

Automated Test Maker の開発と実践への応用

住 政二郎^{*1}, 工藤 多恵^{*1}, キム・ジョリー^{*1}, ローラ・ヒューストン^{*1},
ケント・ジョーンズ^{*1}, ジョシュア・ウィルソン^{*1}
^{*1} 関西学院大学

The Development and Application of ATM

Sei Sumi^{*1}, Tae Kudo^{*1}, Kym Jolley^{*1}, Laura Huston^{*1}, Kent Jones^{*1}, Joshua Wilson^{*1}
^{*1} Kwansai Gakuin University

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、英語教育の改革がはじまった。2017年度からは英語民間試験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。2021年度からは、全ての入学試験にて、英語4技能を評価し、出願資格とする方針が発表された。英語民間試験に準拠しながら学部横断的且つ縦断的に英語教育の成果を評価・検証する環境が整いつつある。しかし、解決すべき課題がある。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、その内容は英語民間試験の評価基準とは一致しない。この課題を解決するために、理工学部では、2015年度より、英語民間試験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）の開発と実践に取り組んできた。

キーワード: 大学英語教育, 外部英語試験, 到達目標型大学英語教育, テスト・プラットフォーム

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、英語教育の外部環境は急速に変化している。その特徴は、英語民間試験に準拠する形で、各学部の英語教育に改革が加えられていることである。しかし、英語民間試験のような集団基準準拠テストと、学部独自の英語教育の成果を測る目標基準準拠テストとは、そもそも性格を異にするものである。2020年度からは、英語科目に関しては民間試験を併用した新しい大学入学共通テストがはじまる。学内でも2021年度から全ての入学試験にて、英語4技能を評価し、出願資格とする方針が発表された⁽¹⁾。今後、ますます、英語民間試験に準拠する形での英語教育の改革の外圧は高まるであろう。

理工学部では、こうした変化を見越し、英語民間試験と学部独自の英語教育の双方の内容と評価基準を加味しながら、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

開発と実践に2015年度から取り組んできた。

2. ATM の開発

ATMは、「多肢選択問題」(Multiple Choice)、「読解問題」(Reading Comprehension)、「CAT」(Computer Adaptive Test)の3種のテストを自動的に生成することができる(図1)。

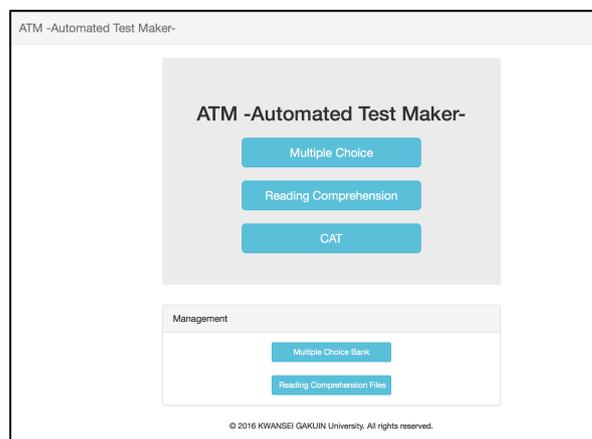


図1 ATM のログイン画面

2.1 多肢選択問題

多肢選択問題は、理工学部 1・2 年生の読解の授業で使う単語集をもとに、アイテムバンクの開発から行った⁽²⁾。出題形式は、TOEIC® Listening and Reading Test の Part 5 を参考にした。現在、アイテムバンクには、約 2,000 問の問題項目がある。

開発した問題項目は項目分析を行い、学期毎に改善を加えている。分析は、古典的反応理論を使い行った⁽³⁾。また、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⁽⁴⁾。

図 2 は、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 まで指定できる。②では、多肢選択問題の出題範囲を指定単語集の単語番号で指定できる。③では、多肢選択問題の出題範囲を指定単語集の Parts 2~5 から、または複数を組み合わせて指定できる。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 形式で出力できる。

図 2 英単語多肢選択問題の出力画面

2.2 読解問題

読解問題の開発は、指定教科書に準拠しながら行った。また、専任教員間で協議を行い、読解の授業で身につけるべき英語力について検討を加えた。

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、複数の素材を通し番号で連結し、解答と解答用紙を含む 1 つの問題として出力できる。

図 3 英文読解問題の出力画面

2.3 CAT

CAT は、受験者の正誤に応じて能力パラメータを推定し、能力パラメータに最適化された問題項目を自動的に出題できる。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できる。

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の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⁽⁵⁾。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⁽⁶⁾。

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

$$\int_{-\infty}^{+\infty} g(\theta_i) L(u_i | \theta_i) d\theta_i$$

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。受験者の能力推定値の結果は、式2を使い線形変化し、KG.ELs（Kwansei Gakuin English Language Scale）として表示し、他の英語民間試験と対照できるようにした。

図4 CATの出力画面

$$KG.ELs = 50 \times \theta_i + 500 \quad (2)$$

3. 読解問題の開発

2019年度からは、読解問題の内容を充実させ、さらに精度を向上させる取り組みを行っている。読解問題はこれまで1年生用の問題しか存在しなかったが、2年生の読解問題開発にも着手した。読解問題の開発にあたっては、専任教員2名と英語母語話者の契約教員4名からなるチームを組織した。

読解問題の開発の際には、専任教員が中心となり、まず、問題カテゴリーを協議した。結果、以下、4種類の問題カテゴリーが決定した。

1. 主題に関するもの
2. 語彙・表現に関するもの
3. 内容理解に関するもの
4. 発展的内容に関するもの

次に、上記の問題カテゴリーに対して、それぞれタ

ーゲットとなる読解スキルと知識を定め、複数の設問タイプを設計した。読解問題を作成する際には、バランス良く設問が配置されることを意図し、また、統一された問題文を使うことにした。こうすることで、1年生から2年生まで、必修の読解の授業では、語彙・表現および内容のレベルはあげながらも、共通した読解スキルと知識を測定できるようにした。

4. 実践への応用と今後の予定

今後は、1年生と2年生の複数クラスで、多肢選択問題の事前テストと事後テストの実施を予定している。事前テストと事後テストの間では、ATMの多肢選択問題またはCATで出力した練習問題を授業で活用し、その効果を検証する予定である。

その他にも期末テスト等で読解問題を活用し、項目分析と困難度パラメータの推定を行う予定である。そして、将来的には、読解問題のCATの開発につなげていきたいと考えている。

読解問題に関しては、テストセットを出力する際に、KG.ELs換算で何点ぐらいのテストセットになるのかを表示できるようにしたいと考えている。こうすることで、GPAを採用しながらも、これまで不統一であったクラス間の評価基準を揃えることが可能になり、また習熟度別クラスにも、よりきめ細やかに対応することが可能になる。

テスト結果に関しても、従来のように、単にテストの結果、つまり、スコアだけをフィードバックするのではなく、KG.ELsを使うことで、学部独自の英語教育内での到達度を示すことができる。また、今後、データ量を増やすことで、KG.ELsの値から各種の英語民間試験のスコアを推定することも可能になる。こうすることで、学部独自の英語教育を行い目標基準テストの要件を満たしながら、集団基準準拠テストの情報も提供することが可能になる。

5. おわりに

英語教育を取り巻く環境は大きく変わりつつある。入試の抜本的改革、民間試験の入試への導入など、その背景には、客観的な指標で英語力を測ろうとする力学が通底する。その影響は、これまで比較的自由に行

われていた学部の英語教育にも及んでいる。民間試験を活用したプレイスメントテストや到達度テスト，こうしたテストと学部独自の英語教育との連携が求められるようになった。

本研究では，民間試験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。ATM をテストプラットフォームとして授業の延長で活用することで，教員個々人で異なる授業アプローチを許容しながらも，到達度を客観的な指標で揃えることができるようになった。未だ実践への応用は部分的であり，不十分ではあるが，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 というテストプラットフォームを介して構築することができたことは，本研究の重要な一歩となった。

参 考 文 献

- (1) 2021 年度 入試改革基本方針について，https://www.kwansei.ac.jp/press/2018/press_20181015_020777.html (2019 年 4 月 5 日確認)
- (2) 河上源一：“カラー版 TOEIC テストに できる順英単語”，KADOKAWA，東京（2011）
- (3) 中村洋：“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”，桐原書店，東京（2002）
- (4) 住政二郎：“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ロジー研究部会 2012 年度報告論集，pp. 83-101（2012）
- (5) 村木英治：“項目反応理論”，朝倉書店，東京（2011）
- (6) 豊田秀樹：“項目反応理論 [入門編] 第 2 版”，朝倉書店，東京（2012）