

# 文学能力モデルに基づいた評価の有用性の検討

相関を用いた文学読解定期試験結果のベイズ推計による分析

西原貴之

はじめに

授業で文学教材を使用した場合には、学習者の文学読解パフォーマンスの評価が不可欠である。近年、新しい文学能力モデルとして、The Innsbruck Model of Literary Competence (以下、IMLC) が提案され (Atler & Ratheiser, 2019)、外国語での文学読解パフォーマンスの評価とその使用 (学習者の到達度の確認やフィードバックの提示) への活用が期待されている。このモデルがこれらの用途に有益に資するためには、IMLC が構成概念としてしっかりと定義され、整理されたものでなければならない (Wiliam, 2010)。本研究では、実際の授業における期末試験のデータを使用し、IMLC がこれらの用途に有益に資するものであるかどうか検討を行う。

すでに Nishihara (2015) は授業の期末試験として実施した文学テストのスコアをもとに調査を行い、IMLC 内の下位能力 (後述) の区分の適切性を支持する証拠を示している。本論文では、Nishihara (2015) のデータに対して従来の仮説検証的な統計手法に変わる方法として近年着目されているベイズ推計 (統計ソフト JASP のベイズ推計プログラムを使用) に基づいて再分析を行った後、学習者の文学能力と関連しつつも大きな相関が期待されない他のパフォーマンス評価指標 (授業の平常点と TOEIC のリーディング・セクションのスコア) との関係を調査し、教室単位の定期試験で学習者の文学読解パフォーマンスを評価する上で、IMLC に基づいた作問は有益に機能するかどうかの証拠を得ることを目指す。本論文は、以下の 2 点を調査する。

- (a) IMLC の下位要素区分は適切なものであるか (ベイズ推計に基づいた Nishihara (2015) のデータの再分析)
- (b) IMLC に基づいたパフォーマンス評価のスコアは、授業の平常点及び TOEIC のリーディング・セクションのスコアと弁別的なものとなっているかどうか

## IMLC (The Innsbruck Model of Literary Competence)

IMLC (Atler & Ratheiser, 2019) は、文学能力を以下の 5 つの要素から構成されると考えている。(1) 以外は文学読解独自の要素と考えられている。

- (1) Reading competence and general linguistic competence in English (processing texts)
- (2) Emphatic competence: “the readers’ ability to personally relate to the characters in a text, their actions and reactions, emotions, thoughts, motives; their ability to connect with a text on an emotional level such as in ‘How would I behave in a similar situation?’ or ‘I think I would feel miserable under these circumstances.’” (pp. 380–381)
- (3) Aesthetic and stylistic competence: “the ability to experience and appreciate a literary text and the literary principles of construction, i.e. the stylistics, it follows.” (p. 381)
- (4) Cultural and discursive competence: “the learners’ ability to identify and work with the specific cultural freight and framing of a text and the discourses these pertain to.” (p. 380)
- (5) Interpretive competence: “the ability to infer meaning from a work of literature.” (p. 382)

データを再分析する Nishihara (2015) は IMLC とほぼ同一のモデルに基づいた研究である。本論文は、IMLC の用語に基づいて結果を報告する。ただし、Nishihara (2015) で取り扱っていない上記 (4) に関するデータは、本研究でも分析対象から除外する (この項目を取り上げていない理由については、Nishihara (2015) を参照されたい)。

## 調査方法

県立広島大学国際文化学科で開講されている「英語VIc」という選択科目を履修している、日本人大学生英語学習者 37 名の期末試験スコアをデータとして使用した。参加者の英語力は CEFR で言うところの A2 から B2 レベルと考えられる。実際、19 名が同時期に TOEIC-IP を受験している。その結果は、582.1 点を平均 (標準偏差 112.1) として、最高得点は 835 点、最低点は 400 点であった。調査参加者、テキスト、授業の内容の詳細は Nishihara (2015) を参照されたい。

本研究では、授業内で扱った Charlotte Gilman の “The Yellow Wallpaper” を元にした設問のスコアをデータとして使用する。設問の文言や形式は、この作品を授業で扱った際のものとは異なるものとした。設問は IMLC に基づき作成した (cultural and discursive competence に関するものは除く)。ただし、emphatic competence に関する設問については、学習者のパフォーマンスの優劣を評価することができなかったため、今回は分析から

除外する。また、これらに加えて文法に関する設問も含めた。結果として、IMLCの(1)、(3)、(5)に関わる設問と文法に関わる設問の4タイプのスコアのみを分析に用いる(以下では、それぞれ「RC & GL」(reading competence and general linguistic competence)、「AS」(aesthetic and stylistic competence)、「I」(interpretive competence)、「G」(grammar)と表記する)。テスト全体の構成、出題した文章、設問の詳細については、Nishihara (2015)と西原 (2022)を参照されたい。なお、設問の指示文は授業時と合わせて英語としたが、指示文の理解で学習者がつまづかないように、極力シンプルな英語で作問した。また、学習者の理解を正確に確認する目的で、解答言語は彼らの母語である日本語を指定した。採点は授業実践者である本研究の著者が1人で行った。

本研究では、調査参加者の(i)本科目の期末試験のスコア、(ii)本科目の平常点(後述、以下「平常点」)、(iii)TOEIC-IPのリーディング・セクションのスコア(以下、「TOEIC-R」)、を分析に用いる。(i)と(ii)のデータについては、調査参加者全員のデータを使用する。(iii)については、授業期間中にTOEIC-IPを受験した19名分のデータを使用する。これらのデータの研究目的での使用は、すべての調査参加者から書面にて許可を得た。

なお、「英語Vlc」の最終授業成績は、出席10%、リーディング課題提出10%(TOEICの設問形式に準じた課題の提出点とその課題のスコア)、Eラーニングの進捗状況10%(TOEICに準じた内容で、どの程度の時間取り組んだかを評価)、小テスト20%(前回の授業で扱った作品の語彙と文法の復習)、期末試験50%、という割合で算出した。分析で用いる(ii)「平常点」とは、受講者の最終成績から期末試験全体のスコアを除いた値である。このデータは、含まれる要素から判断して、学習者の勤勉さと一般的な英語力の指標として機能すると考えられる。

本研究では、(ii)及び(iii)と(i)の各設問タイプの相関関係を見ることで、英語文学読解独自の側面(「AS」と「I」)として測定した設問のスコアが学習者の勤勉さや一般的な英語力の指標となっていないかを確認する。

## 結果

### ● 各設問タイプのスコアの相関分析

本研究の研究課題(a)については、テストで収集した4タイプの設問の相関分析を通して検討を行う。各設問タイプの記述統計は紙面の都合上ここでは割愛する。詳細は西原 (2022)を参照されたい。

本研究では、統計ソフトJASPのベイズ推計プログラムを使用して相関係数及びベイズファクター $BF_{10}$ を算出した。本研究での相関分析はいずれも母相関係数 $\rho$ として、帰無仮説を「 $H_0: \rho=0$ 」、対立仮説を「 $H_1: \rho \neq 0$ 」とした。 $\rho$ の事前分布として、JASPのデフォルトであるベータ分布 $B(1, 1)$ を用いた。各設問タイプ間の相関係数は表2の通りである。表中の( )内はベイズファクター $BF_{10}$ 、[ ]内は相関係数の95%確信区間、を指す。

表1. 各設問タイプのスコアの相関係数

	RC & GL	AS	I
G	.54 (61.3) [.25, .71]	.20 (0.4) [-.13, .48]	.14 (0.3) [-.19, .43]
RC & GL		.23 (0.5) [-.10, .50]	.32 (1.3) [-.003, .57]
AS			.11 (0.3) [-.21, .41]

「G」と「RC & GL」の間で $r=.54$ という強い正の相関( $BF_{10}=61.3$ )が確認されたのに比べて、これらの変数と「AS」、「I」との相関は小さく(最も大きなものでも「RC & GL」と「I」の間に $r=.32$  ( $BF_{10}=1.3$ )の相関係数が確認されたのにとどまった)、かつ、「AS」と「I」の相関も小さかった( $r=.11$ ,  $BF_{10}=0.3$ )。このことから、「AS」と「I」は、文学能力の異なる側面を測定し、かつそれらは「G」や「RC & GL」とも異なる読解の側面(文学読解独自の側面)を扱っていることを支持する結果が示された。この結果は、従来の相関分析に基づいたNishihara (2015)と同じ結果であり、IMLCの下位要素の区分の適切さを支持する結果と言える。

### ● 各設問タイプのスコアと学習者の平常点の相関分析

「平常点」と各設問タイプの相関係数は表2の通りである(記述統計については西原 (2022)を参照)。

表2. 各設問タイプのスコアと「平常点」の相関係数

	G	RC & GL	AS	I
平常点	.46 (11.2) [.15, .67]	.60 (323.8) [.32, .76]	.22 (0.5) [-.11, .49]	.12 (0.3) [-.21, .41]

すでに述べたように、「平常点」には一般的な英語力が関係しており、理論上は「G」と「RC & GL」との高い相関及び「AS」と「I」という文学読解独自の要素との低い相関が予想される。今回の分析結果は、この予想を支持するものであった。「G」と「RC & GL」に比べて「AS」と「I」とは相関が低くなっていた。具体的には、相関係数が「G」とは $r=.46$  ( $BF_{10}=11.2$ )、「RC & GL」とは $r=.60$  ( $BF_{10}=323.8$ )であったのに対して、「AS」とは $r=.22$  ( $BF_{10}=0.5$ )、「I」とは $r=.12$  ( $BF_{10}=0.3$ )となっていた。「AS」と「I」は文学読解独自の指標として機能し

たと言えよう。ここでも IMLC の構成の適切さを支持する結果が得られた。

● 各設問タイプの成績と TOEIC のリーディング・セクションのスコアの相関分析

学期中の TOEIC を受験した 19 名を対象を絞り、各設問タイプの成績と「TOEIC-R」のスコアの関係について検討した。記述統計については紙面の都合上ここでは割愛する。詳細は西原 (2022) を参照されたい。表 3 に相関分析の結果を示す。

表 3. 各設問タイプのスコアと「TOEIC-R」の相関係数

	RC & GL	AS	I	TOEIC-R
G	.64 (16.1) [.22, .83]	.13 (0.3) [-.32, .51]	.13 (0.3) [-.32, .51]	.62 (11.5) [.20, .82]
RC & GL		.09 (0.3) [-.36, .49]	.25 (0.5) [-.21, .60]	.37 (0.9) [-.11, .67]
AS			-.02 (0.3) [-.44, .41]	-.08 (0.3) [-.48, .36]
I				-.06 (0.3) [-.47, .37]

データの数が減ったことにより多少の数値の変動はあるものの、各設問タイプ間の相関関係に関しては、ほぼ表 1 と同様な結果となった (表 3 の分析で用いたデータは表 1 で用いたデータの一部であるため、このことは驚くことではないが)。「G」と「RC & GL」に関しては、 $r = .64$  という強い正の相関 ( $BF_{10} = 16.1$ ) が確認されたのに対して、これらの変数と「AS」と「I」との相関は小さく (最も大きなものでも「RC & GL」と「I」の間に  $r = .25$  ( $BF_{10} = 0.5$ ) の相関係数が確認されたのにとどまった)、かつ、「AS」と「I」の相関も  $r = -.02$  ( $BF_{10} = 0.3$ ) と小さかった。したがって、IMLC の下位要素の構成については再度支持する結果が得られた。

「TOEIC-R」のスコアに関しては、表 2 で見た「平常点」同様に、理論上は「G」と「RC & GL」との高い相関及び「AS」と「I」という文学読解独自の要素との低い相関が予想された。表 3 を確認してみると、「G」とは  $r = .62$  と大きな正の相関 ( $BF_{10} = 11.5$ ) が確認され、この予想を支持する結果が得られた。しかしながら「RC & GL」とは、 $r = .37$  ( $BF_{10} = 0.9$ ) という結果であり、予想に反して低い相関となった。このことについては、TOEIC の中で読む文章と文学教材ではジャンルが大きく異なっており、同じ字義的内容理解であってもその内実は大きく異なっていることが原因かもしれない。このことについては今後さらなる調査が必要である。また、「TOEIC-R」は「AS」と「I」とはほぼ無相関であった (それぞれ  $r = -.08$  ( $BF_{10} = 0.3$ ) と  $r = -.06$  ( $BF_{10} = 0.3$ ))。このことについては予想通りの結果であり、IMLC の適切性を支持する結果と言える。

結論

本研究では、(a) IMLC の下位要素の構成の適切性及び (b) 文学能力との関連はありつつも強くは相関しないと予想される指標との弁別性、の 2 点に関して、支持する結果が得られた。したがって、このモデルに基づく作問について、学習者の文学読解パフォーマンスの評価とその使用 (授業到達度の確認や学習者へのフィードバックなど) への有用性を支持する証拠を得た。しかしながら、今回の調査だけでは、個々の教室を超えて IMLC の有用性を一般化することはできない。今後も様々な環境の中で異なる文学教材、異なる設問形式を多様な学習者に対して与える中で、このモデルに基づく作問の有用性の証拠を集めていく必要がある。さらに、本研究では扱うことができなかった emphatic competence 及び cultural and discursive competence の取り扱いについても今後検討していくことが求められる。

引用文献

Atler, C., & Ratheiser, U. (2019). A new model of literary competences and the revised CEFR descriptors. *ELT Journal*, 73(4), 377–386. <https://doi.org/10.1093/elt/cez024>

Nishihara, T. (2015). Achievement tests for literary reading in general EFL reading courses. In M. Teranishi, Y. Saito, & K. Wales (Eds.), *Literature and language learning in the EFL classroom* (pp. 115–130). Palgrave Macmillan. [https://doi.org/10.1057/9781137443663\\_8](https://doi.org/10.1057/9781137443663_8)

西原貴之 (2022). 『文学能力モデルに基づいた評価の有用性の検討：相関を用いた文学読解定期試験結果のベイズ推計による分析』。第 10 回日本国際教養学会発表資料。 <https://home.hiroshima-u.ac.jp/ntakayk/presentation32-1.pdf>

Wiliam, D. (2010). What counts as evidence of educational achievement?: The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34(1), 254–284. <https://doi.org/10.3102/0091732X09351544>