

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

Simulation for critical thinking adaptive test development

若山 昇^{*1}
Noboru WAKAYAMA^{*1}
^{*1} 帝京大学/CRET

宮澤 芳光^{*2}
Yoshimitsu MIYAZAWA^{*2}
^{*2} 東京学芸大学

梶谷 真司^{*3}
Shinji KAJITANI^{*3}
^{*3} 東京大学

植野 真臣^{*4}
Maomi UENO^{*4}
^{*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

概要：クリティカルシンキングは現代社会を生きるうえで必須の能力になっている。本研究の目的は、クリティカルシンキングの適応型テストの開発のために、シミュレーションを試みるものである。クリティカルシンキングの3尺度（分析、推論、読解）について設問をランダムに選択した場合と、適応型テストで行った場合を比較したところ、適応型テストの方がテスト情報量が大きくなった。この結果、適応型テストの方が短時間で正確に測定可能であること、および本試験の設問項目の有用性が確認された。

キーワード：クリティカルシンキング、適応型テスト、シミュレーション、項目反応理論、試験

1. はじめに

21世紀型スキル⁽¹⁾においてもクリティカルシンキング(以下「CT」という)の重要性が謳われているように、CTは現代社会を生きるうえで必須の能力になっている。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⁽²⁾。

現在CTを評価する試験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するので、多くの項目を出題することは難しい。したがってCTの能力全体を網羅して、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。さらに、CTの重要性の高まりとともに、学生・社会人のCT能力を測定する必要性が高まっている。

本研究の目的は、CTの適応型テスト(以下「CAT」という)開発で作成した試験のシミュレーションを行ない、CATの優位性及び本試験の設問項目の有用性を検討するものである。

2. 項目反応理論、適応型テスト

短時間で効率的に能力を測定するためには、CTにおいても項目反応理論⁽³⁾(以下「IRT」という)によるCATの開発が必要となる。

既存のテストでは、同じテストを受検した者同士でしか能力を比較することができないが、IRTでは、異なる項目で構成されたテストでも同一の尺度上で受検者を評価できることになる。さらにCATでは、受検者の解答履歴を瞬時に分析し最適な設問を提示できることになり、測定精度を落とさず、時間を短縮できる。

3. 研究方法

CTの能力評価には、行動観察や口頭試問などもあるが、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。先行尺度の設問を分析したところ、①分析的思考力、②論理・推論能力、③読解・理解能力の3つの尺度が浮かび上がってきた。

設問の作成の際は、CTの定義にできるだけ沿って、独自に作成したもの及び公務員試験など既存の問題の中からCTに必要な①分析的思考力、②論理・推論能力、③読解・理解能力を測る3尺度で構成した。大学教員かつ博士号を持ち教育に関する研究を行う4人の研究者が議論を重ね、問題を策定した。

各尺度毎に、5問を共通問題とし15問の試験を5～6組作成し、各組毎に50人程度から解答を得た。データをIRTPRO及びSPSSを用い、IRT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した。さらに、シミュレーションでは、能力値の平均0、分散1として1,000人分受検者を発生させ、各尺度を受検したときの、項目数、情報量、能力値の関係を解析した。

4. 結果と考察

試験において出題する項目数が増加するにつれて、テスト情報量が増加する。この変化をCATによる出題項目の選択と、乱数による出題項目の選択を比較した結果を示す(図1)。乱数による情報量変化は比例的な増加となった。一方、CATの情報量は凸型グラフとなり、開始時から1問あたりの増加量が大きくなった。情報量は全項目終了時には、CATでも乱数でも一致することになるが、試験可能な項目数(10～15項目程度)では、CATの方が、かなり情報量が多くなった。CATは、受験者の正誤解答履歴に基づき最も情報量の大きい設問を選んでいくことによる情報量の増加効果が確認された。さらに、想定通りCATの凸型グラフ、乱数の比例型グラフが確認できたことは、テスト項目の内容が有用であることによると考えられた。

加えて、3 尺度について真の能力値別によるテスト情報量を、CAT による項目選択と乱数による項目選択で比較した。項目数を 10 項目から+10 項目ごとに真の能力値とテスト情報量の散布図を分析したところ、3 尺度とも類似の傾向を示した。尺度 1 の分析的能力の例を図 2 に示す。

この結果、どの尺度でも、どの能力値でも、出題項目数によらずに、CAT の項目選択による情報量は乱数による項目選択の情報量以上であった。また、項目数が全体の半分程度(20-30 項目)までは、乱数の情報量は CAT の半分以下の場合がかなり存在している。なお、各尺度の項目数は①54、②71、③54 である。

項目数が十分に多くなると、乱数による選択の情報量は、CAT によるものに接近するものの、乱数では情報量のバラつきがかなり大きい。情報量のバラつきは問題数の増加により、相対的には減少するものの、絶対的には 2-3 程度であった。この結果、CAT を用いないで、画一的な古典的試験を作成し能力を測定すると、情報量のバラつきが大きく、正確に測定することが難しくなることが示唆された。

5. おわりに

これらの結果から、CAT の方が短時間で正確に測定可能であること、さらに設問項目の有用性が確認された。今後は、実際に試験を実施し実データを分析する必要がある。

なお、本研究は科研費(A)15H01772、(C)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

- (1) Griffin, P. et al. : “Assessment and Teaching of 21st Century Skills” Springer, New York, pp.17-66 (2011)
- (2) 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 -クリティカルシンキングに対する志向性と認知欲求の変化から-”, 大学教育学会, Vol.31, No.1, pp.145-153 (2009)
- (3) Robert L. Linn, 池田央ら(訳) “教育測定学” 上下巻, 学習評価研究所 (1992)

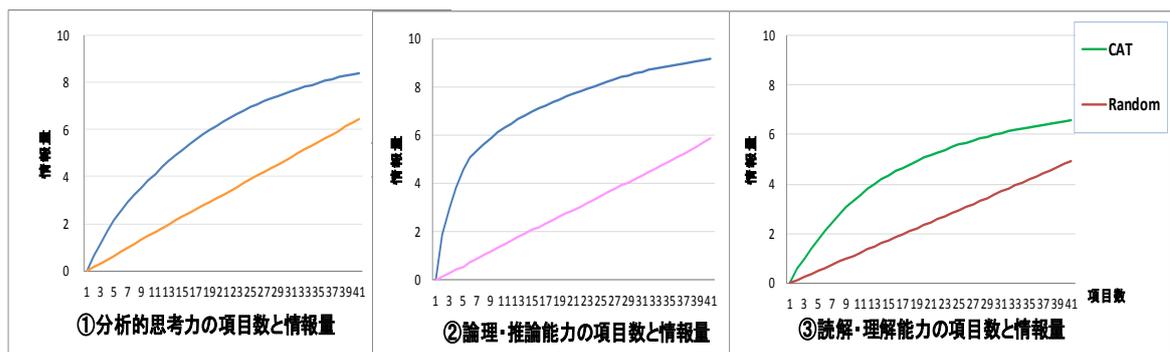


図 1: CAT と乱数により項目を選択との比較, 尺度毎の項目数と情報量の平均

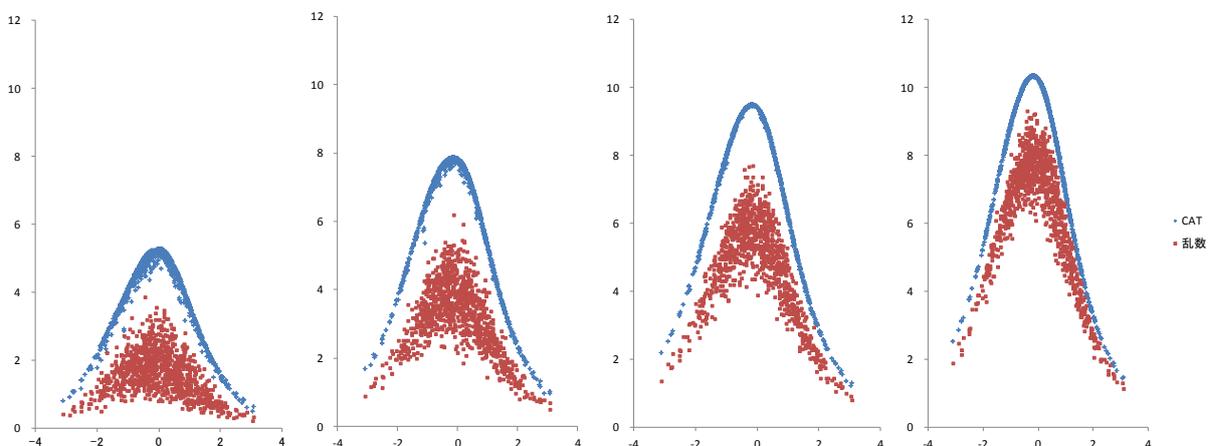


図 2: CAT による項目を選択と乱数による項目選択の比較による, 真の能力値とテスト情報量

(尺度 1: 分析的思考力, 左から 10 項目, 20 項目, 30 項目, 40 項目)