

緒言：ライティング能力の評価

李 在鎬

1. 本特集の企画の趣旨と研究背景

本特集はライティング能力の評価に関する新しい知見を紹介する目的で組まれたものである。本特集の収録論文は、日本語教育支援システム研究会（CASTEL/J）の2021年の大会（CASTEL/J 2021 <http://kaken.sakura.ne.jp/castelj2021/>）のパネル発表「ライティングの評価再考：機械と人間の役割と今後の教育支援」において発表したものがもとになっている。論文化においては、オーディエンスからのフィードバックや発表後の研究成果をできるだけ盛り込むようにした。

以下では、本特集の研究背景として3点を述べる。まず、1.1節では、「キー・コンピテンシー」（Rychenほか 2006）や「21世紀型スキル」（Griffinほか2014）や「Education 2030」に代表される新しい能力観の問題と、言語能力評価におけるパフォーマンス評価の関係性について述べる。1.2節では、ライティング能力を構成する言語能力の問題と、その評価の難しさに関する問題について述べる。1.3節では、本特集で紹介する「ルーブリック評価」や「チェックリスト評価」や「自己評価・ピア評価」、さらには「自動評価」を評価研究の文脈でどう位置付けるべきかについて述べる。

1.1 新しい能力観とパフォーマンス評価

OECDのプロジェクトにおいて提案された「キー・コンピテンシー」（Rychenほか 2006）や「21世紀型スキル」（Griffinほか 2014）、さらには「Education 2030」（OECD 2019）のそのいずれも、現代社会の急速な変化に対応した新しい能力観を示したものである。今世紀、登場したこれらの新しい能力観は、グローバル化とデジタル化によって予測不可能な社会を切り抜くための能力を定義したものであるが、教育のあり方に対しても様々な問題提起をしている。とりわけEducation 2030（OECD 2019）に注目すると、現在の教育の具体的な課題として、1）詰め込みすぎのカリキュラムの改善、2）これからのニーズに応えるカリキュラムの開発、3）知識と技能の相互作用を促進する深い学びの普及、4）新しい学びや教員のアセスメントの実現、5）家庭や地域社会など学校外の学びの促進などが挙げられている。こうした課題を解決するため、言語教育においても「ことばの何を知っているか」という観点より「ことばで何ができるか」という観点が重視されるようになった。このことをもっとも明確に示しているのが、欧州評議会が提案した教育と評価のためのフレームワーク「CEFR（Common European Framework of Reference for Languages）」（西山・大木（編）2021）である。

「CEFR」は非常に多様な目的で利用されてきたフレームワークであるが、Coste（2021: 73）は4つの側面を持っていると指摘している。1）「評価のための能力記述文とレベル

分け)、2)「言語教育活動における行動中心アプローチ」、3)「思考過程を表現する概念としての複言語・複文化主義」、4)「カリキュラムを設計する上でのモジュール性と補完性」である。言語教育と評価においてもっとも直接的な影響を与えたのは、1)と2)であるが、これは、言語学習の目的を行動レベル、すなわち「言語を使って何ができるか」という観点へとパラダイムシフトをさせたものである。このことをより鮮明にさせたのが2018年に公開された「CEFRの補遺版」(CEFR Companion Volume with New Descriptors, <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>)である。

「CEFR補遺版」では、従来の「読む、聞く、話す、書く」の4技能を見直し、7技能として言語能力を捉えている。具体的には、1) Listening comprehension、2) Reading comprehension、3) Spoken interaction、4) Written interaction、5) Spoken production、6) Written production、7) Mediationである。読むことと聞くことで、情報を理解する能力に加え、話すことや書くことでやり取りをする能力、さらには談話文章をクリエイティブに産出する能力などへ細分化されたのである。これにより、言語運用力を中心においた教育と評価のフレームワークとしての「CEFR」の特徴がより明確になったのである。

「CEFR」による行動中心主義的アプローチの流れは、日本語教育の文脈でも確認することができる。例えば、2010年に実施された「新しい日本語能力試験」が良い例になる。1984年に日本語学習者の言語能力を測る目的で作られた日本語能力試験は、社会の変化を受け、2010年に日本語を用いた課題遂行能力を測るテストとして生まれ変わったのである(大隅ほか2009、李2011)。これは、日本語能力試験の英語名である「JLPT (Japanese-Language Proficiency Test)」の「P」の意味が2010年から再定義されたことを意味する。各レベルの文法項目や語彙項目を知っているかどうかではなく、日本語の文法項目や語彙項目を使って、日本語を使ってどんな課題が遂行できるかを問うテストとして生まれ変わったのである。これは、言語能力を行動主義の観点から定義したものとしてみることができる。

以上のように社会の変化に応じた新しい能力観は、言語教育や評価においても運用能力を重視する流れを作った。以下では、「書く」に注目し、どのような能力観や評価の観点があるか、考えていきたい。

1.2 ライティング能力とその評価における課題

「書く」という行為を能力という観点からとらえた場合、その構成要素としては、以下のものが考えられてきた(Weigle 2002: 43)。

1. 言語知識：

1) 文法的知識、2) テキストについての知識、3) 機能的知識、4) 社会言語学的知識

2. 方略的知識：

1) タスクの評価、2) 目標設定、3) 立案、4) 実行の統制

Weigle (2002) は、評価研究の文脈において「書く力」すなわち「ライティング能力」を次のように定義づけている。ライティング能力は、マクロレベルでは「言語知識」と「方略的知識」によって構成されている。「言語知識」とは、ことばに関わる直接的な知識であり、「方略的知識」とは、コミュニケーションの目的を達成するための非言語的知識

である。そして、「言語知識」の中の構成要素としては、4つのものが含まれる。1つ目は、言語の基本単位に関する文法的知識、2つ目は一貫性のあるテキストを作成するために言語的要素を組み合わせていく知識、3つ目は言語の機能や使われ方についての知識、4つ目は社会的な場面での適切性に関する知識である。また、方略的知識にはタスクを評価したり、状況に対するどう応答するかを設定したり、計画を実行するために言語知識を統制したりする知識が含まれている。この方略的知識によって、言語知識と外部の状況や話題などの一般的な知識が結びつけられている。

Weigle (2002) のライティング能力のモデルは、「書く」という行為が持つ複雑性を示唆する。これは、文章という言語単位が持つ複雑性 (佐久間 2003、李 (編) 2019) と「書く」が持つ特殊性が関係している。というのは文章は言語単位の中でもっとも大きな単位であることから、その作成プロセスにおいても、形態論、文法論、意味論などのマイクロな要素と、談話の結束性や一貫性などのマクロな要素が同時並行して働いているからである。Weigle (2002) の言語知識は、文章が持つ言語単位としての複雑性を反映したものである。そして、「書く」という行為そのものに注目した場合、話す行為とは違って、孤独な環境で行われる点があげられる。話す行為の例として、会話を考えてみた場合、(会話の) 相手がいるため、相手から情報をもらったり、不足を補ってもらったりして、全体の言語行為が遂行される。しかし、「書く」の場合は、(読む) 相手は目の前にいないケースが多いため、相手から情報をもらうことはない。Weigle (2002) の方略的知識は、「書く」が持つ特殊性を反映しているものと捉えることができる。

以上のようにライティング能力が持つ複雑さ故に、その評価も容易ではない。伝統的に作文などの評価は、正解が1つにしばられる客観評価ではなく、評価者が目で見、主観的に評価するという方法が用いられる。いわゆる主観評価という方法で、学習者の産出能力が評価されるのである (李 (編) 2015)。こうした主観評価においては、妥当性と信頼性の面で課題が多い。なお、評価研究において妥当性とは、測ろうとするものが適切に測れているのかに関する指標であり、信頼性とは、安定して測れているのかに関する指標である (近藤 2012、西岡ほか (編) 2022)。

主観評価において信頼性が問題になるのは、同じ作文であっても厳しい評価者と甘い評価者で評点が違ってしまふ、という時に起こる。そして、主観評価において妥当性が問題になるのは、評価者間で観点がばらついている時に起こる (野口ほか 2016)。ある評価者は内容の面白さに基づいて評価し、別の評価者は言語的正確さに基づいて評価し、結果的には測定の対象が違ってしまふ、ということもあり得る。主観評価は、評価者の主観によって評価がなされるが、人間の主観というのは、非常にダイナミックであり、一般化・標準化が難しい性質を持っている (伊集院ほか 2020)。そのため、主観に基づく評価には、常に妥当性と信頼性の課題が付きまとう。こうした問題を改善するために、次節で紹介する様々なパフォーマンス評価のツールが提案されている。

1.3 現代的ライティング評価

ライティング評価における信頼性と妥当性を確保するため、2000年以降、教育分野では様々な評価ツールが提案されている。これらは、いわゆる「真正の評価」(西岡ほか (編)

2022)の立場から提案されたものである。本特集で紹介する「ルーブリック評価」(伊集院論文)、「チェックリスト評価」(高野論文)、「自己評価・ピア評価」(安高論文)は、そのいずれも主観評価の妥当性と信頼性の問題を改善するために提案されたものである。同時に、これらの評価ツールは「評価=教師が行うもの」という構図からの脱却を目指したものである。いわゆる「学習者主導型評価」と呼ばれているもので、学習者自身が自分の能力を評価し、管理していくという考え方である。こうした「学習者主導型評価」では、複数の観点から学習者のパフォーマンスを評価することを目指すものであり、学習者と教師が評価の観点を共有することを目指すものである。このような考え方は、学習者の学びを支援するための評価である「形成的評価」(近藤 2012、李(編) 2015)ともオーバーラップするものが多い。

現代におけるライティング評価においても一つ、欠かせない視点として、コンピュータモデルを活用した自動評価(石井・近藤(編) 2020)があげられる。自動評価では、これまで、熟達した教師のみができると考えられてきた作文評価を計算モデルに基づいて行う。この自動評価は、妥当性の問題についてはまだまだ議論の余地があるが、信頼性においては非常に優れていると言える。なぜなら、計算モデルは、同じ文章であれば、誰が、いつ、どこで、評価しても同じ結果を返すからである。本特集の小森論文で紹介する*jWriter*(Lee & Hasebe 2020)は日本語学習者のために作られた自動評価ツールの一つである。自動評価に関する多くのツールは、現時点では発展途上のものであるが、今後、利用可能なデータが増え、機械学習の新しいアルゴリズムが提案されれば、確実に進化できる領域である。従って、現代的なライティング評価を考える上で、自動評価の存在は無視できないものであり、言語教師は、自動評価ツールをどう活用するか考えていかなければならない。

2. 本特集の構成

本特集には、5本の論文が収録されている。1本目の論文は、伊集院郁子氏によるもので、作文評価のためのルーブリックについて具体例を交えながら説明している。とりわけ、ルーブリックの開発手順を詳述しているため、ルーブリック評価の可能性や課題が把握できるものになっている。2本目の論文は、高野愛子氏によるもので、チェックリスト評価を取り上げている。チェックリストを用いた自己修正を実践事例に基づいて紹介しているため、チェックリストが学習者自身による推敲プロセスにどのような影響を与えるのかを明らかにしている。3本目の論文は、安高紀子氏によるもので、アカデミックライティングの教育実践を例に自己評価とピア評価を取り上げている。定量的な分析方法で自己評価とピア評価の傾向を明らかにしている。4本目の論文は、小森和子氏によるもので、自動評価を取り上げている。自動評価による評価結果と言語テストの得点の関連性を統計的な方法で分析した上で、接続表現に対する詳細な分析を行っており、自動評価の有用性と研究利用の可能性を明らかにしている。5本目の論文は、李在鎬氏によるもので、自動評価の研究背景と利用上の課題を俯瞰的に示している。自動評価の現状が把握できると同時に、研究の活性化のために必要なことを提案している。

以上の論考から、それぞれの評価ツールが持つメリットとデメリットが把握できるであろう。教育実践者においては、メリットの部分とデメリットの部分を同時に理解した上で、実際の評価活動において活用していくことが求められる。ライティング評価に限らず、あらゆる評価ツールに関して言えることであるが、万能な評価ツールというもの存在しない。このことを認識した上で、複数の評価ツールを用いて、多様な角度から学習者の言語能力を評価し、教育に結び付けていく姿勢が大切と言えよう。

参考文献

- 石井雄隆・近藤悠介（編）（2020）『英語教育における自動採点—現状と課題』ひつじ書房
- 伊集院郁子・李在鎬・小森和子・野口裕之（2020）「評価コメントに見られる意見文評価の様相—共起ネットワーク及びコレスポネンデンス分析に基づく考察」『第二言語としての日本語の習得研究』23、pp.26-43
- 大隅敦子・谷内美智子・小野澤佳恵・篠崎摂子・浅見かおり・野口裕之・小森和子（2009）「新しい日本語能力試験が目指すもの」日本語教育学会2009年秋季大会シンポジウム【配布資料】https://www.jlpt.jp/reference/pdf/2009_020.pdf
- Coste, D（2021）「CEFRとスイスのアーミーナイフ」、西山・大木（編）『CEFRの理念と現実 理念編 言語政策からの考察』くろしお出版、pp.45-79
- 近藤ブラウン妃美（2012）『日本語教師のための評価入門』くろしお出版
- 佐久間まゆみ（2003）「文章・談話における「段」の統括機能」佐久間まゆみ（編）『朝倉日本語講座7 文章・談話』朝倉書店、pp.91-119
- 西岡加名恵・石井英真・田中耕治（編）（2022）『新しい教育評価入門〔増補版〕—一人を育てる評価のために』有斐閣
- 西山教行・大木充（編）（2021）『CEFRの理念と現実 理念編 言語政策からの考察』くろしお出版
- 野口裕之・李在鎬・小森和子・奥切恵・伊集院郁子（2016）「作文評価の手法を問直す—IRT モデルを用いた尺度値の分析」『第27回第二言語習得研究会（JASLA）全国大会予稿集』p.79
- 李在鎬（2011）「日本語能力試験の挑戦—新しい日本語能力試験を例に」（国際交流基金事業レポート14）『日本語学』第30巻1号、明治書院、pp.95-107
- 李在鎬（編）（2015）『日本語教育のための言語テストガイドブック』くろしお出版
- 李在鎬（編）（2019）『文章を科学する』ひつじ書房
- Griffin, P., McGaw B., & Care, E.（2011） *Assessment and teaching of 21st century skills*,（三宅なほみ監訳、益川弘如・望月俊男編訳、2014『21世紀型スキル—学びと評価の新たな形』北大路書房。）
- OECD（2019）LEARNING COMPASS 2030 (https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_concept_note.pdf)
- Rychen, D.S., & Salganik, L.H. (eds.)（2003） *Key Competencies for a successful life and a well-functioning society*. Cambridge, MA: Hogrefe & Huber Publishers（立田慶裕（監訳）（2006）『キー・コンピテンシー—国際標準の学力をめざして』明石書店）
- Lee, Jae-Ho & Hasebe, Yoichiro（2020）Quantitative Analysis of JFL Learners' Writing Abilities and the Development of a Computational System to Estimate Writing Proficiency. *Learner Corpus Studies in Asia and the World 5*, pp. 105-120
- Weigle, S.C.（2002） *Assessing writing*. Cambridge, UK: Cambridge University Press.

