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

伊集院郁子（東京外国語大学）、小森和子（明治大学）、安高紀子（明治大学）、
髙野愛子（大東文化大学）、李在鎬（早稲田大学）

Rethinking Writing Assessment: Role of Human and Machine Raters and Future Writing Supporting Tools

Ikuko IJUIN, Tokyo University of Foreign Studies　Kazuko KOMORI, Meiji University
Noriko ATAKA, Meiji University　Aiko TAKANO, Daito Bunka University　Jae-ho LEE, Waseda

要旨：客観的かつ効率的に言語能力を評価するために、ICT は有用なツールとなり得るが、日本語教育における作文評価は人間による主観的評価が主流であり、機械による自動評価が利用される機会は少ない。本パネルでは、作文評価における機械と人間の役割について検討することを目的に、*jWriter* による作文評価と言語習熟度との関連性、教師用評価ルーブリックの開発と課題、学習者主導型評価の特徴、学習者用チェックリストを用いた自己修正の 4 つの話題提供を行い、オンライン授業による日本語教育が主流となりつつある中で、今後どのような作文教育・学習支援ツールが期待されるのかについて議論したい。

キーワード：ライティング評価、機械評価、ルーブリック評価、教師評価、学習者評価

1.　パネル発表の趣旨

ライティング能力は、日本語学習者の総合的な日本語運用能力を測るための有力な指標の一つであり、日本留学試験においても「記述」が課されている。作文評価は教師による評価が主流であるが、人による評価は、評価者の属性やビリーフ、個人の心身の状態等に左右されるため、教師の負担を減らし、客観的な指標に基づく評価を可能にする方法として、機械評価が注目されている。大学入学共通テストへの記述式問題の導入に関する議論の中で、南風原（2019）は記述式問題を大規模な共通テストで行うには、ある程度納得のいくような自動採点が条件であると述べており、機械評価への期待は今後も高まることが予想される。また、Covid-19 の影響で日本語教育のオンライン化も急速に進んでおり、場所や時間の制約なく利用できる作文教育・学習支援ツールの必要性も高まっている。

このような状況において、機械と人による作文評価はどうあるべきか、どのような支援ツールが有用かを検討することには意義があると考える。本発表では、4 つの事例報告と参加者との意見交換を通し、今後の作文評価および教育・学習支援のあり方を模索する。

2.　報告①　*jWriter* による作文評価と言語習熟度（李在鎬・小森和子）

2.1.　研究背景

日本語学習者作文評価システム *jWriter* (https://jreadability.net/jwriter/)は、日本語学習者が書いた作文の文章の難易度を、計量テキスト分析の手法を用いて自動評価するシステムである（李他 2019）。難易度の判定には、文字数、文数、段落数、異なり語数、語種別語彙数、レベル別語彙数、語彙の多様性、一文の平均語数等の 39 種の多様な言語統計量が用いられている。また、これらの統計量を用いて推定された作文能力の評定値、およびこの評定値を基にした評価ガイドラインという 5 段階評価（入門～超級）もフィードバックされるように設計されている。なお、作文能力の評定値の推定のための学習データとして「多言語母語の日本語学習者横断コーパス（International Corpus of Japanese as a Second Language: I-JAS）」を使っている(Lee&Hasebe 2020)。

筆者らのグループは、*jWriter* の評価と教師の評価の関係性を検討するべく、学習者が手書きで書いた作文（*N*=30）に対して、教師 3 名が包括的評価、および「内容」、「構成」、「言語」、「誤用」の項目別評価を行った後、*jWriter* による機械評価を施したところ、*jWriter* の評定値は、教師による包括的評価と項目別評価の「言語」との間で、正の強い相関が示されたが、「内容」、「構成」、「誤用」とは無相関であった（小森他 2018）。*jWriter* の作文評定値が「言語」と相関があることが確認できたので、言語知識の産出的側面を見る作文科目の評価に活用できると考え、筆者の教育機関（明治大学）のオンライン・プレイスメント・テスト（以下、OPT）に *jWriter* で評価する作文科目を加えることとした。それまで、OPT は多肢選択式の文字語彙、文法、読解で、言語知識の受容的側面のみを測定しており、採点に時間と労力がかかることから、作文の導入に躊躇していたという背景がある。

そこで、本発表では、これらの言語テストの得点と *jWriter* により判定された作文評定値がどのような関係になっているのかを検討するべく、受験者のデータを分析した結果について報告する。

2.2　調査

2.2.1　調査概要

分析対象は、2018 年 3 月から 2020 年 9 月の間に OPT を受験した 52 名のデータである。

OPT の文字語彙、文法、読解は、それぞれ 50 問、38 問、16 問から成っている。文字語彙と文法は JLPT の N5 から N1 相当の難易度の項目、読解は N3 から N1 相当の難易度の文章である。また、作文は、文字語彙、文法、読解の 3 科目において N3 相当以上の能力があると判定された者のみが受験対象である。作文では「教科書はデジタルブックにするべきだという人がいます。反対に、教科書は今のまま印刷した本の方がいいという人もいます。あなたはどちらの意見に賛成ですか。理由や根拠を示しながら、自分の意見を書きなさい」という課題で、辞書等は参照せずに 30 分以内でコンピュータ上で執筆した。

2.2.2　調査結果

OPT3 科目の平均点は、文字語彙 34.88 点（50 点満点）、文法 25.88 点（38 点満点）、読

解 8.90 点（16 点満点）、3 科目合計 69.67 点（104 点満点）であった。また、作文の文字数や語数等の代表的な言語統計量を *jWriter* で求めたところ、1 編あたり平均 400 文字、約 250 語で書かれており、作文評価レベルは中級から上級であることがわかった。

次に、3 科目の得点と作文評価値との相関の有無を確認したところ、いずれも正の相関が認められ[文字語彙 r=.619, df=52, p<.001，文法 r=.581, df=52, p<.001，読解 r=.468, df=52, p<.001，合計 r=.618, df=52, p<.001]、特に、文字語彙との相関が強いことがわかった。これは、*jWriter* の判定が文字や語を中心とした統計量に基づいていることによるものと考えられる。また、*jWriter* が判定した評価ガイドラインでは、初級 3 名、中級 29 名、上級 20 名となったため、初級の 3 名を除いた中級と上級の 3 科目の得点差を確認したところ、こちらも統計的に有意であった[文字語彙 t(47)=2.632, p<.05, 文法 t(47)=2.662, p<.05, 読解 t(47)=2.462, p<.05，合計 t(47)=2.851, p<.01]。以上のことから、*jWriter* による機械評価は、文字語彙、文法、読解等の総合的な言語知識や能力の推定に有用であることが示唆される。

なお、言語統計量を詳細に見てみると、多くの項目において上級が中級に勝っているものの、接続詞、代名詞、助動詞、形容詞、連体詞等においては差が認められなかった。接続詞は作文の段落間の構成を明示するディスコース・マーカーとして有用であり、形容詞や連体詞は豊かな描写に用いられる品詞である。また、ヴォイスやモダリティに関わる助動詞なども、書き手の視点や態度が現れる重要な言語形式であり、論点の一貫性や主張の強さが反映される。このような言語形式が作文評価レベルに大きな影響を与えないという点を考慮すると、作文の構成、内容、日本語の豊かさ等の側面においては、機械評価ではなく、経験を積んだ教師による評価が必須である可能性が示唆されよう。すなわち、機械評価と人間による評価を住み分け、それぞれの特性を生かし、かつ効率的な評価のあり方考えていくべきであろう。

3. 報告② 作文評価ルーブリックの開発と課題（伊集院郁子）

3.1 研究背景

機械は人間と異なり、処理性能を落とすことなく大量の言語情報を正確に測定できるため、人間による評価と比べて信頼性は高いが、妥当性の観点からは問題視されることもある。そのため、現段階では、入学や卒業に関わるハイステイクス・テストにおいて、作文のようなパフォーマンス評価を機械評価のみで行おうとする動きは見られない。

一方で、人間による評価は、時間的、労力的な負担があり、評価結果が主観に左右されやすいという問題があるが、近年、これらの問題を軽減しつつ、さらに学生への教育的フィードバックを効果的に行う方法の一つとして、ルーブリック評価が注目されている。日本語教育においても教育実践等の知見が蓄積されつつあるが、ルーブリック開発は担当教員の経験や内省に基づくことが多い。本節では、日本の大学教員による作文評価の実証的研究に基づいて開発したルーブリックの報告を行い、今後の課題と発展の可能性を検討する。

3.2　ルーブリック開発の手順

スティーブンス他（2014）によると、ルーブリックは課題、評価尺度（達成レベル・成績評価点）、評価観点（課題が求める具体的なスキルや知識）、評価基準（具体的なフィードバック内容）からなる。本研究では、評価尺度には大学教員による 5 段階評定値、評価観点には大学教員が執筆した評価コメントから抽出したコード、評価基準には評価コメントに特徴的な語や表現を利用し、以下の手順で、現実の評価を反映したルーブリックの開発を目指した。

①日本の大学で 3 年以上の教育経験のある教員 44 名から、日本留学試験の「記述」と同じタイプの作文 30 編に対する評価データ（5 段階評定値と評価コメント）を収集した。

②評価コメントを意味的なまとまりでコード化し、主要 4 コードと作文のレベルや執筆者の母語との関係をコレスポンデンス分析によって明らかにした。（伊集院他 2020）

③評価コメント（「良い点」40,242 字、「悪い点」64,815 字、計 105,057 字（65,088 語））から抽出した評価観点の全体像の分析および評定値 1 から 5 の作文各 3 編に対する評価コメント（計 18680 字、740 文）の共起ネットワーク分析を行った。（Ijuin,2020）

④上記の分析結果を参考に、以下の 3.3 に示す方針でルーブリックの原案を作成し、研究メンバー3 名で 6 編の作文のルーブリック評価を行った。評価が一致しなかった点について、評価基準の文言のわかりにくさの解消や評価尺度の調整を行い、再度評価を行うプロセスを繰り返して最終版（伊集院・髙野 2020, pp.162-163）を完成させた。

3.3　評価観点の選定と評価基準の作成

上記 3.2 に記載した手順②③の分析結果から、評価観点として【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。ルーブリックは実用性を備える必要があるため、評価観点は 6 つか 7 つが最大というスティーブンス他（2014, p.4）等を考慮し、以下のような方針を立てた。

- 【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、これらを主軸に設定するが、【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「根拠」はそのまま評価観点として立て、「根拠」に続いて多かった「論旨」と「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」に大きく二分して立てる。
- 【形式】は下位分類を設けずに一つにまとめる。また、「題」「書き手（の意欲）」「読み手（に対する意識）」は「その他」に入れ込み、加点または減点対象とする。

評価基準に関しては、以下に例示した大学教員の評価コメントのうち、上記の手順③の共起ネットワーク分析で抽出された語（下線の語）や評定値の高低に関わらず出現していた客観性や独創性に関連する表現（□の表現）を取り入れることとした。また、初めに完成させた「3」（「ふつう」）の記述は、評定値「3」の評価コメントに散見された「箇所」「一部」「ところ」等の表現を用い、問題点が局所的に存在することを指摘する評価基準とした。

〈評価基準の参考とした大学教員の評定値と評価コメントの例〉

5 点：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 点：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。

3 点：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。日本語が比較的分かりやすい。

2 点：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 点：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。

3.4　成果物の特徴と今後の課題

このような手順・方針で作成したルーブリックは、与えられた課題に対して主張（意見・立場・提案など）を根拠を挙げて論理的に伝える文章に関し、日本の大学に入学する留学生がどのような力を備えておくべきかを示している。評価観点は【内容構成（「主張」「根拠」「論理構成」）】、【言語表現（「正確さ」「適切さ」）】、【形式】、【その他】から成り、これまでのライティング評価の観点を踏襲する結果となったが、5 段階の評価尺度（「5:すばらしい」「4:よい」「3:ふつう」「2:もう一歩」「1:がんばれ」）の評価基準等が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点に特徴がある。

一方、本ルーブリックは学習者にとって難易度の高い日本語表現で記述されており、英訳版（Ijuin,2020）も備えてあるものの、現状では主として教員用ツールとして位置づけられる。今後、教員の評価結果（点数や評価コメント）を添えたサンプル作文をルーブリックに添えて提供することで、実例に基づいて作文の改善方法を検討するリソースとして学習者の利用にも供することができるのではないかと考えている。また、大学入学後の作文評価はレポート評価が中心となるため、本ルーブリックにテーマ設定や引用の適切さ等の観点を加え、レポート評価用のルーブリックを開発することも課題である。

4.　報告③　論証型レポートに対する学習者の自己評価・ピア評価（安高紀子）

4.1　研究の背景

教育評価において、近年、学習者が主体となって行う学習者主導型評価が注目されており、学習の進度や達成度を判断するための形成的評価として実施されている。学習者が自分自身に対して行う自己評価は、学習成果の進捗状況や達成度の評価を促す（バークレイ・メジャー2020）とされている。日本語教育においても、自己評価や学習者が相互に評価す

るピア評価などが取り入れられており、トムソン木下(2008)は、自己評価の有効な利用は利点が多いとしながらも、学習者自身が自己評価能力を信頼していない場合や馴染みがない場合、成果が上がらないと指摘している。また、市嶋（2009)によると、相互的自己評価活動に対して、学習者は評価の意味付けという点では意義を見出しているが、「学生は評価する力がない」などの不満も同時にあるという。

このように、学習者主導型評価は学習の進捗や達成度を振り返る機会となる一方で、評価の能力に関する懸念もある。学習の途上にある学習者が自己を振り返り、適正に評価することは容易ではないだろう。西岡他（2015）は、他者の具体的なテストや作品に対して評価する経験が、目標である評価基準の理解を促し、学習を評価するメタ認知能力の育成につながると述べており、他者に対して行うピア評価は評価能力を養う有効な手段であると考えられる。そこで、本節では、自己評価とピア評価に焦点を当て、学習者がどのような評価を行っているのか現状を探ることを目的とし、自己評価とピア評価の一致率を指標として分析を試みた。

4.2　調査概要

分析を行ったのは，筆者が担当した学部留学生 1 年生の文章表現（2020 年度秋学期）の授業で学生が作成した論証型レポートに対する自己評価とピア評価（1 回目 37 名、2 回目 38 名）のデータである。学生が自己修正するためには、観点毎に尺度段階の特徴が記述されたルーブリックよりも、チェックすべき点を一覧にした形式の方が利用しやすいと考えたため、本研究では大島他(2005)のレポートチェックシートを基に、評価欄とコメント欄をつけたチェックリスト形式の評価シートを作成し、それを使用した。評価シートの評価の観点は【説得力】【読みやすさ】【明晰さ】【文・語句・表記】【引用の適切性】の 5 つから成る。各観点には、チェック項目が 3～5 項目記載されている。例えば、【説得力】の観点には「主張の展開は適切で説得力があるか」「信頼できる情報やデータを集めているか」などのチェック項目がある。評価は、ABCD の 4 段階とし、観点毎にチェック項目を参照し、達成度が 80%以上は A、70%以上は B、60％以上は C、60％未満は D と評定することとした。

自己・ピア評価の手順は次の通りである。①各自が作成したレポートに対して自分で評価を行う（自己評価）。次に、授業中のピア活動において、②お互いに相手のレポートを読んで評価を行い（ピア評価）、③口頭でレポートに対するコメント、評価のフィードバックを行う。その後、④各自、レポートを見直して自己修正する。翌週、自己修正したレポートに対して、再び①～④を行った。これを合計 2 回実施した。

4.3　結果と考察

自己評価とピア評価がどの程度一致しているか、観点毎に重み付けカッパ係数により、一致率を算出した。

まず、1 回目の評価では、最も一致率が高かったのは【説得力】で、κ ＝.438 であった。その他の 4 つの観点では、【引用の適切性】κ ＝.305、【明晰さ】κ ＝.265、【読みやすさ】κ ＝.262、【文・語句・表記】κ ＝.220 であった。次に、2 回目の評価では、一致率は、【説得力】κ ＝.475、【引用の適切性】κ ＝.475、【文・語句・表記】κ ＝.392、【読みやすさ】κ ＝.389、【明晰さ】κ ＝.183 であった。

これらの結果から、自己評価とピア評価の一致率は 1 回目も 2 回目も全体的にはそれほど高くはないことがわかった。しかし、観点によって、一致率には差が見られた。【説得力】は 1 回目 2 回目とも最も一致率が高く、学生は他の観点に比べ、到達段階を適切に判断し、評価できていると考えられる。一方で、【明晰さ】は 2 回目の評価では他に比べ、一致率が低かった。その要因として、学生はこの観点のチェック項目があまり理解できておらず、評価の難易度が高かったと推測される。例えば、【明晰さ】には「主張には妥当な理由と証拠があるか」というチェック項目があるが、学生には主張に対してどのような理由や証拠であれば妥当と言えるのかといった判断基準が明確でなかった可能性がある。評価が困難な項目については、授業において、具体例を示しながら説明し、実際に判定を行ってみるといったステップを踏んだ指導を行っていく必要があるだろう。

4.4　今後の課題

以上の結果に基づき、今後は評価シートのチェック項目の見直しを行うとともに、評価が難しいと思われる項目については、授業においてモデルを用いた評価練習を行うなどの授業改善を図り、学習者自らが自己の学びの成果や達成度を適正に捉え、評価できるような評価活動の実現を目指したい。

5.　報告④　学習者用チェックリストを用いた自己修正（髙野愛子・李在鎬）

5.1　研究背景

学習者の書いた作文を教師が添削しフィードバックする場合、言語表現の正誤や適切さに関する許容範囲をはじめ、内容構成に関しても教師個人の語感や判断基準に影響されやすい。また、広瀬（2007）が「カンファレンスを行った後、学習者はみなほぼ教師の指摘どおりに原稿を書き直した」と指摘しているように、学習者自身は教師の添削をそのまま受け入れることが多く受け身になりやすい。そこで、学習者が自律的かつ主体的に取り組む姿勢を養成することを目的とし、筆者らが開発したチェックリスト（伊集院・髙野 2020,p.154）を用い、教師に提出する前に学習者自身が自ら修正するプロセスを加えることを試みた。このチェックリストは、「内容構成」「言語表現」「形式ほか」から成っている。

本発表では、1 回目と自己修正後の 2 回目の作文を比較し、*jWriter* によりその相違を分析するのと同時に、学習者自身がどこに注目して修正しているのかについて検討を行い、学習者の自律的な学習を支援する方法と教師による添削と評価の役割を考える。

5.2　調査

5.2.1　調査概要

調査は、筆者が担当した文章表現の授業における学習者の作文 50 編（N1 相当 12 編、N2 相当 38 編）を対象に行った。手順は以下のとおりである。

① 課題の作文を学習者が書く→1 回目の作文

② チェックリストを用いて学習者が自己修正する（修正箇所を記す）→2 回目の作文

③ 1 回目と 2 回目の作文を教師に提出

④ 1 回目と 2 回目の作文を *jWriter* により分析（量的分析）

⑤ 1 回目と 2 回目の作文を目視により分析（質的分析）

学習者自身がどこに注目して修正しているのかについて分析したところ、以下のようなことが明らかになった。

5.2.2　調査結果１（量的分析）

作文の量的分析として、*jWriter* を用いて 1 回目と 2 回目の全作文を分析した。*jWriter* の計算式による作文評価値（李他 2019）を箱ひげ図で示す。

図 1 は *jWriter* の作文評価値で作文の自動評価を行った結果である。作文評価値に注目すると、N1 と N2 レベルにおいて大きな差が見られる。しかし、実線で書かれた 1 回目の評価結果と点線の 2 回目の評価結果に注目すると、N1 集団と N2 集団のどちらにおいてもほとんど差が見られなかった。この傾向は、作文評価の主要指標である平均文長（1 文に見られる形態素数の平均）と漢語率（文章に含まれる漢語の比率）に関しても同様であった。

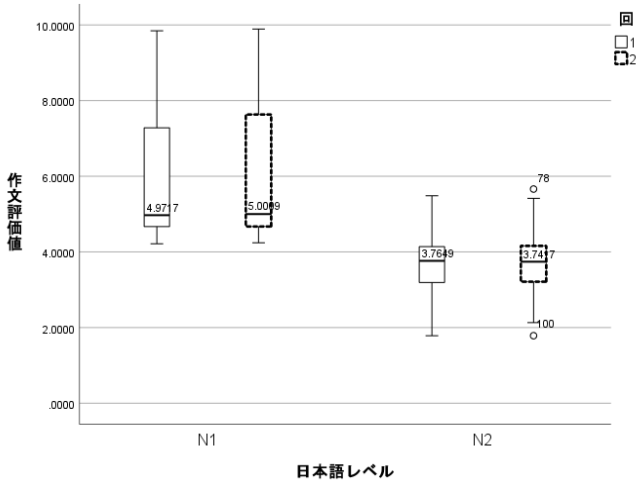

図 1：*jWriter* による作文の自動評価の結果

5.2.3 調査結果２（質的分析）

1 回目と 2 回目の作文を目視で分析したところ、2 回目のほうがよりよい表現になっており、修正が不適切である例はごく少数であった。学習者が注目していた観点は、助詞・副詞・接続詞・指示詞・動詞・文末表現・文体・文の長さ・重複に関し、追加・削除・言い

換えにより修正した言語表現面が中心であった。例えば、小見出し・図表のタイトルを追加した形式面、引用を追加するなど内容構成面に関するものは一例ずつのみであった。

以下は、言語表現面の修正例の一部である。下線は筆者（→学習者による修正の表現）

①助詞：【言い換え】ファッション業界の次に成長も（→が）非常に速いのは

②副詞：【追加】日本において、長時間労働者の割合は年々減少している

③接続詞：【追加】したがって、著作権侵害も少しずつなくなっている。

④指示詞：【追加】このようなコレクションの本は他の本より高いとはいえ

⑤動詞：【言い換え】狂言の中に能が入る（→を入れる）という試みによって

⑥文末：【追加】大学生は（中略）に対する関心が低いことが分かる。

⑦文体：【言い換え】前年に比べて 241 万人増えた（→増加した）

⑧文の長さ：【変更】20 年度の同 4 位は「(中略)」という回答であり、(→ある。)

⑨重複：【削除】音楽が好きな人に対して、CD は精神的~~的~~な高い価値がある。

5.3 考察

本調査で明らかになったこととして、1) チェックリストによる自己修正は量的分析では捉えられないこと、2) 質的分析の結果として、文法や文体などの言語表現面の改善が見られたこと、3)小見出しや図表のタイトルの形式面に関しても部分的な訂正が見られたこと、4) 内容構成面で全体的な構成を入れ替えるものは見られなかったことがあげられる。これは、執筆前にアウトラインの指導を行っているためであると考えられる。

現況において機械評価は言語表現面のみに特化されており、学習者による自己修正も言語表現面が中心である。学習者による自己修正はかなり高度な力を要するものであり必要十分ではないものの、チェックリストが自律的な学習を支援するものであることが明らかとなった。一方で、内容構成面の評価は教師の役割が必要であることが窺われた。

6. まとめ

以上の報告から、*jWriter* を例に機械による自動評価の有用性と限界、教員評価および学習者自身の評価活動の可能性について検討した。機械評価ではフィードバックが難しい内容面や構成面にも意識を向け、大学生の書く力を総合的に向上させるためのツールを提供する方法を検討していく必要があると考える。質疑応答では、教育・学習活動を支援するために機械と人間それぞれに期待することは何か議論したい。

付記：本研究は JSPS 科研費 18K00680 の助成を受けて行われた。

参考文献

ダネル・スティーブンス、アントニア・レビ. 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』 東京: 玉川大学出版部.

エリザベス・F・バークレイ、クレア・ハウエル・メジャー. 2020. 吉田塁（監訳）『学習評価ハンドブック　アクティブラーニングを促す 50 の技法』東京: 東京大学出版.

南風原朝和. 2019.「見失われた記述式の意義：導入の見直しが必要だ」『科学』89(10): 899-904. 東京: 岩波書店.

広瀬和佳子. 2007.「教師フィードバックが日本語学習者の作文に与える影響―コメントとカンファレンスの比較を中心に」『早稲田大学日本語教育研究センター紀要』20: 137-155.

市嶋典子. 2014.『日本語教育における評価と「実践研究」対話的アセスメント：価値の衝突と共有のプロセス』ココ出版.

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: 79 -940. (http://www.lib.kobe-u.ac.jp/repository/81012491.pdf)

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: 26-43. 東京: 凡人社.

伊集院郁子・髙野愛子. 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』東京: アスク出版.

小森和子・李在鎬・長谷部陽一郎・鈴木泰山・伊集院郁子・柳澤絵美. 2018.「教師による評価とコンピュータによる自動評価はどの程度一致するのか―中上級日本語学習者の意見文の評価を対象に―」『2018 年度日本語教育学会秋季大会予稿集』278-283.

李在鎬・長谷部陽一郎・村田裕美子. 2019.「学習者作文の習熟度に関する自動判定と Web システムの開発について」, 李在鎬（編）『ICT×日本語教育』38-53. 東京:くろしお出版.

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: 105-120. (http://www.lib.kobe-u.ac.jp/repository/81012493.pdf)

西岡加名恵・石井英真・田中耕治（編）. 2015. 『新しい教育評価入門―人を育てる評価のために―』東京: 有斐閣

大島弥生・池田玲子・大場理恵子・加納なおみ・高橋淑郎・岩田夏穂. 2005. 『ピアで学ぶ大学生の日本語表現　第 2 版―プロセス重視のレポート作成―』 東京: ひつじ書房.

トムソン木下千尋. 2008.「海外の日本語教育の現場における評価：自己評価の活用と学習者主導型評価の提案」『日本語教育』136: 27-37.

インターネット

jWriter 学習者作文評価システム＜https://jreadability.net/jwriter/＞（2021 年 6 月 7 日）