

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 24 日現在

機関番号：33109

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25370736

研究課題名(和文) コンテンツ・バランスと受験者の教育心理学的側面を考慮したCAT開発

研究課題名(英文) CAT development that takes contents balance and psychological aspects of test takers into the consideration

研究代表者

木村 哲夫 (Kimura, Tetsuo)

新潟青陵大学・福祉心理学部・教授

研究者番号：90249095

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)：測定論的立場に立てば、コンピュータ・アダプティブ・テスト(CAT)の目標正答確率は50%に近づけるべきであるが、受験者の教育心理学的側面を考えるとより高い目標正答確率が望ましい。本研究は、オープンソースの学習管理システム(LMS)上でCATの目標正答確率の調整を行なうモジュール(M-UCAT)を開発し、そのソースを公開した。また、LMSに依存せずにCATをオープンソースで実施する環境(Concerto)とLMS間で、データ連携させる方法を発表し公開するとともに、Concerto上で潜在ランク理論に基づく分析を可能にするR環境のプログラムを発表公開した。

研究成果の概要(英文)：Although the target probability of computer adaptive test (CAT) should be close to 50% from the view point of measurement theory, it should be higher when psychological aspects of test takers are taken into the consideration. This study developed M-UCAT, a module which manipulate target probability of CAT on open-source learning management system (LMS) and opened its sources to public. Additionally, it released data coordination between LMS and Concerto, an open-source platform which enables CAT implementation outside of LMS, and a program on R environment which enables data analysis based on latent rank theory on Concerto.

研究分野：英語教育

キーワード：CAT LMS コンテンツバランス 目標正答確率 教育心理学的側面

1. 研究開始当初の背景

(1) 外国語教育において、1990年代以降の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(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学習管理システム(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となった(Linacre, 2000; Hinkelman & Grose, 2004)。また、CAT開発に欠かせないソフトウェア(項目分析、シミュレーション分析、CATアルゴリズム設計)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(Han, 2012; Magis & Raiche, 2012)。

(2) このような学術的背景の下、「教員・学校間の協働作業による能力記述文との対応付けのあるCAT開発」(基盤C: 22520590)において、小規模であってもCATを開発し実施できることを示してきた(木村・永岡, 2011; Kimura, Ohnishi & Nagaoka, 2012)。実際にLMS上で小規模なCATを実行するプログラムを公開し、国内外の教育機関から利用の申し出が寄せられている。しかしながら、同時に、研究開発を通して、コンテンツバランスの管理の必要性、受験者の教育心理学的側面への考慮の重要性、LMSのバージョンアップへの対応の困難さ、も明らかになった。

(3) コンテンツバランスの管理については、1項目ごとに対処する手法(Kingsbury & Zara(1989)のconstrained CATやVan der Linden(2010)のshadow testsなど)と、複数の項目をまとめて対処する手法(Keng(2008)のtestlet-based CATやvon Davier & Haberman (2012)のmultistage testingなど)が提案されている。受験者の教育心理学的側面については動機づけや学習自己効力感への影響が指摘されている(Tonidandel, 2002; Kimura & Nagaoka, 2012)。測定論的立場に立てばCATの目標正答確率は50%に近づけるべきであるが、受験者の教育心理学的側面を考えるとより高い目標正答確率が望ましい。また、出題項目全体を一律に変えるのがよいのか、部分的に変えるのがよいのか、順序性も合わせて考えるべきなのか、CAT受験者の教育心理学的側面の検討を要するポイントは多岐にわたる。しかしながら、いくつかのシミュレーション研究以外、あまり実証的なデータに基づいた検証は行われていない。LMSのバージョンアップへの対応の問題を回避する方法としては、LMSに依存しないCATを実行するオープンソースConcertoの利用したアプローチ

が有力である(Kimura, et al. 2012)。

2. 研究の目的

(1) 本研究の目的は、英語力を測定するCATのコンテンツバランスの管理や受験者の教育心理学的側面について柔軟にコントロールできるようにするとともに、CATを実施するモジュールをLMSに依存させず、それ自体をオープンソースで開発し、LMSとの間でデータ連携を図る仕組みを構築することによって、LMSのバージョンアップへの対応の問題も解決をはかることである。

3. 研究の方法

(1) 次の3つ研究方法により、先に述べた研究目的の達成を目指す。

コンテンツバランスの管理について、先行研究調査と関連学会やワークショップへの参加により、小規模なCATの場合に最適な方法を検討する。

教育心理学的に最適な項目選択ルールを探るために、CAT全体に対してだけでなく部分的にも目標正答確率の調整が可能な仕組みを開発し、同一の(または等質な)学習者に対して、同一のアイテムバンクを使って、複数の項目選択ルールによるCATを実施し、受験後にアンケート調査を行う。

LMSのバージョンアップへの対応の問題も解決するために、オープンソースでCAT実施する環境(Concerto)の日本語環境へのローカライズを行うとともに、そのアプローチに関するワークショップを国内で開催し、LMSとのデータ連携の仕組みを構築・公開する。

4. 研究成果

(1) コンテンツ・バランスの管理については、先行研究を調べるとともに、CATに関する国際会議やワークショップに参加して情報収集を行なった結果、小規模なテストの場合、1問ごとに項目選択を行なう従来の単問式のCATよりも、複数の設問を束ねていくつかの段階を踏むマルチステージテスト(MST)の方が適していることがわかった(Yan, von Davier & Lewis, 2014)。実際に、近年MSTの実用は増えてきている。

(2) たとえば、図1は一般的に“1-3-3モデル”と呼ばれるMSTのテスト構造を模式図で表したもので、3ステージ7モジュールからなる。第1ステージでは中程度の問題群Midを解答させ、その結果により第2ステージではLow、Mid、Highの3レベルに分岐させる。さらに、第2ステージの結果により第3ステージで解答させる問題群のレベルを決めている。モジュール単位でコンテンツ・バランスを調整することができるので、従来の単問式のCATよりも、コンテンツ・バランスの管理が容易になる。特に小規模CATの場合は、MST方式にした方が、アイテムバ

ンクの項目を無駄なく使える。

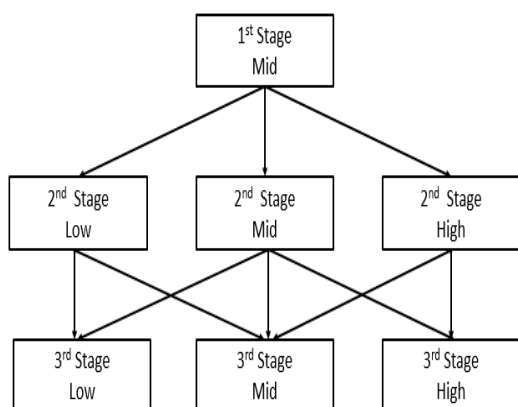


図1 MSTのステージとモジュール

(3) 教育心理学的側面に配慮するために、LMS上でCATの目標正答率の調整ができるようにしたモジュールを改良し、全体だけでなく部分的にも目標正答率を調整できるようにした。デフォルトの目標正答率の他に、テストの序盤S問とテストの終盤T問の目標正答率の調整もできるようになった。この改良されたモジュール UCAT Module for Moodle のソースを公開した。

https://github.com/VERSION2-Inc/moodle_ucat

(4) 科学技術英語の語彙・語句をターゲットとする同じアイテムバンクを使って、【表1】に示すように、目標正答率の異なる2種類のCATを日本人大学生に実施し、その直後に行ったアンケート調査から、日本人大学生のCATに対する印象に関して、次の3点について調べた。

受験者のCATに対する印象は先行研究と同様か

受験者のCATに対する印象は受験者のレベルにより異なるか

受験者は「問題数が多くても易しい問題が出るテスト」と「難しい問題が出て問題数が少ないテスト」のどちらを望むか。

【表1】2種類のCAT

	目標 正答率	出題数	理論上 のSE
CAT1	50%	16	0.5
CAT2	80%	25	0.5

(5) 受験者のCATに対する印象(困難度感や受験後の幸福感)はCAT2の目標正答率を70%にした先行研究と同様であった。

(6) CATの種類(CAT1とCAT2の2水準)と受験者の能力レベル(High群とLow群の2水準)を2つの変動要因として、CATによ

り推定された能力について、2要因の分散分析を行ったところ、受験者の能力レベルについてのみ有意差があり($F(1,320) = 3.24, p < .001$), CATの種類や2つの要因の交互作用について有意差は見られなかった。

(7) HighとLowいずれの受験者群においても、CAT1では実際の正答率よりも正答率感覚の方がやや高い。反対に、CAT2では実際の正答率よりも正答率感覚の方が低い。これは、受験者のレベルに関わらず、CAT1のように難しめのテストにおいては、自分の成績をやや過大評価する傾向が、CAT2のように易しめのテストにおいては、自分の成績を過小評価する傾向があると解釈できる。

(8) 全体として、「問題数が多くても易しい問題が出るテスト」よりも「難しい問題が出て問題数が少ないテスト」を望んでいる学生の割合の方が多く、この傾向は、受験者群Lowにおいて、より顕著であった。

(9) LMSに依存しないオープンソースでCAT実施する環境(Concerto)の日本語環境へのローカライズを行うとともに、Concerto上でCATを実施する手法について、Cambridge University Psychometric Center / Stanford UniversityのMichal Kosinski氏を日本に招き、東京と和歌山でハンズオン・ワークショップを開催した。

(10) LMSのバージョンアップによる影響を排除するために、これまでのCAT開発フレーム(図2)に変わって、LMSに依存しない開発フレーム(図3)を目指し、Concertoで実施したCATの結果を、LMSとデータ連携する方法について発表し公開した。

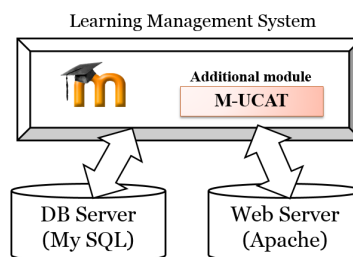


図2 これまでの開発フレームワーク

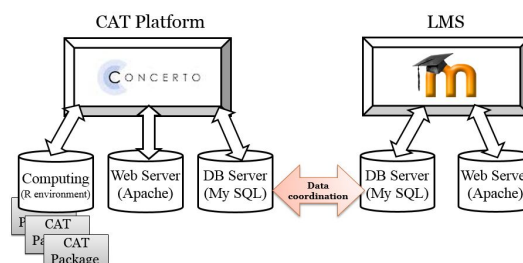


図3 新しい開発フレームワーク

(11) Concerto上で潜在ランク理論に基づく分析を可能にするR環境のプログラムを公表した。 <https://app.box.com/v/R-LRT>

<引用文献>

- Han, K. T. (2012). SimulCAT: Windows Software for Simulating Computerized Adaptive Test Administration. *Applied Psychological Measurement*, **36**(1), 64-66.
- Hinkelman, D., & Grose, T. (2004). Placement testing and audio quiz-making with open source software. *Proceedings of CLaSIC 2004*, 972-981.
- Keng, L. (2008). A comparison of the performance of testlet-based computer adaptive tests and multistage tests. *ProQuest*.
- Kingsbury, G.G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, **2**, 359-375.
- 木村哲夫・永岡慶三 (2011a). Moodle による小規模 CAT 構築に向けて 2: アイテムバンクの統合. *JSET27 講演論文集*, 731-732.
- Kimura, T. & Nagaoka, K. (2012). Psychological aspects of CAT: seeking item selection rules which do not decrease test takers' learning self-efficacy and motivation. *International Association for Computer Adaptive Testing Conference 2012 Australia Conference Program*, 12.
- Kimura, T., Ohnishi, A., & Nagaoka, K. (2012). Moodle UCAT: a computer-adaptive test module for Moodle based on the Rasch model. *The 5th International Conference on Probabilistic Models for Measurement Program*, 83.
- Linacre, J.M. (2000). Computer-adaptive testing: A methodology whose time has come. *MESA Memorandum No 9*.
- Magis, D. & Raiche, G. (2012). Random Generation of Response Patterns under Computerized Adaptive Testing with the R Package catR. *Journal of Statistical Software*, **48**(8), 1-31.
- Tonidandel, S. et al (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, **87**, 320-332.
- Van der Linden, W. J. (2010). Constrained Adaptive Testing with Shadow Tests. In van der Linden, W.J. & Glas, C.A.W. (Eds.) (2010). *Elements of adaptive testing* (pp.231-245). New York: Springer.
- von Davier, A. A. & Haberman, S. (2012). Comparability of Test

Performance and Reported Scores in Multistage Testing. Keynote Speech at IACAT 2012 Conference, Sydney, Australia.

Yan D., von Davier A. A. and Lewis C. (2014). *Computerized multistage testing: Theory and applications*. New York, NY: CRC Press.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計7件)

木村哲夫(2016). コンピュータ適応型テストの心理学的側面: 目標正答率を調整するシステムへの日本人大学生の反応. *統計数理研究所共同研究レポート* 356, 47-56.

木村哲夫(2015). Concerto と Moodle 間のデータ連携. *MoodleMoot Japan 2015 Proceedings*, 38-39.

木村哲夫・登藤直弥・荘島宏二郎(2015). R における潜在ランク理論の推定プログラム. *日本テスト学会第13回大会発表論文抄録集*, 98-99.

木村哲夫(2014). オープンソースウェアによるコンピュータ適応型テスト開発. *統計数理研究所共同研究レポート* 321, 1-14.

木村哲夫・大西昭夫・永岡慶三(2013). Rasch モデルによる小規模 CAT の Moodle への実装. *JSET29 講演論文集*, 761-762

[学会発表](計9件)

Kimura, T. & Koyama, Y. (2015b). Improvement and evaluation of a small scale ESP CAT. *IACAT Conference 2015 Cambridge, UK*.

Kimura, T. & Koyama, Y. (2015a). Implementation of small-scale in-house CAT with corpus-driven lexical item bank for ESP. *FLEAT VI Boston, USA*, <http://sched.co/3Mzo>.

Kimura, T. (2015). Who can be happy with CAT? *PROMS 2015 Fukuoka, Japan*.

Kimura, T. & Kosinski, M.(2014). Startup e-testing with open source: CBT/CAT/MST on Concerto.

[その他]

UCAT Module for Moodle のソース:
https://github.com/VERSION2-Inc/moodle_ucat

R 環境で潜在ランク理論に基づく分析を可能にするプログラム:

<https://app.box.com/v/R-LRT>

6. 研究組織

(1)研究代表者

木村 哲夫 (KIMURA TETSUO)
新潟青陵大学・看護福祉心理学部・教授
研究者番号：90249095

(2)連携研究者

莊島 宏二郎 (SHOJIMA KOJIRO)
大学入試センター・研究開発部・準教授
研究者番号：50360706

(3)研究協力者

永岡 慶三 (NAGAOKA KEIZO)
早稲田大学・人間科学学術院・教授
研究者番号：90127382