

日本語ライティングの評価研究

—文献レビューより「評価の統一」に向けた課題を探る—

安達万里江（関西学院大学）
m-adachi@kwansei.ac.jp

【要約】

本稿は、日本の大学の学部留学生を主に対象としたライティングの評価研究より、評価の統一に向けた課題を探ることを目的としている。その背景には、クラス分け等で行う診断的評価と、教員が行う主観的な評価が一致しなかったことにある。そこで、どのような評価観点において不一致が見られるのかを19本の文献レビューによって探る。その結果、評価者による内容面における評価観の不一致が見られることが明らかになった。

1. はじめに

独立行政法人日本学生支援機構「平成29（2017）年度外国人留学生在籍状況調査結果」（平成29年5月1日公開）によると、日本国内の留学生は267,042人となっており、その中でも大学は、大学院の留学生数と合すると最も多く、123,919人であった。筆者の述べる「日本語学習者」（以下、学習者）とは、上記に含まれる学部留学生（以下、留学生）のことであり、「日本語教育現場」（以下、現場）とは「日本の大学」（以下、大学）とする。

所属する大学・学部によって様々であるが、大学組織の一員として、学生の学力の伸びや学習成果等をデータとして示すことが求められる場合がある。その背景には、文部科学大臣が認証する評価機関の評価を受けることが法律で義務付けられ（学校教育法第110条）、大学の認証評価にあたり、教育の「質の保証」が重要視されるようになったことが挙げられる。そのような環境の下、筆者にとって「何を・どう教えるか」と同様、「どのように評価するか」が課題となっている。中でも、レポートやスピーチといったパフォーマンス課題は、主観的な評価になりやすく、「どのように評価すべきか」が目下の課題である。それらの中から本稿ではライティングの評価について取り上げる。

本稿における「ライティング」とは、日本の大学で「アカデミック・ライティング」として取り上げられるレポート、小論文、作文、エッセイ（意見文・説明文）に着目し、メールや手紙文は取り扱わないこととする。

ライティングのパフォーマンス評価について課題であると考える一例として、クラス分け（プレースメント・テスト）等で行う診断的評価について述べたい。日本語教育機関によって様々であると推察するが、目安として取り上げられるのは、「日本語能力試験(JLPT)」や「J-CAT」のスコアであろう。しかし、その診断的評価の結果が、そのままライティングの能力評価につながるとは言えないだろう。現在、「j-Writer」という作文能力を測定できるインターネットサイトがあるが、上述の診断的評価と比べてどうだろうか。また、一日本語教員の主観的評価とは差があるだろうか。そこで、学習者3名

に同意を得た上で、本稿のために作文を書いてもらった（資料1として末尾に添付）。

その結果を表1に示す。

表1 学習者の診断的評価

	J-CAT	JLPT	jWriter
S1	中級後半	N2合格者	上級
S2	中級後半	N1合格者	中級
S3	中級前半	N2相当	上級

表1から分かるように、J-CATのスコア、日本語能力試験およびj-Writerの判定が「中級」一致していないことが分かる。加えて問題であると考ええる点は、一日本語教員で評価者である筆者自身の主観的評価が3名ともに「中級」であったということである。なぜこのような「評価の不一致」が起こってしまうのだろうか。そこで、本稿では「どのような評価観点において不一致が見られるのか」を問題提起とし、これまでの先行研究より文献レビューを行った上で、新たな研究課題を明らかにすることを目的とする。

2. 先行研究

2.1. 文献の選定基準

現在、日本語教育における評価研究は多種多様にあるが、本稿では以下の条件に絞り、文献レビューを行う。以下「文献の選定基準」を表2に示す。

表2 文献の選定基準

上位条件
日本語教育学界において、アカデミック・ライティングの評価研究を組織的、継続的に行い、かつ汎用性のある評価基準を作成している。
下位条件
1) 評価の対象者が「日本の大学の学部在籍する留学生」であり、評価者が「日本語教員」である。
2) 評価項目ごとの評価観点について調査・分析を試みている。
3) 評価者（日本語教員）だけでなく、被評価者（日本語学習者）にも直接還元できそうな研究である。

表2で上位条件として挙げた文献を①、下位条件として挙げた文献を②とし、下記に示す。本稿では文献レビュー1・2とし、計19本取り上げる。（資料2～4として末尾に添付）

①15本：田中他（1998a）～田中（2017）

②4本：伊集院（2017）、伊集院他（2018）、脇田（2016、2017）

留学生を対象とした「アカデミック・ライティング」の評価研究を概観すると、大きな流れは田中他（1998a）をはじめに、現代では伊集院（2017）、伊集院他（2018）、宇佐美（2014,2016）、李（2017）、脇田（2016、2017）等、「日本語学習者の文章を評価する」評価研究がある。その中でも、日本の大

学に在籍する留学生を対象に、20 数年にわたり日本語のアカデミック・ライティングの評価研究を行ってきた田中の研究論文（共著も含む）は、数多くの他の研究者によって取り上げられている。本発表では、まず、1998 年～2017 年の間に発表された論文・著書計 15 本を挙げ、日本語教育におけるライティングの評価研究を概観する。そして、表 2 文献の選定基準の「下位条件」に当てはまった 2016 年～2018 年の論文 4 本の論文を取り上げ、成果と課題をまとめ、新たな研究課題を示す。

2. 2. 文献レビュー 1（田中他；1998a ～ 田中；2017）

2. 2. 1. 日本語ライティング教育における評価研究のはじまり

まず、上述に田中他（1998a）を「はじまり」と記した理由は、日本語教育学界において「信頼性・妥当性が高く汎用性のある評価基準項目表の作成」を目標に掲げた最初の研究論文だからである。質的・量的にデータを収集し、第二言語としての日本語の作文には「正確さ」「構成・形式」「内容」「豊かさ」の 4 因子が必要であること、評価者である日本語教員・日本語母語話者共に重視しているのは「内容面」であること、さらに、どちらの評価者にとって「いい作文」の決定要因になっている可能性があるのは「内容」「主旨の明確さ」「構成」であると述べている。そして、次の課題を「客観的に図るための基準となるガイドラインの作成」としている（田中他；1998b）。

2. 2. 2. 評価基準・評価法の策定

その後、田中・長阪（2004）は、当時までに公開されている日本語・英語の評価基準のレビューを行い、ライティングの評価においては、「プロンプト（課題文）」「評価方法」「クライテリア（評価カテゴリー・評価項目）」「読み手」が重要な要素であると指摘している。そして、課題として 3 つのクライテリア「内容・タスクの達成」「構成・結束性」「言語能力」を基本とした評価基準の作成と評価の実証が必要であるとし、そのためには多くのデータと組織や研究チームを作って研究を遂行することの重要性を挙げている。また、日本語教育全体で共有できる「マルチプル・トレイト評価法」の採用を提案している（田中；2005）。マルチプル・トレイト評価法には、評価の目的に合わせてトレイト（クライテリア）を自由に設定できるという柔軟性があり、目的に応じて使用できるよう、そのための評価基準の検討、ガイドラインの作成、改良が重ねられた（田中・長阪；2006）。

2. 2. 3. 評価基準・評価法の検証

評価基準の検討は、ワークショップ形式で進められた。1 つの共通した評価基準を一致させること、つまり、個人レベルでの評価のずれについての検討が課題となった（田中他 2009）。これまでの評価研究が真正性は高くとも、人間によるパフォーマンス評価には限界のある事が示唆された。さらに、評価の不一致の根本的な原因を探り明らかにすることによって、評価者の意識を変える必要性、変えられる可能性があることが示された（田中・長阪 2009）。

その対応策として、これまでは評価とアンケート調査による分析が主だったが、評価プロセスや決定要因をプロトコル分析によって明らかにする試みを行っている（田中・坪根；2011）。そこで明らかになったことは、1）日本語教師の good writing 観は一樣ではなく、2）優先順位の重み付けの異なることが一因であること、また、同じ評価者であっても 3）評価時（プロトコル）と評価後（メタ認知）は必ずしも一致せず、評価のプロセスは複雑であることが明らかとなった。good writing がどのようなものであるかを学習者に示すためには、評価者の共通認識が必要であると述べている。

2. 2. 4. テキスト完成と「評価一致」に向けた研究の動向

その後、これまで研究の集大成として、1冊のテキスト『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』が発表された（田中・阿部；2014）。テキストは、留学生のみが対象ではなく、日本語母語話者である日本人大学生・大学院生にも使用されることを想定されている。また、巻末には、トレイト別の評価基準シートが添付されている。他方、「評価一致」に向けた研究が進められた。まず、「言語面」においては、これまでのライティング・フィードバック研究より、すべてのエラーを訂正するのではなく、現在の言語レベルで焦点化した訂正が必要であることが示された（田中；2015）。また、「内容面」「構成面」は、アンケート、プロトコル、クラスター分析の結果、評価者によって異なった基準で評価されていることが明らかになった（坪根・田中；2015）。これまでの研究結果から、評価の不一致をもたらす要因は、1) スクリプト（書き手）の L2 のライティング能力、2) 評価基準や評価方法（評価者おのこの優先順位）、3) 評価者（評価に対する姿勢やライティング観）であると示し、「書き手よりも評価者」によるところが大きいと述べている（田中；2016a）。さらに、評価者による3つの観点と6つの「型」があることを紹介している。また、評価に不一致（ばらつき）が起こった際の対応策について言及されている（阿部・田中；2016）。その傾向として、ばらつきが小さければ「多数派採用」、ばらつきが大きければ「鶴の一声」が多いというパターンが確認された。今後「評価の統一」のプロセスに関する研究が進められることで、評価者の多様な価値観を尊重しつつ行われる評価活動のあり方の解明が新たな課題であると述べている。

2. 2. 5. 近年の研究動向

近年の論文（単著・共著の担当箇所）のテーマは「構成」である（田中；2016b）。日本語ライティング教育において、L1・L2 関係なく「構成」は教えるべきという主張で、例えば日本語母語話者だからと言って、特に規範となるわけではないことを指摘している。また、学習者の母語別の分析も行っている（田中他；2017）。

他方で、共同研究プロジェクトの動向は HP 上で公開されている。2018 年現在は、「日本語ライティング評価の支援ツール開発：「人間」と「機械」による評価の統合的活用」というテーマで、第2言語としての日本語ライティング評価及びライティング教育の支援ツール開発が進められている。では、他の研究者達はどのように評価研究を行っているのか。

続いて、表2 文献の選定基準によって選んだ伊集院（2017）、伊集院他（2018）、脇田（2016、2017）を概観し、最後に、これまでのライティングの評価研究における共通点と今後の課題について述べる。

2. 3. 文献レビュー2（伊集院；2017、伊集院他；2018、脇田；2016、2017）

2. 3. 1. 新たな分析方法から見る評価研究

脇田（2016）では、評価の差異が教員間のみならず、教員・学習者間にもあることに注目し、教員による評価判定の差異を小さくするために、ライティングの成果物に対する評価として、評価観点と段階別の評価基準を明示した「ループリック評価」を数値化し検証している。その結果、教員個人の基準による評価と、ループリック評価においては70%前後が一致したことが分かった。その評価の中で、「文法」の比重が最も軽く、「構成」よりも「内容」（論理の一貫性、独自性）の方が、比重が重いことが明らかになった。また、学習者の自己評価（ループリック評価）が、教員による評価と一致するところが多かったとも述べている（脇田；2017）。

他方、伊集院（2017）は、上述にもある田中他（1998a）、坪根・田中（2015）を先行研究として取り上げ、「よい作文」の条件は一律に規定できるものではないが、近年急速に整備が進んでいる「作文コーパス」や「テキストマイニング」等を用いて作文評価の量的研究を行うことにより、「よい作文」の条件を精緻化していくことの可能性を指摘している。中でも「評価観点」に着目し、大学教員による作文評価の観点を探索的に抽出（M-GTA）した上で、その特徴を分析している（伊集院他；2018）。その結果、評価のコメント量は、「内容」>「言語」>「形式」の順であり、内容的側面が最も重視されていること、評価の上位群では「内容」と「構成」、中位群では「言語」、下位群では「言語」と「形式」に比重が置かれていることが明らかになった。

3. おわりに

以上よりまとめると、「どのような評価観点において不一致が見られるのか」という問題提起に対し、「評価の一致」は困難だが「評価の統一」のためのヒントがいくつかあった。例えば以下の3点が挙げられる。

- ・評価基準シートやフローチャートシートの作成（田中・阿部；2014）
- ・トレイトごとの細かな分析（田中他；2017、伊集院他；2018）
- ・ループリックによって評価を可視化し、学習者側にも研究成果を還元することで授業の改善点を把握する。（脇田；2016）

また、上記表1に示したような「評価の不一致」が起こる原因として、評価者による「内容」面における評価観の不一致があると仮定できる。具体的には、以下のような比重であると考えられる。

評価が高い：内容>構成>言語

評価が低い：言語>構成>内容

しかし、内容面だけに焦点を当てた研究が未だに深められていないことが分かった。今後は、「内容」面での量的・質的な調査・分析が必要であろう。

日本語能力試験やJ-CATでは上級であっても、ライティング能力では中級というような不一致が起こる要因やその改善策も、内容面を探究することで、新たな答えが見つかる可能性があると考えられる。そのためには、日本語教育学界の枠を超えた研究成果を概観し、新たな研究手法の模索が必要であると考えられる。今後も引き続き、ライティングに関する評価研究を行っていきたい。

参考文献

- (1) 阿部新・田中真理(2016) 「グループによるライティング評価における個人評価点の統一パターン」宇佐美洋編『評価を持って町に出よう「教えたこと・学んだことの評価」という発想を超えて』pp.54-69.
くろしお出版
- (2) 伊集院郁子（2017）「作文と評価 日本語教育的観点から見たよい文章」李在稿編 『文章を科学する』pp.38-57,ひつじ書房
- (3) 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」Ishikawa, S. (Ed.).

Learner Corpus Studies in Asia and the World Vol.3. Papers from LCSAW2017. Kobe, Japan: Kobe University.
pp.159-176.

- (4) 宇佐美洋 (2014) 『「非母語話者の日本語」は、どのように評価されているか 評価プロセスの多様性をとらえることの意義』ココ出版
- (5) 宇佐美洋編 (2016) 「『評価を持って町に出よう「教えたこと・学んだことの評価」という発想を超えて』くろしお出版
- (6) GoodWriting.jp <https://goodwriting.jp/wp/> 田中真理 (研究代表者) 「日本語ライティング評価の支援ツール開発: 「人間」と「機械」による評価の統合的活用」科学研究費補助金プロジェクト (研究課題番号: 26284074) (2018年8月21日)
- (7) J-CAT日本語テスト<http://www.j-cat.org/> (2018年8月21日)
- (8) jWriter <https://jreadability.net/jwriter/en> (2018年8月21日)
- (9) 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
- (10) 田中真理、阿部新、影山陽子、佐々木藍子、坪根由香里(2017) 「ヨーロッパ日本語学習者のライティング(エッセイ)分析: 総合的評価とマルチプルトレイト評価結果を参照して」『ヨーロッパ日本語教育論集』22,pp.75-95.
- (11) 田中真理 (2005) 「日本語教育におけるライティング評価」『日本の教育年間 2005年版』くろしお出版
- (12) 田中真理 (2015) 「ライティング研究とフィードバック」大関浩美編『フィードバック研究への招待』pp.107-138. くろしお出版
- (13) 田中真理 (2016a) 「パフォーマンス評価はなぜばらつくのか? アカデミック・ライティング評価における評価者の『型』」宇佐美洋編『評価を持って町に出よう「教えたこと・学んだことの評価」という発想を超えて』くろしお出版
- (14) 田中真理 (2016b) 「日本語教育の立場から L1, L2 双方向から考える日本語アカデミック・ライティング: 「構成」面について」『ESL Writing in East Asia: Practice, Perception and Perspectives』pp.134-153. Shobi Printing Co. Ltd
- (15) 田中真理・坪根由香里・初鹿野阿れ (1998a) 「第二言語としての日本語における作文評価基準—日本語教師と一般人の比較—」『日本語教育』96, pp.1-12.
- (16) 田中真理・坪根由香里・初鹿野阿れ (1998b) 「第二言語としての日本語における作文評価—「いい」

- 作文の決定要因一」『日本語教育』99, pp.60-71.
- (17) 田中真理・坪根由香里 (2011) 「第二言語としての日本語小論文におけるgood writing評価—そのプロセスと決定要因」『社会言語科学』第14巻第1号, pp.210-222.
- (18) 田中真理・長阪朱美 (2004) 「日本語と英語を目標言語とするライティング評価基準の展望：第二言語としての日本語のライティング評価基準作成に向けて」『第二言語としての日本語の習得研究』7号 pp.214-253.
- (19) 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」『世界の言語テスト』pp.253-276, くろしお出版.
- (20) 田中真理・長阪朱美 (2009) 「ライティング評価の一致はなぜ難しいか—人間の介在するアセスメント—」『社会言語科学』第12巻第1号, pp.108-121.
- (21) 田中真理・長阪朱美・成田高宏・菅井英明 (2009) 「第二言語としての日本語ライティング評価ワークショップ—評価基準の検討—」『世界の日本語教育』19, pp.157-176.
- (22) 坪根由香里・田中真理 (2015) 「第二言語としての日本語小論文評価における「いい内容」「いい構成」を探る—評価観の共通点・相違点から—」『社会言語科学』第18巻第1号, pp.111-127.
- (23) 独立行政法人日本学生支援機構 (2017) 「平成29(2017)年度外国人留学生在籍状況調査結果」
https://www.jasso.go.jp/about/statistics/intl_student_e/2017/index.html (2018年8月1日)
- (24) 日本語能力試験<https://www.jlpt.jp/> (2018年8月21日)
- (24) 李在稿 (2017) 『文章を科学する』ひつじ書房
- (25) 脇田里子 (2016) 「ライティング・ルーブリックの実践」『コミュニカーレ』第5号, pp.21-50
- (26) 脇田里子 (2017) 『思考ツールを利用した日本語ライティング』大阪大学出版会

【資料1】課題作文：大学に来てからの生活について

S1

A 大学は私の最初の大学ではないが、この大学に来てから様々な経験がもたらされた。

もっとも深い感慨は悪い経験からもらったかもしれないので、最初にそのことについて説明したいと思う。

A は私の住んでいるマンションからとても遠くて、毎日通学することは時間がかかる。また、大学に入った後、バイトも始めるので、一日は24時がたりないぐらいいいそがしい。そのため、家族や恋人との連絡する時間も少なくなってだんだんストレスが多く蓄積するようになった。また、私は恥ずかしい男の子なので、初期 友達が誰もいなくて心細かった。

しかし、時間が過ぎた後、友達もできてこんな孤独感も下がった。日本語能力や英語能力も改善できたのでちょっとうれしいと感じられた。

今後、また たくさん 苦労やチャレンジを越さなければならないと思う。いつも前方に向かって一所懸命頑張るしかないわけだ。

S2

私は A 大学に来てからまた半年ですが、生活は前より大きく変わったと思っています。

最初は日本の大学で友達ができるとは思いませんし、自分もある距離感を持って他人と話しました。自分と違う文化を持っている人と友達になるのは難しいと思います。だからいつも学校が終わってからすぐ家に帰りました。

しかし、二か月経ったら、私の考えは少く変わってきました。自分と違う文化を持っている人と友達になるのほうの方が面白いです。それはいつも自分がなれるスペースにいるこそ、進めないからだ私と考えています。今は日本人の友達もできるし、外国からの友達もできました。放課後も友達と一緒に帰って、前よりもっと楽しくなりました。

また、前が通っていた日本語学校はほぼ中国や台湾からの留学生で、放課後でも、中国語で話したが、今の学校は毎日違う人と出会って、日本語で話さなければならないので、自分もいつものスペースから独立しましたと思って、少しずつ成長しましたと思っています。

S3

日本で初めて来てからの生活について話をしたいと思います。私は英語ですべての授業を取りましたので、最初はひらがなもカタカナも区別できなかったほど日本語のレベルがひくいでした。

最初10ヶ月くらいは非常に生活するのが大変でした。例えば、一人暮らしをするため、不動産に行って部屋を探した時は、日本語が上手く出来なくて基本的な漢字や文章を読めるのが大変だったので様々な問題が起こりました。なんとなく部屋を探して、契約の時も家賃や水道費やガス・電気代の払い方も知らなかったもので、その払い方を調べる時間が非常にかかり、誰かの補助が必要でした。その上、誰から何の手紙ともらってるのか内容は何かということが理解できなくて無視してしまふもありました。でも、どんどん日本語力が上がると、人たちの会話やニュースの内容を理解できるようになりました。新しい部屋を探す時も前と比べて上手い条件の部屋を見つけて契約したので、さすがに日本語が上手になると生活しやすくなるんだと覚えました。

【資料2】田中論文（共著を含む）文献レビュー

番号	年	単共	タイトル	予備調査	調査方法	評価者(数)	分析方法	学習者のレベル(属性・数)	データ数(元)	プロンプト(課題文)	文章量	結果	課題
1	1998a	共	「第二言語としての日本語における作文評価基準—日本語教師と一般人の比較—」	有	順位付け 質問紙調査	T:日本語教師(71名) G:一般人(71名)	因子分析:評価の観点22項目→4因子抽出「正確さ」「構成・形式」「内容」「豊かさ」	初級—上級(6)	6(112)	日本/東京に来て驚いたこと	400-800字	4因子の抽出:「正確さ」「構成・形式」「内容」「豊かさ」重視 T/G共に「内容面」重視 TはGより「構成・形式」を重視	「豊かさ」に積極的に目を向けた評価項目の検討 信頼性・妥当性が高く、汎用性のある評価基準項目の作成
2	1998b	共	「第二言語としての日本語における作文評価—「いい」作文の決定要因—」	無	順位付け アンケート調査 1・2	T・G各々70名(順位) T・G各々55名(アンケート)1年後	調査1:スピアマンの順位相関係数/フリードマンの検定 調査2:作文ABの使用文数・文型等/評価の平均値と差の検定(t検定)	初級—上級(6)	6・2(112)	日本/東京に来て驚いたこと	400-800字	1) T/G順位付けかなり一致2) 「正確さ」と「読みやすさ」重視の2グループ3) 2) 両方「内容」「主旨の明確さ」「構成」が「いい」作文の決定要因になっている可能性がある	客観的に因るための基準となるガイドラインの作成が必要
3	2004	共	「日本語と英語を目標言語とするライティング評価基準の展望:第二言語としての日本語のライティング評価基準作成に向けて」	無	現在公開されている日本語と英語の評価基準のレビュー		結果:ライティング評価においては「プロンプト」「評価方法」「クライテリアとレベル」「読み手」が重要な要素					課題:「内容・タスクの達成」「構成・結束性」「言語能力」の3つのクライテリアを基本とした評価基準作成、評価の実証、多くのデータが必要のための組織や研究チームを作つて遂行すべき	
4	2005	単	「日本語教育におけるライティング評価」	無	1) ライティング評価基準・方法 2) 現在までに公開されてきた日本語ライティング評価基準について概観 3) ライティング・テスト、評価基準の今後の方向について					今後の方向:1) 日本語教育において現在まず必要とされるのは、日本語教育全体で共有できる「ライティング評価基準」2) マルチプル・トレイト評価法を採用すれば、各機関・コースでそれぞれの目的に合わせてトレイトを調整することも可能3) この新しい「ライティング評価基準」の作成、普及には、日本留学試験とは別の、ライティング能力を測る大規模テストを開発することが最も近道			
5	2006	共	「第二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」	有	評価基準作成過程1:分析的評価法 過程2:評価基準1の見直し(学生のコメントも参考) 評価基準2作成 過程3:評価基準2を学生・日本語教師2名のコメントを参考に改良	日本語教師に渡す評価のセット①ライティング評価のガイドライン②プロンプトと作文③評価基準④プロンプト別・ガイドライン⑤レベル別・サンプル作文⑥評価結果記入法⑦フィードバック票	プロンプト別・評価の対象ライティング JSL大学生43・29 日本人大学生7・7	「ファーストフードとスローフード」「結婚(少子化の問題)」	600字程度	マルチプル・トレイト評価法には、評価の目的に合わせてトレイト(クライテリア)を自由に設定できるという柔軟性がある。(ex. 記述の簡略化)	評価基準は、それぞれの目的に応じたものが必要であるが、本研究では、それらの基本となる「評価基準」の作成を目指している。		
6	2009a	共	「第二言語としての日本語ライティング評価ワークショップ—評価基準の検討—」	有	評価・アンケート調査	元データ(研)4名 ワークショップ参加者(日本語教師定員)20名	クロンバックの α 係数 平均値と標準偏差、相関係数、平均値、最大値、最小値、最頻値、中央値、参考評価点	初級後半(日能試4級) — 上級(日能試1級) 国立大学	各3(76/66)	「ファーストフードとスローフード」 「通学授業と遠隔授業」	600字程度	ワークショップは最低2日必要 1つの共通した評価基準を一致させること(評価者間信頼性を高めること)を目指していかなければならない	個人レベルでのずれについての検討(個人差の是正・減少)
7	2009b	共	「ライティング評価の一致はなぜ難しいか—人間の介入するアセスメント—」	無	アンケート調査・評価者間ミーティング	8名の日本語教師	トレイト(26種トレイト別・平均値、標準偏差、内的一貫性を表す α 係数)、プロンプト、レベル、ライティングの潜在能力、評価者の個人的要因の観点から分析	初級後半(日能試4級) — 上級(日能試1級)	各26	「ファーストフードとスローフード」 「通学授業と遠隔授業」	600字程度	本研究において、真正性は高くとも、人間によるパフォーマンス評価には限界のある事が示唆された。	(評価の)不一致の根本的な原因を探り明らかにすることによって、評価者の意識を変える必要性、変えられ可能性がある。(プロトコル分析中)
8	2011	共	「第二言語としての日本語小論文におけるgood writing評価—そのプロセスと決定要因」	有	総合的評価 アンケート調査	10名日本語教師	1) 上位2小論文の計4論文に絞って分析2) アンケート調査に基づく分析3) プロトコルを中心とした分析	初級後半—上級	各26(127/117)	「ファーストフードとスローフード」「通学授業と遠隔授業」各6編	600字程度	1) 日本語教師のgood writing観は一律ではない2) 主要要素の優先順位重み付けの異なることが一因3) プロトコルとメタ認知は必ずしも一致せず、その評価方法などに関する検討、共通認識が必要	good writingがどのようなものであるかを学習者にしめすためには、今後、その評価方法などに関する検討、共通認識が必要
9	2014	共	『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』	無									対象者:主に日本人大学生1年生向け初年次教育や2年生以降の文章執筆演習などの授業で使用されることを想定。*にも対応可 *卒業論文を書こうとする学部生や修士論文を書こうとしている大学院生 *日本語力に問題ない外国人学生
10	2015a	単	「ライティング研究とフィードバック」	無			最近のライティングFB研究の動向(L2ライティング研究とSLA研究の2つのアプローチ)					・間接よりも直接フィードバックのほうが効果的であると研究のほうが優勢 ・エラーすべてを修正しても徒勞に終わる可能性が高い、訂正する言語項目を決めて、繰り返して訂正し、確実に定着することを旨とした方がよい。	特有、繰り返して出てくるようなエラー、現在の言語レベルで直しておくべきエラーなど、それらに集中した方が効果的、効果的である。(焦点化した訂正)
11	2015b	共	「第二言語としての日本語小論文評価における「いい内容」「いい構成」を探る—評価者の共通点・相違点から—」	有	総合評価 アンケート マルチプル トレイト評価	10名日本語教師	分散分析 クラスター分析(ワード法)クラスター別平均値・標準偏差	初級後半—上級	6(117)	通学授業と遠隔授業	600字程度	「内容」「構成」は評価者によって異なった基準で評価されていることが明らかになった。(表6,7) 「いい内容」の要因:1) 主張の明確さ2) 説得力のある根拠をわかりやすく示すこと3) 全体理解の助けになる書き出し4) 一般論への反論の視点的面白さと例示に関しては評価が分かれた。	(統計的に)示すことができたところのものに示す。評価者は10名で各評価者のグループの人数は少ない
12	2016a	単	「パフォーマンス評価はなぜばらつくのか?アカデミック・ライティング評価における評価者の『型』」	無			・評価の不一致をもたらす要因:A) スクリプトB) 評価基準や評価方法C) 評価者 ・3つの観点(①評価を統一しようとする際に分る傾向②マルチプルトレイト評価をする際に認められる傾向③総合的評価をする際に認められる傾向)から見た評価者が使う6つの「型」(①マイウェイ型②柔軟型③運動型④独立型⑤前提型⑥非前提型)表1					自己の傾向(「型」)や他社の評価についての考え方を認識するためには、評価者間ミーティングが有効である。	評価に向き合う際、自分自身がどのような「型」を使っているのか意識し、自己の評価を見つめ直すことによって、評価の質を高めることが可能となるのではないだろうか。
13	2016b	共	「グループによるライティング評価における個人評価点の統一パターン」	無		2小論文評価(スコア・自由記述) ワークショップ(90分)参加者40名 承諾33名(7G)	個人評価とグループ評価結果の平均値・標準偏差	中上級	2	ファーストフードと家庭料理	600字程度	ばらつきが小さいと「多数派採用」ばらつきが大きいと「鶴の一声」が多いというパターンが見られた	「少数派採用」のトレイト:「読み手」「日本語B:適切な」の考察 「評価の統一」のプロセスに関する研究を進めていくことで、評価者の多様な評価観を尊重しつつ行われる評価活動のあり方を明らかにしていきたい。
14	2016c	単	「日本語教育の立場から L1, L2両方向から考える日本語アカデミック・ライティング:「構成」面について」	無		田中・久保田(2014)、Kubota&Tanaka(2014)「構成」面の分析、田中・迫田・野田(2015)ヨーロッパ言語話者との違い、Tanaka&Kubota(2015)「マクロ構成」のタイプ分類を加えてまとめた論文 (1)マクロ構成(序論・本論・結論)(2)一貫性(序論と結論の呼応)(3)「比較・対照」と「論証」(意見)モードの展開法について検討		J-CAT互換表 上級前半～中級後半	プロンプト1: J51, C53, E44 プロンプト2: J50, C53, E44	「ファーストフードと家庭料理」 「通学授業と遠隔授業」	600-800字程度	L1, L2に関係なく、ライティング教育においては「構成」は教えるべきものであることが確認されたと言えよう。(例)Jは母語話者だからと言って、特に構成面に関しては規範とならない	・「モードによる展開法」については、さらなる研究が必要 ・L1の構成をL2ライティングにどのように生かせるか(英語Ossay writingの構成をどこまで日本語ライティングに生かせるか)要調査・研究
15	2017	共	「ヨーロッパ日本語学習者のライティング(エッセイ)分析:総合的評価とマルチプルトレイト評価結果を参照して」	無		5名の評価者	マルチプルトレイト評価(MT評価)と総合的評価(H評価)の比較を行い、両評価結果において2レベル(L)以上の差があるエッセイを取り上げ検討	J-CAT81・301 SPOT37-80	54編 ドイツ語8名、スペイン語13名、フランス語5名、ハンガリー語9名、トルコ語8名、ロシア語9名、ロシア語母語話者11名	ファーストフードと家庭料理 (I-JASエッセイ)	600-800字程度	全体での一致率(MT=H):46.3%、L3:MT=Hが83.4%、L2、MT=HとMT<H:各40%、MT>H:各20%、L4:MT>Hが77%、L5:MT>Hのものは見られず、MT=TとMT<Hが半々 ヨーロッパ日本語学習者の「目的・内容」における特徴:自身の「意見」を強く主張する姿勢や、「読み手」に対して呼びかけ、議論を促したり論じたりするような姿勢が見られる(田中、2016b)。	ヨーロッパの母語での口頭やライティングのレトリック教育や思考法が関係しているのではないかと推察される。

【資料3】脇田論文 文献レビュー

著者	年	タイトル	概要	研究目的	分析の流れ	分析方法	結果
1 脇田里子	2016	「ライティング・ルーブリックの実践」	学習者のフィードバックの不十分さに対する不満や教員による評価判定の際を小さくするために、ライティングの成果物に対する評価として、評価視点と段階別の評価基準を明示したルーブリック評価を検証する。	ルーブリックによる評価を導入する際の留意点や成果を明らかにすること (表5のライティング・ルーブリックを使用)	(ルーブリックの) 全体評価の数値化 表8レポート評価の方法 教員・学習者による「個人基準による評価」A～Dの4段階評価、ルーブリック、コメントの比較・考察	p38: 数値に置き換え算出	教員評価: 教員個人の基準による評価とルーブリック評価において70%前後が一致した。(概ね相関していると言えよう) 「文法」の比重が最も軽く、「構成」よりも「内容」の方が比重が重かった。 「内容」=「論理の一貫性」「独自性」 学習者評価: ルーブリック評価の方が教員評価と一致するところが多い。
2 脇田里子	2017	『思考ツールを利用した日本語ライティング』	本書は2014年12月に提出した博士論文をもとに、大幅な加筆・削除などの修正を行い、1冊にまとめたもの	リーディングと連携したアカデミック・ライティングの実践研究	第6章: 授業実践の結果と考察 1) 日本語教員のレポート評価 2) 学習者が作成した思考ツール 3) 学習者の自己評価 4) 学習者に対するアンケート	・比較しやすくするため、ルーブリックの「総合」評価を数値化している。 ・学習者が6名、教員が5名と少人数であるため、データの統計処理は行わない。	3) 学習者の自己評価: ルーブリック評価では、実践前より実践後の方が、教員による評価に近い自己評価を下した学習者が多くなった。

【資料4】伊集院論文 (共著を含む) 文献レビュー

著者	年	タイトル	概要	研究目的	分析の流れ	分析方法	結果
1 伊集院郁子	2017	「作文と評価 日本語教育的観点から見たよい文章」	・日本語教育で行われてきた作文研究の振り返り ・作文評価のケーススタディ	日本人大学生と日本語学習者が執筆した意見文の評価の分析を通し、大学教員が共通して「よい」または「悪い」と考える意見文、評価者によって評価が分散する特徴を探る。	①意見文40編の5段階評定の結果 ②評価記述文 ③評価に関するフォロー・アップインタビューおよびディスカッション (計3時間) の記録	①平均値と標準偏差 ②KH Corderを用いて語の使用頻度および共起関係を分析 ③高得点群、低得点群、分散群の特徴を分析	本ケーススタディの日本語教育への応用: 文中・文末の文体の統一、連体修飾節を利用した情報の詳述、結束性をもたらず言語表現や抽象的な語彙の産出、多様な文末表現の使い分けを教えるだけでなく、それらが文章の内容、構成にどのような影響を与えるか意識させることで「よい意見文」につながるものが期待できるだろう。
2 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」	大学教員による作文評価の観点を探索的に抽出したうえで、その特徴を分析	大学教員による文章評価の実態を探る; 留学生を対象とする日本語教育、母語を話者を対象とする文章表現教育のみならず、教員を対象とする評価トレーニングの面においても意義があると思われる。 RQ1: 大学教員44名による評価のコメントから、どのような評価の観点が抽出できるか。 RQ2: 作文の良し悪し (評点の高低) は、評価の観点到影響を与えるか。	①一般公開されている作文コーパスなどから30編の意見文を評価対象データとして抽出 ②大学教員44名に5段階での評定、および評価を決定づけた「良い点」と「悪い点」の記述を依頼 ③評定値の分析結果に基づき、30編を上位群、中位群、下位群に分け、各群から評定値が最も近い日本語母語話者および日本語学習者1ペアずつ、計3ペアの全6編を抽出 ④全6編に付された44名の評価コメントの集約、コード化、分析	評価の観点を抽出 (M-GTA)	・評価の観点は25項目、8カテゴリに集約された。 ・評価のコメント量は「内容」>「言語」>「形式」の順であり、内容的側面が最も重視されていることがわかった。 ・上位群では「内容」と「構成」、中位群では「言語」、下位群では「言語」と「形式」に比重が置かれていた。 ・日本語教員は「言語」と「構成」、専門教員は「形式」により多く着目していた。