学修成果の評価テストの開発：
金言版最小英語テストを用いた試行テストの報告

教養教育センター

小林（こばやし）　昌博（まさひろ）／滝波（たきなみ）　稚子（わかこ）／クリストファー・ハリス

１．はじめに

本年度より実施している学長裁量経費事業において、学生の学修成果を評価するテストの開発・構築は重要である。本年度は、評価テストの第一段階として、岐阜大学の牧秀樹教授が率いるチームによって開発が進められている最小英語テスト（The Minimal English Test、以下 MET）に修正を加えたテストを用いて、本学 1 年生 56 名に対して試行的に英語力の測定を行ったのでここに報告する。学修成果を計測するテストを開発するためには、当テストが学生の習熟度を適切に測定するための信頼性が求められる。その信頼性を確認する一環として、大学入学共通テストのスコアとの相関を調査した。さらに、テストの性質を詳細に調べるために、大学入学共通テストのスコアを説明変数とし、重回帰分析を実施して品詞ごとの説明変数に対する貢献度を調査した。

２．金言版最小英語テストについて

MET（Goto et al (2010)、牧(2018)）は前述の通り、岐阜大学の牧グループにより開発が進められてきたテストである。簡易型日本語運用能力測定試験（The Simple Performance-Oriented Test）から着想を得て、最小日本語テスト（The Minimal Japanese Test、以下 MJT）が作成され、MJT を応用し MET が作られた。2003 年に MET が開発されて以降、小学生版 MET、中学生版 MET、センター試験レベル MET、金言版 MET など、様々なバージョンが作られてきた。本プロジェクトでは大学入学共通テスト英語の得点と統計的に有意な相関が報告されている金言版 MET（以下 kMET）を使っている。kMET の基本的なデザインは、A4 一枚ほどの用紙に 45 文程度の英文が並び、ある一定の間隔で空欄が開けられていて、受験者はその英文を読み上げる音声を聞き、該当する箇所の英単語を埋めていくいわゆる cloze テストである。

３．大学入学共通テストとの相関および重回帰分析の結果報告

MET には用いる英文や受験者のレベルなどによりいくつかのバージョンが存在するが、今回は kMET を用いる。オリジナルの kMET を 2 つ準備し、片方を 5 つの品詞の大学入学共通テストのスコアへの貢献度を調べる重回帰分析用（問題１）に修正して用いた。もう片方を被験者の大学入学共通テストのスコアと当該テストのスコアの相関を見るために用いた（問題２）。なお、問題１を作成するために、オリジナルの kMET を重回帰分析用に修

正するにあたり、問題文を Python の NLPTK ライブラリを用いて各単語に分割し、品詞のタグを割り当て、問題文の中に生じる 5 品詞（冠詞、前置詞、代名詞、動詞、名詞）の単語のリストを作成した。さらにそれぞれの品詞に 12 個の空欄箇所を作成することとし（12×5=60 問）、各品詞に関して等間隔で空欄を設定した。その後、被験者が答えを記入する時間的余裕も考慮に入れ、空欄どうしは少なくとも 3 単語分離れているようにした。さらに 1 文に多くても設問箇所は 2 か所までにするために、それぞれの品詞に対して設問箇所を微調整して問題を作成した。大学入学共通テストを受験している本学の 1 年生 56 名に対して本テストを実施した。なお本調査は、教育支援・国際交流推進機構の非医学系研究倫理審査委員会で承認された実験である。

図 1 は問題 1 における各品詞の平均点である。なお、前置詞を問う問題のうち被験者全員が不正解であった問題が 2 問あり、信頼性係数 ω を求めるために前置詞の問題 12 問から当該の 2 問を除いて分析をした。同様に代名詞の設問にも 1 問、全員が不正解の問題があったため分析より除いた。図 1 より前置詞の平均点が他の品詞の平均点より低いことが見て取れ、今後追加実験等により原因を探す必要がある。

次に、各品詞を独立変数として大学入学共通テストのスコアを説明変数とした重回帰分析を行った。結果を表 1 に報告する。VIF によって多重共線性の可能性を検討したところ、全ての説明変数において 2 未満であった。したがって、多重共線性が生じている可能性は低いと判断できる。モデルの決定係数 R2 は .266 で、自由度調整済み決定係数は .192 とモデルの適合度はあまり高くなかったため、今後の調査では原因を探ってモデルの適合度を上げる必要がある。

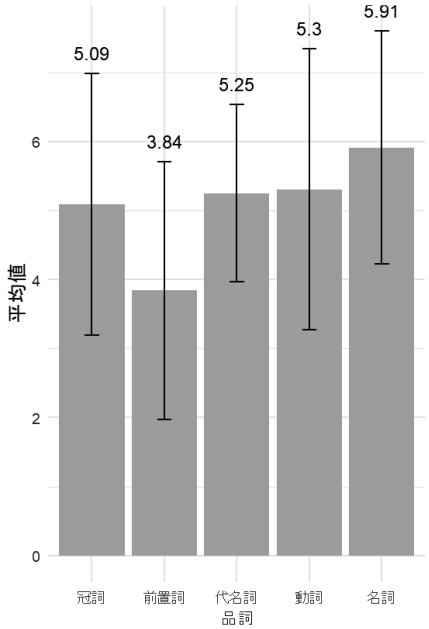

図 1：品詞ごとの平均点

表 1: 重回帰分析の結果

項目	推定値	標準誤差	95% CI
(切片)	82.412	11.919	(58.471, 106.352)
冠詞	0.463	1.496	(2.541, 3.468)
前置詞	1.621	1.587	(-1.566, 4.807)
代名詞	5.018	2.207	(0.585, 9.450)
動詞	1.737	1.463	(-1.200, 4.675)
名詞	-0.785	1.628	(-4.054, 2.485)

$R^2 = .266^{**}$

表 2: 大学入学共通テスト（リスニングおよびリーディング）との相関 (r)

テスト	リスニング	リーディング	L&R 合計
kMET 問題 1	.52***	.25·	.45***
kMET 問題 2	.55***	.34**	.52***
kMET 問題 1&2 合計	.57***	.32*	.52***

$p < .001$ '***'　$p < .01$ '**'　$p < .05$ '*'　$p < .1$ '·'

表 3: 品詞の設問ごととテスト全体の信頼係数 (ω)

冠詞	前置詞	代名詞	動詞	名詞	問題 1	問題 2	問題 1&2 合計
.63	.67	.43	.47	.49	.72	.83	.90

決定係数に関する F 検定は 0.1%水準で有意であったが（$F (5, 50) = 3.623, p < .001$）、5 つの説明変数のうち代名詞の偏回帰係数のみが 5%水準で有意であった。開発チームの岐阜大学牧教授との個人的な意見交換で、重回帰分析をすると前置詞などの機能語の偏回帰係数が高くなる傾向があるということであったので、その点では本学での試行実験でも同様の結果が得られたと言えよう。ただし、その他の品詞は有意水準になく、これも今後の調査の検討課題である。また、これはまだ憶測であるが、機能語の偏回帰係数が高くなる理由としては、本テストが聞き取りをして穴埋めをするタイプのテストであり、発話の特性として機能語は弱く発音され聞き取りが難しいため、英語運用能力の差異をある程度峻別することができるのではないか、と思われる。

次に、大学入学共通テストのスコアと本調査で用いた kMET のスコアの相関（ピアソン）を調べたので、概要を表 2 に報告する。表 2 にあるように、重回帰分析用に作成した kMET（問題 1）と相関分析用に用いたオリジナルの kMET（問題 2）、そして 2 つのバージョンを合わせたもの（問題 1&2 合計）、そしてそれらと大学入学共通テストのスコアとの相関性を調査した。それぞれのバージョンにおいてリスニングとの相関係数が中程度の相関性を示している一方でリーディングに関しては概して弱い相関を示すにとどまっている。これは kMET が聞き取りテストの形式を採用しているためと思われる。リスニングとリーディングの合計スコアとの相関を見ると .52 と中程度の相関性を示している。

最後に今回実施した kMET の信頼性を示すために ω 係数を求めたので表 3 に報告する。5 つの品詞の ω 係数を見てみると、代名詞と動詞、および名詞の ω 係数の値が低めであるため、今回設定された設問項目からどの項目を除外すると値が改善するかを再検討して次の再実験に臨む必要がある。問題 2 の ω 係数は、.83 を超えるため信頼性が確認された。問題の 1 と 2 を合計した場合の ω 係数は設問数の増加から.90 を超えた。

４．今後の展望

今回のパイロットスタディでは、カリキュラムの評価と学修成果の評価用のテストを開発する第一段階として、kMET を本学の 1 年生に対して実施してその結果を調べた。具体

的には被験者が受験した大学入学共通テストのスコアと kMET のスコアの相関を調べた。さらにテスト文に生じる 5 つの品詞の大学入学共通テストのスコアに対する貢献度を調べる重回帰分析を行った。今回は代名詞の偏回帰係数が有意であった。これは回帰式において機能語の貢献度が高くなる傾向があるという先行研究の知見と一致する。相関係数に関しては、大学入学共通テストのリーディングセクションよりもリスニングセクションの方が kMET との相関が高いことが示された。これは、kMET が聞き取りをして単語を埋める形式のテストであることが理由の一つであると考えられる。大学入学共通テストに対しては、kMET の合計のスコアは中程度の相関があることがわかった。

最後に今後の課題と方向性を述べたい。まず、信頼性をさらに保証するために、テスト項目の見直しをしてさらなる修正を加えた後に、追加実験を実施して評価用のテストの開発を進める予定である。特に重回帰分析におけるモデルの適合度がまだ低いということと大学入学共通テストのスコアや他の外部試験との相関性もさらに高める必要があることなど課題が残る。それと同時に、本事業の目的は本学の英語カリキュラムの評価および学生の学修成果の評価であるため、テスト開発の方法については検討する必要があるだろう。

参照文献

Goto, Kenichi, Hideki Maki, and Chise Kasai (2010). “The Minimal English Test: A New Method to Measure English as a Second Language Proficiency,” *Evaluation & Research in Education.* 23(2), pp.91–104.

牧秀樹（2018）『The Minimal English Test （最小英語テスト）研究』，開拓社，東京.

謝辞：

今回のテストに参加してくださった被験者の学生の皆様と MET を用いたテスト開発に協力してくださった岐阜大学の牧秀樹教授に感謝致します。