



# HOKKAIDO UNIVERSITY

Title	自己報告に基づく学習成果の間接評価における妥当性の検証 : 英語系学科所属学生の英語運用能力に関する事例から
Author(s)	浅賀, 圭祐; Asaka, Keisuke; 松本, 博文 他
Citation	高等教育ジャーナル : 高等教育と生涯学習, 29, 69-79
Issue Date	2022-03
DOI	<a href="https://doi.org/10.14943/J.HighEdu.29.69">https://doi.org/10.14943/J.HighEdu.29.69</a>
Doc URL	<a href="https://hdl.handle.net/2115/84351">https://hdl.handle.net/2115/84351</a>
Type	departmental bulletin paper
File Information	HighEdu_29_069.pdf



# Validity of Indirect Assessment of Learning Outcomes Based on Self-Reports: A Case Study of English Language Proficiency in an English Language Department

Keisuke Asaka,<sup>1)\*</sup> Hirobumi Matsumoto<sup>2)</sup> and Shigeyuki Hidai<sup>2)</sup>

1) Higher Education Development Center, Research Institute, Tamagawa University

2) Department of English Language Education, College of Humanities, Tamagawa University

## 自己報告に基づく学習成果の間接評価における妥当性の検証 —英語系学科所属学生の英語運用能力に関する事例から—

浅賀 圭祐<sup>1)\*\*</sup>, 松本 博文<sup>2)</sup>, 日躰 滋之<sup>2)</sup>

1) 玉川大学学術研究所高等教育開発センター

2) 玉川大学文学部英語教育学科

*Abstract* — In this study, we investigated the validity of indirect assessments of English language proficiency in an English language department by considering consistency between student self-reports and IELTS scores. The former data were obtained from a survey of freshmen conducted by the Institutional Research Consortium of Japanese Universities. Regarding English language proficiency, we focused on five categories—Listening, Reading, Spoken Interaction, Spoken Production and Writing. The consistency was evaluated by the bias in student self-reports by comparison with converted values of the IELTS scores in terms of the Common European Framework of Reference for Languages (CEFR), as well as by correlation. As a result, all five correlations were found to be statistically significant (medium correlation for all categories except Writing). Furthermore, it was shown that receptive and productive skills had almost no bias and non-negligible bias, respectively. In conclusion, this study partially supported the validity of indirect assessment of learning-outcomes based on self-reports.

(Accepted on 20 January 2022)

### 1. 研究の目的と背景

大学教育において内部質保証システムを構築するためには、学修成果を把握・評価し、可視化する必

要がある。その手法は直接評価と間接評価に大別することができる。松下 (2016) によれば、直接評価とは「学習者の知識や能力の表出を通じて—『何を知り何ができるか』を学習者自身にやってみさせる

\*) Correspondence: Higher Education Development Center, Research Institute, Tamagawa University, Machida 194-0041, Japan

E-mail: asaka@lab.tamagawa.ac.jp

\*\*\*) 連絡先: 194-8610 東京都町田市玉川学園 6-1-1 玉川大学学術研究所高等教育開発センター

ことで行われる評価」であり、間接評価とは「学習成果についての学習者の自己報告を通じて『何を知り何ができると思っているか』を学習者自身に答えさせることによって行われる評価」である。具体的には、GPA や個別または標準テスト、レポート、ポートフォリオなどが直接評価に分類され、学生調査や授業評価などが間接評価に分類される。これらの有効活用に基づく学修成果志向の教育の推進は、昨今の大学教育改革の潮流の一つと言える。

間接評価の活用方法には複数の考え方がある。1 つには学習成果の評価に至るまでのプロセスの評価に間接評価を用いるというものがある。山田(2012)は、「『直接評価』と『間接評価』はどちらが優れているのかという二項対立で捉えるのではなく、組み合わせることで教育評価の精緻化が進展できる」とした上で、学習成果の評価には直接評価、そしてそれに至るまでのプロセスの評価には間接評価が適切であると述べ、プロセス評価としての間接評価の重要性を指摘している。本人にしか知り得ない学修行動や知識・能力変化の実感、満足度などは自己報告に基づき把握するしかなく、このような直接・間接評価の組み合わせによる学修成果の可視化は理想的と言えるだろう。ただし、この方法では評価すべき学習成果を直接評価により十分に捉えられることが前提となるが、学士力などを構成する「高次の統合的な能力」(斎藤 2019)を直接的に評価することは、近年精力的に研究され発展が見られるものの容易ではない。

一方で、間接評価を学習成果の評価に用いる試みもある。学生調査では修得した知識・能力の程度を尋ねる項目が設定されていることが少なくない。このような項目は先のプロセス評価の考え方では学習過程における知識獲得や能力変化の実感を尋ねるものという位置づけになるが、これを学習成果そのものと捉えることも一つの活用方法として考えられる。ただし、一般的には、この方法は機関・教育プログラムレベルの集合的・総括的な評価として位置付けられるものであり、必ずしも個々の学生の学習成果を間接評価により捉えようとするものではない。それでも、この方法は内部質保証の推進などの目的には適うものであり、前述のように評価すべき学習成果を直接評価により捉えることが容易ではな

い場合には、調査実施の簡易さも考慮し現実的な方法と言える。

しかし、特に後者のように学習成果の集合的な評価に用いる場合には、学生の自己認識に頼る間接評価の信頼性・妥当性に疑念が生じるのは避けられず、その検証が課題となる。本研究はそのうちの妥当性に着目し、直接評価と間接評価の整合性、つまり学生の自己報告の正確性の検証を試みたものである。

一般に、真の値を想定し得る対象の測定・評価には誤差が含まれる。誤差にはランダム誤差(統計誤差)とバイアス(系統誤差)の2種類があり、どちらも本研究のテーマである妥当性を損なう要因となる。バイアスが含まれない場合、データの平均値はサンプルサイズの増加に伴い真の値に収束するが、バイアスが含まれる場合にはそうはならない。従って、集合的な評価への活用を想定した場合には、特にバイアスの有無が重要となる。間接評価に含まれるバイアスの例としては、印象操作や自我防衛を動機とする「社会的望ましきのバイアス」や、ダニング=クルーガー効果(Kruger and Dunning 1999)として知られる認知バイアスが挙げられる(Cole and Gonyea 2010; 松下 2017)。

米国では古くから間接評価の妥当性について議論されており、研究成果の一定の蓄積がある。その内容は、自己報告のGPAと実際のGPAとの高い相関を示す報告(Kuncel et al. 2005)から学習成果の自己報告の妥当性に懐疑的な論調(Porter 2013)まで多岐に渡る。Pike(2011)はそれまでの間接評価の(基準関連)妥当性に関する研究を総括し、自己報告とその基準となる指標に正の相関があることが示されているがその程度は一般にあまり大きくない、と述べている。また、松下ら(2014)は米国における直接評価と間接評価の相関に関する研究を概観し、「その結果には多様性があり」、「直接指標と間接指標の結果の相関の有無や強さをめぐっては、いまだ諸説ある状態」としている。しかしながら、このように様々な見解がある中で、達成度の低い学生ほど自己報告が不正確なことが多くその場合には過大報告の傾向にあることが多くの研究から支持されており、主要な知見の一つと言えるだろう(Kuncel et al. 2005; Cole and Gonyea 2010)。

国内ではまず、2012年度採択の文部科学省大学間

連携共同教育推進事業「教学評価体制（IR ネットワーク）による学士課程教育の質保証」において、英語力調査に関する報告がある。そこでは、一大学におけるデータから、「リスニング」と「リーディング」に対する直接評価と間接評価の相関係数としてそれぞれ 0.360 と 0.380 という値が示されている<sup>1</sup>（北海道大学高等教育推進機構 2014）。また、大学教育学会の課題研究「学士課程教育における共通教育の質保証」（2013～2015 年度）において間接評価の妥当性に関する議論がある。その中で山田(2016)は直接評価と間接評価の相関関係を調べ、読解力を問う英語のクイズの正答数と英語のリーディングのレベル別の自信の程度に大まかな正の相関が見られることを報告している。ただし有意な相関が報告されたのは複数ある項目の中で1つだけで、その相関係数も 0.164 とそれほど大きな値ではない。そして、斎藤ら(2016)は初年次教育科目におけるルーブリックを用いたパフォーマンス評価から、教員による直接評価と学生の自己報告の相関分析において、ほぼ無相関～弱い相関という結果が得られたことを報告している。高等教育における学修成果の可視化を目的とし間接評価の妥当性について検証した国内での目立った研究は、筆者の知る限り上記に留まる<sup>1</sup>。従って、国内では当該テーマに関する研究の蓄積がほとんどない状態と言える。

以上より、この問題については米国では議論と知見の蓄積がある程度進んでいるが、国内での検証はほとんど行われていない状況にあるため、その検証には意義がある。そこで本研究では、英語運用能力に着目し、ほぼ同時期に行われた学生調査と外部試験の結果を用い、特定の一学科におけるデータを事例とし、間接評価の妥当性を検討・検証した。英語運用能力に着目した理由は、端的に言うとするに標準化された試験に基づく直接評価データの入手が比較的容易だからである。しかしながら、もちろん英語運用能力についての学習成果の可視化も重要であるが、本研究の先に見据えているのは直接評価が容易ではない他の能力・技能の間接評価についての研究・開発である。従って、本研究は特定の一学科における英語運用能力においてどのようなバイアスが見られるのかを特に明らかにし、学習成果の間接評価の妥当性に対する知見を得ることを目的とする。

後述するように、本研究では英語運用能力を構成する各技能に対し妥当性の検証を行うと共に、ヨーロッパ言語共通参照枠（Common European Framework of Reference for Languages：CEFR）との対応関係を用いることで、相関分析に留まらずバイアスの有無や程度の分析を行っている。従って、前述の英語力調査報告や山田(2016)と比較し、より詳細な検証をリスニングやリーディング以外の技能も含め行っている点に本研究の新規性がある。

## 2. 研究の方法

### 2.1 使用した指標

本研究では間接評価として大学 IR コンソーシアムの一年生調査に含まれる英語運用能力に関する調査項目を用いた。その妥当性を検証するために基準とする指標には、直接評価である IELTS（International English Language Testing System）を用いた。

一般社団法人大学 IR コンソーシアム（以下、コンソーシアム）は相互評価に基づく大学教育の内部質保証の推進を目的として設立され、学生生活や学修状況について尋ねる学生調査（一年生調査および上級生調査）のマネジメント等を行っており、2021 年 7 月現在、国内の 60 大学（国立 8 大学、公立 8 大学、私立 44 大学）が加盟している（一般社団法人大学 IR コンソーシアム 2021）。その学生調査の間 8<sup>2</sup> の 5 項目は、英語運用能力に関するルーブリック型の調査項目となっている（北海道大学高等教育推進機構 2017）。これらの項目は CEFR に準拠しており、基礎段階の言語使用者にあたる A1 と A2、独立した言語使用者にあたる B1 と B2、熟達した言語使用者にあたる C1 と C2 といった 6 段階の言語運用能力に対応した選択肢となっている（順に難易度が上がる）。ただし、会話力のみ、A1 および A2 がそれぞれ A1.1 と A1.2、A2.1 と A2.2 に細分化され、8 段階の選択肢となっている。表 1 にその詳細を示す<sup>3</sup>。

IELTS はブリティッシュ・カウンシル、IDP：IELTS オーストラリア、ケンブリッジ大学英語検定機構の 3 団体により共同運営されている国際的な英語運用能力測定試験である。世界 140 カ国、11,000

表 1. 大学 IR コンソーシアム一年生調査問 8 の選択肢

CEFR との対応	聞く力	読む力	会話力	表現力	書く力
A1	はっきりと、ゆっくり話してもらえれば、聞き慣れた語やごく基本的な表現を聞き取れる。	掲示やポスター、カタログなどの中をよく知っている名前、単語、単純な文を理解できる。	A1.1 決まった言い回しを使って自己紹介をしたり、相手の趣味を尋ねたりできる。 A1.2 家族や身の回りのことについて、簡単な質問なら聞いたり答えたりできる。	住んでいるところ、また、知っている人たちについて、簡単な語句や文を使って表現できる。	お祝いのメッセージなど、短い簡単な葉書を書くことができる。
A2	最も頻繁に使われる語彙や表現を理解することができる。	ごく短い簡単な文章や、簡単に短い個人的な手紙は理解できる。	A2.1 短い社交的なやり取りができる。ひとりで会話を続けにくい、相手の助け舟で、身近な話題について話し続けられる。 A2.2 準備をすれば、日常的でなじみのある話題について、簡単な言葉を使ってまとまりのある会話ができる。	家族、周囲の人々、居住条件を簡単な言葉で説明できる。	簡単に短いメモやメッセージ、短い個人的な手紙なら書くことができる。
B1	身近な話題について、明瞭で標準的な話し方の会話なら要点を理解することができる。	日常語や、自分の知っている分野の文章なら理解できる。簡単に個人的な手紙を理解できる。	日常生活に直接関係のあることや個人的な関心について、準備なしで会話ができる。	簡単な方法で語句をつないで、自分の経験や出来事、夢や希望、目標を語るすることができる。	身近で個人的に関心のある話題を書くことができる。個人的な手紙で経験や印象を書くことができる。
B2	テレビのニュースや時事問題、標準語の映画ならほとんど理解できる。	現代の問題についての記事や報告が読める。現代文学の散文は読める。	身近な話題の議論に積極的に参加し、自分の意見を説明できる。	興味関心のある話題について、明瞭で詳細な説明ができる。	興味関心のある話題について、明瞭で詳細な説明文を書くことができる。
C1	特別な努力なしにテレビ番組や映画を理解できる。	複雑な文章を理解できる。自分の関連外の分野の専門の記事も理解できる。	社会上、目的・場面に合った言葉遣いができ、自分の考えや意見を正確に表現できる。	複雑な話題を、一定の観点を展開しながら、適切な結論でまとめあげることができる。	手紙やエッセイ、レポートで複雑な主題を扱うことができる。
C2	母語話者の速いスピードで話されても、どんな種類の話し言葉も難なく理解できる。	抽象的で複雑な文章など、あらゆる形式で書かれた言葉を容易に読むことができる。	いかなる会話や議論でも努力しないで加わることができる。	論理的な会話で聞き手に重要点を把握させ、記憶にとどめさせることができる。	論理的に事情を説明し、複雑な内容の手紙、レポート、記事を書くことができる。

以上の機関で認定されており (British Council 2021)、英語圏への留学時に求められる英語運用能力の証明として世界で広く活用されている。日本国内ではブリティッシュ・カウンシルと日本英語検定協会により共同運営されており、大学入試や英語教員採用試験などでも加点の際の参照基準といった形で活用され、ここ 10 年ほどで受験者数が伸びている (日本英語検定協会 2018)。

試験はスキルごとに Listening, Reading, Writing, Speaking と分かれており、Speaking の試験を対面で行うことに特徴がある。英語運用能力はスキルごとおよび総合ともに 1.0 から 9.0 まで 0.5 刻みのバンド・スコアとして表され、数値が高いほど英語運用能力が高いことを示す。1.0 および 1.5 は非ユー

ザー、2.0 および 2.5 は一時的なユーザー、3.0 および 3.5 は非常に限定的なユーザー、4.0 および 4.5 は限定的ユーザー、5.0 および 5.5 は中程度のユーザー、6.0 および 6.5 は有能なユーザー、7.0 および 7.5 は優秀なユーザー、8.0 および 8.5 は非常に優秀なユーザー、9.0 はエキスパートユーザーとされる。日本の受験者の平均スコアは総合スコア 5.8 で (日本英語検定協会 2018)、英語圏の大学への入学には一般的に総合スコアで 6.5 前後を求められることが多い。

本研究では一年生調査問 8 の各項目と IELTS の各技能を表 2 に示す通り対応させた (Council of Europe 2001)。会話力と表現力はどちらも Speaking と対応し、検証には両方で同一の Speaking スコアを

表 2. 一年生調査問 8 (5 項目) と IELTS (4 技能) の対応

一年生調査問 8	IELTS
聞く力	Listening
読む力	Reading
会話力	Speaking
表現力	Speaking
書く力	Writing

用いた。本研究ではこれらの各項目について検証を行った。以下では、各項目のペアをそれぞれ「聞く力×Listening」のように表記する。なお、一年生調査問 8 の各項目では①入学時と②現在（調査参加時現在という意味）のそれぞれの能力を回答することになっているが、本研究では後者を用いた。

## 2.2 使用データ

本研究で用いたデータは、私立 A 大学 B 学部英語系学科 1 年生の、2010 年代の連続する 5 年分の、一年生調査問 8 の回答データおよび IELTS の 4 技能 (Listening, Reading, Speaking, Writing) のスコアデータである (N=207)。当該学科は英語運用能力を重点的に学ぶ学科であり、一般的な学生と比べ英語運用能力への関心が高い学生が在籍していると考えられる。英語運用能力の水準は、次章の図 1.1, 1.2 に示されている通り、技能により多少の差はあるが概ね 3.5~5.0 に多くのデータが見られ、限定的ユーザーから中程度のユーザーとされる水準となっている。

A 大学では大学 IR コンソーシアムの学生調査を全学的に行っているが、IELTS 試験は基本的に当該学科でしか実施していない。この試験は 2 年次以降に予定されている留学のための準備の一環として 1 年次に実施するものである。そのため、本研究で使用したデータは当該学科 1 年生のものとなっている。

また、各実施年において一年生調査と IELTS はいずれも秋季から冬季にかけて行っており、実施時期の違いは 1, 2 か月程度である。A 大学では一年生調査を Web 上で行っており、問 8 の実際の回答画面では選択肢に A1, A2 などの CEFR との対応を示す表示はなく、選択肢に紐づいたラジオボタンをクリックする形式となっている。そのため、調査回答

時に IELTS のスコアを知っていたとしても、スコアと選択肢の内容との対応を理解することは困難であり、IELTS のスコアの認識と調査回答のタイミングが前後することの影響はないと考えられる<sup>4</sup>。

## 2.3 検証方法

本研究では、一年生調査問 8（以下ではその回答データを自己報告データと呼ぶ）の妥当性を検証するため、データの散布図を確認した上で、①自己報告データと IELTS スコアデータの相関と② IELTS スコアを基準としたときの自己報告の不正確さ（以下ではこの不正確さをずれと呼ぶ）の分布とバイアスを調べた。

①については、ピアソンの積率相関係数とポリコリック相関係数を指標とした。対象となる両変数の相関分析にはピアソンの積率相関係数を用いるのが一般的と思われるが、両変数は厳密には順序尺度であることから、これら 2 つの相関係数を用いた。ポリコリック相関係数とは、両変数が順序尺度であり、かつ背後に想定される確率分布に正規性を仮定できる場合に適した相関係数として知られているものである。両変数はこの条件に当てはまると考えられる。また、ピアソンの積率相関係数に対しては無相関検定を行った。

②におけるずれの評価には、CEFR との対応関係を用いた。前述したように一年生調査問 8 の選択肢には CEFR の A1 から C2 までの基準が使用されている。CEFR と様々な英語運用能力測定試験との対応関係については多様な議論があるが、目安となる対応関係は試験の実施団体等により公表されている。IELTS についても目安となる CEFR との対応関係が示されている。表 3 はその対照表である (British Council et al. n.d.; 文部科学省 2018)。従って、一年生調査問 8 の選択肢と IELTS のスコアとの間に対応を付けることができる。この対応関係から、個々のデータについて、IELTS スコアを正確な習熟度と見なしたときの学生の自己報告のずれを見積もることができる。そして、このずれの分析から、自己報告に含まれるランダム誤差およびバイアスの有無や程度を評価することができる。本研究では CEFR の基準 1 つ分の過大報告を +1 のずれとし、

表 3. CEFR の基準と IELTS スコアの  
対照表

CEFR	IELTS
C2	8.5~9.0
C1	7.0~8.0
B2	5.5~6.5
B1	4.0~5.0
A2	2.5~3.5
A1	1.0~2.0

そのずれの平均をバイアスの指標とした。

①は学生の自己報告の相対的な妥当性の評価に関連し、②は絶対的な妥当性の評価に関連するものと考えられる。いくら相関係数が大きくても全体的にバイアスが掛り過大あるいは過小報告の傾向にあることも考えられるので、相対的・絶対的の両方の妥当性について調べることが重要であろう。なお、相関係数や誤差は、A1 から C2 までの選択肢にそれぞれ 1 から 6 までの値を割り振り計算した。ただし、会話力では A1.1 と A1.2 を A1, A2.1 と A2.2 を A2 とそれぞれ見なした。

### 3. 結果

自己報告データと IELTS スコアデータの各ペアの散布図を図 1.1 および図 1.2 に示す。各ペアの相関係数を表 4 に示す。各項目の自己報告のずれの分布を図 2 に示す。自己報告のずれの平均、標準偏差、そして平均の 95%信頼区間を表 5 に示す。各散布図において中央値の位置を白色の×印で示している。また、表 3 に示した対照表により CEFR と IELTS の基準が対応する点（交点）を長方形の枠で囲っている。つまり、枠内にあるデータ、枠の右側にあるデータ、枠の左側にあるデータがそれぞれ正確な自己報告（ずれ=0）、過大報告（ずれ>0）、過小報告（ずれ<0）を行った学生のデータである。例えば聞く力×Listening の散布図において枠内にあるデータ数は 108 である。従って、全体の中で正確な自己報告を行った学生の割合は  $108 \div 207 = 0.522$  である。この数字が図 2 の聞く力のグラフのずれ 0 の棒の数値に反映されている。

①については、すべてのペアで統計的に有意な、弱～中程度の相関が認められる結果となった。その中で書く力×Writing のみ相関係数が約 0.15 と他の

ペアと比較し目立って弱い相関となった。

②については、項目により傾向が異なる結果となった。他項目との比較において、聞く力のずれの平均は負の向きに小さな値、読む力ではほぼ 0 に近い値、会話力では負の向きに大きな値、そして表現力と書く力では正の向きにやや大きな値となった。

### 4. 考察

①についての結果から、学生の自己報告と IELTS のスコアには弱から中程度の相関が認められた。これは第 1 章で述べた 2014 年の英語力調査報告で示された相関係数と同程度の水準である。従って、本研究の結果は、聞く力と読む力に関しては先行研究と同程度の相関を再現するものであり、会話力と表現力についてはそれらと同程度、書く力についてはそれらに比べやや低い水準にあることを示唆する結果と言える。ただし、あくまで弱～中程度の相関であり、散布図から分かるように個々のデータの散らばり具合は決して小さなものではない。従って、統計的に有意な相関が示されたものの、この結果は、達成度の一般的な指標として学生の自己報告を用いることは正当化できるが試験のスコアに代えて特定の自己報告を用いることは正当化できないとする Pike (1996) の見解を支持するものと考えられる。

②についての結果から、各項目におけるバイアスの有無と程度について検討する。まず聞く力についてはわずかに過小報告のバイアスが見られる。ただし、この程度のバイアスであれば、学習成果の集合的な評価への活用を疑問視するほどのものではないだろう。次に、読む力については信頼区間が 0 を含むことからバイアスの存在は認められないと結論付けることができる。しかし、会話力については明らかな過小報告、表現力と書く力については明らかな過大報告のバイアスが見られる。従って、本研究の結果は、受容的技能（リスニングおよびリーディング）に該当する項目にはバイアスがほとんど見られず、産出的技能（スピーキングおよびライティング）に該当する項目については過大または過少報告のバイアスが見られるということを示すものとなっている。

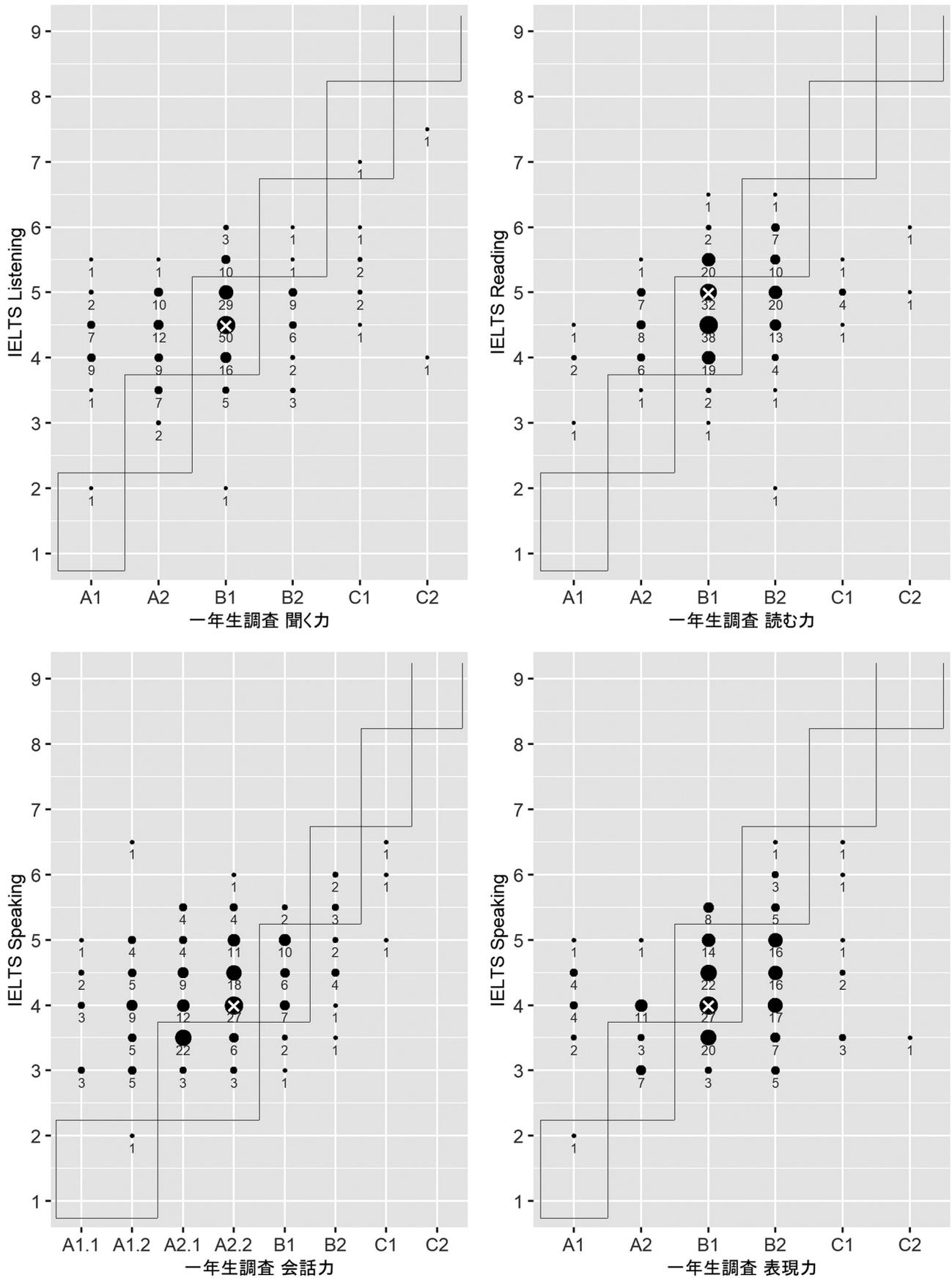


図 1.1. IELTS スコアと一年生調査回答結果の散布図 (N=207, グラフ内の数字は人数, ×印は中央値の位置を示す)

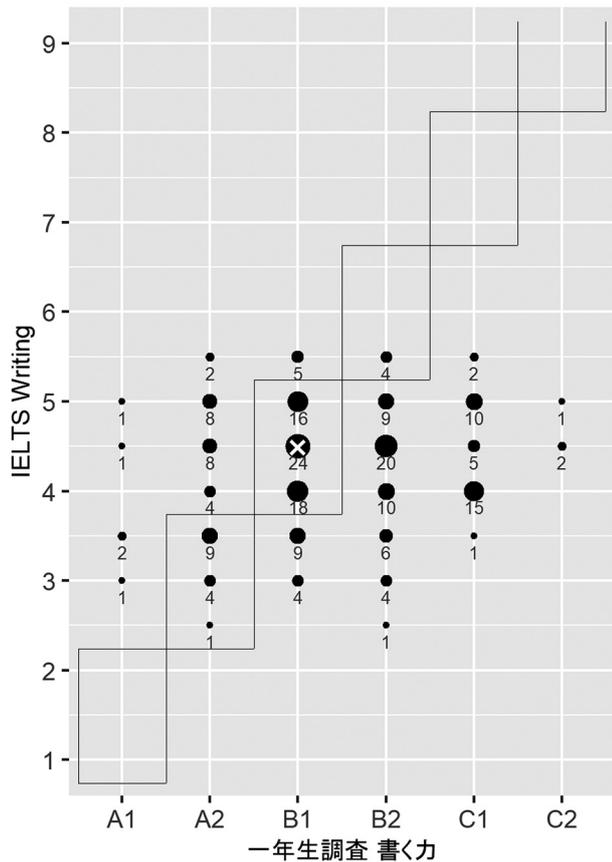


図 1.2. (図 1.1. の続き)

ただし、バイアスの存在がただちに学生の自己報告を学習成果の評価に用いることが不適切であるということを意味する訳ではない。バイアスに規則性・法則性が見いだせるのであれば、自己報告の結果からバイアスの分を調整して間接評価を行うことはできるだろう。

以上より、特に受容的技能の相関の水準とバイアスの程度を踏まえ、本事例に基づく分析結果は学習成果の間接評価の妥当性を部分的に支持するものと結論付けられる。

本研究の結果のみから産出的技能に見られたバイアスの要因について議論を深めることは難しいが、考えられる要因について言及したい。まず、学生の英語運用能力に関する学修履歴の中で、評価を受ける機会や言語活動の経験の少なさがバイアスの要因となった可能性が挙げられる。文部科学省 (2019) によると、パフォーマンステストの実施状況として高等学校で「話すこと」「書くこと」の両方の評価を行っている割合は 4 割に満たない状況であり、産出的技能のパフォーマンステストの実施率の低さが課

表 4. 自己報告データと IELTS スコアデータの相関 (N=207)

	ピアソンの積率相関係数	ポリコリック相関係数
聞く力×Listening	0.355**	0.357
読む力×Reading	0.299**	0.346
会話力×Speaking	0.387**	0.407
表現力×Speaking	0.259**	0.273
書く力×Writing	0.151*	0.154

\* p<0.05, \*\* p<0.01

表 5. 自己報告のずれの平均の推定 (N=207)

項目	平均	標準偏差	平均の 95%信頼区間
聞く力	-0.208	0.945	[-0.341, -0.075]
読む力	0.037	0.847	[-0.087, 0.165]
会話力	-0.720	0.853	[-0.846, -0.593]
表現力	0.372	1.00	[0.235, 0.509]
書く力	0.541	1.07	[0.399, 0.683]

題として指摘されている。そして、指導と評価の一体化という観点から、評価を実施する機会の少なさは授業での学習者の言語活動の少なさにも関連していると考えられる。そのため、産出的技能は評価の機会や言語活動の経験が少なく、それがバイアスにつながった可能性が考えられる。しかしながら、この要因からはなぜ会話力で過少報告となり、表現力と書く力で過大報告となったかは説明できない。

ただし、会話力に見られるバイアスの要因としては、他項目と異なり会話力のみで選択肢の数が多く設定されていたことが影響した可能性に留意する必要があるだろう。つまり、選択肢設計の意図としては A1 と A2 をそれぞれ 2 つに分け細分化したということなのだが、回答者には 8 つの選択肢が等間隔に見え、相対的な位置関係から水準を判断し回答した可能性が考えられる。従って、この項目のバイアスは他の項目とは異質のものであり、他項目と同列に並べ考察することが適切ではない可能性がある。

また、本研究の結果には専門性が関与した可能性も考えられる。本研究は英語系学科における英語運用能力に関する検証である。英語系学科において英語運用能力は言わば専門分野であり、学生は日頃から自身の英語運用能力の反省を意識的に行っていると考えられる。このことが本研究で示された部分的な妥当性につながった可能性が考えられる。しかし、それが自己報告の正確さではなく、むしろ過大あるいは過少報告のバイアスにつながった可能性も

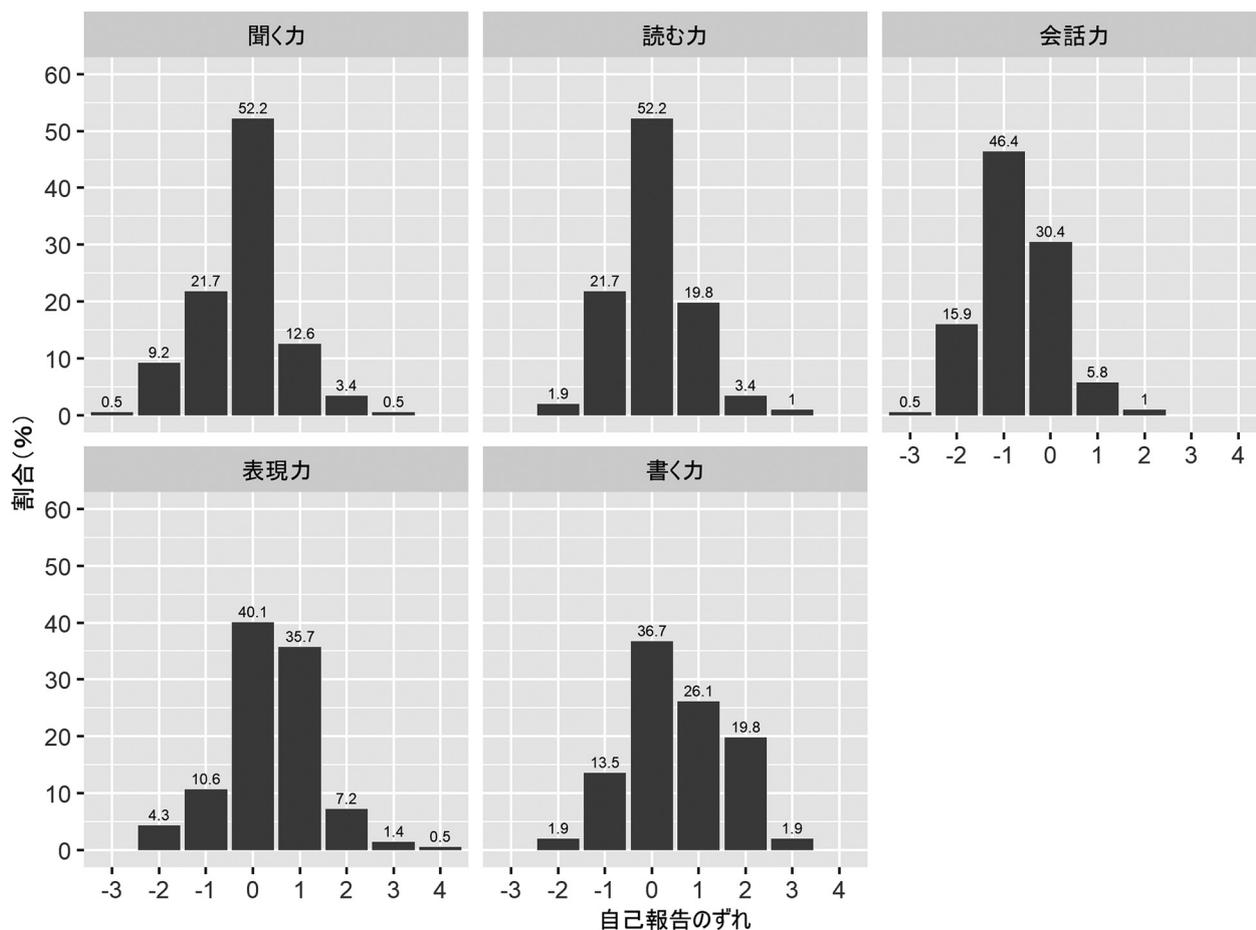


図2. 自己報告のずれの分布 (N=207)

否定できるものではない。また、当該学科では、1年次に英語運用能力の向上を主な目的とする科目が他学科に比べそれほど多い訳ではないものの、2年次から3年次にかけ約9か月間の留学プログラムが予定されており、特に産出的技能の向上が入学時から強く動機づけられている。そのような能力向上への意識が実際の能力向上に先行する形でバイアスの要因になったとしても不思議はないだろう。しかしながら、本研究の結果のみからこのような専門性に起因するバイアスについて厳密に議論することはできないため、非英語系学科所属学生を対象とした事例との比較を踏まえ検証する必要がある。

最後に、他の能力・技能を含めた学習成果の間接評価の妥当性に対する本事例からの示唆について考察する。先に述べたように、学生の学修履歴において評価を受ける機会や活動経験の少なさがバイアスの要因となった可能性がある。この文脈での評価とは直接評価を意味する。また、評価を受ける機会の差は、評価を実施することの難易の差と考えられる。

そのため、もし評価を受ける機会の少なさがバイアスの要因であるならば、学習成果の間接評価の意義は限定的と言わざるを得ない。直接評価が容易ならば間接評価により学習成果を評価する必要はあまりないからである。

一方で、活動経験の少なさがバイアスの要因であるならば、経験を増やすことでバイアスを減らすことができ、間接評価の妥当性はコントロール可能となることから、間接評価の意義は大きいと言えるだろう。

また、受容的か産出的かということ自体がバイアスの有無や程度に関連している可能性もある。その場合は、評価対象の能力・技能についての調査項目が受容的か産出的かということでバイアスのある程度見積もることができ、学習成果の間接評価を行う上で有用な知見となるだろう。

## 5. まとめと展望

本研究では学習成果の間接評価における妥当性を検証するため、特定の英語系学科における英語運用能力を事例とし、学生調査の英語運用能力に関する5項目の回答と、直接評価の一種であるIELTSの4技能スコアとの整合性を分析した。具体的には、相関分析および自己報告に含まれるバイアスの評価を行った。バイアスの評価はCEFRとの対応関係に基づき行った。

その結果、相関についてはすべての項目で統計的に有意な相関が認められ、書く力を除く4項目で中程度の相関が示された。そしてバイアスには項目により異なる傾向が見られた。受容的スキルにはほとんどバイアスが見られなかった一方で産出的スキルにはバイアスが見られ、会話力には過小報告、表現力と書く力には過大報告のバイアスが見られた。このことから、本事例に基づく分析結果は学習成果の間接評価の妥当性を部分的に支持するものと結論付けた。そして、バイアスの要因として考えられる可能性について検討し、それを踏まえた学習成果の間接評価の妥当性への示唆について考察した。

本研究を足掛かりに、今後も同様の検証が望まれる。その際に着目すべき点として次の3点を挙げる。第一に、他学部・学科における英語運用能力に関する検証である。第4章で述べたように、内容が自身の専門分野であることに起因するバイアスの存在が考えられるが、これについては他学部・学科での検証との比較から論じることが可能となる。第二に、英語運用能力以外の能力・技能についての検証である。特に、受容的か産出的かによるバイアスの違い、そしてバイアスが見られる場合にはその傾向の一貫性に興味を持たれる。第三に、自己報告の仕方による妥当性の違いである。本研究では間接評価としてルーブリック型の調査項目を用いたため、選択肢が文章で詳細に示されていた。しかし、知識・能力の程度や変化を尋ねる調査項目では「大きく増えた」「増えた」といった単純な選択肢からなるリッカート尺度が用いられることが一般的であり、その場合、第1章で述べた認知バイアスの影響を強く受ける可能性がある。これらの点に着目した検証が今後の課題である。

## 注

- 1 この調査報告ではリスニングとリーディングの他にも「やりとり」「話すこと」「ライティング」の3項目について分析されているが、これらの項目の直接評価指標にはリスニングとリーディングの試験のトータルスコアが用いられており、参考程度の位置づけとなっている。
- 2 当該調査項目の項目番号は2018年の調査までは問11だったが、本稿では調査年に依らず問8で統一した。
- 3 表1では一年生調査に倣いA1から順に選択肢を示した。
- 4 実際のところは、大半の学生はIELTSの結果を知る前に調査に回答したと考えられ、調査回答時にIELTSの結果を知っていた可能性があるのは一部の学生に限られる。

## 文献

- 一般社団法人大学IRコンソーシアム(2021),「当法人について」, <https://irnw.jp/about-us> (2021年8月23日閲覧)
- 斎藤有吾・小野和宏・松下佳代(2016),「パフォーマンス評価における教員の評価と学生の自己評価・学生調査との関連」,『日本教育工学会論文誌』40(Suppl.), 157-160
- 斎藤有吾(2019),『大学教育における高次の統合的な能力の評価 量的 vs. 質的, 直接 vs. 間接の二項対立を超えて』, 東信堂
- 日本英語検定協会(2018),『IELTS白書:2018年度版』, 日本英語検定協会
- 北海道大学高等教育推進機構(2014),『IRネットワーク報告書2013「英語力調査報告」』, <https://8gp.high.hokudai.ac.jp/data/report2013.pdf> (2021年12月3日閲覧)
- 北海道大学高等教育推進機構(2017),『IRネットワーク報告書2016「最終年度事業成果報告」』, <https://8gp.high.hokudai.ac.jp/data/report2016.pdf> (2021年8月23日閲覧)
- 松下佳代・小野和宏・斎藤有吾・白川優治(2014),

- 「学士課程教育における共通教育の質保証—直接評価と間接評価の開発と統合について—」, 『大学教育学会誌』 36 (2), 17-21
- 松下佳代 (2016), 「アクティブラーニングをどの評価するか」, 松下佳代・石井英真編『アクティブラーニングの評価』, 東信堂, 3-25
- 松下佳代 (2017), 「学習成果とその可視化」, 『高等教育研究』 20, 93-112
- 文部科学省 (2018), 「各資格・検定試験と CEFR との対照表」, [https://www.mext.go.jp/b\\_menu/houdou/30/03/\\_icsFiles/afieldfile/2019/01/15/1402610\\_1.pdf](https://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afieldfile/2019/01/15/1402610_1.pdf) (2021 年 8 月 20 日閲覧)
- 文部科学省 (2019), 「令和元年度「英語教育実施状況調査」概要」, [https://www.mext.go.jp/content/20200715-mxt\\_kyoiku01-000008761\\_2.pdf](https://www.mext.go.jp/content/20200715-mxt_kyoiku01-000008761_2.pdf) (2021 年 8 月 17 日閲覧)
- 山田礼子 (2012), 「学士課程教育の質の改善に向けてのプロセス評価の役割と限界」, 『大学教育学会誌』 34 (2), 6-12
- 山田礼子 (2016), 「共通教育における直接評価と間接評価における相関関係」, 『大学教育学会誌』 38 (1), 42-48
- British Council. (2021), Take IELTS, <https://takeielts.britishcouncil.org> (2021 年 8 月 20 日閲覧)
- British Council. IDP: IELTS Australia. and Cambridge Assessment English. (n.d.), IELTS in CEFR scale, <https://www.ielts.org/about-ielts/ielts-in-cefr-scale> (2021 年 8 月 20 日閲覧)
- Cole, J. and Gonyea, R. (2010), “Accuracy of Self-reported SAT and ACT Test Scores: Implications for Research,” *Research in Higher Education* 51, 305-319
- Council of Europe. (2001), “*Common European Framework of Reference for Languages: Learning, teaching, assessment*,” Cambridge University Press
- Kruger, J. and Dunning, D. (1999), “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments,” *Journal of Personality and Social Psychology* 77 (6), 1121-1134
- Kuncel, N. Crede, M. and Thomas, L. (2005), “The Validity of Self-Reported Grade Point Averages, Class Ranks, And Test Scores: A Meta-Analysis and Review of the Literature,” *Review of Educational Research* 75 (1), 63-82
- Pike, G. (1996), “Limitations of Using Student’s Self-Reports of Academic Development as Proxies for Traditional Achievement Measures,” *Research in Higher Education* 37 (1), 89-114
- Pike, G. (2011), “Using College Student’s Self-Reported Learning Outcomes in Scholarly Research,” *New Directions for Institutional Research* 150, 41-58
- Porter, R. (2013), “Self-Reported Learning Gains: A Theory and Test of College Student Survey Response,” *Research in Higher Education* 54, 201-226