

JF 日本語教育スタンダードに基づいたパフォーマンス評価と 日本語能力試験の合否判定との関係

—国際交流基金研修参加者を対象とした試行調査—

国際交流基金 日本語事業運営部

1. はじめに

JF 日本語教育スタンダード（以下、JFS）は、「相互理解のための日本語」を理念に 2010 年に開発された、日本語の教え方、学び方、そして学習成果の評価のし方を考えるためのツールである。JFS は CEFR（Common European Framework of Reference for Languages）に基づいており、これを利用することで日本語能力の熟達度を他の言語と共通の尺度で評価することが可能となる。

一方、日本語能力試験（以下、JLPT）は 1984 年に開始され、2011 年には世界で 60 万人以上が受験した、世界最大規模の日本語試験である。2010 年に「課題遂行のための言語コミュニケーション能力」の測定を主眼とした改定が行われている。

このように、JFS と JLPT は異なる経緯で作成されており、相互に関連づけられたものではない。また、JLPT が「言語知識(文字・語彙・文法)」「読解」「聴解」という 3 つの能力を、多枝選択方式の筆記試験によって一斉に測る大規模試験であるのに対して、JFS は「産出(話す・書く)」「受容(聞く・読む)」「やりとり(話す・書く)」の各々をパフォーマンスによって個別に評価するための基準である。JFS と JLPT は日本語習熟度を異なる側面から評価するものであり、一方が他方の代用をする関係にはない。

しかし、日本語教育界およびその周辺では、JLPT でレベルの認定を受けた学習者について、一般的には JFS の基準でどの程度のパフォーマンスが期待できるのか(その逆はどうか)を求める声が高い。そこで、今回、国際交流基金（以下、基金）の日本語国際センター（以下、NC）、関西国際センター（以下、KC）、日本語試験センター（以下、TC）の 3 センター合同で調査チームを設置し、基金の研修参加者を対象に JFS に基づいたパフォーマンス評価¹と JLPT を実施し、その結果を照らし合わせることで JFS と JLPT との関係を探るための試行調査を行った。

2. 調査概要

2.1 調査対象

調査対象者は、NC で行われている研修Ⅰ、研修Ⅱ、KC で行われている研修Ⅲの研修参加者である。各研修の概要は以下の通りである。

¹ 本稿では、特定の課題を課してパフォーマンスを表出させるのではなく、各調査対象者の教室内外のパフォーマンスを総合的に評価することを「パフォーマンス評価」と呼ぶ。

表 1 各研修の概要

	研修Ⅰ	研修Ⅱ	研修Ⅲ
調査対象者数 (出身国数)	49名 (30か国)	15名 (3か国)	32名 (30か国)
研修期間 (実施時期)	約6か月 (10月～3月)	約2か月 (1月～3月)	約8か月 (10月～6月)
日本語の授業内容	Aコース：映像視 聴・総合日本語・文 法・読解作文 Bコース：文法演 習・作文・口頭表 現・読解	総合日本語・文法	文法・会話・聴解・ 漢字・読み書き・発 音・スピーチとプレ ゼンテーション・語 彙・読解・ニュース 聴解・社交会話
属性	日本語教師	日本語教師	外交官・公務員
想定される JFS のレベル	A2～B2	A2～B1	A1～A2

2.2 調査方法

調査は、上記の調査対象者に対して実施した JFS パフォーマンス評価と、JLPT の可否判定結果を照らし合わせる方法で行った。調査対象者には研修期間中に JLPT の調査用試験を受験してもらい、その後1週間以内に、研修で授業を担当している講師がパフォーマンス評価を行った。以下、研修Ⅰの参加者に対して実施した調査を調査①、研修Ⅱを調査②、研修Ⅲを調査③と呼ぶ。

3. 調査の実施

3.1 JFS に基づいたパフォーマンス評価の実施

3.1.1 パフォーマンス評価票の作成

パフォーマンス評価票は、JF Can-do の「産出（話す・書く）」、「やりとり（話す・書く）」の中から選び、基金内外の協力者による一次選別、調査チームによる二次選別を経て作成したものに、パフォーマンス評価実施に先立って行なわれた評価シミュレーション（3.1.2 参照）の際のコメントを反映させて完成させた。調査の対象となる各研修の参加者は、それぞれ想定されるレベルが異なるため、3種類のパフォーマンス評価票を用意した（表2）。各パフォーマンス評価票の Can-do 項目数は異なるが、各レベルで扱う Can-do は共通である。なお、調査には JF Can-do（A1 から B2 までを提供）を用いたことから、C レベルは扱っていない。

表2 パフォーマンス評価票の構成

	調査① (研修Ⅰ)	調査② (研修Ⅱ)	調査③ (研修Ⅲ)
A2	20 項目	20 項目	20 項目
B1	20 項目	20 項目	—
B2	20 項目	—	—
合計	60 項目	40 項目	20 項目

表3は、調査①で使用した評価シートの一部である。

表3 パフォーマンス評価票の例

【話す】

4：できる 3：難しいがなんとかできる 2：あまりできない 1：できない

No	Can-do	研修参加者 A	研修参加者 B	研修参加者 C
1	郷土料理について、特徴や店の評判などの詳しい情報を、友人に質問したり、答えたりすることができる。			
2	旅行中、盗難などの問題が生じたとき、警察（や大使館）などに、盗まれたものやその状況などについて簡単に説明し、相談することができる。			
3	友人に自分の出身地など、よく知っている町を案内するとき、名所や名物などを短い簡単な言葉で紹介することができる。			
この研修参加者の「話す」レベルは…？ (A2, B1, B2 のいずれかを記入)				

【書く】

1	自分が買った日用品や食品などについての報告や感想を、ある程度詳しく（ネットショッピングのサイトなどに）書き込むことができる。			
2	(消費者アンケート調査の集計作業などをしてもらう) アルバイトを雇うため、目的や予算など、その必要性を明確にした簡潔な起案書を書くことができる。			
3	必要に応じて確認することができれば、環境問題に関する一般的な意識調査アンケートの質問をだいたい理解し、答えを書くことができる。			
この研修参加者の「書く」レベルは…？ (A2, B1, B2 のいずれかを記入)				

【総合】

この研修参加者の産出能力（「話す」「書く」総合）は…？ (A2, B1, B2 のいずれかを記入)				
--	--	--	--	--

3.1.2 パフォーマンス評価の実施

パフォーマンス評価は、調査対象者 1 名に対し、当該研修参加者を指導してきた 2～3 名の講師が個別に評価することを原則とした。講師には、調査の趣旨説明及び評価の観点や基準をすり合わせるワークショップ及び評価シミュレーションを事前に行った。また、講師には、評価を行うにあたって、以前に受験した JLPT の合否情報を調査対象者に訊ねないこと、スピーチや作文などの成果物を本調査前に講師間で共有しておくことを依頼した。

評価は、後述の JLPT 調査用試験の終了後 1 週間以内に開始し、評価する調査対象者の人数に応じて回答期間を設けた。結果は調査対象者ごとにまとめ、各 Can-do の 4 段階評価及び「話す」、「書く」、両技能の「総合」それぞれのレベル評価を一覧表にした。

同じ調査対象者を担当した講師間の評価結果をすり合わせるために、パフォーマンス評価終了後、ディスカッションを行った。ディスカッションは、各自が記入した評価結果を確認する照合作業で終わらず、パフォーマンス評価票の Can-do をなぜそのように評価したのか理由を述べ合いながら丁寧に進めた。このディスカッションを通して各講師の評価が変更された場合は一覧表を修正し、最終的に調査対象者それぞれの評価を決めた。

3.2 JLPT の実施

TC が準備した JLPT の調査用試験を表 4 の通り実施した。受験するレベルは、過去の JLPT の受験歴、日本語学習歴に照らして調査チームが決めた。試験の内容は年 2 回実施される JLPT と同等のもので、実施条件、採点や合否判定の方法も実際の JLPT と同様に行った。

表 4 JLPT 調査用試験の実施概要

	調査① (研修Ⅰ)	調査② (研修Ⅱ)	調査③ (研修Ⅲ)
試験実施日と 会場	2月5日(日) NC	2月22日(水) NC	2月10日(金) KC
受験者数の レベル別内訳	N1 18名 N2 20名 N3 11名	N4 15名	N5 32名

4. 調査結果

4.1 全体概要

調査①②③の結果を、表 5 にまとめた。横軸は、3.1 で述べた講師による JFS の評価である。今回の調査では、前述のように「話す」、「書く」、両技能の「総合」の 3 つのレベル評価を行ったが、ここでは「総合」の評価結果を用いた。「総合」の評価は、「話す」、「書く」それぞれの評価と、互いに整合性があることを確認した。また、各講師が総合評価の前に行った Can-do 各項目の 4 段階評価と全体

のレベル評価の整合性についても確認した。

縦軸は、JLPT のレベルを示している。それぞれのセルの●が合格者、○が不合格者の人数を表している。JLPT のあるレベルにおける不合格者のレベルを正確に判断するためには、本来、一つ下のレベルも改めて受験させ、その合否を判定する必要があるが、今回の調査ではそこまでは実施していない。しかし、調査対象者の過去の JLPT 受験歴等から、それぞれのレベルの不合格者が、二つ下のレベル以下であることはないと判断されたため、表では、合格者と不合格者の行を連続させて整理することとした。

表 5 JFS の総合評価と JLPT の合否

	A1	A2	B1	B2	
N1			●	●●●	調査①
			○○○○○ ○○	○○○○○ ○○	
N2			●●●●●●● ●	●●	調査①
		○○○	○○○○○ ○○○○		
N3		●●	●●●●●●● ●		調査①
		○○○			
N4	●●●●●	●●●●●●● ●●●●●●●			調査②
	○				
N5		●●●●●●● ●●●●●●● ●			調査③
	○○○○○ ○○○○○	○○○○○ ○○○○○ ○○			

4.2 調査①

表5を見ると、調査①においては、JLPTのレベルがN3からN1へ上がっていくにつれて、JFSの評価もA2からB2へと上がっている、というゆるやかな連関が見られる。

まず、N1に合格した4名のうち、3名がB2と評価されている。今回の調査①で用いた評価票のレベルがA2～B2であったため、総合評価もB2レベルまでしか出されなかったが、この3名は、講師のディスカッションの中で、C1の可能性も指摘されていた。合格者の残りの1名はB1と評価されているが、全体としてJLPTのN1合格者は、JFSではB2以上と評価される可能性が高いと考えられる。

次に、N2合格者を見ると、B1と評価された者が多いが、B2と評価された者もいる。一方、N3合格者は、やはりB1の評価を受けた者が多かったが、A2の評価を受けた者もいた。このことから、JLPTのN2合格者はJFSではB1あるいはB2、JLPTのN3合格者はJFSではA2あるいはB1と評価される可能性が高いと考えられる。

JLPTでは、各レベルで理解できる日本語は「基本的な日本語(N5, N4)」→「日常的な場面で使われる日本語(N3)」→「幅広い場面で使われる日本語(N2, N1)」のように広がっていくとしている。そして、N3は、「基本的な日本語」から「幅広い場面で使われる日本語」への橋渡しとなるレベルとして、2010年に新設されたレベルである。一方、JFSでも、A1・A2が「基礎段階の言語使用者」であるのに対して、B1・B2は「自立した言語使用者」と定義されており、扱える話題も、「ごく身近なもの」から「自分の専門などある程度よく知っているもの」に広がるレベルである。

今回の調査①の結果では、N2やB1を中心にしてみると、JLPTで測っている「受容」を中心とした能力でも、JFSで評価した「話す」と「書く」の総合としての「産出」能力でも、いずれか一方だけが特に高いケースは見られなかった。基礎段階を超えた学習者の日本語能力は、「受容」「産出」ともに向上していることがうかがえる。

4.3 調査②および調査③

同じく表5を見ると、調査②と調査③においては、JLPTのN4、N5、JFSのA1、A2の間には明確な連関が見られなかった。調査②ではA2と評価された者全員がN4に合格しているが、調査③ではA2と評価された者の半数がN5に不合格となっている。また、調査②ではA1と評価された者の大部分がN4に合格しているが、調査③ではA1と評価された者全員がN5に不合格となっている。このように、同じJFS評価であっても、N4およびN5のJLPT合格状況において、逆転している部分がある。

その原因として、JFSのA2がA1に比して幅広いことから、A2と評価された者の中にもJLPTで測る言語知識や受容能力の開きがあった可能性が考えられる。その他、以下のようないくつかの要因が推測される。

調査③は8か月のコースの中盤で行われた。国内で、ある程度長期のコースを行う場合、参加者が日常的に日本語環境にさらされ、また日本人との交流の機会が増えるために、実践的な産出に慣れていることが考えられる。また、このコースが外交官のための研修であったことから、この調査対象者

は、その職業上、パフォーマンス力に長けており、産出能力が特に高かったと推測できる。

他方、調査②では、N4 に合格した者のうち、4 名が A1 と評価された。これは、調査③とは逆に、JLPT で測った日本語力が N4 レベルであったのに対し、パフォーマンスでは相応の評価が得られなかった者が存在したことを示している。これには、まず、調査②の研修が 2 か月という短い期間だったことで、調査時点でも、彼らがそれぞれの国で学んできた学習スタイルを持ち、言語知識や受容能力の方に長けているというアンバランスな日本語能力が影響を及ぼした可能性が考えられる。また、調査②の研修は教師研修であり、その研修目標は日本語力の向上だけでなく教授法等にも重きを置いて、外に出かけて日本語環境に触れる機会も③の研修に比べると少なかったということも影響している可能性がある。

5. おわりに

今回の調査は、基金の 2 つの附属機関で行われている 3 種類の研修の参加者を対象に、それぞれ、いろいろな制約の中で行われた一事例である。

調査①の結果からは、基礎段階を超えた学習者の JFS の評価と JLPT の合否にゆるやかな連関が見られることがわかった。しかし、本調査では、調査対象者の国籍は多様であったものの、日本語教師という同じ職業に携わる者が半年間等しい環境で同じ研修を受けていたという背景から、日本語能力が比較的均質であったと考えられ、今回のデータが、国内外すべての日本語教育現場に当てはまるとは言えない。

一方、基礎段階の学習者を調査対象者とした調査②と③の結果からは、JLPT で測った言語知識や受容能力を中心とした能力と、JFS で評価した「話す」と「書く」の総合としての産出能力に、明らかな連関は見られなかった。それぞれの調査対象者の特性や参加したコースの特徴等に影響されたことも考えられるが、このような技能別能力の不安定さは基礎段階の学習者の特徴と捉えることもできるだろう。

今後、国内外の多くの現場で検証やデータ蓄積を重ねることで、JFS のパフォーマンス評価と JLPT の連関の様相をより明らかにしていきたい。