィクションの散文でレトリック(修辞)と議論を区別する力		1	1
(b) 提示された情報を説明するため最善の前提と仮説を認識する力		1	1
(c) 変数間の関係を推定し、解釈する力	1	1	1
(d) 提示された情報に基づいて有効な結論を引き出す力	1	1	
Halpern Critical Thinking Assessment (HCTA)			
(a) 言葉による推論スキル	1	2	1
(b) 議論と分析スキル	1	2	1
(c) 仮説検証に資する思考スキル	1	1	
(d) 確実性と不確実性の利用	1	1	
(e) 意思決定と問題解決のスキル	1	1	
Watson-Glaser Critical Thinking Appraisal tool (WGCTA)			
(a) 推論	1	2	1
(b) 仮定を認識	1	2	1
(c) 演繹法		1	
(d) 解釈	1	1	1
(e) 議論を評価	1	2	1
Watson-Glaser II: Short form: 30min			
(a) 仮定を認識する力	1	2	
(b) 議論を評価する力		2	1
(c) 結論を導く力	1	2	1

3. 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^{(10)~(13)}。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、同一尺度上に評価可能
- (2) データの中に欠損値を推定可能

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。

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

ただし、

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

能力パラメータ θ_j において識別力パラメータ α_i (discrimination parameter) と難易度パラメータ β_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 α_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 θ_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 α_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 α_i の値が大きいと、図 1 の IRF の曲線は立ってくる。

同様に式(1)の β_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 θ_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 β_i が左から ($\beta_i = -1$, $\beta_i = 0$, $\beta_i = 1$) の項目反応関数を示す。難易度パラメータ β_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⁽¹³⁾。

$$I(\theta_j | \alpha_i, \beta_i)$$

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

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 θ_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。

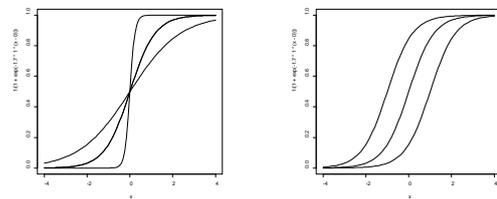


図 1a: 能力値による正解率の変化 a 図 1b: 能力値による正解率の変化 b

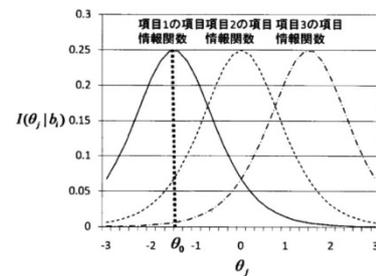


図 2 項目情報関数例

4. 項目反応理論に基づいた開発尺度の分析

4.1 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、2014年7月~2015年12月に、東京及び近郊の大学生736人を対象として、試験を実施した。18歳から概ね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、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。

クリティカルシンキングの項目では考える能力を測定するため、前提にグラフや表を用い、本文において状況を説明したりするので、項目の設問自体が長い、いわゆる「大問」の場合がある。これらを理解し的確に解答するには、受検者は考える時間が十分に必要とされるので、全体の項目数には制限されることが多い。しかし、本研究では尺度の信頼性を高めるべく、設問を短い「小問」として、項目数を確保することとした。

また、等化の推定のための共通項目を⁽¹⁴⁾、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組ずつ行い合計632人のデータを得た。別途週末に55人ずつ(3人が重複)2日で107人が集まり①分析と③読解の両方に設問、加えて②推論の共通項目に解答してもらった(表3)。これらで得られたデータを、IRTPRO3、SPSS20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。

4.2 分析結果と考察

個人の能力における各尺度間の相関(表4)と尺度間の散布図(図3a～3c)は以下のとおり。各尺度毎の困難度・識別力の散布図を以下に示す(図4a～4c)。さらに、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(表5、図5a～5c)。また、各尺度の固有値プロットを図6a～6cに示し、各尺度のテストごとの信頼性係数(アルファ係数)を求め⁽¹⁵⁾表6に示す。

各尺度(①分析的思考力、②論理・推論能力、③読解・理解能力)の相関は、.423～.570となり、想定よりも小さくなった。散布図を見ても各図で2元性以上は確認されず、偏りもそれほど大きくはなかった。困難度・識別力の散布図は、想定通りであり困難度0付近を中心に左右対称で正規分布のような形状が確認された。被検者数が必ずしも十分に大きくはないこともありテスト情報量(I)と標準誤差(S.E.)は限界があるものの概ね想定される範囲におさまった。

尺度1及び尺度3は、主に週末に学生が集合して1日で試験を受けてもらったが、尺度2は主にいわゆる個別の授業でデータを収集するため対象人数も多くなった。尺度2ではグラフには若干の歪みが見られ、誤差は小さく(0.24)となったの

は、このデータの収集方法の差異によるものと考えられる。

また、信頼性係数で、尺度2が最も小さくなったのは、尺度1及び尺度3は、週末に学生が集合してもらったので金銭的なインセンティブがあったが、尺度2は個別の授業でデータを収集するためそれがなく、その差異による可能性が考えられる。尺度3の読解・理解の作問の際には、あえてなるべく広範囲にわたる分野から、さまざまな種類の文章の項目を選んだが、学生は一般に得意分野を有し、比較的領域固有性があると想定されるので、結果的に尺度3の信頼性係数に跛行が見られたと考えられる。なお、固有値プロットでは、どの尺度も第1と第2因子間で、大きく半減しており1因子の可能性が示唆された。これらのことから、尺度の信頼性が確認できた。

さらに、「平均的な大学生の能力値 Θ は、大学の研究者より低く、平均的な高校生より高い」という仮説を立てた。2016年9月から同年12月に尺度1を例にして実装実験を行った。平均的なレベルの高校生25人及び平均的なレベルの大学の学生64人に加えて、分析するには必ずしも十分な人数とはいいきれないが大学研究者10人に受検してもらった。

得られた能力値 Θ について、分散分析したところ、群間の能力値の差が有意で($F(2,94)=35.21, p<.001$)あったことから、多重比較(TukeyのHSD)を行ったところ、高校生と大学生及び高校生と大学研究者とは共に $p<.001$ の有意差が、さらに大学生と大学研究者では $p<.05$ で有意差があった(図7)。このことによって尺度1の妥当性が示唆された。

さらに、(1)項目策定時にクリティカルシンキングの定義に沿って、選定していること、(2)相関、散布図の分析において概ね独立的であること、(3)尺度1のみではあるが、大学研究者、大学生、高校生の能力値に差があることから、総合的に鑑みるに妥当性に問題はないと考えられた。

表3 尺度毎の試験問題と受検者の概要

○: おおむね50人以上

被検者	試験 問題数	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	2F	2共通	3A	3B	3C	3D	3E
授業 632人	10	10	10	10	10	5	10	10	10	10	10	10	5	10	10	10	10	10	5
週末 55人			○	○	○	○							○						○
週末 55人		○	○	○	○	○							○	○	○	○	○	○	○

注: 授業と2つの週末において3人のダブリがある

表4 個人能力の尺度間相関

	Θ 1	Θ 2	Θ 3
Θ 1			
Θ 2		.568**	.570**
Θ 3			.423**

** : $p < .01$

表5 アイテムバンクの情報量と標準誤差

	情報量(max)	Θ (I=max)	標準誤差
尺度1	11.65	-0.20	0.29
尺度2	17.44	0.20	0.24
尺度3	10.15	-0.70	0.31

表6 各テストの信頼性係数

	尺度1	尺度2	尺度3
A	0.506	0.485	0.365
B	0.735	0.508	0.618
C	0.750	0.479	0.630
D	0.717	0.353	0.598
E	0.580	0.513	0.388
F	---	0.684	---
平均	0.658	0.504	0.520

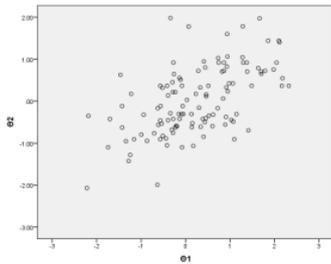


図 3a ①分析的思考力と②論理・推論能力

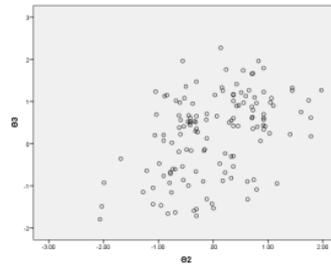


図 3b ②論理・推論能力と③読解・理解能力

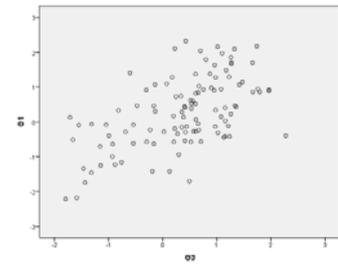


図 3c ③読解・理解能力と①分析的思考力

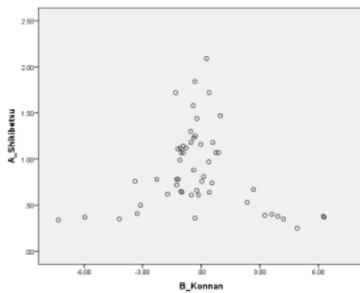


図 4a ①分析的思考力の困難度と識別力

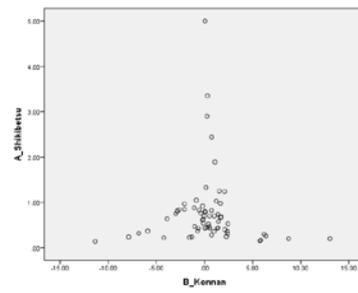


図 4b ②論理・推論能力の困難度と識別力

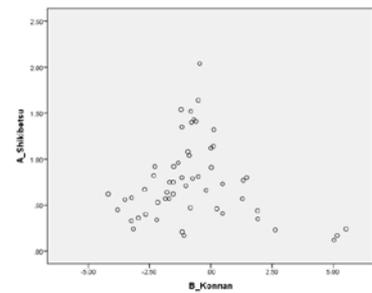


図 4c ③読解・理解能力の困難度と識別力

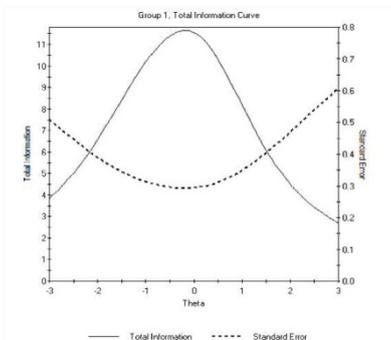


図 5a ①分析的思考力の情報量と標準誤差

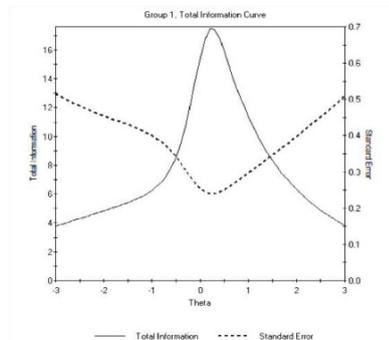


図 5b ②論理・推論能力の情報量と標準誤差

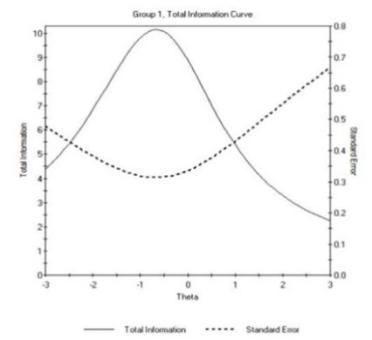


図 5c ③読解・理解能力の情報量と標準誤差

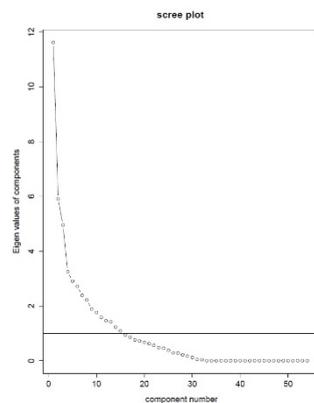


図 6a ①分析的思考力の固有値プロット

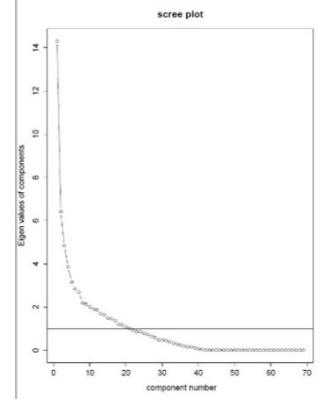


図 6b ②論理・推論能力の固有値プロット

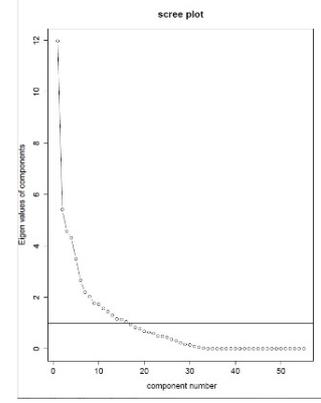


図 6c ③読解・理解能力の固有値プロット

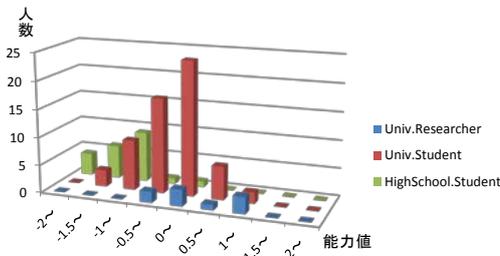


図7 ①分析的思考力の能力値と度数

5. おわりに

開発する尺度を、下位の3尺度、①分析的思考力、②論理・推論能力、③読解・理解能力から構成したところ、下位尺度の相関はそれほど大きくはなく、固有値プロット、困難度・識別力、情報量から、各尺度は1因子であり統計的に独立で、信頼性が確認された。さらに、尺度1を例に平均的な高校生、大学生及び大学研究者の間で能力値に差があることなどから、この尺度の妥当性が確認された。

今後は、クリティカルシンキングの能力評価において3尺度を検討したが、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。

謝辞

本研究の一部は、科学研究費助成事業 基盤研究(A)番号 15H01772, 基盤研究(C)番号 15K01088, 及び、CRET(教育テスト研究センター)の助成を受けている。

参考文献

- 若山昇: “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, Vol.31, No.1, pp.145-153 (2009)
- 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回大会発表論文抄録集, pp.58-61, (2016)
- 若山昇, 梶谷真司, 渡辺博芳, 赤堀侃司: “クリティカルシンキング教育の現状と課題 —大学における授業実践者の視点から—”, 帝京大学ラーニングテクノロジー開発室年報, Vol11, pp85-94, (2014)
- 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, 23(2), pp43-50, (2016)
- Ennis, R. H.: “Critical Thinking Assessment: Theory into Practice”, 32(3), pp179-186 (1993)
- Lee, Hee-Sun; Liu, Ou Lydia; Linn, Marcia C.: “Validating measurement of knowledge integration in science using multiple-choice and explanation items”, Applied Measurement in Education, 24(2), pp115-136, (2011)
- Zahner, D. : “Reliability and validity-CLA+”, Council for Aid to Education, http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf (参照日 2017.02.06)
- Liu, Ou Lydia; Frankel, Lois; Roohr, Katrina Crotts: “Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment”, Research Report. ETS RR-14-10, Educational Testing Service (2014)
- 若山昇: “誰でもわかるクリティカルシンキング”, 北樹出版 (2013)
- 植野真臣, 永岡慶三: “e テスティング”, 培風館(2009)
- 豊田秀樹: “項目反応理論 [入門編] 第2版”, 朝倉書店 (2012)
- 宮澤芳光, 植野真臣: “適応型テストを用いた携帯型観光・学習ナビゲーションシステム”, 教育システム情報学会誌, 29(2), 110-123, (2012)
- 宮澤芳光, 植野真臣: “テスト情報量と移動距離を最適化するモバイル・テスト・システム”, 電子情報通信学会論文誌 D, Vol.J98-D, No.1, pp.30-41, (2015)
- 野口裕之, 大隅敦子: “テストの基礎理論”, 研究社 (2014)
- Robert L. Linn (原著, 編), 池田央ら(訳): “教育測定学”, 学習評価研究所 (1992)