

金言版最小英語テストに基づくアセスメントテストの開発

—— Cloze テストの品詞別重回帰分析 ——

小林昌博^{*}

滝波稚子[†]

クリストファー・ハリス[‡]

キーワード：テスト開発、最小英語テスト、重回帰分析、相関分析

1 はじめに

かねてから英語教育の現場において、カリキュラムの到達目標を明確にする必要性が主張されており、したがってカリキュラム自体の有効性を検証するために、学修者の習熟度や到達度を測るアセスメントテストを開発することは喫緊の課題となっている。本稿では、習熟度を計測するテストを開発することを目指して、まず最小英語テスト（Goto, Maki and Kasai, 2010）に基づく聴解問題形式の Cloze テストを作成してその有効性を検証することとした。まず、学修者の習熟度の指標として大学入学共通テストのスコアを利用し、そのスコアと本研究で開発した聴解問題形式のテストのスコアの相関性を調べたところ、中程度の相関性が確認された。さらに、空欄に設定する単語の品詞をどのように統制するべきかを調べるために、内容語として「名詞」「代名詞」「動詞」、機能語として「前置詞」「冠詞」の5種類を選び、これらを独立変数に、大学入学共通テストのスコアを目的変数として重回帰分析を行った。その結果、前置詞、動詞、代名詞の3品詞が正の影響を持っていることが確認された。今後は、本研究の知見を踏まえて、さらに他の形式も取り入れたアセスメントテストの開発を進める予定である。

2 調査の目的と背景

大学の英語教育カリキュラムにおいて、各科目の学修成果の明確化とその測定方法の開発は重要な課題である。また、プレイスメントテストなどを目的とした習熟度テストの開発も同様に重要となる。教育機関の現場では、学修成果を計測するアセスメントツールとして TOEIC や TOEFL などの外部試験を利用することが多いが、外部試験の利用は試験のデザインがカリキュラムの目的に必ずし

^{*} 鳥取大学 教育支援・国際交流推進機構 教養教育センター 兼務教員（高等教育開発センター 教授）

[†] 鳥取大学 教育支援・国際交流推進機構 教養教育センター 助教

[‡] 鳥取大学 教育支援・国際交流推進機構 教養教育センター 講師

も適合しない、試験時間が長く実施のための時間を確保するのが難しいといった問題が指摘されている (Goto, Maki and Kasai, 2010)。さらに、受験料が高額であることなどの予算的な問題も存在する。一方、独自にアセスメントテストを開発する場合、テストの信頼性や妥当性を考慮する必要があり、作成したテストを改善していくプロセスが必要不可欠となる。このような背景から、本研究では、受験者に負担をかけないように短時間で終了する、英語学習者の習熟度を計測するアセスメントテストを開発することを最終的な目的とし、まず、岐阜大学の牧秀樹教授のグループが開発している金言版最小英語テスト (牧, 2020) を修正したテストを作成・実施して、大学入学共通テストのスコアと本研究で開発したテストのスコアとの相関性、そして大学入学共通テストのスコアを目的変数に、5つの品詞を独立変数にした重回帰分析を行い、作成したテストの信頼性と妥当性を調べることとした。

3 テストの実施と分析

3.1 テストと実験のデザイン

前節で述べたように、既存の外部試験をそのままアセスメントテストとして利用するには、いくつかの問題が指摘されている。そのような背景から、岐阜大学の牧秀樹教授の研究グループは、約5分間で実施可能な最小英語テスト (The Minimal English Test) を開発し、その有効性などの検証を行っている (牧, 和佐田, 橋本, 2003; 牧, 2018)。最小英語テストは、cloze テスト形式の聴解問題であり、45文程度の英文が提示され、一定間隔で空欄が設けられた構成となっている。受験者は英文音声聞きながら、空欄に該当する単語を記入する形式をとる。

最小英語テストには複数のバージョンが存在し、TOEIC や TOEFL iBT など外部試験のスコアとの相関関係に関する研究が行われていて、その相関性が報告されている (陳 & 牧, 2023)。本研究においても、金言版最小英語テストを基にしたオリジナルのアセスメントテスト開発を行い、テスト作成に着手した。具体的には、金言版最小英語テストから45文を選定し、各文に1つの空欄を設けた聴解テストを作成した。学習者の習熟度を予測する上でどのような単語を空欄に設定すべきかを検討するため、空欄に設定する単語の品詞を統制し、内容語として「名詞」「代名詞」「動詞」、機能語として「前置詞」「冠詞」の5種類を選定した。それぞれの品詞から9語ずつ、合計45箇所の空欄を設けた。

開発したテストの有効性を調べるために、まず、本研究のテストのスコアと被験者の大学入試共通テストのスコアの相関性を調べることにした。さらに大学入学共通テストのスコアを目的変数として設定し、独立変数である5種類の品詞がスコアにどの程度の説明力を持つかを検証した。なお、各文に1箇所のみ空欄を設定したため、全体で45文・45箇所の穴埋め問題となっている。牧らのグループが開発したテストでは、1文に複数の穴埋め箇所が存在するが、本研究の事前調査 (小林, 滝波, ハリス, 2024) では、45文で60箇所の穴埋め問題を用いたところ、被験者の正答率が低く、問題数が多すぎて解答が困難である可能性が示唆された。そのため、本研究では45文・45箇所に問題数を減らしたデザインを採用した。また、比較のためにオリジナルの金言版最小英語テスト (66箇所の穴埋め問題) を同時に実施し、その分析結果も報告する。本研究のテストは、大学の1年生および2年生469名を対象に実施したものである。複数の英語教員で採点を行うため、統一した採点基準を

設定した。文頭の単語が小文字で記されている場合は誤答としないが、それ以外のスペルミス（例えば、過去形や複数形の「-s」の脱落など）はすべて誤答とすることとした。本研究で開発したテストもオリジナルの金言版最小英語テストも一問一点で採点した。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の承認を受けて実施した。

3.2 テストの結果と分析

本節では、前述のテストの結果の分析を述べる。まず、本研究で作成したテスト（45点満点）の受験者の平均点は26.87点（SD=6.66）であり、オリジナルの金言版最小英語テスト（66点満点）の平均点は34.73点（SD=9.30）であった。この二つの平均点に差があるかを検討するために、まず、異なる二つのテストのスコアを標準化し、対応のあるt検定を行った。本研究で開発したテスト（Q1）とオリジナルの金言版最小英語テスト（Q2）のスコア分布の箱ひげ図は図1となる。その結果、平均差は -3.71×10^{-16} （95%CI[-0.063, 0.063]）であり、統計的有意差は認められなかった（ $t(469) = -1.16 \times 10^{-14}$ 、 $p = 1.00$ ）。95%信頼区間が0を含んでいるので、平均値に有意差がない可能性が強いこと、同一被験者が受験したスコアであること、また二つのテストは異なるテストではあるが、開発したテストは金言版最小英語テストに基づいていることを考えると統計的有意差がないことは妥当であると言える。

次に、本研究で開発したテストにおいて5品詞に関して9問ずつの設問を作成したので、それぞれ

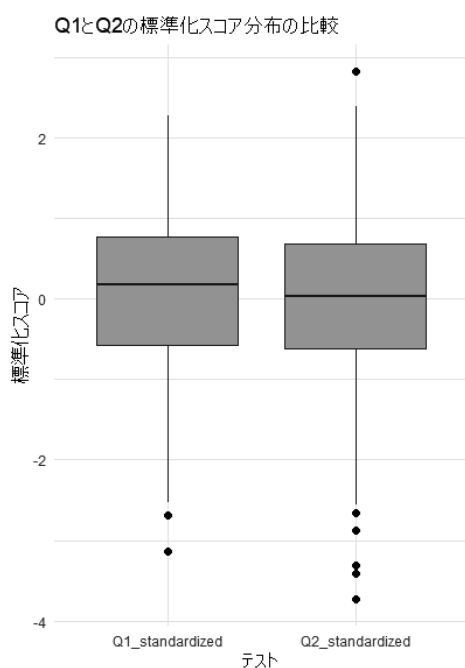


図1 本研究のテスト（Q1）とオリジナルの金言版最小英語テスト（Q2）のスコア分布

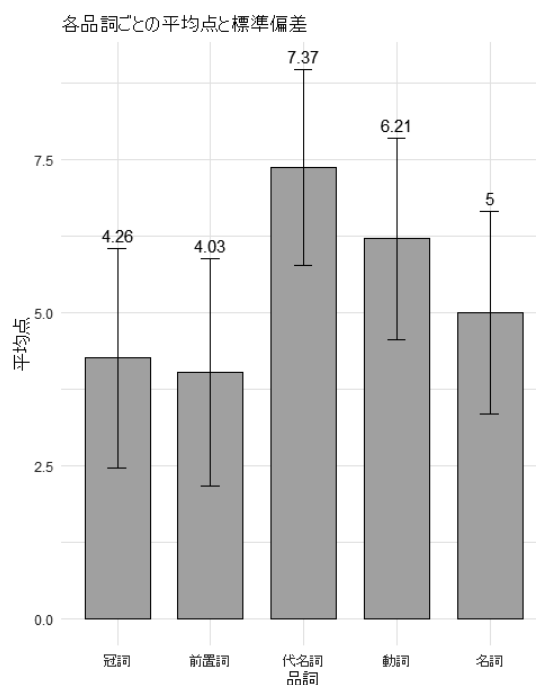


図2 本研究のテストにおける各品詞ごとの平均点

表1 重回帰分析の結果

変数	係数	標準誤差	t 値	p 値	95%CI
(定数項)	74.4523	3.4567	21.54	<0.001	[67.654, 81.251]
前置詞	1.5353	0.6745	2.28	0.023	[0.216, 2.855]
動詞	2.7051	0.6912	3.91	<0.001	[1.192, 4.218]
名詞	1.4423	0.7401	1.95	0.054	[-0.026, 2.911]
冠詞	0.9747	0.7032	1.39	0.170	[-0.418, 2.367]
代名詞	3.5736	0.7894	4.53	<0.001	[2.037, 5.110]

の品詞ごとの正解数の平均を調べた(図2)。冠詞は4.26 (SD=1.79)、前置詞は4.03 (SD=1.86)、代名詞7.37 (SD=1.59)、動詞6.21 (SD=1.64)、名詞が5.00 (SD=1.65)となった。今回のテストでは、正答率が最も高かったのは代名詞で、次いで動詞、最も低かったのは前置詞であった。本研究の事前調査である小林他(2024)では、代名詞と名詞の順位が逆転しているものの、それ以外の順序は一致しており、先行研究と同様の傾向が確認された。動詞や名詞といった内容語は比較的明瞭に発話される一方で、冠詞や前置詞などの機能語は弱く発話される傾向がある。このため、聴解問題形式のテストにおいて、図2に示すような結果が得られたと考えられる。なお、開発したテストの信頼性を確認するために、Chronbachの α 係数を求めた。 α 係数は.84であり、尺度の内的一貫性が高いことが示された。全体の ω 係数は、.85であり全体の信頼性が高いことが示された。

次に、本研究のテストと被験者の大学入試共通テストの相関に関して報告する。標準化する前のスコアにおける相関係数は $r = .54$ ($p < 0.01$, 95%CI[0.47, 0.60])となり、中程度の相関があることがわかる。図3は双方のテストを標準化した後の散布図となる。今回の実験では、被験者の英語習熟度が比較的均一であったために中程度の相関性ということであったと思われる。今後の展開としては、様々な習熟度の被験者に対してテストを実施して相関性を確認することが挙げられる。なお、本研究のテストと大学入試共通テストのリーディングパートの相関係数は $r = .39$ ($p < 0.01$, 95%CI[0.31, 0.46])で、リスニングパートとの相関係数は $r = .53$ ($p < 0.01$, 95%CI[0.46, 0.59])となり、本研究のテストは大学入試共通テストのリスニング試験のスコアとの相関の方がリーディングパートより大きくなった。これは本研究のテストが聴解問題形式であることから、実験前から予想されており、先行研究の結果とも一致する。本研究のテストと大学入試共通テストのリーディングパートおよびリスニングパートの散布図をそれぞれ図4と5に示す。

5つの品詞をコントロールして穴埋め問題としてテストを作成しているため、それぞれの品詞が大学入試共通テストのスコアに対してどれだけ説明力を持つかを調べるために、5つの品詞を独立変数として、大学入試共通テストを目的変数とした重回帰分析を行い、結果を表1にまとめた。多重共線性に関しては、VIFがすべての変数において2以下であるため、問題にならないレベルであると言える。構築したモデルの決定係数 R^2 は.3015であり、中程度の説明力を持っていると言える。調整済み R^2 も.2939と近似しており、モデルの適合度が確認できる。モデル全体の有意性を評価するために用いたF検定に関しては、F値が40.05、対応するp値が<0.01であることから、少なくとも一つの独立変数が目的変数に対して有意な影響を与えていることがわかる。残差の分析については、最小

標準化されたQ1テストと大学入学共通テストのスコアの相関

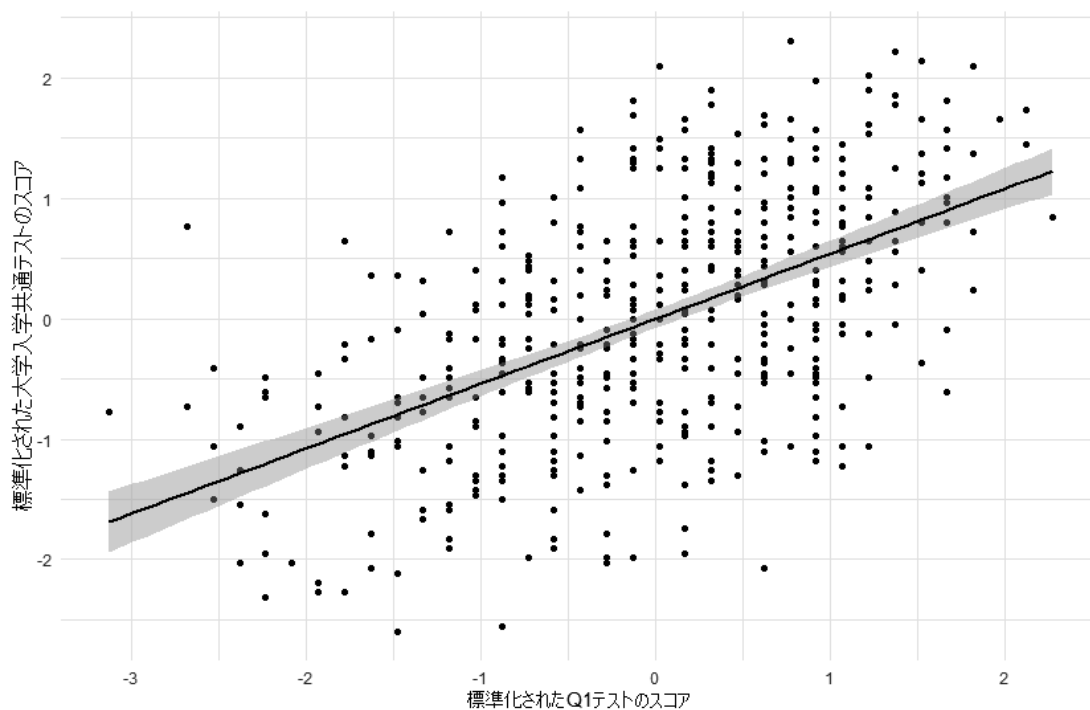


図3 本研究のテストと大学入試共通テストの相関

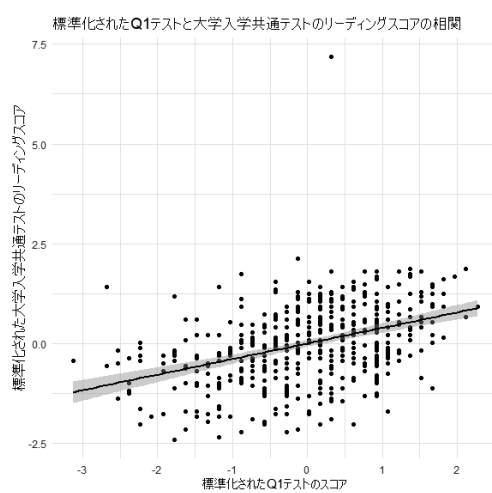


図4 本研究のテスト(Q1)と大学入試共通テストリーディングパートとの相関

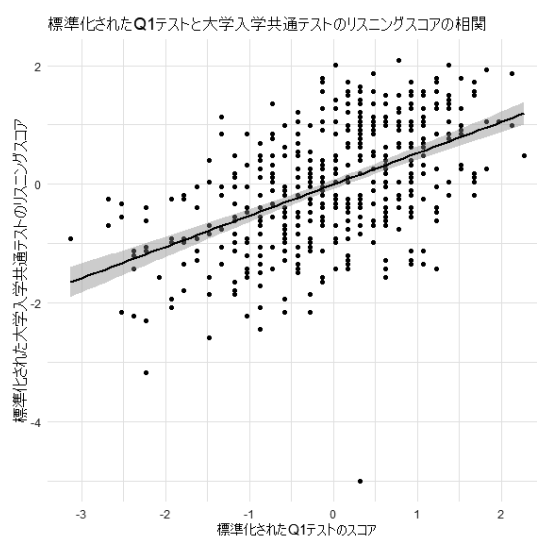


図5 本研究のテスト(Q1)と大学入試共通テストリスニングパートとの相関

値が-59.15、最大値が 57.21、中央値は 0.20 で第一四分位数が-14.623、第三四分位数が 13.87 となっており、残差の分布はおおむね対称であり、モデルの適合度は良好と言える。次に、各々の品詞について見てみると、前置詞、動詞、代名詞の 3 品詞において統計的に有意な影響を与えていることがわかった。名詞に関しては、 $p = .054$ で境界的に有意であった。特に動詞と代名詞は強い正の影響を持つことが確認された。前置詞や代名詞などの機能語において、有意な影響が観察されることは先行研究と同じ結果であったが（陳 & 牧, 2023）、動詞においても強い正の影響があることは、その原因も含めて今後さらに検証をすすめる予定である。

4 おわりに

本研究では、「金言版最小英語テスト」に基づき、英語習熟度を測定するための聴解問題形式のアセスメントテストを開発した。開発したテストの有効性を検証するため、469 名の被験者を対象にテストを実施し、その結果を分析した。テストの有効性を確認する一環として、被験者の大学入試共通テストのスコアとの相関性を調査したところ、相関係数は.54 であり、中程度の相関が確認された。さらに、本研究の聴解問題形式のテストにおいて、どの品詞を設問として設定すれば習熟度を予測する精度が高いかを明らかにするため、大学入試共通テストを目的変数とし、前置詞、動詞、名詞、冠詞、代名詞を独立変数とした重回帰分析を実施した。その結果、前置詞、動詞、代名詞の 3 品詞が統計的に有意な正の影響を持つことがわかった。

最後に、今後の研究の方向性について述べる。「最小英語テスト」の開発背景には、受験料や試験時間の確保など、受験者だけでなく試験実施側の負担を最小化する必要性がある。本研究では、最小限のコストで最大限の成果を得るという方向性を維持しつつ、聴解問題形式のテストが学習者のリーディング力やライティング力の測定には限界があることを考慮し、他の形式を取り入れたテスト開発も検討する必要がある。さらに、習熟度を測定するテストに加え、教育カリキュラムの評価を目的とした到達度テストの最適な形式についても、将来的な研究課題として取り組む予定である。

謝辞

本研究は、JSPS 科研費 22K00762 および 21K00760 の助成を受けて実施されました。本研究におけるテスト開発に際し、「最小英語テスト」の利用を快諾してくださった岐阜大学の牧秀樹教授に深く感謝申し上げます。また、被験者として参加してくださった学生の皆様にも心より感謝申し上げます。さらに、テストの採点作業にご協力いただいた鳥取大学の重松恵梨氏およびマイケル・ムラー氏にも、厚くお礼申し上げます。

参考文献

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.

陳韻 雯 & 牧秀樹. 2023. 「MET 60/kMET 得点と TOEFL iBT 得点との相関—初期研究—」『中部地区英語教育学会紀要』 52, pp.1-6.

小林昌博, 滝波稚子, クリストファー ハリス. 2003. 「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」. 『鳥取大学教育支援・国際交流推進機構 大学教育研究年報』 29. pp.89-92.

牧秀樹. 2018. The Minimal English Test (最小英語テスト) 研究. 開拓者. 東京.

牧秀樹. 2020. 金言版 最小英語テスト (kMET) ドリル. 開拓者. 東京.

牧秀樹, 和佐田裕昭, 橋本永貢子. 2003. 「最小英語テスト (MET)：初期研究」. 『英語教育』 53(10). pp.47-50.