

作文評価ルーブリックの開発と課題

伊集院 郁子

要 旨

作文評価は、評価の観点や尺度が評価者の主観に左右されやすく、学生へのフィードバックにも時間と労力を伴う。この問題を改善する方法として、ルーブリック評価が注目されている。本研究では、大学教員から収集した作文評価データに基づいて評価の共通点と相違点を見出し、帰納的にルーブリックを作り上げるという手法で小論文評価用ルーブリックを開発した。収集データのうち、5段階評定値を評価尺度、評価コメントから抽出したコード・カテゴリーを評価観点、評価コメントに特徴的な語や表現を評価基準に用いて第一版を作成した後、大学教員による試用を経て、評価基準の文言と尺度の調整、補足資料の追加を行った。ルーブリック評価には多くの利点があるが、より客観的な評価を効率的に行うためには、「言語表現」と「形式」を自動評価と連動させ、教師は「内容構成」と独創性や意欲に関わる観点を中心にルーブリック評価ができるようになることが望まれる。

キーワード

作文評価 ルーブリック開発 小論文 評価観点 評価基準

1. 研究の背景と目的

現在、大学教育の質保証と向上を図る目的で、様々な取り組みが推進されている。授業の内容や方法、成績評価基準等の可視化もその取り組みの一つであり、中央教育審議会(2012)は、成績評価の際の具体的な測定方法としてルーブリックや学修ポートフォリオを挙げている。知識を問うテストによる評価ではなく、知識を使って何ができるかを測るパフォーマンス評価は、評価の客観性の担保や学生へのフィードバックの方法に困難を伴うが、この問題を改善する方法の一つとして、ルーブリック評価の有効性が注目されている。文部科学省(2021)は、学部段階における成績評価基準で一部の科目をルーブリックにより明示している大学数について、2015年には12.7%のところ、2018年には28.2%と大きな進展が見られたと報告しており、ルーブリック評価は未だに広く利用されているとは言えないものの、徐々に浸透しつつあることがわかる。

ルーブリックとは、「パフォーマンスを通して到達度を評価する際に用いられる評価基

準」であり、「評価の観点と到達度をマトリックスの形で組み合わせ、到達度に関しては順序尺度で表される段階と、文章で記述される具体的パフォーマンスが、一体になって表される点に特徴がある」（野口・大隅 2012, p.357）。大規模テストでは、パフォーマンスを総体として評価する包括的尺度のルーブリック¹が活用されることがあるが、Weigle (2002, p.121) は、ライティング能力を測る観点は習得レベルが一様ではないため、L2作文の評価では分析的尺度の方がより適していること、分析的尺度の方が包括的尺度より信頼性が高いことを指摘している。また、パフォーマンスを観点ごとに個別に評価する分析的尺度であれば、学生に診断的フィードバックをすることが可能となり（Kroll 1998, p.228）、教育的効果も期待できる。

しかし、分析的尺度のルーブリックにおいて、評価のためにどのような観点を立てるべきか、尺度ごとの評価観点の特徴をどのように記述すべきかといったことを検討するには多くの時間と労力を要する。今後、汎用性が高く、だれでも自由に改訂して利用できるルーブリックの共有が進めば、ルーブリック評価を取り入れやすくなるのではないだろうか²。

本稿では、日本語教育・学習を支援するツールの一つとして共有を目指す「小論文評価用ルーブリック」の開発について報告する。このルーブリックは、主に日本語学習者を対象とするアカデミック・ライティングの入り口として位置づけられる「小論文」または「意見文」と呼ばれるタイプの作文の評価に用いるものであり、大学教員による評価の実証的研究に基づいて開発された点に特徴がある。

以下、2節でルーブリック開発に関する先行研究を概観し、3節で本研究におけるルーブリック開発のプロセスとその成果物の特徴について詳述し、4節でルーブリック評価の課題を検討する。

2. 先行研究

日本の大学におけるルーブリック評価は、上述のとおり徐々に広がっており、本研究の焦点であるライティング教育に関しても、英語教育や初年次教育、理系の実験レポート等で先行研究が見られる。先行研究のテーマも、ルーブリックを用いた教育効果や課題、ルーブリックを用いた評価の検証、ルーブリック評価のあり方に関するメタ分析等、多岐にわたるが、ルーブリックの開発過程そのものに着目し、詳述している研究は数少ない。

ルーブリックの開発過程は、理論的には以下の3つの方法が考えられる。

- ① 既存のモデル・ルーブリックを改訂し、自身の実践現場のニーズにあったものを作る方法。その具体的なプロセスは以下の②を適用できる。（スティーブンス・レビ 2014）
- ② 課題を作成した目的等を振り返り、課題の具体的内容と学習目標を焦点にリストを作成し、リストの項目をグループ化して見出しを付ける。その過程で得られた評価観点と評価基準を表形式にまとめる。（スティーブンス・レビ 2014）
- ③ 高評価の作文をもとに特徴をリストアップし、その後複数の作文を採点して得点順に並べ、それぞれの作文の特徴をリストアップする。同じ観点でも質的に異なる

るものの特徴を記述し、尺度（レベル）ごとにそれぞれの記述をまとめる。（伊東 2022）

ここでは便宜的に①モデル・ルーブリックに基づく方法、②教師経験の振り返りに基づく方法、③成果物（作文）の分析に基づく方法に分類するが、現実のプロセスは①と②と③を部分的に、または全て組み合わせて作成する場合や、作成後に検証まで行ったうえで修正する過程が加わる場合等、多様である。また、教師が単独で作成する場合や複数名で共同して作成する場合、さらには学習者自身が単独または複数で作成する場合も想定される。

上記の①②③を組み合わせたルーブリック開発の一例として、渡邊（2018）は、授業で使用していたオリジナルの簡便チェック表で実際に学生の文章を評価しながら、評価途中で見出された問題点を洗い出し、各項目の細分化や整理をして初年次生を対象としたライティング科目のルーブリックを作成している。その際、形式はVALUEルーブリック³を参考にしつつ、利用対象（初年次生）と利用目的（評価とフィードバックの迅速化）に鑑み、評価観点と尺度は簡略化し、評価基準の記述は詳細にしたという。また、第一段階のルーブリックを作成した後、4名で33名分の提出物の評価を行って評価結果のばらつきを分析した結果、記述内容の具体化や評価項目の対応関係の厳密化等の修正が必要であることが明らかになったと報告している。

留学生を対象としたライティング授業での利用を目的に開発されたルーブリックとしては、脇田（2016）や山同他（2017）、三井他（2019）等がある。脇田（2016）は、日本人学生を対象とした松下他（2013）⁴のライティング・ルーブリックをモデルとし、日本語学習者用に「文法・語彙」の観点を加え、記述を短く平易な表現に書き換えて作成し、授業実践を通してその成果や留意点を報告している。また、山同他（2017）は、先行研究のほか、授業で学習者に配布する冊子に記載されているレポートの目標を参考にしている。レポート添削を経験したことのある教員からの意見をもとに、打ち合わせを重ねてルーブリックを作成した後、5名の教員によるレポート2編に対する評価のずれを分析し、ルーブリックの改善を行っている。三井他（2019）も同様に、同じ大学の別のレベルの日本語クラスで使用されているルーブリックを参考に、評価基準を調整してルーブリックを作成し、教員と受講生の評価のずれを分析している。

これらの先行研究は、レポート用のルーブリックであるが、坪根他（2021）には、ナラティブ作文の評価項目と評価基準の開発に関する報告がある。はじめに、ナラティブ・ディスコースの先行研究や既存の評価基準を参考に原案を作成し、その後日本語母語話者を対象とした評価及び日本語教師のナラティブ作文の評価に関する調査を実施し、評価基準を修正してルーブリックを作成するという手順を踏んでいる。

なお、日本語教育全体で共有できるような汎用的な評価基準の作成を目指した研究としては、田中・長阪（2006）が先駆的である。成果物の最新版の名称は「Good writingのための評価基準：トレイト別・基準説明」（田中・阿部 2014, pp.172-173）であり、評価尺度ごとの評価基準が示されていない点で、一般的なルーブリックとは異なるものであるが、先行研究のライティング評価基準・評価表のレビューを踏まえて原案を作成し、日本語学習者の自己評価と教師評価の結果との比較や、学生からのコメントを受けて項目を見直し

たり基準説明を充実させたりする過程を経た後、再度、学生及び日本語教員からコメントを得て改良を加え、評価基準を完成させている。

このように、ライティング評価のためのツールの開発方法はさまざまであるが、いずれも既存のモデル・ルーブリック等を参考に、目的や対象に応じて修正を施す過程を経ていく点が共通しており、作文の評価データを広く収集し、データから見出された特徴に基づいて帰納的にルーブリックを開発した先行研究は見当たらない。本研究では、日本国内の多機関の大学教員から、大学生が実際に執筆した作文に対する評価データを収集し、データから大学教員に共通する特徴を見出してルーブリックを開発するという手法を用いる。これにより、日本の大学への進学を計画している学習者や日本の大学で学ぶ学習者にとって目安となるようなモデル・ルーブリックの開発を目指す。

3. 本研究におけるルーブリック開発の手順

本節では、ルーブリック開発の具体的な手順について、伊集院・小森・奥切（2018）、伊集院・李在鎬・小森和子・野口裕之（2020）、Ijuin（2020）からも引用しつつ報告する。

3.1 全体的なプロセス

ルーブリック完成までの全体的な流れを以下に示す。①から④が評価研究のプロセス、⑤以降がルーブリックの原案作成から調整のプロセスに当たる。

- ① **評価データの収集**：日本の大学で3年以上の教育経験のある教員44名（日本語教育を専門とする教員22名及びその他の人文社会系の専門教員22名、所属先は24機関）から、30編の作文（執筆者の母語は日本語、中国語、韓国語、英語）に対する評価データ（5段階評定値及び作文の「良い点」・「悪い点」に関する評価コメント）を収集
- ② **分析1**：5段階評定値の量的分析及び評価コメント全体（「良い点」40,242字、「悪い点」64,815字、計105,057字（65,088語））の共起ネットワーク分析【伊集院他（2020）】
- ③ **分析2**：評価コメントのコード化・カテゴリー化⁵及び主要4カテゴリーと作文のレベル（評定値によって3群に分類）、執筆者の母語、評価者の専門との関係をコレスポネンス分析によって視覚化【伊集院他（2020）】
- ④ **分析3**：評価コメントから抽出した評価観点の全コードの出現数の量的分析、及び各評定値（「5：非常によい」～「1：悪い」）を付与された作文各3編（計15編）に対する評価コメント（計1,740文）の共起ネットワーク分析【Ijuin（2020）】
- ⑤ **ルーブリック（原案）の作成**：上記②③④の分析結果をもとに、3.2に示す方針に従って研究代表者がルーブリックの原案を作成。その後、研究グループのメンバー3名での話し合いによる調整に基づき原案完成
- ⑥ **ルーブリック（第一版）の作成**：研究メンバー3名により、作文6編に対してルーブリック評価を実施。その後、評価のずれが生じた要因を探るため、評価基準

の記述と作文の特徴に関して話し合いを重ね、評価基準のわかりにくさの解消や評価尺度の調整を実施。調整したルーブリックを用いて再度評価を行うプロセスを繰り返し、第一版（伊集院・高野 2020, pp.162-163 「小論文評価用ルーブリック」）を完成

- ⑦ **第一版の試用データの収集**：日本の大学で3年以上の日本語教育経験のある教員10名から、6編の作文に対し、5段階評定による包括的評価及びルーブリック評価、ルーブリックに対するコメント（使用感や改善すべき点）を収集
- ⑧ **分析4**：上記⑦のデータを用い、評価者間で評価が分かれる観点がなにか、分かれる場合はなぜか、包括的評価とルーブリック評価の結果に一貫性があるか等を分析【分析結果の一部は、伊集院・李・小森・高野・野口（2021）で発表済み】
- ⑨ **ルーブリック（第二版）の作成**：上記⑧の分析結果をもとに、第一版の評価基準及び尺度を調整し、「ルーブリック（第二版）」（巻末資料）及び参照用の補足資料を作成

以上の手順で、評価の実態を反映したルーブリックを開発した。なお、現在は、公開形態についての検討を行っている段階である。

3.2 ルーブリック作成の基本方針

ルーブリックの形式については、Stevens and Levi (2013)、ステイーブンス・レビ (2014) に従い、課題 (Task Description)、評価尺度 (Scale)、評価観点 (Dimensions)、評価基準 (Description of the Dimensions) からなるルーブリックを作成することとした。

本研究における各構成要素の位置づけは、以下のとおりである。

課題：日本留学試験の「記述」と同様に、与えられた課題に関し、主張（意見・立場・提案など）を根拠を挙げて論理的に伝えることを目的とする文章を書くというタスクで、文章の種類としては小論文または意見文と呼ばれるものである。

※大学での使用を想定した場合、レポート用のルーブリックが必須となるが、レポートは小論文に「問いの設定」と「客観的な根拠の収集（及びその引用）」が加わったものと捉え（伊集院・高野 2020, p.11）、まずは小論文用のルーブリックを作成することとした。

評価尺度：分析1の結果、44名による30編の評定値の平均は3.14となり、「3」を「よくも悪くもない平均（ふつう）」として据えることが妥当であると判断したため、5段階評定値（「非常によい（5点）」「よい（4点）」「ふつう（3点）」「あまりよくない（2点）」「悪い（1点）」）をそのまま用いることとした。ただし、ルーブリック上の評価尺度に添える文言は、学生にそのまま返却することを想定し、「すばらしい」「よい」「ふつう」「もう一歩」「がんばれ」の5段階とした。

評価観点：分析1・2・3を参考に、大学教員が記述した評価コメントから抽出したコード・カテゴリーを用いて評価観点を立てた。具体的な選定方法は3.2.1に詳述する。

評価基準：分析1・2・3を参考に、評価コメントに特徴的な語や表現を用いて記述した。具体的な記述方法は3.2.2に詳述する。

3.2.1 評価視点の選定

大学教員が記述した評価コメントから評価視点を抽出するために、評価対象の作文も参照しつつ、コンテキストに基づいて意味的なまとまりごとに評価コメントをリライトし、コードを生成していく作業を行った（伊集院他 2018）。さらにいくつかのコードからカテゴリを生成する作業を繰り返した結果、最終的に、評価視点として【内容（論旨、立場、根拠、状況把握、譲歩、反論、問題提起）】、【言語（表現力、文体、文法、語彙、接続、表記、誤用）】、【構成（段落間、段落内、バランス、立場、根拠、状況把握、譲歩、反論、問題提起）】、【形式（文字数、原稿用紙、書字、句読点）】、【題】、【書き手】、【読み手】が浮かび上がった（伊集院他 2020）。【内容】と【構成】は特に密接な概念であるため、分類の判断が難しいコメントも散見されたが、研究の段階では便宜上、「コーディング基準表」（表1に一部抜粋）に従って厳密に分類した。

表1 「コーディング基準表」抜粋（【内容】【構成】の一部を抜き出し）

カテゴリーコード	コードの定義	評価コメントの記述例	
		「良い点」	「悪い点」
内容	論旨	・要点が明確で、一貫している ・説得力がある	・内容を整理しきれていない ・整合性が弱い
	立場	・立場を明確に示している ・設問に簡潔に答えている	・作文課題「新聞・雑誌は必要か」に対して明確な答えがない
構成	段落間	・文章全体の構造が明確 ・序論・本論・結論のまとめ方が簡易でわかりやすい	・構成が明確でない ・段落間の関係がわかりにくい
	段落内	・各パラグラフで、一つの話題を詳しく説明している	・一段落一主文が守られていない

注：「コーディング基準表」は膨大な資料であるため、紙幅の都合上、極一部を抜粋したが、それぞれの定義には以下のような補足説明や参照用のキーワードが設けてある。

- ・【内容】と【構成】：「何を述べているか（内容）」ではなく「どこに配置しているか（内容の配置・配分）」に着目したコメントは、【内容】ではなく【構成】とする。
※「論理展開」に関するコメントは【内容：論旨】か【構成：段落間】か判断が難しいコメントも見られるが、論理展開の巧みさ、深さ、一貫性にフォーカスしたものは【内容：論旨】、「序論・本論・結論の流れがスムーズ」のように構成を意識した用語が見られるものは【構成：段落間】とする。
- ・【内容：論旨】：【内容（立場、根拠、状況把握、譲歩、反論、問題提起）】を包括する上位概念である。
※例えば「比較検討の仕方がよい」「よく分析できている」等、【内容：根拠】とは断定しきれない包括的なコメントも含める。
- ・【内容：立場】：「主張」という用語は、評価コメントの中で多数出現するが、【内容：立場】にコーディングする際の「主張」は、課題文に対する回答（本研究の課題文の場合は、新聞の必要性に関する執筆者の立場表明）という意味で、狭義に捉える。
※例えば「インターネットの速報性、新聞・雑誌の信憑性を主張している」というコメントにおける「主張」は【内容：根拠】の主張である。

続いて、ルーブリックを教育で利用する際の実用性を考慮し、研究によって抽出された上記の評価観点の整理・統合を行った。「評価観点は6つか7つが最大」（ステイブンス・レビ 2014, p.4）という指摘に従い、以下の方針で評価観点を立てた結果、最終的に評価観点は表2のとおりとなった。

- ・【内容】【言語】【構成】【形式】のカテゴリーに属するコードが全体の88.4%を占めるため（伊集院他 2020）、これらを主軸に設定する。ただし、【内容】と【構成】は密接に関連し合うため、教育利用を目的とするルーブリックで厳密に切り分けることはせず、【内容構成】としてまとめる。さらに、【構成】の下位コードはまとめて「論理構成」とする。
- ・コード数の最も多かった【内容：根拠】はそのまま評価観点として立て、次いで多かった【内容：論旨】と【内容：立場】をまとめて「主張」という評価観点とする。
- ・【言語】は着眼点ごとに細かく立てるのではなく、「文法」「語彙」「表記」などの「正確さ」と「表現」「文体」などの「適切さ」に二分して立てる。
- ・【形式】は下位分類を設けずに一つにまとめる。
- ・【題】【書き手（の意欲）】【読み手（に対する意識）】は、評価者によって全く言及しない場合もあり、着目度の個人差が大きいため、「その他」としてまとめ、加点または減点対象とする。

表2 ルーブリックの評価観点および観点の説明

評価の観点		観点の説明
内容構成	主張	主張（課題に対する書き手の回答）が容易に伝わるか。
	根拠	主張を支える根拠に説得力があるか。
	論理構成	全体の論理構成が明確で、段落間、段落内の情報が整理されているか。
言語表現	正確さ	日本語表現（文型、文法、語彙、表記）に誤用がないか。
	適切さ	アカデミック・ライティングにふさわしい表現（「ふつう体」、「かたい表現」、多様な表現）が適切に用いられているか。
形式	書式	書式など（フォーマット、文字数、原稿用紙の使い方など）の指定が守られているか。
その他		○加点対象 独創的な視点が盛り込まれていて読み手の興味を引くなど、書き手の意欲や工夫が伝わる。 ○減点対象 タイトルが本文の内容と合っていない、手書きの場合に極端に読みづらいなど、読み手を混乱させる要因がある。

3.2.2 評価基準の作成

評価基準は、5つの尺度の記述の整合性や一貫性を保ちつつ、いかにレベルの違いを記述し分けるかが重要であると考え、時間をかけて検討した。はじめに、作文30編の評定値の平均「3」（「ふつう」）の記述を確定させ、続いて「5」と「1」、最後に「4」と「2」

の評価基準を記述した。44名の教員が「3」と評価した作文の評価コメントを分析したところ、表3に示すように「一部」や「箇所」「ところ」等の表現（表3・4の_____の表現）を用い、問題点が局所的に存在することを指摘する評価コメントが散見されたため、これらの語を用いて表4のように記述した。

表3 大学教員が「3」と評定した作文に対する評価コメントの実例

カテゴリー：コード	評価コメント
【構成：段落間】	一部に構成上の問題があり、段落分けがうまくできていない。
【言語：表現力】	表現の選択が時々不自然である。
【言語：文体】	文体が不統一なところがある。
【言語：接続】	一部接続詞が不適切で、結束性に問題がある箇所がある。

表4 ルーブリックの評価尺度「3」に該当する評価基準の実例

評価観点	評価基準
内容構成（主張）	一読しただけで課題に対する主張は伝わるが、読み手が容易に理解できない内容も一部含まれる。
内容構成（論理構成）	序論・本論・結論に分けられるが、バランスの悪い点や情報が飛躍または重複している部分もある。
言語表現（正確さ）	不正確な日本語表現が部分的に見られる。

その他の尺度の評価基準に関しても、大学教員の評価コメントのうち、共起ネットワーク分析で抽出された語（表5の_____の語）や評定値の高低に関わらず、広く出現していた客観性や独創性に関連する表現（表5の□□の表現）を取り入れて記述した。

表5 評価基準の参考とした大学教員の評価コメントの実例

評定値	評価コメント
5：非常によい	主張の根拠が明確であり、説得力がある。論理構成・言語表現ともに非常によくできている。
4：よい	構成がわかりやすく、性質の異なる理由を3つ挙げられたのがよかった。最初に意見が述べられた上で、理由が記載されており読みやすい。
3：ふつう	独創的な反面、主張がやや主観的になっており、そう思わない人を納得させる根拠に乏しい。一部接続詞が不適切で、結束性に問題がある箇所がある。日本語が比較的わかりやすい。
2：あまりよくない	語彙や文法に不正確なところがある。話し言葉の使用、文体が不統一、タイトルが平凡。
1：悪い	語彙・表現・文法が不正確である。根拠が主観的で、信頼性に欠ける。字数が少ない。

3.3 ルーブリックの試用結果に基づく調整

ルーブリック完成の最終段階として、3.1のプロセス⑦⑧⑨で示したとおり、本研究グループ以外の大学教員10名から、6編の作文に対する5段階評定による包括的評価とルーブリック評価のデータを収集し、本ルーブリックの有効性や改善すべき点について検討を行った。

試用データの分析結果の詳細については別稿に譲るが、ルーブリックの評価観点6項目のCronbachの α 係数を求めたところ、「内容：主張、根拠、論理構成」「言語：正確さ、適切さ」「形式：書式」の順に0.93、0.93、0.95、0.93、0.96、0.62となり、「形式」を除いて評価者間の内的整合性が高いことが確認できた。「形式」は、作文によっては評価者間の評価のばらつきが大きかったため、その要因を探ったところ、指定の文字数にどの程度不足が見られる場合に減点すべきかに関し、評価者間の考えが大きく分かれていることがわかった。また、「形式」と「その他」（加点または減点対象）については尺度の調整を行い、配点を低くした方が実際の評価結果を反映できることが明らかとなった。さらに、試用に関するコメントを受け、評価基準の明確化が望まれる点について想定問答を準備する、複数の評価観点に関わる特徴をどのように扱うか対応事例を示す等の工夫を施すことで、利用者の不安を解消できる可能性が見出された。

4. ルーブリック評価の課題と展望

本稿では、大学教員による実証的な評価データに基づいて開発された小論文評価用ルーブリックについて、その開発プロセスの詳細を述べ、開発の理念と特徴を明示した。試用した教員からは「感覚に合っている評価基準」、「シンプルかつ分かりやすい区分け」、「評価を行うのに適切な項目数」等のコメントが得られ、また分析の結果からも、本ルーブリックは概ね問題なく機能していることが確認できた。ルーブリック評価の際に難しかった点や判断に迷った点についてもコメントが得られたため、これらについては上記に示したような改善を加え、作文サンプルとその評価結果の実例や評価手順の具体例を参照用の補足資料として提供することで、さらに利便性を図ることができると考えている。

最後に、教師側の評価の利便性に関し、ルーブリックの開発過程で多くの教員から得られたフィードバックと開発者自身によるルーブリック評価の実践から、残されている課題と今後の展望をまとめたい。

ルーブリックは、評価の客観性を担保する手段の一つとして有効であるが、当然のことながら、評価者の主観を完全に排除することは不可能である。本研究で収集したデータにも、「正確さ」における誤用はどこまでを指すのか、「文字数は何字からを不足とみなすのか」等、ルーブリック評価の際の評価者の迷いや不安が垣間見られるコメントが残されている。評価基準の「不正確な日本語表現が多数見られる」という場合の「多数」や「全体を通して日本語表現がほぼ正確に用いられている」という場合の「ほぼ」がどの程度を表すのかは、評価者の判断によって揺れが生じる点であり、教員個人の経験年数やビリーフによっても異なるだろう。このような評価の揺れについては、教師トレーニングや事前の評価のすり合わせによって、教員間の合意を形成しておくことも可能であるが、将来的

にはルーブリック評価と自動評価を連動させ、教員とコンピュータが連携して評価することが望ましいと考える。

例えば、言語表現の「正確さ」は、機械が誤用の数とその重篤度の判定を担い、「適切さ」も同様に、アカデミック・ライティングに不適切なスタイルが混在している割合や、使用語彙の難易度と豊かさに基づく判定を自動処理できるようになることを期待したい。「書式」に関連する文字数の判定についても、指定の文字数に1割以上不足している場合は評価を下げる（伊集院他 2020）といった処理であれば、実現の可能性が高いと思われる。

ルーブリック評価の観点のうち、教師による評価の強みが生かせるのは「内容構成」及び「その他」の独創性や書き手の意欲等に関する加点や減点の判断である。教育に携わる教員であれば、個々の学生の学習背景や言語能力の特徴が把握できるため、既存の知識や言語リソースを生かして意欲をもって書き上げたか、オリジナリティのある知見を取り入れているかといったことを評価に加えることも可能であろう。

このように、作文評価の際、ルーブリック評価に自動評価を連動させることが可能となれば、教師の評価の効率性に寄与するだけでなく、内容を掘り下げ、構成を磨くためのフィードバックに注力し、教育効果を高めることが期待できる。今後、機械評価との連動を視野に入れたルーブリック評価の開発が必要となると考える。

付記 本研究はJSPS 科研費18K00680、19H01273、22H00667の助成を受けて行われました。作文評価にご協力くださった皆さまにお礼申し上げます。

注

- 1 TOEFL iBT®のWRITING Rubricsや独立行政法人日本学生支援機構（JASSO）が実施する「日本留学試験」の日本語記述採点基準等がある。
- 2 日本高等教育開発協会「ルーブリックバンク」や関西大学ライティングラボ「学びの方向性と達成度を判断する評価ツール（ルーブリック）」が参考になる。
- 3 VALUE ルーブリックの詳細は、Rhodes (2010) 及びRhodes and Finley (2013) に詳しい。これらの資料によると、全米カレッジ・大学協会（Association of American Colleges and Universities, AAC&U）が「リベラル教育とアメリカの約束」（Liberal Education and America's Promise, LEAP）構想の一環として2007年からVALUE（Valid Assessment of Learning in Undergraduate Education）プロジェクトに着手し、2010年時点で“Written Communication”を含む15種のルーブリックを公開した。その後、2013年に“Global Learning”が新たに加わり、現在16種のVALUE ルーブリックがAAC&Uのウェブサイトで開催されている。VALUE ルーブリックをモデル・ルーブリックとして利用する先行研究も多い。
- 4 脇田（2016）が参照した松下他（2013）も、VALUE ルーブリックを参照して作られたとのことである。（松下佳代・高橋雄介他（2013）「VALUE ルーブリックの意義と課題—規準とレベルの分析を通して—」『第17回大学教育研究フォーラム発表資料』）
- 5 コード及びカテゴリーは、伊集院他（2018）の予備調査では25コード・8カテゴリーであったが、その後、伊集院他（2020）で33コード・10カテゴリーに分類された。このうち、3つのカテゴリー（「完璧」「丁寧でない」等の全体的印象を述べた【印象】、評価者の評価の意図が判断できない【分類不能】、「特になし」というコメント【なし】）は、評価基準とは無関係の項目として、ルーブリック開発の際には検討対象外とした。

参考文献

- 伊集院郁子・小森和子・奥切恵 (2018) 「大学教員によるライティング評価の観点を探る」『*Learner Corpus Studies in Asia and the World*』3、pp.159-176
- 伊集院郁子・高野愛子 (2020) 『日本語を学ぶ人のためのアカデミック・ライティング講座』アスク出版
- 伊集院郁子・李在鎬・小森和子・高野愛子・野口裕之 (2021) 「作文評価のための教師用ルーブリックの作成と試用」『2021年度日本語教育学会秋季大会予稿集』 pp.267-272
- 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ダ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
- 伊東祐郎 (2022) 『日本語教育よくわかる評価法』アルク
- 関西大学ライティングラボ「学びの方向性と達成度を判断する評価ツール (ルーブリック)」
<<https://www.kansai-u.ac.jp/ctl/labo/useful/rubric.html>> (2022年8月31日)
- 田中真理・阿部新 (2014) 『Good Writingへのパスポート—読み手と構成を意識した日本語ライティング』くろしお出版
- 田中真理・長阪朱美 (2006) 「第2言語としての日本語ライティング評価基準とその作成過程」国立国語研究所 (編) 『世界の言語テスト』くろしお出版、pp.253-276
- 中央教育審議会 (2012) 「新たな未来を築くための大学教育の質的転換に向けて—生涯学び続け、主体的に考える力を育成する大学へ (答申)」<https://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afidfile/2012/10/04/1325048_1.pdf> (2022年8月31日)
- 坪根由香里・トンプソン美恵子・影山陽子・数野恵理 (2021) 「第二言語としての日本語ナラティブ作文の評価基準とルーブリックの開発」『大阪観光学研究論集』21、pp.85-94
- 独立行政法人日本学生支援機構 (JASSO) 「日本留学試験」日本語記述採点基準
<https://www.jasso.go.jp/ryugaku/eju/about/score/_icsFiles/afidfile/2021/03/04/kijyutu_saitenkijyun.pdf> (2022年8月31日)
- 日本高等教育開発協会「ルーブリックバンク」<<https://www.jaedweb.org/rubricbank>> (2022年8月31日)
- 野口裕之・大隅敦子 (2012) 「テスト・評価」近藤安月子・小森和子 (編) 『研究社日本語教育事典』、p.357
- 三井一巳・鄭在喜・藤田百子・吉田好美 (2019) 「レポート作成におけるルーブリック評価の再考—教員と受講生の評価視点のずれとレポート産出の変化からの考察」『早稲田日本語教育実践研究』7、pp.15-22
- 文部科学省高等教育局 (2021) 「令和元年度の大学における教育内容等の改革状況について (概要)」<https://www.mext.go.jp/content/20211104-mxt_daigaku03-000018152_1.pdf> (2022年8月31日)
- 山同丹々子・高橋雅子・伊藤奈津美・藤本朋美・安田励子 (2017) 「ルーブリック作成と評価視点の「ずれ」の分析—上級前半レベルのレポート課題」『早稲田日本語教育実践研究』5、pp.123-130
- 脇田里子 (2016) 「ライティング・ルーブリックの実践」『コミュニカール』5、pp.21-50
- 渡邊淳子 (2018) 「ライティング指導のためのルーブリック評価表開発—均質なパフォーマンス評価の試み」『熊本保健科学大学研究誌』15、pp.101-107
- Ijuin, Ikuko (2020) Developing an Essay-Writing Rubric for Learners of Japanese: Based on the Analysis of Writing Assessment by Japanese University Faculty. *Learner Corpus Studies in Asia and the World* 5, pp.79-94
- Kroll, Barbara (1998) Assessing Writing Abilities. *Annual Review of Applied Linguistics* 18, pp.219-240
- Rhodes, T. L. (Ed.) (2010) *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Association of American Colleges and Universities.
- Rhodes, T. L., & Finley, A. (2013) *Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment*. Association of American Colleges and Universities.
<<https://commission.fiu.edu/helpful-documents/competency-based-courses-degrees/using-the-value-rubrics-for-improvement.pdf>> (2022年8月31日)

- Stevens, D. D., & Levi, A. J. (2013) *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning* (2nd ed.). Sterling, VA: Stylus Publishing. (ステイブンス、ダネル・レビ、アントニア (2014) 佐藤浩章 (監訳) 井上敏憲・俣野秀典 (訳) 『大学教員のためのルーブリック入門』玉川大学出版部)
- TOEFL iBT® WRITING Rubrics <https://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf> (2022年 8月 31日)
- VALUE Rubrics <<https://www.aacu.org/initiatives/value-initiative/value-rubrics>> (2022年 8月31日)
- Weigle, S.C. (2002) *Assessing writing*. Cambridge, UK: Cambridge University Press.

(いじゅういん いくこ 東京外国語大学大学院国際日本学研究院)

巻末資料：小論文評価用ルーブリック（第二版）

小論文：与えられた課題に関し、主張（意見・立場・提案など）を根拠を挙げて論理的に伝えることを目的とする文章

評価の観点/観点の説明	すばらしい（5点）	よい（4点）	ふつう（3点）	もう一步（2点）	かんばれ（1点）
主張 主張（課題に対する書き手の回答）が容易に伝わるか。	課題に対し、明確で一貫した主張が論理的に展開して、深い洞察が容易に理解できる。	課題に対し、明確で一貫した主張が展開して、全体の内容が容易に理解できる。	一読しただけで課題に対する主張は伝わるが、読み手が容易に理解できない内容も一部含まれる。	一読しただけで課題に対する主張が伝わらず、読み手が理解できない内容も含まれる。	課題を正確に捉えていないか、または課題に対する回答が伝わらない。
根拠 主張を支える根拠に説得力があるか。	客観的かつ具体的な根拠が過不足なく挙げられ、反論も想定されていて非常に説得力がある。	客観的または具体的な根拠が複数挙げられていて、説得力がある。	根拠が複数挙げられているが、客観性や具体性に欠けるため、やや説得力が弱い部分もある。	根拠の多くが主観的または表面的で主張のサポートとして弱い。そのため、説得力に欠ける。	根拠が主張のサポートになっていないため、説得力がない。
論理構成 全体の論理構成が明確で、段落間、段落内の情報が整理されているか。	序論・本論・結論がバランスよく配置され、情報のつながりも明確である。	序論・本論・結論がバランスよく配置され、情報のつながりもほぼ問題がない。	序論・本論・結論に分けられるが、バランスの悪い点や情報量が飛躍または重複している部分もある。	序論・本論・結論のバランスが悪いか、または情報のつながりに問題がある。	序論・本論・結論のバランスおよび情報のつながりに大きな問題がある。
正確さ 日本語表現（文型、文法、語彙、表記）に誤用がないか。	全体を通して日本語表現が正確に用いられている。	全体を通して日本語表現がほぼ正確に用いられている。	不正確な日本語表現が部分的に見られる。	不正確な日本語表現が多数見られる。	全体を通して日本語表現が不正確である。
適切さ アカデミック・ライティングにふさわしい表現（「ふつう体」、「かたい表現」、多様な表現）が適切に用いられているか。	全体を通してアカデミック・ライティングにふさわしい表現が適切に用いられている。	全体を通してアカデミック・ライティングにふさわしい表現がほぼ適切に用いられている。	アカデミック・ライティングにふさわしくない表現（「ていねい体」、「やわらかい表現」、単調で推論的な表現）が部分的に見られる。	アカデミック・ライティングにふさわしくない表現（「ていねい体」、「やわらかい表現」、単調で推論的な表現）が多数見られる。	全体を通してアカデミック・ライティングとして適切ではない。
書式 書式など（フォーマット、文字数、原稿用紙の使い方など）の指定が守られているか。					書式などが守られていない。
その他	独自の視点が盛り込まれていて読み手の興味を引くなど、書き手の意欲や工夫が伝わる。				
その他	タイトルが本文の内容と合っていない、書き手の場合極端に読みづらくないなど、読み手を混乱させる要因がある。				

★採点方法

配点は、各教育実践の内容に合わせて適宜変更してご利用ください。基本の評価方法としては、以下の(1)(2)の手順を想定しています。

- (1) 【内容構成】「言語表現」のそれぞれの観点に関し、「すばらしい=5点」「よい=4点」「ふつう=3点」「もう一步=2点」「かんばれ=1点」で評価し、【形式】に関しては「ふつう=3点」「もう一步=2点」「かんばれ=1点」で評価する。
- (2) 【その他】に当てはまる場合、最大2点まで加点または減点する（満点30点）。