

学位論文要約

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

## 第1章 序論

### 1.1 本研究の背景と目的

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

### 2.1 属性の異なる評価者による日本語作文評価

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（池田, 2012 ; 田中・坪根・初鹿野, 1998）、及び評価者の母語（胡, 2014 ; 近藤, 2014 ; ツOWN, 2007 ; 長谷川, 2008 ; 村上, 2001）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（胡, 2014 ; 田中ほか, 1998 ; ツOWN, 2007）、評価項目についての判断基準が異なること（近藤, 2014 ; 長谷川, 2008 ; 村上, 2001）がわかる。しかし、なぜそのような違いが生じたかについては、ほとんど述べられていない。また、先行研究における非日本語母語話者評価者は、いずれも日本語教師経験のない学生であり、日本語非母語話者教師による評価は取り上げられていない。

### 2.2 評価者の作文評価行動

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。一方、日本語の作文に対する評価者の評価行動を取り上げる研究は宮島（2006）と宇佐美（2014）しかなく、日本語教師の評価行動を取り上げるものはない。

### 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

文レベルの誤用に対する評価を取り上げる先行研究の結果から、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（趙，1991；Hughes & Lascaratou，1982；Hyland & Anan，2006；Schmitt，1993）。また，NS 教師と NNS 教師の誤用の重篤度の評定にも違いが見られ，英語学習者の誤用について 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る（Hughes & Lascaratou，1982；McCretton & Rider，1993；Porte，1999；Schmitt，1993；Sheorey，1986）。

作文全体に対する評価を取り上げる先行研究の結果から，NS 教師と NNS 教師の評価の相違について一致した見解はいまだ得られていない。教師の評価の理由を調べた Shi（2001）は，NS 教師と NNS 教師の評価の着目点が異なると報告している。一方で，総合的評価の評定値と分析的評価の評定値の関係を分析した Chiang（2003）と Rinnert and Kobayashi（2001）は，NS 教師と NNS 教師の評価の着目点が類似していることを明らかにしたが，教師が総合的な評価尺度を用いて評定する際の根拠は異なっていた。また，分析的な評価尺度については，Kobayashi and Rinnert（1996）と Hamp-Lyons and Zhang（2001）が構成（「導入」など）や内容（「説得力」など）に対する NS 教師と NNS 教師の評定値が異なっていることを報告している。また，その違いが生じる原因として，評価者の母語のレトリックパターンと文化的背景が異なることを挙げている。評価の厳しさについては，評価に熟練し，外国語能力が母語話者に近い評価者が総合的な評価尺度を用いて評定する場合，NS と NNS の評価の厳しさに違いは見られなかった（Johnson & Lim，2009）。一方で，外国語教師が分析的な評価尺度を用いて評定する場合，NS 教師より NNS 教師のほうが全体的に厳しいことが報告されている（Lee，2009；Marefat & Heydari，2016）。しかし，その違いについて先行研究から一致した結果は得られなかった。評価の厳しさは教師の所属する教育環境に大きな影響を受ける可能性がうかがえる。

#### 2.4 先行研究の問題点と本研究の課題

ここでは，先行研究が残した課題を3点にまとめて論じる。

1つ目の課題は，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。教師の母語という属性を扱った作文評価については，英語教育のものがほとんどで，日本語教育のものはほとんど見られない。海外における日本語教師，特に学習者数の多い中国国内で日本人日本語教師が不足している現状から，中国人日本語教師による評価に関する研究を進めることは急務だと言える。2つ目の課題は，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。これまでの先行研究から，NS と NNS による評価の違いが報告されている。しかし，なぜそのような違いが生じたかについてはまだ明らかにされていないと言いがたい。3つ目の課題は，日本語作文の評価モデルが解明されていない点である。先行研究から，NS 教師と NNS 教師による評価の厳しさ，評価項目の重要度に対する認識，評定する際の着目点そのものに違いが見られた。しかし，用いられた総合的・分析的な評価尺度による評定値間の関係は，まだ十分に分析されていない。

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、総合的な評価尺度の評定（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）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 の作文評価の厳しさは異なっており、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）

#### 3.1 目的

調査 I は JST と CST が作文評価する際の評価行動を比較するために、発話思考法を用いた質的調査を実施した。

#### 3.2 方法

まず、評価対象の 4 本の作文と評価尺度（総合的な尺度 1 項目、分析的な評価尺度 20 項目）を選定し、評価表を作成した。次に、発話思考法を用いて JST10 名と CST10 名に作文を評価させ、その過程を録音した。録音したものを文字化し、JST と CST の評価行動を抽出した。調査 I では、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。

#### 3.3 結果と考察

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、JST も CST も「日本語」の運用に注目して作文を読んでいると予測できるが、CST の出現率のほうがやや高いという結果になった。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。「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ため、ここではその他のカテゴリーの分析結果を示す。

「日本語」の評価行動の下位カテゴリーの結果から、全体の傾向として、評価者は「日本語：正確さ」と「日本語：修正」を多く行っていることがわかる。しかし、「日本語：修正」については、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。JST と CST の「日本語」に属する評価行動に違いが見られた原因として、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。日本語が母語である JST にとって日本語の修正は自動化できており、作文を読みながら、不適切な言語形式を修正することは比較的簡単だと考えられる。一方、日本語が外国語・第二言語である CST にとって、言語形式の指摘はできても、随時修正を行い、発話することは難しかったと推測できる。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。CST は、中国人日本語学習者によくみられる誤用を心得ていることから、「日本語：正確さ」を多く行っていると考えられる。一方、本研究は評価者に作文の訂正フィードバックを求めているにもかかわらず、JST の「日本語：修正」がより多く見られた。書き手と言語的背景が異なる JST にとって、作文の言語形式が不適切な箇所から、書き手の表現意図を推測することは難しいと考えられる。

「内容」の評価行動の下位カテゴリーの結果から、JST も CST も「内容：サポートの質」と「内容：内容の確認」についての評価行動が多いことがわかった。しかし、「内容：サポートの質」において、CST より JST のほうが出現率が高い。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。書き手が主張のサポートとして用いている材料は、中国語母語話者が論述する際に一般的に用いるものが多い。そのため、サポートの質に着目する CST が少なかったと言える。それに対し、作文のサポートとして用いられている材料は、JST にとって予期しなかったものが多いため、サポートの質についての発話が多かったと考えられる。

「構成」の評価行動の下位カテゴリーの結果から、JST も CST も「構成：結束性」と「構成：段落の適切性」についての評価行動が多いことがわかる。JST・CST ともに、文のつながり、段落レベルの構成に多く注目している。しかし、CST も JST も文と文のつながりにかかわるミ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る。JST と CST の「構成」に関する評価行動の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。学習者の作文について、CST が中国語のレトリックパターンに多く接触し違和感を感じないのに対し、JST はマクロ構成の適切性を判断しながら読んでいることが予想される。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。ミ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。「日本語」に多く注目している CST はマイクロ構成に注意を多く向けていると推測できる。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。作文の内容理解を重視する JST は、作文のマクロ構成により多く注意し、その適切性を判断しながら作文を読んでいると想定される。

次に、「評定」段階の評価行動を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 JST より「日本語」について厳しく評価していることがわかる。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。

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。JST は CST より「内容」について厳しく評価していることがわかる。「文章全体を通して主張に一貫性があるか」については、JST6 名がマイナス評価をしているのに対し、CST6 名がプラス評価をしている。プロトコル例から、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、マイナス評価している評価者は、JST7 名に対し CST2 名であった。プロトコルを見ると、JST からは作文Ⅱの「最後の部分の分かりにくい」とのコメントがあった。一方、CST からも、そのようなコメントは見られたが、数としては少なかった。日本語の作文は、最後の部分で全体のまとめと、書き手の主張が書かれていることが多い。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る。

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。「指示詞が適切使われているか」については、CST5 名がマイナス評価をしているのに対し、JST 全員がニュートラルまたはプラス評価をしている。具体的なプロトコルを見ると、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、JST と CST の評価行動に違いが見られた。プロトコル例から、CST は文の中の語彙の使い方や文の接続の正確さに注目しており、また、そのような不適切な箇所が一つだけでも、評定値を大きく下げる場合があることがわかる。それに対し、JST は複数の箇所を判断の材料とし、それらを総合的に評価しているようである。JST のプロトコルから、つながりが不適切な箇所があっても、作文全体の出来具合を考え、比較的高い評定値を与えるという寛容な態度が見られた。

調査Ⅰの結果から、JSTは意味理解を重視しているのに対し、CSTは言語形式の正確さを重視していることが推測できる。また、JSTとCSTの言語的背景、文化的背景が異なることが、評価行動に違いが生じた原因として分析した。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、発話思考法では「日本語」に関する評価行動が多く出現するが、「内容」「構成」に関する評価行動が比較的出現しにくい。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）

### 4.1 目的

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ために、質問紙法による量的調査を実施した。また、作文評価において重視する項目、作文指導経験による評価の厳しさについても検討する。

### 4.2 方法

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。質問紙法による量的調査を実施した対象者はJST88名、CST84名であった。

### 4.3 結果と考察

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（田中ほか、1998）とほとんど一致していた。そのため、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。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。因子分析の結果、第2因子「正確さ」はCST・JSTともに共通して見られたが、その他の因子パターンは予想とは異なっていたため、仮説「JSTとCSTの因子パターンは異なっている」は支持された。

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る。「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る。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る。この結果から、仮説「JST と CST が作文評価をする際、総合的な評価尺度の評定値を予測する要因が異なる。」が支持された。そもそも、JST と CST の因子パターンが異なっていたため、当然の結果であるが、「全体」について評価する際、JST も CST も内容と日本語の正確さを根拠としているという類似点も見られた。また、総合的な評価尺度の評定にあまり影響を与えないものとして、JST の「結束性」と CST の「構成」があった。

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項目について回答してもらった。具体的には、評価者に作文を評価する際に、「最も重視するカテゴリー」と「最も難しいカテゴリー」について、それぞれ「内容」「構成」「日本語」の中から選択してもらった。カイ二乗検定を行った結果、「最も難しいカテゴリー」のみ有意であった。この結果から、仮説「作文評価において、JST と CST が重視する項目、評価が困難な項目は異なっている」は部分的に支持されるという結果となった。「評価が難しいカテゴリー」に関しては、「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。

フェースシートに記入された JST 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 の因子得点は、作文指導経験による差が比較的小さく、因子による差が比較的大きいことがわかった。本調査の JST の評価は作文指導経験による影響が小さいが、因子による影響が大きく、「結束性」について寛容的であることがわかる。CST の因子得点は、因子による影響が大きく、「正確さ」の評価が厳しいことがわかる。また、作文指導経験による差も見られ、「内容」について作文指導経験の短い CST が厳しい。この結果から、仮説「JST と CST の作文評価の厳しさは異なっており、その評価には作文指導経験が影響している」は部分的に支持された。JST は「結束性」について寛容的であるのに対し、CST は「正確さ」について厳しいことがわかった。作文指導経験が与える影響は、CST による「内容」の評価にしか見られなかった。

## 第5章 結論

### 5.1 本研究のまとめ

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」であり、具体的な研究課題は以下の2点であった。

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。

1-1 については、先行研究の知見から、作文を読む際に JST は内容、構成に多く注目するのに対し、CST は言語形式に多く注目するという結果を予測していたが、予測が支持された。「読む」段階における評価行動の傾向から、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。

1-2 については、先行研究の知見から、評価項目について評定する際に、JST と CST の言語形式、内容、構成について判断基準が異なり、JST は内容、構成について厳しく評価し、CST は言語形式について厳しく評価するという結果を予測していたが、予測が部分的に支持された。言語形式に関する項目を評定する際、CST が JST よりも厳しく評価するのに対し、内容に関する評価項目を評定する際、JST は CST より厳しく評定しており、JST と CST による判断が異なることがわかった。しかし、予測と異なり、構成については JST より CST のほうが厳しいことがわかった。マイクロ構成 (i.e. 接続詞、指示詞) は言語形式の正確さと強く関わっていることから、CST はより厳しく評価していることが推測できる。

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」であり、具体的な研究課題は以下の4点であった。

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 の作文評価の厳しさは異なっており、その評価には作文指導経験が影響しているかどうか。

2-1 については、JST と CST の因子パターンは異なっていると仮説を立てたが、仮説が支持される結果となった。JST の分析的な評価尺度の評定値から「説得力のある構成」「正確さ」「結束性」という3つの因子が、CST の分析的な評価尺度の評定値から「構成」「正確さ」「内容」という3つの因子が得られた。CST の評価構造は従来の日本語の作文の評

価基準と類似しているが、JSTの3つの因子すべてに「構成」がかかわっていたことがわかる。

2-2については、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）の評定を予測する要因が異なると仮説を立てたが、仮説が持される結果となった。JSTの場合、「全体」の評価を予測するには、「説得力のある構成」「正確さ」が、「読み手意識」の評価を予測するには、「説得力のある構成」が挙げられている。CSTの場合、「全体」「読み手意識」とも評価を予測するには、「内容」「正確さ」が挙げられている。ただし、「全体」について評価する際、JSTもCSTも内容と日本語の正確さを根拠としているという類似点も見られた。

2-3については、作文評価において、JSTとCSTが重視する項目、評価が困難な項目は異なっていると仮説を立てたが、仮説が部分的に支持されるという結果となった。「作文を評価する際に、最も重視するカテゴリー」に関しては、JSTとCSTの意識に差は見られなかったが、「評価が難しいカテゴリー」に関しては、「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。

2-4については、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と仮説を立てたが、仮説は部分的に支持された。JSTは「結束性」について寛容的であるのに対し、CSTは「正確さ」について厳しいことがわかった。しかし、作文指導経験が与える影響は、CSTによる「内容」の評価にしか見られなかった。

調査Ⅱは、JSTとCSTの評価モデルの違いを明らかにし、両者の評価に与えるそれぞれの潜在的な要因を解明した。

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評価行動の実態、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の評価行動の傾向を踏まえ、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。

#### 5.2 日本語教育における作文教育への示唆

(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。

#### 5.3 今後の課題

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の効果は先行研究で示唆されているが、まだ解明されたとはいえない。今後、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) 今後の日本語の作文教育において、内容とマクロ構成を統合した項目を含む評価規準を作成し、使用する必要性について述べた。しかし、そのような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) 本研究で明らかとなった JST と CST の評価の違いはそれらの作文の質を反映していることも考えられる。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だろう。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

- 池田隆介 (2012) 「作文の評価基準-日本語教師と一般日本語母語話者の比較-」『基盤教育センター紀要』 12, 21-40, 北九州市立大学
- 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
- 胡芸群 (2014) 「メールに対する読み手の評価-読み手の属性による評価の観点の違い-」『一橋大学国際教育センター紀要』 5, 81-91
- 近藤行人 (2014) 「ウズベキスタン出身の留学生と現役日本語教師が持つ文章観の検討」『名古屋学院大学論集 言語・文化篇』 2, 597-109
- 田中真理・坪根由香里・初鹿野阿れ (1998) 「第二言語としての日本語における作文評価基準-日本語教師と一般日本人の比較-」『日本語教育』 96, 1-12
- 趙南星 (1991) 「韓国人の日本語学習者の誤りの評価-日本語話者と韓国語話者による誤りの重み付け-」『日本語と日本文学』 15, 19-30, 筑波大学国語国文学会
- ツォウンティマイ (2007) 「依頼のEメールに対するベトナム人・日本人の評価意識比較-『対人配慮』に注目して-」『日本言語文化研究会論集』 3, 169-196
- 長谷川哲子 (2008) 「日本語学習者による非母語話者の作文に対する評価について」『大阪産業大学論集 人文・社会科学編』 4, 91-100
- 宮島良子 (2006) 「日本語教師を目指す上級日本語学習者は日本語作文をどのように評価するのか」名古屋大学大学院修士学位論文
- 村上京子 (2001) 「作文の評価の信頼性と評価者の特性-日本人母語話者と非母語話者の評価の比較-」『名古屋大学日本語・日本文化論集』 9, 47-69
- 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.
- Hamp-Lyons, L., & Zhang, B. W. (2001). World Englishes: Issues in and from academic writing assessment. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes*. 101-116. Cambridge: Cambridge University Press.
- 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, H., & Rinnert, C. (1996). Factors affecting composition evaluation in an EFL context. *Language Learning*, 46, 397-437.
- 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.
- Shi, L. (2001) . Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18, 303-325.