

作文評価における日本語教師の評価ポリシーの多様性 —ホリスティック評価の観点から—

安芝恩 (なし) †

Diversity of Japanese Language Teachers' Assessment Policies in Writing Assessment: From the Viewpoint of Holistic Scoring

Jieun Ahn (Not affiliated)

要旨

本研究は、日本語教師が日本語学習者の作文をどのように評価しているのかという評価者の評価ポリシーを明らかにすることを目的とするものである。特に、評価者一人ひとりにおける評価ポリシーは存在するのか、作文評価において評価ポリシーは一貫性が保たれているのかに注目した。日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文) を対象とし、日本語教師 12 名に Google フォームを用いた作文評価をしてもらった。評価方法については、全体的・総合的に評価してもらうため、ホリスティック評価を採用し調査を行った。その結果、日本語教師は、明示的な認識があるとは言えないものの日本語学習者の作文に対して各自の評価ポリシーを持っており日本語学習者の作文によってその評価ポリシーは個人の中でも変動することが明らかになった。評価者個人が持っている評価ポリシーの多様性を認めつつ、安定した評価への必要性があることが示唆された。

1. はじめに

1.1 研究背景

評価には、「評価のばらつき・ゆらぎ」(宇佐美 2014)がある,または、「評価の不一致をもたらす要因」(田中ほか 2009, 田中 2016)があるという「評価の多様性」が報告されている。さらに、「評価価値観の変容や自己調整,といったことも研究の対象として取り上げられなければならない」(宇佐美 2016)という指摘があるように,評価は、「変動するもの」としての評価価値観として捉えられているために,複数の評価者による評価には,評価の不一致が生まれるものである。

最近,複数の評価者による評価を一致するための試み(阿部ほか 2016)や日本語の学習者と教師のための Web システムである jWriter, 読み手と構成を意識した日本語ライティングである GoodWriting Rater といった評価ツールが開発されている。

しかし,日本語教育現場では,いまだ評価者による作文評価が行われており,日本語の添削に焦点が置かれている現状にある。そのため,作文の構成や内容よりは文法の正確性が問われる傾向がある。おそらく作文評価は添削という認識が評価ポリシーの根底に位置付けられているだろう。

そこで,本研究では,日本語教師は,日本語学習者の作文をどのように評価しているのかという評価者の評価ポリシーを明らかにすることを目的としている。本研究の調査結果から,

† parkhc2015@gmail.com

評価者個人が持っている評価ポリシーの多様性を認めつつ、その評価ポリシーにおける「安定した評価」への必要性が示唆されることに意義があると思われる。

1.2 用語定義

では、本研究で用いる「評価ポリシー」および「作文」の定義について説明する。「評価ポリシー」は、宇佐美(2014)での「評価」の用語定義を援用し、「主体がもつ内的・暗黙的な価値観に基づいて、対象についての情報を収集し、主体なりの解釈を行ったうえで、価値判断を行うまでの一連の認知プロセス。またその結果として得られる判断」と定義する。さらに、宇佐美(2016)で示されているように、「価値観に基づく判断」であり、価値観の相違によって判断も、それに先立つ情報収集や解釈のありようも変わってくる」と捉える。「作文」については、日本語学習者が4コマイラスト(安 2017)を見て各コマの内容について描写したストーリー文を「作文」と定義する。日本語学習者から収集された作文のスタイルは、登場人物の会話文に注目した「会話型」と叙述形式の「説明型」が混在している。

2. 先行研究

2.1 先行研究の概観

では、「評価者による評価」および「評価ツールによる評価」についての従来の研究について述べる。

評価ポリシー(例:ループリック評価)における理解が、評価者それぞれ異なる解釈をしている(藤浦ほか 2018)ことが報告されている。田中(2016)によると、各評価者におけるポリシーについては、「評価基準やガイドライン等に柔軟に対応する人となかなか自身の評価ポリシーを譲らない人がいる」と指摘されている。一方、宇佐美(2016)では、「評価」に対する明示的な認識の不在が指摘されている。これは、評価基準が設けられていても各評価者における評価ポリシーが存在している(田中ほか 2009)ことが影響されていると考えられる。研究は背景でも述べたように、「評価のばらつき・ゆらぎ」(宇佐美 2014)は、「評価の不一致をもたらす要因」(田中ほか 2009, 田中 2016)である、評価ポリシーに対する明示的な不在から起因している可能性がある。

近年は、「評価者による評価」から発生する「ばらつき」「ゆらぎ」といったものを減らすため、「GoodWriting Rater」や「jWriter」のような評価ツールが開発されている。これらの評価ツールは、それぞれの評価ポリシーに基づき評価していることが特徴である。

「GoodWriting Rater」は、ホリスティック評価とマルチプルトレイト評価によって評価されるツールであり、矢部ほか(2019)によると、「人間」と「機械」による評価の統合的活用」の一部として開発された評価ツールであると言及されている。文字数400~1600字の作文が評価の対象となっている。また、「jWriter」は、総合的評価によって評価されるツールであり、「日本語学習者の作文テキストを入力すると、推測される作文力の到達レベルを5段階で判定」できる。最小入力文字数は300文字で、600字前後でもっとも良い判定ができると言及されている。このような総合的評価について、占部(2007)では、妥当性があるとは言えないものの実用性は高いという利点があると指摘されている。一方、評価者による評価のばらつき・ゆらぎを防げるものの、より良い判定のための文字数制限があるため、全ての日本語教育現場での作文評価に活用できるとは限らない。

以上を踏まえ、日本語教育現場では、「評価者による評価」が求められている現状で、評価者一人ひとりにおける評価ポリシーがあるのか、その評価ポリシーは作文評価において評価の

対象が変わっても一貫性が保たれているのかといったことについてはまだ明らかにされていない。

2.2 研究課題

そのため、以下の3つの研究課題を設定し調査を行った。

【課題 1】日本語学習者の作文に対する日本語教師の評価には、評価者それぞれの間にどのような違いが見られるのか

【課題 2】日本語教師は、日本語学習者の作文に対する評価において、どのような評価ポリシーを持っているのか

【課題 3】評価ポリシーは、評価者個人の中で一貫性が保たれているのか

課題 1 は、評価者による評価の不一致といった評価ポリシーの多様性を見るために、課題 2 は、評価者個人における明示的な評価ポリシーの存在有無を確認するために、課題 3 は、評価者個人において評価ポリシーが存在する場合、評価対象の作文すべてにおいて、その評価ポリシーは一貫性のある働きをするのかを確認するために設定したものである。

3. 研究方法

3.1 コーパス

本研究に用いられたコーパスは、現在構築中の「日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文)」である。4 コマイラスト(安 2017)を用いて、日本語学習者に描写ストーリー文を作成してもらったものである。次の図 1 のように、日本語学習者から収集された作文のスタイルは、登場人物の会話文に注目した「会話型」と叙述形式の「説明型」が混在している。

A さん：お腹がすいたなあ。あ、前のコンビニへ買いに行きます。
スタッフ：いらっしゃいませ。
A さん：食べ物を買いました。これください
スタッフ：はい、わかりました
A さん：すみません。現金がないので、カードではらえますか。
スタッフ：すみませんが、ここで現金だけです。
A さん：わかりました、待ってください、ATMで現金を出してはらいます。
スタッフ：はい、わかりました。

・「会話型」作文の例

会社から家に帰る途中で、急に何か
甘い物を食べたくなった私は汐井公園の近くにあるコンビニに寄りました。コンビニに入って、色々なお菓子があって、みんなは美味しそうだった。どれにしようかなあと迷っちゃた。
やっぱり私の大好物のどら焼きが一番いいと思って、三つのどら焼きにしました。
会計するとき、財布が家に忘れてしまったことを気がついた。
どうしようかなあ、とても恥ずかしくて、顔が真っ赤になった
仕方がない、ATMからお金を引き出そうと思った私は引き出して、払いました。

・「説明型」作文の例

図 1 日本語学習者 13 名による作文の例

また、コーパスの構築は、表 1 の日本語学習者 13 名の協力を得て行われた。協力した日本語学習者の国籍は、表 1 に示した通りでベトナム・中国・マレーシア・ミャンマー・インドネシアである。日本語の関連資格 (JLPT, J-TEST)については、いずれかの資格を保持してい

る日本語学習者が 10 名、いずれの資格も保持していない日本語学習者が 3 人である。

表 1 日本語学習者の属性

No.	性別	国籍	年齢	来日時期	日本語学習機関	JLPT	J.TEST
作文 1	男	ベトナム	20	2020 年 3 月	日本語学校と独学	N4	ない
作文 2	男	中国	24	2019 年 1 月	独学	ない	A-C レベル
作文 3	女	中国	33	2020 年 11 月	日本語学校	ない	D-E レベル
作文 4	女	中国	33	2020 年 12 月	日本語学校	ない	ない
作文 5	男	ベトナム	30	2013 年 7 月	日本語センター	N1	A-C レベル
作文 6	男	ベトナム	20	2020 年 11 月	日本語学校	ない	A-C レベル
作文 7	女	中国	25	2020 年 11 月	日本語学校	N5	ない
作文 8	男	マレーシア	24	2020 年 10 月	日本語学校	N4	ない
作文 9	未回答	ミャンマー	24	2020 年 10 月	日本語学校	ない	ない
作文 10	男	インドネシア	18	2020 年 11 月	日本語学校	N5	ない
作文 11	女	ベトナム	26	2019 年	大学	N1	ない
作文 12	男	ベトナム	20	2021 年 1 月	日本語学校	N5	ない
作文 13	女	ベトナム	29	2021 年 1 月	日本語教室	ない	ない

3.2 調査対象および調査期間

調査対象者は、日本語教師 12 名であり、属性については表 2 に示した通りである。調査は、2021 年 3 月～4 月に日本語教師 12 名に、3.1 で紹介した「日本語学習者 13 名による作文（4 コマイラストの描写ストーリー文）」の 13 作文の評価を依頼し行われた。日本語教師（日本語教師 4 を除く）は、日本語専攻/副専攻・420 時間養成課程修了、日本語教育能力検定試験合格等の日本語教育関連資格を保持している。

日本語学習者のレベルが作文評価に影響されることを排除するため、3.1 の表 1 日本語学習者の属性については、伏せた状態で日本語学習者の作文のみで評価してもらった。

表 2 調査対象者の属性

No.	指導経験	指導期間	指導機関	日本語教育関連資格	性別	年齢
教師 1	あり	4 年～5 年	日本語学校	大学副専攻	女	20 代後半
教師 2	あり	11 年	日本語学校	420 時間養成課程修了、日本語教育能力検定試験合格	女	50 代半ば
教師 3	あり	3 年	高校、日本語教室	中等学校の教員資格（日本語、英語）	女	40 代前半
教師 4	あり	10 年	日本語学校	大学院専攻、日本語教育能力検定試験合格	男	30 代後半
教師 5	あり	1 年	大学	なし	男	30 代後半
教師 6	あり	6～7 年	日本語学校	420 時間養成課程修了	女	30 代後半
教師 7	あり	1 年 6 ヶ月	日本語学校	420 時間養成課程修了	男	40 代半ば
教師 8	あり	3 年	日本語学習塾、大学（ア	420 時間養成課程修了、	女	20 代後半

			シスタント),ボランティアセンター	大学院専攻		
教師 9	あり	18 年	日本語学校,短期大学	420 時間養成課程修了,日本語教育能力検定試験合格,大学院専攻	女	50 代
教師 10	あり	7 年	日本語学校	420 時間養成課程修了	女	30 代後半
教師 11	あり	10 年以上	日本語学校,技能実習,専門学校	大学副専攻,日本語教育能力検定試験合格	女	30 代後半
教師 12	あり	29 年	日本語学校,大学,企業	420 時間養成課程,日本語教育能力検定試験合格	女	50 代後半

3.3 調査および分析方法

本研究の調査は,日本語教師 12 名に「日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文)」の 13 作文を評価してもらい,その結果は Google フォームを用いて収集した。ホリスティック評価の観点から,図 2 のように総合的に評価してもらい,気づいた点についてのコメントの記入に加え,5 段階評価 (わるい 1 ~ よい 5) のスコアをつけてもらった。作文に対するコメントについては, KH-Coder3(樋口 2020)を用いた計量テキスト分析を,評価のスコアについては,Excel のピボットテーブルを用いて分析を行った。図 2 に,作文 6 に対する日本語教師 4 のコメントを例で示してある。

本研究で採用しているホリスティック評価は,一般的に全体的・総合的評価 (holistic scoring) と知られており,占部 (2007)によると,「評価者個人の判断をもとにして全体評価として 1 つのスコアをつける方法」であると定義されている。また,「代表的なものとして, TOEFL の writing section での評価があり, 6 段階の全体的評価になっている」ことが紹介されている。

文のスタイルがバラバラ、長さは十分、語彙力は初級後半位、状況の理解はできていると思う。イラストの状況描写よりも、自分の日本語力を最大限に表出したい学生ではないだろうか。評価を4にしたいと思ったが、イラストの状況描写と関係のない部分を除外して考えてみると作文のレベルが高いとは思えない。

上記の作文の評価をお願いします。*

わるい 1 2 3 4 5 いい

 ○ ○ ● ○ ○

図 2 作文の 5 段階評価

4. 調査結果および考察

4.1 評価ポリシーのカテゴリー

まず,日本語学習者の作文に対する日本語教師 12 名のコメントの傾向を見るために, KH-Coder3 で分析を行った。その結果,図 3 のように,【作文,状況,描写,ストーリー,印象,コマ,話】という 7 つのカテゴリーが抽出された。日本語教師による作文評価において,本研究での作文が,4 コマイラストの描写ストーリー文であるという共通認識が働いて図 3 のようなカテゴリーが抽出されたと考えられる。

樋口(2020)によると,出現頻度の多い語は,大きく描画されている語であり,共起の程度は,太い線で描画されており共起の程度が強いことを意味している。

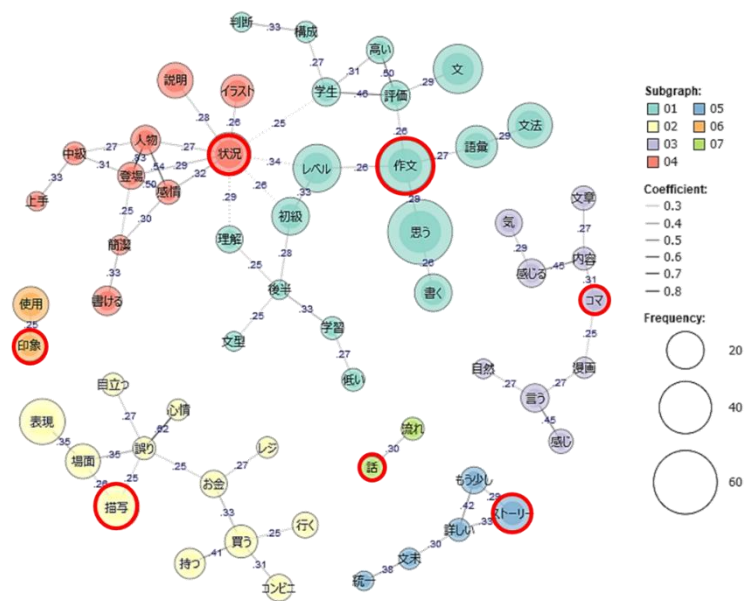


図 3 作文評価コメントの共起ネットワーク図

作文ごとのコメントの抽出語を表 3 に示した。作文それぞれに対する評価ポリシーは異なっているものの、日本語教師による評価コメント全体の結果から抽出された【作文,状況,描写,ストーリー,印象,コマ,話】の7つのカテゴリーがメインとなっていることが見て取れる。また,紙面の都合上,各作文に対するコメントの共起ネットワーク図は割愛するが,作文それぞれにおいて日本語教師の評価ポリシーは一貫性が保たれているとは言えない結果となった。

表 3 作文に対するコメントの抽出語

No.	抽出語
作文 1	レベル、作文、表現、ストーリー、過去、書く等
作文 2	語彙、初級、コマ、文法、簡潔等
作文 3	状況、描写、日本語等
作文 4	会話、勘違い、普段、ストーリー等
作文 5	作文、レベル、初級、状況等
作文 6	状況、イラスト、統一、ストーリー、使い方等
作文 7	文、会話、ストーリー、イラスト
作文 8	作文、文法、文型、描写、場面、使い方等
作文 9	会話、文、コマ、説明等
作文 10	間違い、文法、状況、レベル、描写等
作文 11	会話、文法、描写、展開、場面、使い方等
作文 12	会話、状況、敬語、描写、文等
作文 13	作文、文法、説明、状況、話等

ここで、注目したい箇所は、赤い字で示した「会話」であるが、日本語教師のコメントのうち、「会話」についてのコメントが 24 件見られた。「会話」につながるコメントは、ポジティブな評価とネガティブな評価に分かれていることが分かった。おそらく「作文」に対する日本語教師の用語定義がそれぞれ異なっており、「会話型」の作文を柔軟に受け入れないことが影響されていると言える。日本語学習者の作文が、「会話型」か「説明型」かが、評価ポリシーに影響され、スコア（5 段階評価）の判定に影響を与えており、評価者の間に評価ポリシーのズレが存在(藤浦ほか 2018)していると考えられる。おそらく、「会話型」の作文スタイルをどのくらい容認してくれるかがスコア判定の判断の材料となっていると窺える。一方、イラストを、描写前提での作文という共通認識から、【状況、描写、ストーリー、コマ】のような語が抽出されたと考えられる。

4.2 評価ポリシーの多様性

では、4.2 では、【課題 1】日本語学習者の作文に対する日本語教師の評価には、評価者それぞれの間にどのような違いが見られるのかについて見ていきたい。

評価者による評価の不一致といった評価ポリシーの多様性を見るために、日本語教師による作文評価のスコアをまとめた。日本語教師 1~12 による「日本語学習者 13 名による作文（4 コマイラストの描写ストーリー文）」の 13 作文に対するスコア（5 段階評価）を表 4 に示した。同じ作文に対して日本語教師の評価スコアは、同じ場合もあればそうでない場合もあることが確認できた。

表 4 日本語教師による作文評価のスコア

No.	教師 1	教師 2	教師 3	教師 4	教師 5	教師 6	教師 7	教師 8	教師 9	教師 10	教師 11	教師 12
作文 1	3	3	2	3	2	3	3	4	3	3	3	4
作文 2	4	2	2	2	3	2	3	4	3	3	2	3
作文 3	1	2	2	2	2	2	4	4	3	3	2	3
作文 4	4	1	2	2	2	1	3	2	4	3	3	3
作文 5	2	2	1	1	3	2	3	2	3	3	1	2
作文 6	4	5	2	3	4	4	2	4	5	4	4	4
作文 7	1	1	1	1	1	1	2	1	2	3	1	1
作文 8	4	3	2	4	2	4	3	3	3	3	1	3
作文 9	5	2	2	2	3	1	4	4	5	3	3	3
作文 10	2	3	2	3	2	3	3	3	4	3	3	3
作文 11	4	3	3	2	3	1	3	3	5	4	4	5
作文 12	4	2	2	3	4	1	3	4	4	3	3	3

また、同じ作文を提示して評価を依頼したものの、評価ポリシーにおける理解が評価者それぞれ異なる解釈をしている(藤浦ほか 2018)可能性が窺えた。図 4 は、表 4 に基づき作文の評価スコア（5 段階評価）の揺れ幅をグラフで示したものである。作文 10 に対しては、スコア（5 段階評価）2~4 に分布しており揺れ幅 3 を見せているものの、スコア（5 段階評価）3 をつけた日本語教師が 8 名ということからある程度日本語教師による評価ポリシーが一致している可能性が見受けられる結果となった。

一方,作文 8 に対しては,スコア (5 段階評価) 1~4 に分布しており揺れ幅 4 を見せている。さらに,作文 9 に対しては,スコア (5 段階評価) 1~5 に分布しており,評価ポリシーの不一致が最も目立つ揺れ幅 5 を見せている。

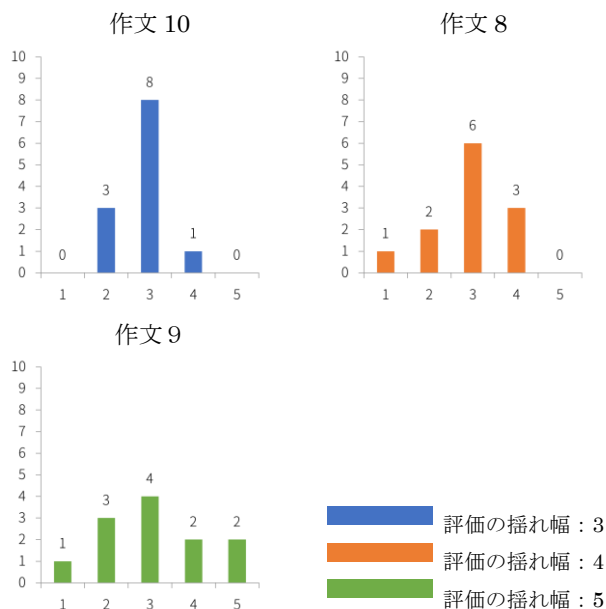


図 4 作文評価スコアの揺れ幅

特に,4.1 と 4.2 の調査結果の評価スコアとコメントを比較した結果, スコアの判定が高評価の場合は, [ストーリー,説明,状況] 等のストーリーの拡張に焦点が置かれている一方,低評価の場合は, [文法,語彙,初級,レベル] 等のような日本語の間違いに焦点が置かれており正確性が問われていることが窺えた。日本語学習者の日本語レベル等の属性については,日本語教師 12 名に伏せた状態で調査を行ったが,作文を評価する際は,日本語学習者の作文から改めてレベルを類推し作文の評価を行っていることが特徴であると言える。

4.3 評価ポリシーの明示的な認識

続いて,4.3 では【課題 2】日本語教師は,日本語学習者の作文に対する評価において,どのような評価ポリシーを持っているのかについて見ていきたい。評価スコア (5 段階評価) と評価コメントとの比較から分析を行った。

評価ポリシーがあると考えられる調査対象者は,日本語教師 4・10・11 であり,日本語教師 4 の場合は, [日本語学習者のレベルの想定,状況説明,文法] に焦点が置かれていることが分かった。一方,日本語教師 10 の場合は [ストーリー,使い方] に,日本語教師 11 の場合は [文法,話の流れ,日本語学習者の書いた文の意図] に注目していることが確認された。

詳しくは,表 5 の日本語教師 4 のコメントを参照されたいが,日本語教師 4 のコメントから,作文評価において事前に評価ポリシーが提示されていなくても作文評価に対する個人の評価ポリシーを持っていることが分かった。また, [日本語学習者のレベルの想定,状況説明,文法] のように具体的な言及をしていることから,ある程度安定した個人の評価ポリシーに基づき評価が行われたと考えられる。

表 5 日本語教師 4 のコメント

No.	コメント (※原文ママ)
作文 1	レベルは初級後半、作文の構成は文同士のつながりがよくはない、文法の間違いや語彙の特徴から中国の学生なのかなと思う、状況理解して説明していると思う。
作文 2	初級であって作文の構成を判断できるほどの長さではなく、状況理解しているのかどうか判断できるほどの語彙が見られない、4 コマ漫画の各コマを説明すると言うより 4 コマ漫画全体の流れを要約したような感じ。使える語彙だけを使ってなるべく安全に（間違いを避けて）作文を済ませた感じ。
作文 3	文法や状況描写においては初級後半だろうと思うが、三つの文のスタイルがそれぞれで一致しないのは、文型や語彙に対する理解度の低さが表れているのだろうと思う
作文 4	4 コマ漫画と言うこともあり、セリフを書いている感じ。状況を文章で説明することの難しさから、会話文の羅列に逃げている可能性もあると考える。作文と言うより普通の会話活動の文字化に近い。正確でない表記は普通の会話の中での耳コピーに原因があるのではないかと思う。
作文 5	状況理解よりも自分が正確に使える自信がある語彙の使用を優先した作文だと思う。手書きの作文ではない分、漢字の変換ミスをどのように捉えるべきなのかは迷う（作文のレベルの判断基準にすることを迷う）。
作文 6	文のスタイルがバラバラ、長さは十分、語彙力は初級後半位、状況の理解はできていると思う。イラストの状況描写よりも、自分の日本語力を最大限に表出したい学生ではないだろうか。評価を 4 にしたいと思ったが、イラストの状況描写と関係のない部分を除外して考えてみると作文のレベルが高いとは思えない。
作文 7	初級前半の学生と会話をしているような作文。文の長さが足りないため作文の構成、文法、語彙等については判断できる素材がない。ただこれが全力の作文であると仮定して判断すると評価は 1。
作文 8	初級後半以降の学習者が、新しく習得した文型を無理に使った感じがする。しかしそこを間違っているのが残念。流れが自然ではないが「そして」を使って文を繋げようとしているところは、これが作文であると言うことを意識している表れだと思う。文法的な間違いさえなければより評価は高かったと思う。
作文 9	4 コマ漫画を見て会話文を書くのは自然な流れかもしれないが、やはり作文の評価としては、叙述文を書くことの難解さから学生が逃げているのではないかと言う疑念を持ってしまう。文法や語彙だけを考えると初級前半の会話文レベルだと思う。
作文 10	状況描写・作文と言う意識がしっかりある学生だと思う。漢字の変換については手書きではないので考慮しないことにしたい。学生が自分が使える語彙や文法を優先していると言うよりはイラストの内容やその状況描写を優先して、それらを説明するために頑張っている様子がうかがえる。文法的な間違いや語彙のレベルの問題から評価はそこまで高くはならない。
作文 11	かっこつけている。バイト先・日本での生活で耳コピーした日本語を自慢している。その割に文法的正確さや語彙のレベルはそこまで高くない。会話文であることを除いても、作文としての評価は 1・2 レベル。
作文 12	状況描写は少なく、イラストの状況でよく使われる会話文をつないでいる感じ。ただ他の学生より会話文の正確性は高い。敬語もちゃんと使っている（状況にあってはいるが）。語彙や文法の理解度は高いが、その運用能力がまだ備わっていない学生ではないかと思う。初歩的な間違いが少ない分評価を高くしたいとも思ったが、状況を描写した作文になっているのかと思うとそうではないので評価を低くした。
作文 13	状況描写ではなく日記のように作文をしているが、スタイルが首尾一貫しているなのでそこを評価したい。作文の構成としては文全体がまとまりがあって、ちゃんと作文として成立し

ていると思う。完璧な作文ではないが他の学生と相対的に高いレベルにあると思う。

一方、上述した日本語教師 4・10・11 以外の日本語教師のコメントは、気づいた点に焦点が置かれているコメントとなっており、個人における評価ポリシーが存在するかどうかも確認ができなかった。そのため、個人における評価ポリシーがあるとは言えない結果となっている。

4.4 一貫性が問われる評価ポリシー

4.3 では、個人における評価ポリシーがあるケースとあるとは言えないケースが確認されたが、4.4 ではより詳しく、【課題 3】評価ポリシーは、評価者個人の中で一貫性が保たれているのかについて考察する。

表 6 は、日本語教師 1 がスコア（5 段階評価）4 判定をした作文に対する評価コメントである。同じ 4 というスコアではあるが、「描写できていないコマがある。文の作り方は良いが、レベルはそれほど高くない。助詞の使い方が間違っているのが残念。」というネガティブな評価をしている。一方、「ことばは間違えているが、きちんと描写できている。耳で聞いたことばをそのまま覚えているため、正しい表記ができていない。レベルは高め。」とポジティブな評価をしている。

表 6 同じスコアに対する日本語教師 1 の評価コメント

スコア	コメント（※原文ママ）
4	描写できていないコマがある。文の作り方は良いが、レベルはそれほど高くない。助詞の使い方が間違っているのが残念。
4	ことばは間違えているが、きちんと描写できている。 耳で聞いたことばをそのまま覚えているため、正しい表記ができていない。レベルは高め。
4	多くのことばを知っていて、ある程度使えるのでレベルは高め。 会話が得意な印象。主語が私になっているのが気になる！
4	きちんと描写できている。 不自然な部分はあるが、少し練習すればもっと良い文が書けそう。
4	日本での生活が 1 年以上？ 日常良く使う言い回しを知っている。
4	敬語がきちんと入っている。 アルバイトでレジをしたことがあるかも。きちんと描写できている。
4	文をきちんと組み立てることができている。 分かりやすい。

このことから、日本語教師 1 は、個人における評価ポリシーがあるとは言えないと考えられる。つまり、個人における評価ポリシーが確立されていないため、スコアの判定においても揺れが生じていると判断できる。おそらく同じ 4 というスコアではあるものの、その都度気づいた点に焦点を当て判断したため、評価コメントにも揺れが見られたと思われる。ほかの日本語教師の評価においても、このような傾向が見られたことから、評価ポリシーは個人の中でも評価対象の作文が変わることによって、個人における評価ポリシーも変動していく（一貫性が保たれているとは言えない）ものであることが分かった。

5. おわりに

以上の調査結果から、評価者である日本語教師に作文評価における明示的な評価ポリシーへの認識があるとは言えないものの、日本語学習者の作文に対する認識していない何らかの各自の評価ポリシーを持っており(宇佐美 2016)、日本語学習者の作文によってその評価ポリシーは個人の中でも変動していく可能性があることが明らかになった。本研究は、日本語教育に携わっている者に日本語学習者の作文評価に対する評価において、評価者個人が持っている評価ポリシーの多様性を認めつつ、その評価ポリシーにおける「安定した評価」(宇佐美 2016)への認識を提示し、その必要性が示唆されたと考える。評価ポリシーが提示されない場合のある日本語教育現場でも、評価者による「安定した評価ポリシー」に基づき「安定した評価」が行われることにより、妥当性のある作文評価につながるものと期待される。

今後は、日本語教師からの意見を踏まえ Google フォームの評価シートを改善していきたい。特に、調査対象者からの意見の中で「普段は日本語学習者の作文用紙をめくりながら評価(添削)していたが、それができなかったため、評価に一貫性が保たれていない」や「評価基準がないから評価しにくい」といった意見があったため、より明確な指示の必要性が確認できた。また、Google フォームによる作文評価に加え、フォローアップ調査を実施し、評価ポリシーの変動についての実証研究を行いたい。

文 献

- 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」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 安芝恩(2017). 「第二言語環境における中級及び上級日本語学習者の言い換えストラテジーについて」九州大学大学院地球社会統合科学府地球社会統合科学専攻修士学位論文.
- 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慈義―』ココ出版.
- 宇佐美洋(2016). 「人間探求のための「評価」、という新しい視点」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 占部昌蔵(2007). 「項目応答理論を応用した英作文評価者トレーニングの有効性について」『STEP BULLETIN』19回, pp.14-22.
- 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(<特集>言語・コミュニケーションの学習・教育と社会言語科学-人間・文化・社会をキーワードとして-)」『社会言語科学』12巻1号, pp.108-121.
- 田中真理(2016). 「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」―」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 樋口耕一(2020). 『社会調査のための計量テキスト分析―内容分析の継承と発展を目指して―第2版』ナカニシヤ出版.
- 藤浦五月・宇野聖子・小針奈津美・坂井菜緒・柴田幸子・服部真子・中川純子・長松谷有紀(2018). 「初年次レポート・ライティング教育における共通ルーブリック使用時の評価のズレに関する研究」『武蔵野大学しあわせ研究所紀要』1号, pp.3-24.
- 矢部玲子・桐山聡(2019). 「「人間」と「機械」による評価の統合的活用に関する検証―「GoodWriting Rater」を用いた留学生対象の日本語文章作成指導に基づいて―」『情報教育』1巻, pp.1-7.

関連 URL

jReadability PORTAL 日本語の学習者と教師のための Web システム <https://jreadability.net/>
GoodWriting.jp 読み手と構成を意識した日本語ライティング <https://goodwriting.jp/wp/>