

自動テスト生成システム（ATM）の開発と実践への応用

住 政二郎（理工学部・研究代表者）
工 藤 多 恵（理工学部）
乗 次 章 子（関西学院大学非常勤講師）
山 脇 野 枝（関西学院大学非常勤講師）

要 旨

関西学院大学は、2014年に「スーパーグローバル大学創成支援」に採択された。それ以降、学内では英語教育に関するさまざまな施策が実施されている。2017年度からは外部テストを活用した全学規模のプレイスメントテスト、習熟度別クラス編成、到達度テストが導入された。こうした変化は、学部横断的且つ縦断的に客観的指標で英語教育の成果を評価・検証するために重要である。しかし、外部テストの導入にはデメリットもある。外部テストの結果の取り扱いについて明確な方針はない。多様な学部テストの活用が学内に混乱をもたらす可能性もある。また、各学部は専門性を踏まえた特色のある英語教育を行っており、外部テストの評価指標とは必ずしも一致しない。特に理工学部は、各学科の専門性や大学院への進学を念頭に、科学技術英語に特化した英語教育を行っているためその影響は大きい。こうした課題を背景に、2017年度高等教育推進センター共同研究助成を受け、外部テストの内容と学部独自の教育内容の双方を加味しながら、英語教育の成果を客観的指標で評価・検証できる到達目標型英語教育の実現を目指し、ATMの開発と実践への応用を行った。

1. はじめに

スーパーグローバル大学創成支援事業の採択以降、多様な改革が英語教育には加えられている。全学規模のプレイスメントテスト、入門英語クラス、そして到達度テストの導入は、その主たるものである（関西学院, 2016）。こうした改革は、客観的な指標に基づき英語教育の成果を検証し、改善するための基盤として重要なものであろう。

しかし、多くの課題も残されている。例えば、学部の英語教育の内容と外部テストの測定内容の違いである。他にも、外部テストの結果のみが強調され、学部の英語教育の成果が軽視される可能性もある。今後、外部テストによる成果主義の圧力が過度に高まった場合、学部の英語教育が外部テストに取って代わられる可能性もある。特に理工学部への影響は大きい。理工学部では、各学科の専門性や大学院への進学を念頭に、科学技術英語に特化した英語教育を行っている。

その内容は、外部テストが想定する構成概念とは大きく異なる。

こうした課題を解決するために、2017年度高等教育推進センター共同研究助成を受け、外部テストの内容と学部独自の教育内容の双方を加味しながら、英語教育の成果を客観的指標で評価・検証できる到達目標型英語教育の実現を目指し、自動テスト生成システム (Automated Test Maker、以下 ATM) の開発と実践への応用を行った。

2. ATMの開発

2.1 ATMの概要

ATMは、データベースに保存されたテスト項目を使って、「英単語多肢選択問題」(Multiple Choice)、「英文読解問題」(Reading Comprehension)、そして、受験者の正誤に応じて問題を出题する「CAT」(Computer Adaptive Test)の3種のテストを自動的に出力することができる。ATMは、サーバー上で稼働し、担当教員はIDとパスワードを使い、各種テストを出力して授業内外の学習や期末考査等で活用することができる。図1は、ATMのログイン画面である。

テスト項目は恒常的に改善され、また、新しいテスト項目も随時追加されている。すべてのテスト項目には、等化作業を加え、調整された困難度パラメータが付与されている。困難度パラメータが付与されたテスト項目を使いテストを出力し、それを期末考査で活用することで、習熟度別クラス編成の受験者能力の違いに対応することができる。また、同一科目・複数クラスが開講される英語科目で、教員個人で異なる授業アプローチを許容しながらも、学習到達度を客観的な指標で揃えることができる。テスト結果は、独自のスコア・スケール (Kwansei Gakuin English Language Scale、以下 KG. ELs) に換算される。これは、テスト結果から学生の能力パラメータを推定し、その値を線形変化させたものである。KG. ELs を共通尺度として採用することで、学生個人と学部独自の英語教育の到達度を指標化し、同時に外部テストの結果との比較検討が可能になる。

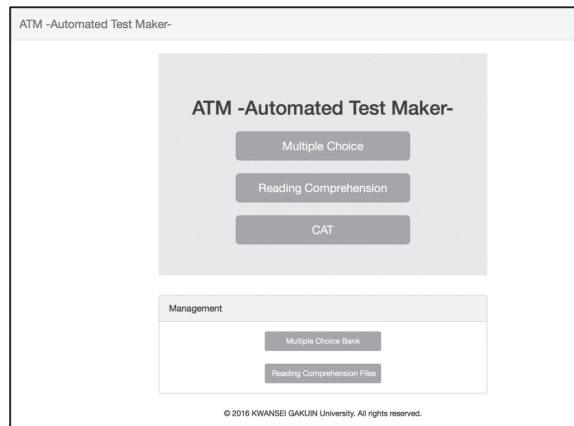


図1 ATMのログイン画面

2.2 英単語多肢選択問題

ATMに実装した英単語多肢選択問題は、アイテムバンクの開発作業から行った。開発素材に

は、河上（2011）『カラー版 TOEIC にできる順英単語』を活用した。これは、理工学部1・2年生の英文読解（必修科目）の授業で、副教材として使われているものである。

多肢選択問題の出題形式は、TOEIC® Test の Part 5 を参考にした。問題文および選択肢には、大学英語教育学会が定める基本語リスト JACET8000 の内、Level 4（大学受験、大学一般教養初級）までの語彙を使うことにした。現在、アイテムバンクには、約2,000問の問題項目がある。

開発した問題項目には、項目分析を加えた。分析は、古典的反応理論を使い、テスト毎に標準困難度適切度、標準項目弁別力適切度、標準実質選択肢適切度、標準適切度合計の観点から考察を加え（中村，2002）、ラッシュモデルの推定結果（*outfit, infit*）と合わせて改善を加えた（住，2013）。

図2は、英単語多肢選択問題の出力画面である。図2の①では、出題する多肢選択問題の数を1～100まで指定することができる。初期設定は20問で、出力結果は20問毎にA4の用紙1ページに収まるように設計されている。②では、多肢選択問題の出題範囲を指定単語帳の単語番号で指定することができる。③では、多肢選択問題の出題範囲を指定単語帳の Part 2～5 から、または複数を組み合わせて指定することができる。④では、共通項目の指定などを考慮し、特定の項目をカンマ区切りで指定することができる。以上、上記①～④の機能を組み合わせ、約2,000問のアイテムバンクから自由に多肢選択問題を PDF 形式で出力することができる。

The screenshot shows a web form titled "PDF出力" (PDF Output). It contains several input fields and a list of options:

- Title:** A text input field with a note "(Use alphabets and numerics only. Not required.)" and a small icon.
- Question number:** A dropdown menu currently showing "20", with a circled 1 next to it.
- Item No.:** A range selection field (e.g., "1 ~ 10") with a circled 2 next to it.
- Part:** A list of options: "Off", "Part02(501 - 1000)", "Part03(1001 - 1515)", "Part04(1516 - 2025)", and "Part05(2026 - 2535)". A circled 3 is next to the list.
- Include:** A text input field with a note "(Specify by comma separator. ex) 123,456,567" and a circled 4 next to it.
- Keyword:** A text input field.
- PDF生成:** A button at the bottom left.

図2 英単語多肢選択問題の出力画面

2.3 英文読解問題

読解問題の開発は、主に科学技術に関する題材を取り扱う指定教科書に準拠して行った。さらに、コンテンツを追加するためにVOA (Voice of America) の素材を使って読解問題を開発した。読解問題を開発する際には、担当者間で理工学部の学生に身につけて欲しい読解力について協議を行い、テスト仕様を決定した。また、複数の外部テストのテスト形式を検討し、本文の内容に

即して、豊富な問題形式を持つ IELTS の出題形式を参考にすることにした。VOA の素材を選定する際には、教科書本文とのバランスを考慮し、約800字で、Flesh-Kincaid Grade Level 6-8 の基準を定めた。読解問題に関しても、語彙・文法問題と同様にラッシュモデルと古典的反応理論を使った項目分析を行った。

図3は、英文読解問題の出力画面である。①のデータベースより読解問題を選択し、②に追加して、③のボタンをクリックすることで、複数の素材を組み合わせた読解問題（解答と解答用紙を含む）を出力することができる。

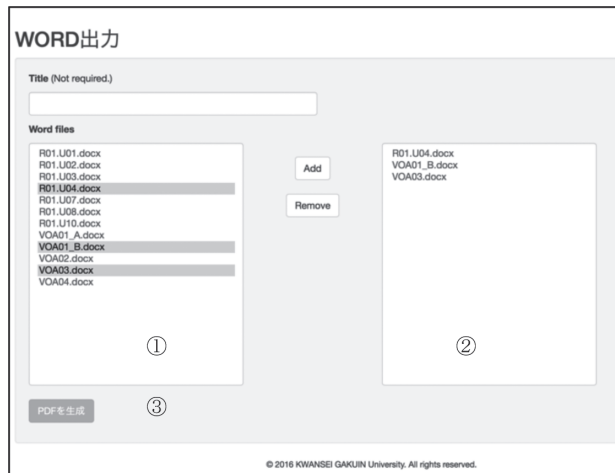


図3 英文読解問題の出力画面

2.4 Computer Adaptive Test

CATは、困難度パラメータの付与された問題項目を学生が受験し、正誤に応じて学生の能力パラメータを推定することができる。また、推定された能力パラメータから、学生の能力に最適化された問題項目を自動的に出題することができる。英単語多肢選択問題の機能では、全員に同じ問題を提供する一方で、CATでは、学生の単語力に応じた問題を出題することができる。多肢選択問題の機能の弱点を補い、学生の日頃の英単語学習に役立ててもらうことを意図している。

CATの開発は、多肢選択問題で開発したテスト項目を等化することから行った。等化は、共通項目を含む複数のテスト・ブロックを出力し、テスト結果からテスト間の共通項目の回帰直線を求めて行った。現在、等化作業とCATへの実装は、1年生の学習必須単語約1,000問のみ完了している。

CATの開発には、ラッシュモデルを採用した。受験者の能力推定にはベイズEAP (expected a posteriori) 推定法を、項目選択にはフィッシャー情報量を用いた。ベイズEAP推定法(式1)は、受験者の回答パターンが全問正解あるいは全問不正解の場合でも受験者能力の推定が可能で、推定作業も短時間で済むことからCATに適しているとされている(村木, 2011)。受験者の能力値の分布には事前確率として標準正規分布を仮定した(豊田, 2012)。

$$\hat{\theta}_i = \int_{-\infty}^{+\infty} \theta_i g(\theta_i) L(u_i | \theta_i) d\theta_i / \int_{-\infty}^{+\infty} g(\theta_i) L(u_i | \theta_i) d\theta_i \quad (\text{式 1})$$

図4は、CATの出力画面である。①で問題数を指定し（初期設定30問、最大50問）、②で出題範囲を指定する。

図4 CATの出力画面

図5は、出力されたCATの開始画面である。テストは、オンライン、または画面を印刷して配布し、QRコードをスキャンすればスマートフォンでも受験することができる。受験者能力の推定結果は、式2を使い線型変化し、KG.ELsとして表示するようにした。テスト結果は、図6のように表示され、教員側は受験結果を一元管理することができる。

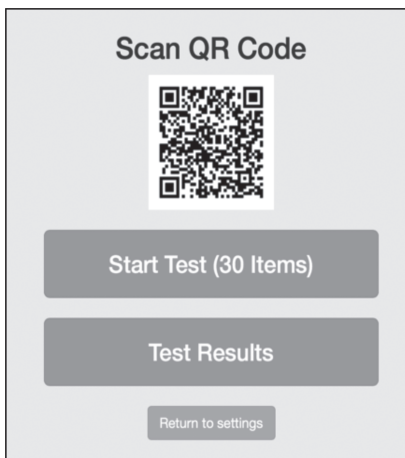


図5 出力されたCATの開始画面



図6 CATの受験結果の表示画面

$$KG.ELs = 50 \times \theta_i + 500 \quad (\text{式 2})$$

3. 英語教育への応用

ATMの各機能について、英語教育への応用を行い、その効果を検証した。英単語多肢選択出力機能については、事前・事後テストを行い、ATMで出力した小テストの継続利用が学生の単語力向上にもたらす効果について検証した。英文読解問題出力機能については、1年生を対象とする英文読解の2017年度前期・期末考査で利用し、その感想を担当教員から収集した。

3.1 英単語多肢選択問題について

2017年度前期、読解の授業を受講する1回生268名(11クラス)を対象に、事前・事後テストを行った。事前・事後テストは、ATMの英単語多肢選択問題で出力した100問で行った。事前テストは4月に、事後テストは7月に実施した。事前・事後テストとも同じテストを使用した。事前・事後テストの間には、10問ずつの多肢選択問題をATMで出力し、毎回の授業で活用した。表1と図7は、事前・事後テストの結果である。表1は、記述統計量の一覧で、図7は、事前・事後テストのバイオリンプロットを示している。結果より、授業時間を使った継続した学習が、学生の英単語力の向上に寄与していることが伺える。

表1 事前・事後テストの結果

<i>N</i> = 268	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
事前	62.90	13.29	18	93
事後	73.27	12.08	33	100

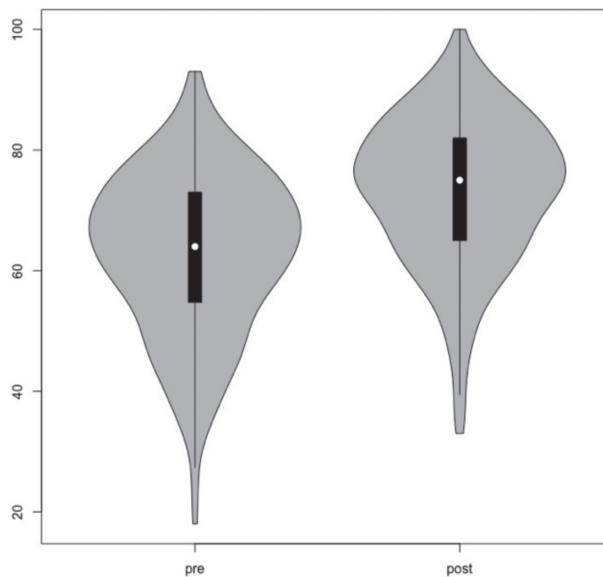


図7 事前・事後テストの結果

3.2 英文読解問題について

英文読解問題については、1年生が受講する英文読解（27クラス）で、ATMで出力した読解問題を使用して期末考査を行った。その後、担当教員（8名）へのアンケート調査を行った。

アンケート調査は、6件法（1：「まったくそう思わない」～6：「まったくそう思う」）を採用した。また、自由記述の欄を設けた。

「共通読解問題の利用は、各クラスの授業到達度をそろえるために有益だ」という質問に対しては、4（1名、12.5%）、5（3名、37.5%）、6（4名、50.0%）という結果だった。「共通読解問題の利用によって、テストの作成負担が軽減された」という問いに対しては、5（1名、12.5%）、6（7名、87.5%）という結果だった。「共通読解問題の利用によって、教えるポイントが予め明らかなので、今後は、授業内容も良い方向に変わる」という問いに対しては、4（2名、25.0%）、5（5名、62.5%）、6（1名、12.5%）という結果だった。以上の結果から、ATMの導入が、担当者のテスト作成負担を大幅に軽減し、クラス毎の到達度を揃えるために有益であった様子がうかがえる。

自由記述で特徴的であったのは、予めテストがあることで到達度が明確となり、教授内容の改善にATMが有用である、という意見だった。また、多様な読解問題を用意することによって、どのような点で指導が不足しているのか、各クラスを比較して検証できるようになったとする意見も聞かれた。こうした意見は、担当クラス数の多い非常勤講師から多く聞かれた。自由記述からは、単に統一テストを導入して学習到達度を揃えるだけでなく、測定と評価を連動させ、授業にフィードバックすることの重要性が明らかになった。

4. まとめ

本研究では、外部テストと学部独自の英語教育の内容を加味しながら、客観的指標で英語教育の成果を評価・検証できる到達目標型英語教育の実現を目指しATMを開発し、英語教育へ応用した。ATMを活用し、困難度の調整されたテスト項目から語彙問題や読解問題を自動的に生成できる環境を整備することによって、英語教育の成果を客観的指標で評価・検証できる体制を整えることができたことは、本研究の大きな成果である。また、問題作成に伴う教師の負担を大幅に軽減すると共に、現在は読解の授業に限ってではあるが、英語教育の到達度を揃えることが可能になった。英語科目は同一科目で複数クラスが開講されるという他の科目にはない特徴を持つ。同じシラバスと教科書を使いながらも、到達度を揃えることがこれまで極めて困難であった。担当教員間でアイテムバンクを構築し、テスト項目の改善を続け、また、その結果を授業に還元する再帰的な授業改善サイクルをATMを介して構築することができたことは、学内に外部テストが導入され、教育の質保証と説明責任が求められる中で重要な成果となった。

謝辞

本研究は、2017年度 関西学院大学高等教育推進センター 共同研究助成を受けたものである。

参考文献

- 河上源一 (2011). 『カラー版 TOEIC テストにできる順英単語』 東京：KADOKAWA.
- 関西学院大学 (2016). 「中期計画の取り組み」, http://www.kwansei.ac.jp/kikaku/kikaku_009760.html
- 村木英治 (2011). 『項目反応理論』 東京：朝倉書店.
- 中村洋一 (2002). 『テストで言語能力は測れるか：言語テストデータ分析入門』 東京：桐原書店.
- 住政二郎 (2013). 「ラッシュモデルの導出」『外国語教育メディア学会関西支部メソドロジー研究部会2012年度報告論集』, 83-101. Retrieved from http://www.mizumot.com/method/2012-07_Sumi.pdf
- 豊田秀樹 (2012). 『項目反応理論 [入門編] 第2版』 東京：朝倉書店.