

早稲田大学審査学位論文
博士（人間科学）
概要書

潜在ランク理論を用いたコンピュータ適応型テスト
のためのアルゴリズムの提案と実装

Proposition and Implementation of Algorithm for
Computer Adaptive Test Based on Latent Rank Theory

2013年1月

早稲田大学大学院 人間科学研究科
木村 哲夫
KIMURA, Tetsuo

研究指導教員： 永岡 慶三 教授

近年、教育へのコンピュータあるいは IT 利用は急速に進展している。コンピュータあるいは IT を利用したテストによる学習評価には、学習者と教育者の両者に様々な面で利点がある。これまで大規模な開発と実施でないと、コンピュータが困難度を調整して出題するコンピュータアダプティブテスト (computer adaptive test: CAT) を導入することは不可能であると考えられてきた。しかし、Moodle に代表されるオープンソースのラーニング・マネジメント・システム (learning management system: LMS) が登場したことと、パーソナル・コンピュータの飛躍的な処理能力向上により、小規模であってもコンピュータを利用したテストの実施が可能な時代となった。また、CAT 開発に欠かせないソフトウェアも、オープンソースとして利用できるものが増えてきている。

本研究の目的は、理論と実践の両面から CAT について検討を加え、新規のアルゴリズムを提案するとともに、オープンソースを利用してそれを実装するシステムを開発し、そのシステムを英語教育へ適用することによって検証することである。本研究では、1 つの教育機関の 1 学年分の学習者を対象とする小規模な CAT 開発を念頭における。小規模な CAT とは、受験者数が 200 名前後である場合を想定したものである。

より具体的には、次の 3 点について、理論と実践の両面から論じることが本研究の目的である。

1. テスト理論の変遷と CAT の根源をふりかえり、小規模 CAT 開発に適したテスト理論と CAT アルゴリズムを検討・提案する。
2. CAT 開発に利用可能なオープンソースにはどのようなものがあるのか整理し、実際にそれを利用して小規模 CAT を実装するシステムを開発する。
3. 開発したシステムを英語教育へ適用することにより検証するとともに、CAT の受験結果に診断的情報を付加する方法を検討する。

これまで大規模な開発と実施でないと CAT を導入することは不可能であると考えられてきたが、本研究によって、オープンソースを利用して小規模であっても CAT を実装するシステムを開発することができることを示すことで、個人あるいは教員・学校間でも CAT を開発・実施される機会が増えることが期待される。また、テスト結果に診断的情報を加えることは、テスト実施者の説明責任を果たす上でも重要な課題である。

理論的な側面では、理論的な側面として、潜在ランクに基づくコンピュータ適応型テスト (LRT-CAT) に関して次の 2 つの提案を行った。

- (1) CAT を実装するためのアイテムバンクを構築する際に、望ましくない項目を LRT の IRP 指標を使って除去する方針
- (2) LRT-CAT アルゴリズム: 項目選択ルールと終了条件

項目除去に関する方針はヒューリスティックなもので、より多くの分析を重ねて修正する必要があるかもしれない。LRT-CAT アルゴリズムは初めて提案されたもので、シミュレーション研究と実テストの両面から、さらに考察を加えることで、システムはより機能の向上を続けて進展している。

大学入学生の英語基礎力を測定する小規模な CAT を開発するために RM と LRT に基づいて行った一連の実践的研究から、次の 5 つのことを示すことができた。

- (3) CAT を実装するためのアイテムバンク構築は、小規模であっても、RM または LRT に基づき行えること
- (4) RM に基づく項目分析も、LRT に基づく項目分析も、オープンソースを利用して十分行えること
- (5) CAT を実装するシステムをオープンソース LMS である Moodle を使って開発できること
- (6) シミュレーションにより、用意できたアイテムバンクを使って CAT を実装した場合、どの程度の結果が得られるか、アイテムバンクのどこに弱点があるかを把握できること
- (7) 用意できたアイテムバンクを使って実際に CAT を実施し、どのレベルの項目の使用頻度が高くなるか(どのレベルの項目を今後追加すべきか)を把握できること

また、CAT 受験者に対するアンケート調査からは、次の(8)～(10)の3つことを明らかにすることができた。

- (8) 先行研究で指摘されていた通り、通常の CAT の項目選択ルールでは、ほとんどの受験者がテストを難しいと感じ、多くの者が受験後に落ち込んでしまうこと
- (9) 目標正答確率を通常の CAT の項目選択ルールの 50%から 70%に変更すると、難しいと感じる受験者は少なくなり、受験後に落ち込む者も少なくなること
- (10) 目標正答確率を通常の CAT の項目選択ルールの 50%から 70%に変更しても、実施項目数を 16 から 20 に増やすことで、同じかそれ以上の精度で CAT が終了すること

診断的情報を CAT の結果に付加する方策については、LRT における受験者の潜在能力のとらえ方であるランクメンバーシッププロファイル(rank membership profile, RMP)を利用することと、能力記述文(can-do statement, CDS)を使った自己評価の試みから、次の 2 つのことを示した。

- (11) LRT に基づき分析した場合、潜在ランクだけでなく RMP を示すことで、変化の状況をより細かく示すことができること
- (12) CDS を使って自己評価することは、自分の能力を過大評価(あるいは過小評価)していることを気づかせることができること

本研究によって、段階評価に適した新しいテスト理論である LRT-CAT アルゴリズムが提案され、オープンソースを使った CAT 開発の道筋を示せたことで、一人でも多くの方が CAT というツールを使って教育測定をできるようになれば幸いである。規模の大きさにかかわらず、CAT を自分の手で開発し実行できる時代になったのである。これから CAT を開発しようとする方は、ぜひ受験者の心理的側面にも配慮した CAT を目指していただきたい。“Happy CAT for everyone”これがこれまで筆者の研究モットーであり、これからも変わることはない。