

学位論文要旨

日本語学習者の作文に対する評価
— 日本人教師と中国人教師の比較 —

広島大学大学院 教育学研究科

文化教育開発専攻 日本語教育学分野

方 正

I 論文題目

日本語学習者の作文に対する評価－日本人教師と中国人教師の比較－

II 論文構成（目次）

第1章 序論

- 1.1 本研究の背景と目的
- 1.2 本研究の構成

第2章 先行研究と本研究の課題

- 2.1 属性の異なる評価者による日本語作文評価
- 2.2 評価者の作文評価行動
- 2.3 母語話者教師と非母語話者教師による作文評価の比較
 - 2.3.1 誤用に対する評価の比較
 - 2.3.2 作文全体に対する評価の比較
- 2.4 先行研究の問題点と本研究の課題

第3章 日本語母語話者教師と中国語母語話者教師の評価行動（調査Ⅰ）

- 3.1 目的
- 3.2 方法
 - 3.2.1 評価用作文
 - 3.2.2 評価尺度の項目選定
 - 3.2.3 評価者
 - 3.2.4 手続き
 - 3.2.5 分析方法
- 3.3 結果と考察
 - 3.3.1 「読む」段階の評価行動
 - 3.3.2 「評定」段階の評価行動
- 3.4 まとめ

第4章 日本語母語話者教師と中国語母語話者教師の評価モデル（調査Ⅱ）

- 4.1 目的
- 4.2 方法
 - 4.2.1 評価用作文と評価尺度
 - 4.2.2 評価者

- 4.2.3 手続き
- 4.2.4 分析方法
- 4.3 結果と考察
 - 4.3.1 因子パターン
 - 4.3.2 総合的評価に影響を与える要因
 - 4.3.3 評価項目の重視度と困難度
 - 4.3.4 作文指導による評価の厳しさ
- 4.4 まとめ

第5章 結論

- 5.1 本研究のまとめ
- 5.2 日本語教育における作文教育への示唆
- 5.3 今後の課題

参考文献

資料

第1章 序論

学習者のパフォーマンスを評価するためには、評価者が介在しなければならない。評価者が加わることによって、評価は必然的に評価者の主観的な要素を反映するものになり、評価の信頼性と妥当性に影響を与えることが多くの先行研究によって指摘されている (Johnson & Lim, 2009 ; McNamara, 1996 ; McNamara, 2000)。本研究は評価者の母語の違いによる評価の相違を検討する。質的調査では、評価者が実際にどのように作文を評価しているのかを調べるため、発話思考法を用いる。量的調査においては、多くの先行研究で扱われてきた日本語作文の「説得力」「分かりやすさ」といった評定に用いる個別の評価尺度だけでなく、総合的な評価尺度とそれに影響を与える分析的な評価尺度を全体的に取り上げる。2つの調査結果により、日本人日本語教師（以下、JST）と中国人日本語教師（以下、CST）の評価における問題点を発見し、その改善策を提案することが可能となる。

第2章 先行研究と本研究の課題

第2章では、日本語教育における属性の異なる評価者による評価を取り上げた研究、英語教育における評価者の作文評価行動を取り上げた研究、母語話者教師と非母語話者教師による作文評価の比較した研究をまとめ、その課題を考察した。

日本語の作文評価研究では、これまで評価者の教師経験の有無、及び評価者の母語が評価に影響する要因として分析されてきた。先行研究の結果から、教師経験、または母語の異なる評価者が作文を評価する際に重視する項目や、評価項目についての判断基準が異なることがわかる。評価者の作文評価行動に関する先行研究は、英語の作文を対象とするものが多く、未熟な評価者は熟練した評価者より言語使用に多く注目し、厳しく評価していることがわかる。それに対し、熟練した評価者は作文の全体的な要素をより重視し、作文全体の言語運用、内容、構成に注意を向けていることがわかる。母語話者教師と非母語話者教師による作文評価の比較した研究の結果から、母語話者（以下、NS）教師は理解が可能かどうかを重視するのに対し、非母語話者（以下、NNS）教師は文法規則の違反を基準とする傾向がある。また、NS教師よりNNS教師のほうが全体的に厳しく評価する傾向がうかがえた。

先行研究が残した課題は3点にまとめられる。

- (1) 非母語話者教師による日本語の作文評価研究が行われていない点である。
- (2) 多くの先行研究が評価者のつけた評定値だけを用いて分析している点である。
- (3) 日本語作文の評価モデルが解明されていない点である。

以上のような先行研究の課題を踏まえ、本研究では、以下のように研究課題を設定した。

- 1) JSTとCSTによる作文の評価行動にどのような違いがあるのか。
 - 1-1 作文を読む際に、JSTとCSTの評価行動にどのような違いがあるのか。
 - 1-2 評価尺度について評定する際に、JSTとCSTの評価行動にどのような違いがある

のか。

- 2) JST と CST による作文の評価モデルにどのような違いがあるのか。
 - 2-1 JST と CST の因子パターンが異なっているかどうか。
 - 2-2 JST と CST が作文評価をする際、総合的な評価尺度（作文全体に対する印象、読み手を意識しているかどうかに対する印象）の評定値を予測する要因が異なっているかどうか。
 - 2-3 作文評価において、JST と CST が重視する項目、評価が困難な項目が異なっているかどうか。
 - 2-4 JST と CST の作文評価の厳しきは異なっており、その評価には作文指導経験が影響しているかどうか。

第3章 日本語母語話者教師と中国語母語話者教師の評価行動（調査Ⅰ）

第3章では、JST と CST が作文評価する際の評価行動の違いを明らかにすることを目的とした調査Ⅰの結果をまとめた。調査では、発話思考法を用いて JST10 名と CST10 名に4本の作文を評価させ、その評価行動を抽出した。評価者の評価過程を「読む」段階と「評定」段階の2つの段階に分け、質的に分析した。

まず、「読む」段階の評価行動を、評価者が注目する作文の要素によって「自己モニタリング」「内容」「構成」「日本語」の4つの上位カテゴリーに分け、次に、この4つの上位カテゴリーをさらに下位カテゴリーに分けた（以下、下位カテゴリーを「日本語：修正」のように示す）。上位カテゴリーの分析結果から、CST の「日本語」の出現率のほうがやや高いということがわかった。複数の JST が「作文の分かりやすさに支障をきたさない場合、日本語の正確さを特に重視しない」と述べていた。それに対し、CST は「作文を評価する際、日本語も内容も重視する」と回答しており、「誤用を容認する」という回答は得られなかった。先行研究においても、NS 教師より NNS 教師のほうが文法を重視しており（Lee, 2009 ; Marefat & Heydari, 2016）、母語話者が「理解可能」を評価の基準としていることが報告されている（Hughes & Lascaratou, 1982 ; Hyland & Anan, 2006 ; Schmitt, 1993）。本調査の結果は先行研究と一致していると言える。

下位カテゴリー「日本語：修正」は、CST よりも JST の出現率の平均値が高く、ばらつきも大きかった。CST は日本語の不適切な箇所を指摘することにとどまっているのに対し、JST は作文の日本語の不適切な所を指摘した上で修正することがわかった。その原因として、以下の2点が考えられる。1点目は、JST と CST の日本語に関する知識が異なっているという点である。2点目は、CST が中国語母語話者であり、作文の書き手と日本語の学習経験を共有しているという点である。従来の研究では、正確さの評価については教師の母語による評定値の厳しきの違いしか論じられてこなかったが（Kobayashi, 1992 ; Lee, 2009 ; Marefat & Heydari, 2016）、両者の評価行動を質的にみることで、正確さに対する評価行動の違いが明らかになった。また、「内容」の下位カテゴリー「内容：サポートの質」は、CST より JST

のほうが出現率が高いことがわかった。JST と CST の内容に対する評価行動の違いは、JST と CST の文化的背景に起因すると考えられる。「構成」の評価行動の分析結果からは、CST も JST も文と文のつながりにかかわるマイクロ構成を重視している一方で、JST は作文全にかかわるマクロ構成をも重視していることがわかった。英語 NS 教師と英語 NNS 教師（日本語母語）による作文評価を比較した Conner-Linton (1995) は、作文を評価する際に重視する項目として、英語 NS 教師が作文全体レベルでの構成を挙げているのに対し、NNS 教師は段落レベルでの構成を挙げていることを指摘しており、本研究と類似した傾向を示していると言える。マイクロ構成は文と文のつながりを対象としているため、「日本語」（接続詞や指示詞の適切な使用）と密接にかかわっている。それに対し、「首尾一貫性」などの作文のマクロ構成は書き手の表現意図を理解するために重要である。JST と CST の「構成」に関する評価行動に違いが生じた原因として、以下 2 点が挙げられる。1 点目は、両者の母語のレトリックパターンが異なることである。2 点目は、JST と CST が着目する作文の側面が異なることである。最後に、「自己モニタリング」については、JST も CST も全体の評価行動を占める割合が低く、個人差も大きいことがわかった。

「評定」段階の評価行動は、評価項目別に分析した。JST と CST の評定値の差が大きい作文の項目について、評価者が評定値をつける際のプロトコルを取り上げ、その差が生じた原因について考察した。分析の結果、作文Ⅱの「日本語」に属するいずれの評価項目においても、JST より CST の評定値が低いことがわかった。CST は不適切な語彙使用の箇所をその頻度にかかわらず低く評定していた。それに対し、JST は日本語が間違っている箇所を指摘しても、作文全体の日本語の質を考え、低く評定していなかった。NS 教師と NNS 教師の評価を比較した多くの先行研究は、「言語形式」について NNS 教師のほうが厳しいと結論づけている (Hughes & Lascaratou, 1982 ; Lee, 2009 ; Marefat & Heydari, 2016 ; McCretton & Rider, 1993 ; Porte, 1999 ; Schmitt, 1993 ; Sheorey, 1986)。本研究の結果は先行研究と一致していると言える。作文Ⅲの「内容」に属する評価項目においては、CST より JST の評定値が低い傾向がみられた。「文章全体を通して主張に一貫性があるか」については、JST は「作文Ⅲは主張がない」と述べているのに対し、CST は書き手の表現意図をまとめており、「作文Ⅲは主張がない」との発言が見られなかった。「内容が分かりやすいか」については、JST は作文の最後に主張が明確に書かれていないことをマイナス評価していることがわかった。作文Ⅱの「構成」に属する評価項目については、JST より CST の評定値のほうが低いという傾向がみられた。「指示詞が適切に使われているか」については、CST は具体的な箇所を言及せず、マイナス評価をすることや、指示詞の使用が不適切な箇所だけ列挙することが見られた。一方、JST は指示詞の使用が適切な箇所も不適切な箇所も挙げており、作文全体における指示詞の使用を考慮していることがわかる。「段落と段落のつながりがスムーズであるか」、「文と文のつながりが自然であるか」についても、類似した傾向が見られた。

以上の調査Ⅰの結果から、JST, CST の評価行動の実態の一部が明らかとなった。しかし、調査Ⅰの協力者は JST, CST それぞれ 10 人であったため、結論を一般化するために、評価

者の人数を増やす必要がある。また、発話思考法という手法に限界があり、調査Ⅰでは評価者が言語化できるような評価の着目点や評価の厳しさの一端は明らかになったが、JST と CST の評価行動に影響を与える潜在的な要因については明らかにすることができなかった。こうした調査Ⅰの課題を解決するため、調査Ⅱでは質問紙法を用いて量的調査を行った。

第4章 日本語母語話者教師と中国語母語話者教師の評価モデル（調査Ⅱ）

第4章では、JST と CST の評価行動に影響を与える潜在的な要因を解明し、両者の評価モデルを比較することを目的とした調査Ⅱ（質問紙法による量的調査）の結果をまとめた。対象者は JST88 名、CST84 名であった。調査Ⅱでは、調査Ⅰで使用した4本の作文の中から、JST と CST による評定値の差の傾向を最も反映している作文Ⅱと作文Ⅲを選定した。また、調査Ⅰのフォローアップインタビューで得られた評価者の意見に基づき、評価尺度を修正した。総合的な評価尺度2項目（「全体」と「読み手意識」）、分析的な評価尺度21項目を選定した。

作文2本に対する分析的な評価尺度21項目の平均値を用いて、JST と CST 別に探索的因子分析を行った。CST の分析的な評価尺度の評定値から得られた3因子は、先行研究をもとに予め想定していた評価項目の構造とほとんど一致していた。第1因子は「構成」、第2因子は「正確さ」、第3因子は「内容」と命名した。一方、JST の分析的な評価尺度の評定値から得られた3因子は、先行研究や CST の結果とは異なっていた。第1因子は、作文の内容の説得力、論理性、独自性だけでなく、それらを支える構成にかかわる項目を含んでいたため、「説得力のある構成」と命名した。坪根・田中（2015）は日本語教師にとって「いい内容」「いい構成」がどのようなものかを調査し、「内容」と「構成」（特にマクロ構成）が密接にかかわっていることを示唆した。本研究で得られた「内容」と「マクロ構成」を統合する「説得力のある構成」という潜在因子は日本語が母語である JST の総合的な評価観点を反映していると考えられる。第2因子は、予め想定していた日本語に関する評価項目を含んでおり、CST の第2因子とほぼ同じ因子パターンだったため、「正確さ」と命名した。第3因子は、作文全体の構成というより、段落間、或いは段落内の構成にかかわる項目を含んでいたため、「結束性」と命名した。

次に、JST と CST による総合的な尺度の評定値を、因子得点により、どの程度正確に予測できるかを明らかにするために、総合的な尺度の評定値を目的変数とし、因子得点を予測変数として重回帰分析を行った。JST の評定値を用いた重回帰分析の結果、「全体」の得点の予測には、「説得力のある構成」の因子得点が最も大きな影響を与えており、「正確さ」の因子得点はその次に影響を与えていることがわかった。「読み手意識」の得点の予測には、「説得力のある構成」の因子得点が大きな影響を与えていることがわかった。CST の評定値を用いた重回帰分析の結果、「全体」・「読み手意識」とも、評定値の予測には、「内容」の因子得点が最も大きな影響を与えており、「正確さ」の因子得点はその次に影響を与えていることがわかった。総合的な評価尺度の評定値と分析的な評価尺度の評定値の関係について

て、重回帰分析を用いて分析した研究は Chiang (2003) と Rinnert and Kobayashi (2001) があるが、どちらの研究も本研究と異なる結果を示している。しかし、Chiang は作文評価において文法面と結束性のどちらが作文を評価する際に影響するかについて調べるもので、内容に関する評価項目は取り上げられていない。また、Rinnert and Kobayashi の NNS 教師の 9 割近くは日本の大学で英語を教えており、また半分以上が日本で 6 年以上居住していた。Rinnert and Kobayashi の NNS 教師の所属する教育機関、EFL 環境が、彼らの評価に影響を与えていると考えられる。

評価者の日本語の作文評価に関する考え方を調べるため、2 本の作文について評価してもらった後、JST と CST の作文評価において重視する項目と困難だと思うカテゴリについて回答してもらった。回答についてカイ二乗検定を行った結果、評価が困難なカテゴリのみ有意であった。最も評価が困難なカテゴリは「内容」と答えた JST が有意に多く、「内容」と答えた CST が有意に少なかった。書き手とは文化的背景、及び母語のレトリックパターンが異なる JST にとって、作文の内容を理解することが比較的難しいと考えられる。

最後に、JST、CST の作文指導経験に関する回答に基づき、JST と CST をそれぞれ作文指導経験が長いグループと作文指導経験が短いグループに分け、記述統計量による比較を行った。JST のデータについては、作文指導経験の長さによる因子得点の差はみられず、「結束性」を寛容に評価する傾向があることがわかった。一方、CST のデータでは、作文指導経験による因子得点の差が見られ、作文指導経験が長い CST よりも短い CST のほうが「内容」を厳しく評価していることがわかった。また、CST は作文指導経験の長さにかかわらず「正確さ」の評価が厳しいこともわかった。先行研究では、評価経験の異なる評価者による評価の厳しさに違いがあること示唆されている (Cumming, 1990 ; Barkaoui, 2010 ; Weigle, 1999 ; Wolf, 2006)。しかし、多くの先行研究における未熟な評価者は外国語教育を専攻とする学生であり、教師経験がなく、本研究の日本語教師である評価者とは異なる。それに加え、先行研究での評価者の数は少なく、評価経験が評価に与える影響を明らかにするために、今後量的な調査を増やしていくことが望ましい。

第 5 章 結論

調査 I で検討した研究課題は、「JST と CST による作文の評価行動にどのような違いがあるのか」であった。CST より JST のほうが内容、マクロ構成に多く注目しているのに対し、JST より CST のほうが言語形式に多く注目していることがわかった。また、言語形式、構成について、CST が JST よりも厳しく評価するのに対し、内容について、JST は CST より厳しく評定していることがわかった。調査 II で検討した研究課題は、「JST と CST による作文の評価モデルにどのような違いがあるのか」であった。JST は内容とマクロ構成を統合した「説得力のある構成」という潜在的な要因の影響を受けているのに対し、CST は従来の日本語の作文の評価基準と類似していることがわかった。

本研究の意義は、以下の 4 点である。

(1) 非母語話者教師による日本語の作文評価の実態を明らかにした。

これまでの日本語の作文評価では、日本語教師、一般日本人、または日本語学習者による評価が取り上げられ、非母語話者教師を取り上げる研究はない。本研究では CST を取り上げ、JST と CST による作文評価の違いを明らかにした。今後、日本語の作文評価研究の領域がさらに拡大し、推進されることが期待できる。

(2) 日本語母語話者教師 (JST) と非母語話者教師 (CST) の評価行動の傾向を明らかにした。

本研究により、JST・CST が作文を読み、評価項目について評定する際の実態、及び評価行動における両者の違いが示された。これらは、今後の作文評価研究、大規模テストにおける評価者トレーニング、実際の教育現場における教師の評価活動に一つの有益な情報となる。

(3) 日本語母語話者教師と非母語話者教師の評価モデルを提示した。

JST と CST の評価活動の傾向を踏まえ、量的調査を行った結果、JST と CST の評価モデルがそれぞれ示された。日本語教育分野において、教師の作文評価モデルを取り上げる研究が少ないうえ、これまでは、評価尺度に対する重視度だけが分析の対象となっていた。本研究で評定値を用いて分析することにより、JST と CST の評価モデルを提示することができた。

本研究の結果から、今後の日本語の作文教育について以下の 3 点を提言する。

(1) 本研究の知見を作文テストの評価者トレーニングに活用することである。

(2) 実際の教育現場で行われる学生の作文能力を把握するための診断的テスト、或いは教師が学生にフィードバックを与える作文評価の場合、評価者が自身の評価を内省し、調整する必要があることである。

(3) 本研究で探索的に明らかとなった評価モデルを作文の評価規準に活用することである。

本研究の限界は以下 2 つが挙げられる。

第一に、本研究の調査 I では発話思考法を用いて JST と CST の評価行動を調べたことである。評価行動は互いに関連しあっているものが多いのに加え、本研究は母語の異なる評価者を対象としている。発話思考法で収集した発話を評価行動別に分類することに限界がある。

第二に、本研究では、日本語教育歴、作文評価の経験、作文評価の能力 (expertise)、所属している教育機関と教師のポジションなどの要因が作文評価に与える影響を明らかにすることはできなかった。

今後の課題は以下 3 つが挙げられる。

(1) 評価者トレーニングを実施することによって、評価者の評価における問題点、或いは不一致な点は解決できるかについてさらに調べる必要がある。

(2) 内容とマクロ構成を統合した項目を含む評価規準を使用する場合、JST または CST による評価の信頼性、妥当性について検証しなければならない。

(3) 今後は本研究とは作文の側面の質が異なる作文を対象とする必要がある。また、ナラティブ（物語文）、説明文などの意見文以外のジャンル、または、書き手の文化的背景が表れにくい課題の場合においても、JST と CST の評価に本研究と同じ違いが見られるか、今後検証する必要がある。

引用文献

- 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』18（1）, 111-127
- Barkaoui, K.（2010）. Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7（1）, 54-74.
- Chiang, S.（2003）. The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System*, 31, 471-484.
- Conner-Linton, J.（1995）. Cross cultural comparison of writing standards: American ESL and Japanese EFL. *World Englishes*, 14, 99-115.
- Cumming, A.（1990）. Expertise in evaluating second language compositions. *Language Testing*, 7, 31-51.
- Hughes, A., & Lascaratou, C.（1982）. Competing criteria for error gravity. *ELT journal*, 36（3）, 175-182.
- Hyland, K., & Anan, E.（2006）. Teachers' perceptions of error: The effects of first language and experience. *System*, 34（4）, 509-519.
- Johnson, J. S., & Lim, G. S.（2009）. The influence of rater language background on writing performance assessment. *Language Testing*, 26（4）, 485-505.
- Kobayashi, T.（1992）. Native and nonnative reactions to ESL compositions. *TESOL Quarterly*, 26（1）, 81-112.
- Lee, H.（2009）. Native and nonnative rater behavior in grading Korean students' English essays. *Asia Pacific Education*, 10, 387-397.
- Marefat, F., & Heydari, M.（2016）. Native and Iranian teachers' perceptions and evaluation of Iranian students' English essays. *Assessing Writing*, 27, 24-36.
- McCretton, E., & Rider, N.（1993）. Error gravity and error hierarchies. *IRAL-International Review of Applied Linguistics in Language Teaching*, 31（3）, 177-188.
- McNamara, T.（1996）. *Measuring second language performance*. London: Addison Wesley Longman.
- McNamara, T.（2000）. *Language testing*. Oxford: Oxford University Press.
- Porte, G.（1999）. Where to draw the red line: Error toleration of native and non-native EFL faculty.

- Foreign Language Annals*, 32 (4) , 426-434.
- Rinnert, C., & Kobayashi, H. (2001) . Differing perceptions of EFL writing among readers in Japan.
The Modern Language Journal, 85, 189-209.
- Schmitt, N. (1993) . Comparing native and nonnative teachers' evaluations of error seriousness.
Japanese Association of Language Teachers: Journal, 15 (2) , 181-191.
- Sheorey, R. (1986) . Error perceptions of native-speaking and non-native-speaking teachers of ESL.
ELT journal, 40 (4) , 306-312.
- Weigle, S. C. (1999) . Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, 6 (2) , 145-178.
- Wolfe, E. W. (2006) . Uncovering rater's cognitive processing and focus using think-aloud protocols.
Journal of Writing Assessment, 2 (1) , 37-56.