

論文

日本語教育における評価研究の変遷と課題

制度が規定する評価から、実践を起点とした評価、
思想としての評価へ

市嶋 典子*

概要

本稿では、日本語教育学会の学会誌『日本語教育』に掲載された評価研究の変遷を追ひ、問題点を考察した。分析の結果、評価研究全体としては、「測定」的、「査定」的の評価が主流を占めていること、OPI や Can-do statements といった「目標到達性の把握」的評価は、背景にある理念と乖離した形で、その枠組のみが表層的に受容されている傾向が明らかになった。これらの問題点を踏まえ、制度に規定される評価から、実践を起点として制度を改革していく評価、思想としての評価への転換を提起した。

キーワード

「実態把握」的評価、教育観、実践を起点とした評価、思想としての評価

1. 日本語教育における評価¹の問題

日本語教育においては、国家政策（留学生受け入れ 10 万人計画・難民の受け入れ等）を背景に、1980 年代から学習者の多様化傾向が見られるようになってきた。その状況を踏まえ「内容中心」「学習者中心」が議論され

* 秋田大学国際交流センター (ichis@gipc.akita-u.ac.jp)

1 本論文では、評価、測定、テスト、アセスメントという言葉を一般的な意味で使用するときは、カギ括弧をつけずにそのまま記述しているが、先行文献の中でカギ括弧がついて表記されている「評価」「測定」「テスト」「アセスメント」について引用する際は、カギ括弧をつけて記述した。

(石井, 1989; 岡崎, 岡崎, 1990), 同時に 90 年代から学習・教授形態にも徐々に変化が生じた。このような流れに伴って, 評価のあり方を再考する必要性が生まれてきた。2008 年に日本語教育学会より発行された『日本語教育』136 号では, 評価の特集が生まれ, 従来の評価の在り方を大きく問い直す主張がなされている。その流れの一つは能力主義を重視した結果主義的評価から, 学習者一人ひとりの学びの過程を重視したプロセス的評価へという変化である。一方で, 庄司, 今村, 楠本, 三枝, 村上 (2008) は, 日本語教育における評価に関する教育現場での取り組みや, そこで得られた具体的な知見は, 日本語教育関係者の中で広く共有されているとは言えない現状にあり, とかく評価の方法や信頼性といった教育技術の問題に関心が偏りがちであるという問題を指摘している。

そもそも「評価」という概念は, 英語の「エバリュエーション (evaluation)」が翻訳されたもので, 戦後, 教育学の分野で使用されるようになったものである。この「エバリュエーション」という概念を提起したのは, タイラー (Tyler, R. W., 1902-1994) である。田中 (2008, p. 26) は, タイラーの提起した「エバリュエーション」の概念の特徴を, ①評価の規準は, 教育目標である。②教育目標は, 高次の精神活動を含む重要な目標郡を含むべきである。③教育目標は, 生徒に期待される行動で記述すべきである。④目標実現の度合いを知るための多様な評価方法が工夫されるべきである。⑤もし, 目標に未到達の子どもがいる場合には, 治療的授業が実施されるべきである。⑥以上のことはカリキュラムや授業の改善につながる。⑦以上のことは, 実践家と研究家の協力によって行われることが望ましい, とまとめている。これら 7 点からは, 教育目標の明確化という視点を重視し, 学習者一人ひとりの学びを視野に入れながら, 授業を改善するプロセスとして評価をとらえていることが分かる。

このように, 評価の概念は, 本来, 教育目標, 教育活動と一体化したものと捉えられていた。評価は, 教育の思想が具現化されたものであり, イデオロギー色の強いものであるといえる。しかしながら, 日本語教育の評価研究の分野においては, 評価方法や技術の問題に関心が偏りがちで, それぞれの評価が, いかなる理念のもと, どのような方法で行われているのか。教育活動とどのような関係があるのか。そこにどんな問題性が隠されているの

かといった事々については、十分な説明がなされてこなかった。

そこで、本稿では、日本語教育における評価の変遷を追い、それぞれの評価研究の基底にある目的や理念を考察する。その上で、評価研究の傾向を明らかにし、そこにどのような問題点が内在しているのかを明らかにする。

2. 分析方法と分析の視点

日本語教育における評価研究の変遷を追うために、日本語教育学会発行の学会誌『日本語教育』（以下『日本語教育』）1号（1962年発行）から146号（2010年発行）の全論文から、評価に関する論文を全て抽出し、内容を分析した。『日本語教育』を分析対象とした理由は、日本語教育学関係の雑誌の中で戦後いち早く刊行され、現在まで継続的に発行されている雑誌であり、評価の過去から現在までの歴史的動向を把握できると判断したためである。抽出の際には、論文カテゴリー、タイトル、サブタイトル、キーワード、要旨に評価という単語や、測定、テスト、アセスメントなど、評価に類する語を含むものを抜き出した。さらに、抽出された論文を、梶田（1983）が提起した①「実態把握」的評価、②「目標到達性の把握」的評価、③「測定」的、「査定」的評価の基準に従い分類し、評価研究の全体的な傾向を明らかにした。

- ① 「実態把握」は、問題となる領域あるいは側面に関してできるだけ多くの情報を集めようというものであって、目標到達性の吟味、対象自体の持つ価値についての値踏みなどといったことには必ずしもこだわらない。学習者一人ひとりの持つ主観の世界をよりよく理解するために個別的面接を行ったり、作文を書かせたりするというもので、志向する目標によって、情報収集の領域や収集すべき情報の内容が定められてくる。「実態把握」的評価は、数値的な結果ではなく、学習者個々の主観的な学びのプロセスを重視し、学習者の学びの実態を様々なアプローチで把握しようとする点が特徴的である。
- ② 「測定」的な性格を持つものは、学習者の諸特性を数値的に表示しようというもので、標準学力テストや知能検査などがこれに当たる。このような測定的評価の場合には、評価内容の標準化と数値化にあ

たって、どのような論理と基準が採用されるかが重要な問題になる。

- ③ 「目標到達性の把握」は、設定された教育目標、あるいは教授・学習目標群を、それぞれの学習者がどのように達成しているか表示しようというものであり、最終的な目標到達点と目標志向的活動の開始点とを結ぶ順次の達成系列上に、それぞれの学習者が示す教育成果を位置づけようとするものである。
- ④ 「査定」的な性格を持つものは、学習者の現状について、何らかの基準に基づき、その価値を妥当な形で値踏み、あるいは表示しようとするものであり、各種の表彰や資格認定などがこれに当たる。

本研究では、評価研究の傾向を明らかにするために、上記の梶田(1983)の分類基準、①「実態把握」②「測定」的な性格を持つもの③「目標到達性の把握」④「査定」的な性格をもつものに沿って、先行研究を分析した。なお、本論文では、分類にあたり、上記の4つの観点を3つにまとめ、①「実態把握」的評価、②「目標到達性の把握」的評価、③「測定」的、「査定」的評価と表記した。「測定」的な性格を持つもの、「査定」的な生活を持つものは、ほぼ概念が重なるため、1つにまとめた²。

なお、分析対象とした論文は、①～③に該当する論文に限定し、評価について論じていても、学習者へのアンケートやコメントによって明らかにした授業への評価や、日本語母語話者による日本語学習者の発話への評価といったような論文は、今回の分析対象に含めなかった。

分析の結果、49本の論文が抽出された。さらに、これら49本の論文を、時期区分ごとに分類し、評価研究の変遷を浮かび上がらせ、問題点を示した。

2 梶田(1983)は、4種の基本的性格類型は、必ずしも個々の具体的評価活動と1対1の対応関係を持つわけではないとし、例えば、知能検査が「測定」的な性格を持つことは言うまでもないにしても、教育活動との関連において、現実にはこれが「査定」的な意味を持つものとして利用される場合も少なくないとする。例えば、日本語教育における評価としては、日本語能力試験が該当する。日本語能力試験は、「測定」的な性格を持つものであると同時に、「査定」的な性格も持っている。

3. 分析結果

分析の結果、以下のような評価研究の傾向が明らかになった。

表1 日本語教育における評価研究の全体的な傾向

	年代	1964～ 1977年	1985～ 1987年	1994～ 1997年	2000～ 2010年	計
「実態把握」的評価		0	0	1	8	9
分「目標到達性の把握」的評価		0	1	1	6	8
類「測定」的「査定」的評価		8	14	4	4	30
その他 (②③)		0	0	0	2	2

* その他 (②③) は、②「目標到達性」把握の評価、③「実態把握」の評価についてどちらも言及されていたものを意味する。

3. 1 「測定」的、「査定」的評価

3. 1. 1 1960年～1970年代：海外の標準テストの受容とテストの妥当性の模索

1960年代には、それまでのGT法（文法訳読法）が実用的ではないという批判から、口頭練習中心のオーディオリンガル法が登場し、隆盛を極めた。オーディオリンガル法は、文法積み上げ式の構成シラバスを基本とし、形式的な正確さを重視した。このような日本語教育の流れに伴い、評価研究においても、日本語の基礎や正確性を図るためのテストのあり方が模索されるようになった。

評価に関する論文は、1964年から掲載され始めた。1964年から1977年までは、③「測定」的、「査定」的評価のみが抽出された。日本語教育学会の機関誌『日本語教育』4・5号（1964）では、「テスト」の特集号を行っている。日本語教育学会・編集委員会（1964）は、日本語教育機関から提供された試験問題を、出題意図と形式の両面から分類整理して、学習者の学習成果が反映されるような形式を整えた試験問題の作成や教育方法の樹立のための手掛かりを示している。また、『日本語教育』32号では、「評価と標準テスト」という特集が組まれている。この中では、評価の手段として、様々な標準テストが紹介されている。例えば、吉川（1977）は、国費留学

生現地採用試験のあり方について、具体的なテスト問題を示しながら説明している。さらに、高見沢(1977)は、米国国務省日本語研修所で行われているテストの方法を紹介している。『日本語教育』55号では、古川(1985)が、TOEFLを中心としたアメリカにおける英語標準テストを紹介し、三枝(1985)は、Goethe-Institutの統一試験の構成と評価方法を示している³。これらの論考は、選抜のための標準テストを評価方法として重視し、日本語教育への応用を示唆している点で共通している。

さらに、テストの妥当性について検討したのが、植松(1977)である。植松は、様々なテストをテストすることによって、それぞれのテストが言語能力を測定する器具として適正であるかどうかを考察した。植松は、絶対評価の妥当性を問題視し、より信頼性のある評価の方法が必要であるとした。具体的には、学生個々の能力差を統計学的処理から得られた数値の差を基に割り出し、この数値の差を基準にして、変数としてのテスト間の差異を比較して、その妥当性を測っている。

このように、海外の標準テストを日本語教育へ応用していこうとする流れが生まれ、さらに、評価方法としてテストの妥当性、信頼性が重視されるようになっていった。しかし、古川(1985)が、「もちろん日本語教育のためにはこうした標準プレースメントテストの重要性は否めないが、なんのために日本語教育をするのかを考えると問題は広がっていく」(p. 128)と述べていることから分かるように、日本語教育において、様々な標準テストの方法が取り入れられ、その妥当性も問われ始めた流れの中で、標準テストの目的や理念をどこに据えるのが問題視されていた。

3. 1. 2 1980年代～2000年代：言語能力を測る客観的指標としての日本語能力試験

日本語学習者の増加に伴い、日本語教育の向上と日本語学習者の学習意欲の増進を目的に、日本語能力を測定し、認定するための日本語能力試験が、私費外国人留学生統一試験から日本語の科目を分離独立させて実施されることになり、1984年12月8日に国内と国外ではじめて一斉に行われるよう

3 他にも、H. C. カンケンブッシュ(1977)が、南オーストラリア州における日本語の大学入学資格試験について報告している。

になった（日本語教育協会・調査普及課，1986）。これに伴い、『日本語教育』58号では、「能力とその測定」という特集が生まれ、様々な日本語能力試験に関する論文が掲載された。掲載されている論文では、日本語能力試験の実施の経緯や手順、試験の構成及び認定基準等を説明している。また、具体的な日本語能力試験の分析を通して、試験の問題点を示し、問題の解決法や試験の向上を示唆している。これらの論文からは、当時、発足したばかりの日本語能力試験の意義を広く認知させ、さらに、具体的な分析を通して、試験問題の精度を上げていく研究の流れを作っていこうとする意図がうかがえる。この特集の中では、言語能力とは何か、言語能力の評価というものがあるのか、どのようなものであるべきかが考察されている。例えば、林（1986）は、言語能力を測るための評価の客観化の必要性を主張している。具体的には、言語能力を「あらゆる知的能力の基礎にある記号操作能力の重要部分を占め、あらゆる知的能力に関係するものである」とし、日本語能力を「日本語を母語としない人の能力」としてとらえ、「日本人が、それらの人々に対して期待するところの言語能力」とであると定義している（p. 2）。また、能力試験の目的として、能力の項目のなかで、具体的事項が身につけているか、どの程度に学習が到達しているか、どのような領域で、どのような程度に、日本人との共同生活ができるか、その保証を与えようとするのか、を挙げている。そして、試験問題の事例が、蓄積され、常に検討されることが望ましいとする。

これらの主張からは、学習者は、日本人と共生するために日本人が必要であると考える能力を習得することが必要で、この能力の到達度を測るために日本語能力試験を機能させようとしていること、さらに、この試験をより客観的な指標として精緻化させていこうとする意図がうかがえる。日本語能力試験は、幅広い層の人の日本語能力を測定し、認定することを目的とし、判定結果は、現在でも、就職や進学、昇給などに大きな影響を与えている。一方で、日本語能力試験1級の高得点者であっても、大学で求められる日本語の運用能力が十分でないケースが見られるという問題点が度々、指摘されるようになっていった（例えば、嶋田，2005）。

3. 2 「目標到達性の把握」的評価⁴

3. 2. 1 1980年代～2000年代：コミュニケーション能力測定のための ACTFL-OPI

知識の多寡ではなく、日本語運用能力を問うことの必要性は、1980年代に指摘されるようになった。上野、古川、田中（1986）は、『日本語教育』58号の中で、英語、ドイツ語、フランス語の能力試験を紹介している。中でも、英語の能力試験では、コミュニケーション・アプローチが注目され、言語能力の観点として、言語そのものから言語運用力に力点が移っていったことを指摘している。日本語教育においても、いかにして学習者にコミュニケーション能力を育成するのかという問題が注目されるようになり、コミュニケーション・アプローチをはじめとした言語教育実践が盛んになっていった。このような流れの中で、牧野（1987）は、ACTFL-OPI（ACTFL Oral Proficiency Interview：以下 OPI）による言語能力基準を提示し、4技能別教育目標、能力テスト、教材との関連を概観している。この試験は、従来のテストでは測りえなかった言語の運用能力を測定するものとして紹介されている。OPI の言語能力基準では、常識的な「聴」、「話」、「読」、「書」の4つの能力をまず認め、それぞれの能力を機能、コンテキスト、正確さの3面から分析的に評価をする。牧野（2000）は、このような言語能力基準に基づく自然な言語伝達の実践を中心にする言語能力中心のアプローチを「言語能力主義」と呼んだ。そして、「語学教育哲学」⁵としての言語能力主

4 「目標到達性の把握」的評価としては、以下に挙げる ACTFL-OPI, Can-do statements の他にも JSL バンドスケールも挙げられる。川上、高橋（2006）が主張しているように、JSL バンドスケールの目的は、日本語能力の判定のみにあるわけではない。判定結果を基に、教師同士が協働的に言語発達について理解を深め、実践の方向性や教育指導的観点を模索するための「手がかり」や「めやす」としての役割が大きい。また、JSL バンドスケールは、主に提唱者の川上郁雄教授の研究室の院生や修了生によって運用、論文化されており、後述する ACTFL-OPI, Can-do statements のように枠組のみが広く流布しているような状況にはない。

5 牧野（2000）は、語学教育哲学としての言語能力主義を、OPI の背景となる思想として、6つの観点から説明している。①言語運用能力は言語を越えて普遍的に評価できるという思想：言語は、その構造の単位と単位の結合の仕方や意味構造のパラダイムが共通しているため、それとの関係で運用能力の一般的基準をたてて、それを使っ

義の考え方とその能力基準の浸透と実践の必要性を主張している。

このような牧野の主張に呼応するように、OPIを参考にした論文が多く見られるようになってきた。例えば、許(2000)は、自然発話における日本語学習者による「テイル」の習得研究をOPIの発話資料に基づいて明らかにしている。また、スニーラット(2001)は、OPIの発話資料を用い、中国語、韓国語、英語母語話者の自然発話から条件表現を取り出し、日本語学習者による習得順序を明らかにしている。また、横山、木田、久保田(2002)は日本語能力試験とOPIとの相関関係を明らかにし、岩崎(2002)は日本語能力簡易試験(SPOT)の得点とOPIの査定による発話能力との関係を報告している。これらは、OPIの手法を用いて、研究内容の妥当性やテストの信頼性を明らかにすることが目的となっており、OPIの思想との関連は言及されていない。

鎌田(2009)は、OPIの問題点を次のように5つの観点にまとめている。
①インタビューという固定した対話場面のもとで展開された言語活動であるため、ロール・プレイ以外、別の難易度を持つ多種多様な「場面」を成立させられない。②「機能・タスク」の遂行能力、「場面・内容」の処理能力、「正確さ」生成能力の3つの要素の区別が曖昧。③能力評価が「正確さ」の表面化した「テキスト・タイプ」の分析に偏りがち。④OPIにおける「文

て能力を評価できる。②言語はコミュニケーションの手段であるという言語道具観の思想：70年代から台頭してきたコミュニカティブ・アプローチの思想に基づいており、機能論的な言語観に立っている。③脱文化の思想：英語の Speaker-Talk(話し手中心の話し方)と日本語の Listener-Talk(聞き手中心の話し方)のような話し方のスタイルの違いなどは、文化的なことだから除外するという思想。④半デジタル化の思想：言語能力の総体は点数化できないという思想、話す能力を幅でとらえていて、デジタル化していない。レベルとレベルの間に顕著な断絶はなく、連続性があるという思想。⑤個人的事情無視のきびしい思想：脱文化の思想と近似しており、受験者の性格とか個人的状況とかはまったく無視して、発話されるスピーチサンプルの総合的機能性、タスク性だけに着眼して能力のレベルを決める。⑥2重役割(dual role)の思想：テスターはテスターであるとともによき会話者でなければならないという思想。日本語の教師という態度を一切捨て切って、「普通の」日本人として受験者とできるだけ自然に会話する。牧野はこれら6つの思想を流動的に進化していくものとしてとらえていくべきであると主張している。

化」の扱いが不透明。⑤固定した「場面」で行われており、「総合的タスク遂行能力」を測ると言いつつ、一部の能力しか見ていない。⑥複言語主義的発想の欠如：Superior の設定である educated speaker の意味は何かという批判を受け、その表現は外されているが、「正確さ」の記述は母語話者の「負担」や「違和感」を軸にしており、欧州共通参照枠の複言語主義的発想はない。

OPI は、このような批判を受けながらも、現在も継続的にテスター教師育成や開発が進められており、その枠組は、様々な教育機関の評価システムの中に浸透している。しかし、言語能力の複雑な要素を、項目化、固定化して測定することの限界は乗り越えられていない。

3. 2. 2 2000年代：言語運用能力の自己判定のための Can-do statements

上述したように、日本語教育において、日本語能力試験の認定基準や出題基準、OPI の判定尺度と基準などが、日本語能力の熟達度を示す枠組みとして機能してきた。国際交流基金では、日本語能力試験の改定と平行し、Can-do statements の開発に取り組んでいる。日本語能力試験では、合格や得点が試験の結果として得られるが、それだけでは、受験者の現実場面での言語運用能力は判断できない。このような問題を乗り越えるために、Can-do statements が注目されるようになってきた。Can-do statements とは、学習者に具体的な言語行動場面を記述した短い文章を提示して、それに対して「できる」「できない」を自己評定により回答させる質問調査である（島田、三枝、野口、2006）。日本語以外にも、英語能力検定試験が「英検 Can-do リスト」を、TOIEC が、「TOIEC Can-do Guide」を公開している。日本語教育学会・試験分析委員会（1999）によると、日本語に関しては、日本語能力試験の妥当性の一環として、1997年に日本語能力試験の受験者を対象に Can-do statements の調査が行われるようになったということである。テストの得点解釈のための Can-do statements は、「合格者が自己評価で「自分はこの項目に書かれたことができる自信がある」と答えたものを統計的な手法を使って分析し作成され、テスト結果から「実際の言語使用場面で何ができるか」を受験者に Can-do statements で分かりやすく示すこと」（塩沢、石司、島田、2010、p. 24）を目指している。島田、野口、谷部、齋藤（2009）は、Can-do statements を利用して、日本の大学と海外協定

校の日本語科目の能力水準の対応づけを行い、技能の習熟度における相違点を検討している。また、山本（2008）は、CEFR に準拠した日本語スタンダードを考察し、Can-do 導入をはじめとする日本語学校の取り組みをモデルとして示している。これらは、Can-do によって、日本語能力判定を透明化、明確化し、組織内でモデル化、システム化をはかっている点が共通している。

Can-do statements は「ヨーロッパ共通参照枠」（以下 CEFR）に導入されたことにより、多くのテスト開発機関や教育機関の評価システムに多大な影響を与えることになった。一方で、細川（2010）は、CEFR の Can-do リストは、語彙・文法の確定を目的としたものではなく、社会での個人の行動を前提として考えられたもので、自分や自分の家族のことから始まり、次第に自分を取り巻く社会へと視野を拓けていく過程がリストに見られるとする。つまり、Can-do リストから、必要な文型や語彙を作成し、それを教育内容とするような動きは、CEFR の理念に逆行するものであると言える。日本語教育における Can-do statements は、CEFR の複言語・複文化の理念⁶とは、乖離した形で、その枠組のみが表層的に流布し、様々な教育機関の中でシステム化されている現状にある。

6 複言語・複文化主義の定義としては、以下の CEFR の定義がしばしば引用される。コミュニケーションのために一つ以上の言語を使い、間文化的やりとりに参加する力であり、そこでは人間が、いくつかの言語においてそれぞれ様々な程度の言語実力を有し、またいくつかの文化での経験をも有する社会的主体として見なされている。この力は、別個の諸能力を縦に並べたり、横に並べたりしたものとしてはみなされず、むしろ使用者がその都度引き出すことができる、複雑で、複合したとさえもいえる存在としてみなされている。（Council of Europe, 2001, p. 168；訳文は、柳瀬, 2007, p. 1による）

The ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent has proficiency, of varying degrees, in several languages and experience of several cultures. This is not seen as the superposition or juxtaposition of distinct competences, but rather as the existence of a complex or even composite competence on which the user may draw. (Council of Europe, 2001, p. 168)

3. 3 「実態把握」的評価

3. 3. 1 1990年代～2000年代：学習者が主体的に自らの学びを評価する 自己評価・ポートフォリオ評価・相互自己評価

言語教育において、学習者の多様化と、それぞれのニーズに対応するもの、学習者に自律を促す指導法として、1990年代以降、自己評価が注目されるようになってきた。自律学習が注目されるようになった背景は、「学習者の多様化」に伴い、学習者の多種多様なニーズに答えるという社会的な要請があった。小山（1996）は、「日本語教育に実際に携わってきた経験上感じられることは、全ての学習者が初めから自律的であるわけではなく、また中には、そのような学習形態を採用すること自体、教師の責任放棄と考える学習者も少なからず存在する」（p. 91）と述べ、学習者をより主体的で自律的な学習へと導くためには、学習者自身が自己の学習目標や学習のプロセスを意識化できることが必要であると主張している。また、自己評価の信頼性は、個々の学習者の学習体験といった主観的な要因に負うところが大きく、一回限りの調査でどのような数値が出たかが問題なのではなく、長期的なスパンでその意味を考えていくことの必要性を述べている。また、トムソン木下（2008）は、海外の日本語教育の現場において、とかく学習者の成績付けのための評価になりがちであるのに対して、学習のプロセスに着目し、学習者のオートノミーを育成することを目指した学習者主導型の評価を提案している。トムソン木下は、日本語教育における評価には、「日本語」に関する目標と「教育」に関する目標があるとし、「教育」の観点からは、教師が学習者の言語習得の一部分に一方的な評価を下すものではなく、言語習得の過程での教師と学習者の相互作用としての評価であり、学習者の現実を反映した評価であるべきだと主張している。また、学習の結果だけではなく、学習プロセスを重視し、学習者と教師の相互作用性の中で育まれる学習者主導型の評価の意義を示している。このように、自己評価においては、教師だけではなく、学習者も評価行為の主体として捉えられていること、学びのプロセスや学習者の自律性が重視されていることが分かる。

横溝（2001）は、「測定」的、「査定」的評価の代替的な評価として、ポートフォリオ評価法に注目した。ポートフォリオ評価の理論的根幹には、1930年代にアメリカにおける教育学の分野で注目された、ジョン・デュー

イの思想がある。1980年代のアメリカでは、従来の「測定」的「査定」的評価によって成績を決めることへの疑問が持たれ始めた。そして、学習者が自らの経験を自分の中で咀嚼し、振り返ることの重要性を主張したデューイの思想（1938/2004）が注目されるようになった。横溝（2002）は、こうしたデューイの思想を具現化した学習者参加型の評価への教育関係者の注目は、アメリカから日本へと波及していったと指摘している。日本語教育においては、川村（2005）が、ポートフォリオ評価を作文クラスにおいて実践し、学習者の学習過程の詳細な記述・分析を試みている。具体的には、「振り返り」と自主学習を記録する「ログ」の2つの内省活動を取りあげ、学習を記録し内省する活動の中で学習者がどのような学びの文脈をたどるのかを分析している。その上で、内省活動の有効性やポートフォリオ評価の目指す学習者の学びに沿った学習支援について考察している。そして、自律性育成を目的とする内省活動を進めていくには、学習者自身が学習に対して問題意識を持ち、内省をする動機や必要性をしっかりと認識することが不可欠であると主張している。

また、市嶋（2009a）は、学習者の主体的な相互行為により、問題意識と論旨が明確なレポートを書くことを目標とした日本語教育実践の評価活動として、相互自己評価を行っている。市嶋は、日本語教育活動、学習活動のプロセスの総体を相互自己評価と位置づけ、このプロセスを経ることによって、自己と他者の表現に責任持てるような主体的な力が育成されることを目指している。

これらの「実態把握」的評価は、言語能力の普遍性だけでなく、動態性、関係性に注目し、学習者の自律性、主体性や学びのプロセスを重視している。さらに、「測定」的、「査定」的評価とは異なり、評価行為の主体として、教師のみならず、学習者をも位置づけており、評価と教育活動とが密着に結びついている点が特徴的である。一方で、このような「実態把握」的評価のあり方は、教師の共感を呼ぶけれども、組織の制度との兼ね合いで、日々の実践に移すことは難しいと考えられていることも指摘されている（Organisation de coopération et de développement économiques [OECD], 2005/2008；武，市嶋，金，中山，古屋，2007）。

表2 日本語教育における評価に関する論文

番号	掲載号	年	著者名	タイトル	分類
1	4・5	1964	興水実	言語テストの種類と作り方	③
2	4・5	1964	編集委員会	实例からみた日本語の試験問題－出題意図と形式による分類－	③
3	32	1977	武井一美	私費留学生統一試験について	③
4	32	1977	吉川武時	国費留学生現地採用試験（日本語）について	③
5	32	1977	植松清	テストを「テスト」する	③
6	32	1977	高見沢孟	米国国務省日本語研修所の場合	③
7	32	1977	石垣貴千代	初級日本語の文型テストについて－書き試験－	③
8	32	1977	H.C. カッケン ブッシュ	南オーストラリア州における日本語の大学入学資格試験について	③
9	55	1985	古川ちかし	アメリカにおける英語標準テスト－TOEFLを中心に－	③
10	55	1985	三枝令子	Goethe-Institutの統一試験の構成と評価法	③
11	58	1986	林大	日本語能力の評価について	③
12	58	1986	大坪一夫	「外国人のための日本語能力試験試行試験」について	③
13	58	1986	山田正春	第1回日本語能力試験の実施について	③
14	58	1986	調査普及課	日本語能力試験について	③
15	58	1986	植松清, 吉中由紀子	「日本語能力試験」を受験者の側から見る	③
16	58	1986	和栗雅子	「日本語能力試験」を受けさせる側として	③
17	58	1986	仁科喜久子, 中山史恵, 川島至	「結果を参考にする側から見た日本語能力試験（統一試験・日本語）」	③
18	58	1986	上野田鶴子, 古川ちかし, 田中望	外国語としての言語能力試験の紹介	③
19	58	1986	清地恵美子, 西口光一, 松本隆	発話能力テストの現状と問題点	③
20	58	1986	小宮さなえ, 久野寛之, 村岡英裕, 柳沢好昭	外国語習得適正テスト－その日本語教育への応用－	③
21	58	1986	石田敏子	英語・中国語・韓国語圏別日本語学力の分析	③
22	61	1987	牧野成一	ACTFL 言語能力基準とアメリカにおける日本語教育	②
23	63	1987	菊池康人	作文の評価方法についての一試案	③

24	83	1994	中島和子, 桶谷仁美, 鈴木美和子	年少者のための会話力テスト開発	③
25	85	1995	フォード丹羽順子, 小林典子, 山本啓史	「日本語能力簡易試験 (SPOT)」は何を測定しているか-音声テープ要因の解析-	③
26	88	1996	横溝紳一郎	オーラル・プロチーブメント・インタビュー-プロフィシエンシーとアチーブメント統合の試み-	②
27	88	1996	小山悟	自律学習促進の一助としての自己評価	
28	94	1997	大坪一夫	評価法	③
29	94	1997	プロティゼ・E・ウッドフォード	ETS テスト開発について	③
30	104	2000	許夏珮	自然発話における日本語学習者による「テイル」の習得研究-OPI データの分析結果から-	②
31	107	2000	横溝紳一郎	ポートフォリオ評価と日本語教育	③
32	111	2001	スニーラット・ニンジャロンスック	OPI データにおける「条件表現」の習得研究-中国語, 韓国語, 英語母語話者の自然発話から-	②
33	113	2002	横山紀子, 木田真理, 久保田美子	日本語能力試験と OPI による運用力分析-言語知識と運用能力との関係を探る-	②
34	114	2002	岩崎典子	日本語能力簡易試験 (SPOT) の得点と ACTFL 口頭能力測定 (OPI) のレベルの関係について	
35	116	2003	庄司恵雄, 青山眞子, 金沢眞智子, 伊藤裕郎, 野口裕之	大規模口頭能力試験開発に関する基礎的研究-発話標本採取法の検討-	
36	117	2003	村上隆	テストはなぜ完全なものになり得ないのか	
37	125	2005	川村千絵	作文クラスにおけるポートフォリオ評価の実践-学習者自己評価に関するケーススタディー	
38	135	2007	野口裕之, 熊谷龍一, 大熊敦子	日本語能力試験における級間共通尺度の試み	
39	136	2008	春原憲一郎	技術研修生のための日本語研修における評価の観点-「技術研修生」が突きつける評価の課題群-	
40	136	2008	上農正剛	聴覚障害児の言語獲得と日本語能力評価をめぐる諸問題	
41	136	2008	トムソン木下千尋	海外の日本語教育における評価-自己評価の活用と学習者主導型評価の提案-	

42	136	2008	山本弘子	日本語学校から見た評価の観点の見直し－ヨーロッパ共通参照枠の視点から－	③
43	136	2008	根岸雅史	英語教育における最近の評価の動向	② ③
44	136	2008	廣瀬香恵	日本留学試験「記述問題」におけるトピックの影響	③
45	138	2008	青木直子	日本語を学ぶ人たちのオートノミーを守るために	
46	141	2009	松見法男, 福田倫子, 古本裕美, 邱愈琰	日本語学習者用リスニングスパンテストの開発－台湾人日本語学習者を対象とした信頼性と妥当性の検討－	
47	141	2009	島田めぐみ, 野口裕之, 谷部弘子, 斎藤純男	Can-do statements を利用した教育機関相互の日本語科目の対応づけ	
48	142	2009	市嶋典子	相互自己評価活動に対する学習者の認識と学びのプロセス	
49	144	2010	藤森弘子	高度専門職業人養成課程における日本人学生と留学生の協働作業及びピア評価の試み	

4. 制度が規定する評価から、実践を起点とした評価、思想としての評価へ

以上のように、日本語教育における「評価」に関する論文を、①「実態把握」的評価、②「目標到達性の把握」的評価、③「測定」的、「査定」的評価、という基準に沿って分類し、その特質を検討してきた。評価研究全体としては、①「実態把握」的評価が9本、②「目標到達性の把握」的評価が8本、③「測定」的、「査定」的評価が30本、抽出された。これらの結果からも、日本語教育における評価研究は、全体的に③「測定」的、「査定」的評価が主流であることが指摘できるが、2000年以降、「実態把握」的評価や「目標到達性」把握的評価に分類される論文も数を増やしてきたことも明らかになった。

「測定」的、「査定」的な性格を持つ評価としては、標準テストや日本語能力試験が挙げられる。日本語教育における「評価」は、1960年代に、日本語の正確性や基礎力を重視する日本語能力観に基づき、これらを測るための

テストを「評価」として位置付けることから始まった。このようなテストによって、学習者の言語習得に関わる要素を、限定的に判定するための一試案を示した。その後、客観的かつ正確に言語能力を測定するための評価方法のあり方が注目され、評価方法としてテストの妥当性、信頼性が問われるようになっていった。また、1970年代から1980年代にかけて、海外の多様な標準テストが注目されるようになっていった一方、標準テストの目的をどこに据えるのが問題視された。

「目標到達性の把握」的評価としては、OPI, Can-do statements が挙げられる。知識の多寡ではなく、日本語運用能力を問うことの必要性は、1980年代から指摘されるようになり、いかにして学習者にコミュニケーション能力を育成するのかという問題が注目されるようになっていった。このような流れの中で、牧野(1987)が従来のテストでは測りえなかった言語の運用能力を測定するものとしてOPIを提唱した。この牧野(1987)の主張に呼応するように、OPIを参考にした論考が多く見られるようになってきた。しかし、これらの多くは、OPIの手法を用いて、研究内容の妥当性を示すことが目的となっており、教育的観点に欠如したものであった。また、OPI自体も、言語能力という複雑な要素を、固定化、限定化、項目化して測定することの限界や矛盾を乗り越えられてはいないこと、複言語主義的な観点に欠けている点が課題として挙げられる。さらに、Can-do statementsについては、CEFRの理念と乖離し、評価方法としてその枠組のみが表層的に受容されている現状が問題として指摘できる。一方で、これらの「測定」的、「査定」的評価や「目標到達性の把握」的評価は、評価方法の中心に据えられ、多くの教育機関でシステムとして機能していることも事実としてある。

「実態把握」的評価としては、ポートフォリオ評価、自己評価、相互自己評価が挙げられる。これらの評価は、言語能力の普遍性だけでなく、動態性、関係性に注目し、学習者のオートノミーや学びのプロセスを重視している。さらに、梶田が定義した「実態把握」的評価に新たな特徴を加えるとするならば、評価行為の主体として教師のみならず、学習者を置き、評価と教育活動とが密着に結びついているという点が挙げられる。一方で、武ほか(2007)は、現実には、能力評価を基にした教育が現存する以上、客観的な日

本語能力を測定するための具体的な方法こそを問うべきであるという議論が、日本語教育実践の現場の教師の間でよくなされることを指摘している。佐藤（1997）は、扱う問題の解決の多くが科学的な知見や合理的な技術の「確実性」で基礎付けられている他の専門職と比較して、教師の仕事はそのほとんどが「不確実性」によって支配されていると述べている。そして、普遍的で万能なプログラムを追い求めたり、見える結果を至上のもとして、学びの結果を過剰にテストで測定したり言語化させたりするのも、この「不確実性」による不安から派生する教師の心性の行動であるとする。佐藤の主張は、武ほか（2007）によって指摘された、日本語教師の測定評価の現状肯定を志向する傾向と重なる。しかし、「能力を確実に測る」といったときの、能力とはいかなるものなのか。テストによって測られている能力は、本当に確実なものといえるのだろうか。春原（2008）が、「測定できること、評価できることが当然だと思いなす傾向があるが、例えば言語能力をはかっているというとき、はかっているのは実ははかれるところしかはかっていないということは忘れられがちである」（p. 5）と述べているように、日本語能力の全貌を確実に把握することは不可能であると考えられる。現実には、能力評価を基にした制度が現存する以上、能力を測定するための具体的な方法や実践のあり方こそを問うべきであるという議論に対しては、その現状認識そのものを問い直すことが求められる。また、その問い直しは、自らの実践の内部で行われるべきものであるといえる。代替案を外側に求め、枠組のみを無批判に受容する安易性こそが問題視されるべきである。

タイラーの提起したエバリュエーション（評価）の原点に立ち戻れば、評価は、教師としての教育目標や哲学を起点として構築されるべきである。いかなる教育目標のもと、どのような日本語教育実践を行うのか、それを実現するためにはどのような評価が考えられるのか、というように、実践における教育目標を起点に評価を考えていく必要がある。制度が評価を規定するのではなく、自らの実践を起点とした評価によって、制度を改革していく視座が重要になるといえる。評価の固定的な枠組そのものを批判的に見るまなざし、そのまなざしを持ちつつ実践に参与し、枠組を揺り動かしながら、自身の思想を評価に具現化していく試みが求められる。日本語教育実践の内部から生成された評価の理念と方法を考察することによって、様々な「実態把

握」的評価を「机上の空論としての評価」から「現場に根付く評価」へ、「思想としての評価」へと発展させていかなければならない。筆者自身、そのための教室デザインのあり方を模索し（市嶋，金，武，中山，古屋，2008；市嶋，2009b），「対話的アセスメント」という評価の考え方とアプローチを提言した（市嶋，2010）。「対話的アセスメント」とは，日本語能力の知識や技能を問う評価ではなく，日本語を含む，言語活動全体を，教師と学習者が主体的，相互的に価値付けていくホリスティックな評価活動である。「対話的アセスメント」においては，評価の目標や基準そのものを教師と学習者が互恵的に創造していく。時には，評価の概念や自身の教育観，学習観を改変し，編み直すことが迫られる。この編み直しのプロセスを創出し続けることによって，教師にとっての言語教育の信念を，学習者にとっての言葉を学ぶことに対する信念を育んでいく。教師と学習者が言語教育や言葉を学ぶことへの固有の信念を育み，これらを共有していくプロセスこそが「対話的アセスメント」であり，また，「対話的アセスメント」は，筆者の言語教育への思想が具現化されたものであるといえる。言語教育における評価に重要なのは，このような教師一人一人の信念や思想であり，思想は，個々の教師に求められると同時に，教育機関として，さらには日本語教育学という研究領域にも問われるべきものだろう。

文献

- 石井恵理子（1989）．学習のとらえ方と教室活動『日本語教育論集』6，1-18，国立国語研究所．
- 市嶋典子，金龍男，武一美，中山由佳，古屋憲章（2008）．学習者が評価する日本語教育実践——自律学習に向かって，細川英雄，ことばと文化の教育を考える会（編）『ことばの教育を実践する・探求する——活動型日本語教育の広がり』（pp. 123-141）凡人社．
- 市嶋典子（2009a）．相互自己評価における学習者の認識と学びのプロセス『日本語教育』142，134-144．
- 市嶋典子（2009b）．プロセスの評価，主体的評価はどのような授業設計で可能か——学習者と教師が共に評価について考える意味をめぐって『WEB版リテラサイズ』6(2)，11-20．<http://literacies.9640.jp/vol06.html>

- 市嶋典子 (2010). 日本語教育における「対話的アセスメント」の実践研究.
佐藤慎二, 熊谷由里 (編) 『アセスメントと日本語教育——アセスメントの多様化とその実践へ向けて』 (pp. 125-152) くろしお出版.
- 岩崎典子 (2002). 日本語能力簡易試験 (SPOT) の得点と ACTFL 口頭能力測定 (OPI) のレベルの関係について 『日本語教育』 114, 100-105.
- 上野田鶴子, 古川ちかし, 田中望 (1986). 外国語としての言語能力試験の紹介 『日本語教育』 58, 111-126.
- 植松清 (1977). テストを「テスト」する 『日本語教育』 32, 15-34.
- 岡崎敏雄, 岡崎眸 (1990). 日本語教育学会 (編) 『日本語教育におけるコミュニケーション・アプローチ』 凡人社.
- 梶田淑一 (1983). 『教育評価 [第二版補訂版]』 有斐閣.
- カッケンブッシュ, H. C. (1977). 南オーストラリア州における日本語の大学入学資格試験について 『日本語教育』 32, 35-46.
- 鎌田修 (2009). ACTFL-OPI における“プロフィシエンシー” 『プロフィシエンシーと日本語教育』 (pp. 3-20) ひつじ書房.
- 川上郁雄, 高橋理恵 (2006). JSL 児童の日本語能力の把握から実践への道すじ——新宿区大久保小学校の実践をもとに 『日本語教育』 128, 24-35.
- 川村千絵 (2005). 作文クラスにおけるポートフォリオ評価の実践——学習者の内省活動に関するケーススタディ 『日本語教育』 125, 126-135.
- 許夏珮 (2000). 自然発話における日本語学習者による「テイル」の習得研究——OPI データの分析結果から 『日本語教育』 104, 20-29.
- 小山悟 (1996). 自律学習促進の一助としての自己評価 『日本語教育』 88, 91-103.
- 三枝令子 (1985). Goethe-Institut の統一試験の構成と評価法 『日本語教育』 55, 133-145.
- 佐藤学 (1997). 『教師というアポリア——反省的实践へ』 世織書房.
- 塩澤真季, 石司えり, 島田徳子 (2010). 言語能力の熟達度を表す Can-do 記述の分析——JF Can-do 作成のためのガイドライン策定に向けて 『国際交流基金日本語教育紀要』 6, 23-39.
- 嶋田和子 (2005). 日本留学試験に対応した日本語学校の新たな取り組み——

- 課題達成能力の育成をめざした教育実践『日本語教育』126, 45-54.
- 島田めぐみ, 三枝令子, 野口裕之 (2006). 日本語 Can-do statements を利用した言語行動記述の試み——日本語能力試験受験者を対象として『世界の日本語教育』16, 75-88.
- 島田めぐみ, 野口裕之, 谷部弘子, 斎藤純男 (2009). Can-do statements を利用した教育機関相互の日本語科目の対応づけ『日本語教育』141, 90-100.
- 庄司恵雄, 今村和宏, 楠本徹也, 三枝令子, 村上京子, 136 号特集ワーキンググループ (2008). [特集] 教育現場から問い直す「評価」について『日本語教育』136, 1-3.
- スニラット・ニンジャロンスック (2001). OPI データにおける「条件表現」の習得研究——中国語, 韓国語, 英語母語話者の自然発話から『日本語教育』111, 26-35.
- 高見沢孟 (1977). テストと評価——米国国務省日本語研修所の場合『日本語教育』32, 35-46.
- 武一美, 市嶋典子, 金龍男, 中山由佳, 古屋憲章 (2007). 活動型日本語教育における評価の意義について考える——「誰が」「何のために」「どのように」評価するのか『WEB 版日本語教育実践研究フォーラム報告』
<http://www.nkg.or.jp/kenkyu/Forumhoukoku/kk-Forumhoukoku.htm>
- 田中耕治 (2008). 『教育評価』岩波書店.
- 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.
- 日本語教育学会試験分析委員会 (1999). 『Can-do-statements 調査報告』国際交流基金.
- 日本語教育学会編集委員会 (1964). 実例からみた日本語の試験問題——出題意図と形式による分類『日本語教育』4+5, 13-59.
- 日本語教育協会調査普及課 (1986). 日本語能力試験について『日本語教育』58, 39-52.
- 林大 (1986). 日本語能力の評価について『日本語教育』58, 1-9.
- 春原憲一郎 (2008). 技術研修生のための日本語研修における評価の観点——「技術研修生」が突きつける評価の課題群『日本語教育』136, 4-16.
- 古川ちかし (1985). アメリカにおける英語標準テスト——TOEFL を中心

に『日本語教育』 55, 119-132.

- 細川英雄 (2010). 議論形成の場としての複言語・複文化主義——言語教育における海外理論の受容とその文脈化をめぐる。細川英雄・西山教行 (編)『複言語・複文化主義とは何か——ヨーロッパの理念・状況から日本における受容・文脈化へ』 (pp. 148-159) くろしお出版.
- 牧野成一 (1987). ACTFL 言語能力基準とアメリカにおける日本語教育『日本語教育』 61, 49-62.
- 牧野成一 (2000). OPI の思想とその可能性『月刊日本語』 9, 56-59.
- 柳瀬陽介 (2007). 複言語主義 (plurilingualism) 批評の試み『中国地区英語教育学会研究紀要』 37, 61-70.
- 山本弘子 (2008). 日本語学校から見た評価の観点の見直し——ヨーロッパ共通参照枠の支店から『日本語教育』 136, 38-48.
- 横溝紳一郎 (2001). ポートフォリオ評価と日本語教育『日本語教育』 107, 105-114.
- 横溝紳一郎 (2002). 学習者参加型の評価法。細川英雄 (編)『ことばと文化を結ぶ日本語教育』 (pp. 172-187) 凡人社.
- 横山紀子, 木田真理, 久保田美子 (2002). 日本語能力試験と OPI による運用能力分析——言語知識と運用力との関係を探る『日本語教育』 113, 43-52.
- 吉川武時 (1977). 国費留学生現地採用試験 (日本語) について『日本語教育』 32, 7-14.
- Council of Europe (2001). *Common European framework of references*. Cambridge: Cambridge University Press.
- Organisation de coopération et de développement économiques (2005). *L'évaluation formative: Pour un meilleur apprentissage dans les classes secondaires*. Paris: OECD Publishing. (OECD 教育研究革新センター (編) (2008). 有本昌弘, 小田勝己, 小田玲子, 多々納誠子 (訳)『形成的アセスメントと学力——人格形成のための対話型学習をめざして』明石書店.)
- Dewey, J. (1938). *Experience and education*. Macmillan. (デューイ, J. (2004). 市村尚久 (訳)『経験と教育』講談社.)