

複数の評価方法による児童生徒の日本語能力評価の可能性 —SPOT-Y, DLA, JSL バンドスケールを組み合わせた評価方法の活用—

キーワード

児童生徒, 日本語能力評価, SPOT-Y, DLA, JSL バンドスケール

河野あかね・白田千晶・三木杏子・秋本瞳

1. 背景

日本語指導が必要な児童生徒数は年々増加の一途をたどっている。

文部科学省（2024a）によると、「日本語指導が必要な児童生徒数⁽¹⁾は、69,123人で前回調査より10,816人増加」しており、外国籍、日本国籍共に増加している。それに伴い、「(1)学校において特別な配慮に基づく指導を受けている児童生徒の人数及び割合」、および「(2)特別な配慮に基づく指導を受けている児童生徒に占める『特別の教育課程』による日本語指導を受けている人数及び割合」も増加している。

小中学校では2014年度（平成26年度）から、高等学校では2023年度（令和5年度）から、日本語指導が必要な児童生徒に対する「特別の教育課程」による日本語指導が位置づけられた。文部科学省は、2014年に「外国人児童生徒のためのJSL対話型アセスメントDLA」⁽²⁾を作成し、活用を推奨してきた。しかしながら文部科学省（2024b）によると、日本語指導が必要な児童生徒等の判断基準等については、「(イ)児童生徒の学校生活や学習の様子から判断している」が最も多く、次に「(ウ)児童生徒の来日してからの期間を対象基準にしている」「(ア)DLAや類似の日本語能力測定方法により、判定している」が続く結果となっており、JSL評価参考枠や客観的な言語測定基準で判定されることは、まだまだ少ない状況にあることが明らかとなった。つまり、日本語指導が必要な児童生徒が増加している現状を受け、児童生徒の日本語能力を評価⁽³⁾する方法として文科省がDLAを推奨しているが、十分に活用されていないと言える。

さらに、同じく文部科学省（2024b）によると、「特別の教育課程」による指導を実施していない場合の理由としては、「(ア)日本語と教科の統合的指導を行う担当教員がいないため」「(イ)日本語指導における『特別の教育課程』で行うための教育課程の編成が困難であるため」「(エ)拠点校への通級などのための学校間の連携体制が整っていないため」等、物理的な問題が理由にある一方で、「(ウ)個別の指導計画の作成や学習評価が困難なため」「(オ)該当する児童生徒本人または保護者が希望しないため」「(カ)校内に日本語指導における『特別の教育課程』の対象児童生徒がいないと判断するため」などの児童生徒の日本語能力を適

切に評価するための課題も多く見受けられる。

児童生徒の日本語能力を適切に評価するための方法の選択は、学習支援や個別の指導計画の成否に直結するため非常に重要である。文部科学省（2014）は、DLAの「信頼性について」では、実施者の質問や応答の仕方によって、また、実施するための環境要因によって、児童生徒の反応が異なる場合があったり、実施者の採点や評価の仕方が主観的になる傾向もあったりして、結果における信頼性の確保はむずかしいこと」を挙げており、「信頼性を高めていくためには、実施者向けの訓練を充実させることが必要」であると指摘している。

2. 目的

現在、児童生徒の日本語能力を評価する方法には、主要なものとして、①年少者用 SPOT-Y (Simple Performance-Oriented Test for Young Learners) (以下、SPOT-Y) ②外国人児童生徒のための対話型アセスメント DLA (Dialogic Language Assessment for Japanese as a Second Language) (以下、DLA), ③JSL バンドスケール (以下、バンドスケール) の 3 つが存在している。

これまで、3種類の日本語能力の評価方法が各現場で単一で使用されてきたが、対象となる児童生徒の背景や発達段階、評価の実施環境、また評価実施者の背景などにより、日本語能力を包括的に把握することが難しい場合があった。

筆者が所属するインターナショナルスクールでは、筑波大学と連携して年少者用 SPOT-Y の開発、試用に取り組み、完成後も 1 年に 1 度継続的に実施し、酒井（2017, 2018）や河野（2017a, 2017b）らが実施時の様子や結果について報告している。酒井他（2015）では、SPOT-Y について、「10 分間以内という短時間で集団での実施が可能であり、かつ、教師に日本語教育の専門的な知識がなくとも実施できることは、受験者にも実施者にも負担が少ない」という長所がある一方で、「対面式の診断や測定と比べて、測ることのできない部分も多い」という短所があることが示唆された。そのため、SPOT-Y の実施後に、「精度の高い調査を必要とする児童・生徒を見つけ出す使い方も有効であろう」と提案しているが、実際に SPOT-Y と他の日本語能力の評価方法を併用した調査結果や実践報告は見受けられない。

この点について、河野他（2025）は、児童生徒に対し、SPOT-Y に加え、DLA および JSL バンドスケールを使用した日本語能力の評価を実施し、3種類の日本語能力の評価方法の関係性について検討した。3種類の評価方法の是非や優劣を問うのではなく、あくまでそれらの関係性を明らかにすることにより、それぞれの特徴を活かして、より児童生徒の現状を多角的かつ客観的に把握することが可能であることが明らかになった。そして、指導計画立案の参考や教材などの検討資料となったり、他の教師との日本語教育観の理解を深め、児童生徒の学習活動や教師の協働的な実践を作り出す助けとなったりする可能性を示した。

本研究では、代表的な 3 種類の児童生徒の日本語能力の評価方法である SPOT-Y, DLA, バンドスケールを取り上げ、それぞれの特性と有効性を改めて整理し、状況に応じた適切な組み合わせによる具体的な使用の可能性を検討する。そして、児童生徒の日本語指導における

るより効果的な評価を目指すことを目的とする。

3. 3種類の児童生徒の日本語能力の評価方法の概要

以下、酒井他（2015）、酒井（2018）、文部科学省（2014）、川上（2020a, 2020b）、河野他（2025）をもとに、3種類の児童生徒の日本語能力の評価方法の概要についてまとめる。

3-1 年少者用 SPOT-Y (Simple Performance-Oriented Test for Young Learners)

SPOT は Simple Performance-Oriented Test の略であり、SPOT-Y は年少者版として作成されたものである。自然な発話速度で読み上げられた文から、空欄に入る平仮名 1 文字を聞き取って即時的に解答する単純な形式である。聴取の認知的な仕組みを利用して即時的処理能力を測り、日本語運用力を間接的に推定する。空欄部分は「生活言語」「JFL 教科書日本語」「学習言語」として学ぶと考えられる文法項目となっており、各問題間は無関係で独立している。学習段階および発達段階に応じて A1, A2, B1, B2, C1, C2 の 6 種類があり、いずれも 30 問、5 分程度の短時間で一斉に実施でき、信頼性のある結果が得られる。実施者の事前訓練等も不要である。一方で、児童生徒の性格が結果に影響する場合や、学習障害等があつて困難な場合などもある。客観的かつ容易に日本語能力を診断できること、複数の実施報告から児童生徒が実施に対して肯定的に捉えていることが特徴である。

3-2 外国人児童生徒のための対話型アセスメント DLA (Dialogic Language Assessment for Japanese as a Second Language)

DLA は「対話型」を基本としている。日本語能力を含めた児童生徒の実態等を把握するための「診断的評価」、児童生徒の学びやつまずきなどを把握するための「形成的評価」、児童生徒の将来の見通しについて検討するための「総括的評価」がある。対話を重視しながら児童生徒の最大の認知活動を引き出し、実施過程そのものを児童生徒の能力を伸ばす学びの機会として捉え、同時に指導者が児童生徒に向き合う大切な機会であると捉えるところに特徴がある。実施者や実施環境によって、児童生徒の反応が異なる場合や、実施者の評価の仕方が主観的になる傾向があるため、実施者は事前の訓練が必要である。評価結果をもとに個別の指導計画を立てたり、弱点補強や学習進度を調整したりするために有効である。

3-3 JSL バンドスケール

JSL バンドスケールは「観察」を基本としている。小学校低学年、小学校中高学年、中学校、高等学校の発達段階ごとに、聞く、話す、読む、書くの 4 技能について、初級から日本語母語話者に近いレベルまで幅広く診断する。教師と児童生徒のやり取り、児童生徒同士のやり取り、タスクへの理解の過程など、日常的な授業実践における児童生徒の行動や理解から見られることばの力を見立てて、バンドスケールと照らし合わせながら、日本語の発達段階を確認していくことを特徴としている。各段階における日本語能力に合った指導をした

り、授業活動や教材を決定したり、活動を通じて日本語能力の伸びを見通したりすることが可能である。

3-4 3種類の評価方法の特色

3種類の評価方法は、いずれも児童生徒の発達段階を踏まえたものであり、学校における学習を前提とした児童生徒の日本語能力の習得と向上を目指すための評価方法であることは共通している。そのため、それぞれの特性を活かし、目的や学校の状況、児童生徒および実施者の背景などに応じて使い分けたり組み合わせたりして使用することが可能である。また、効果的に使うためには、使い分けや組み合わせの検討が重要である。

表1：3種類の児童生徒の日本語能力の評価方法と特色

	SPOT-Y (SPOT: Simple Performance-Oriented Test, Y: Young Learners)	外国人児童生徒のための JSL対話型アセスメント DLA (DLA: Dialogic Language Assessment)	JSL バンドスケール
共通点	児童生徒の発達段階を考慮、学校現場での利用および活用を想定、評価・測定結果を支援に活かすことが目的		
概要	<ul style="list-style-type: none"> 多肢選択式 聴取の認知的仕組みを利用し、自然な速度で読み上げられた文から、平仮名1文字を聞き取って即時に回答する単純な形式 単純な機械的な採点方法で実用的なテスト 短時間に集団で一斉実施が可能 即時的処理による客観的かつ容易な日本語力の診断 日本語運用力のおおよそのレベルを間接的に推定 児童生徒の発達認知段階に合わせた教材を基盤に問題を出題 実施者の事前訓練等は不要 	<ul style="list-style-type: none"> 1対1の対話型アセスメント 対話を重視しながら児童生徒の最大の認知活動を引き出す テストの実施過程が児童生徒の能力を伸ばす学びの機会 指導者が児童生徒に向き合う大切な機会（対話重視） 日頃の学習の成果や支援活動に必要な情報の取得 実施者の事前訓練が必要 年齢にともなう認知力の発達段階を考慮 母語、年齢、入園年齢、滞在年数（四大要因）による影響を日本語能力が受けけることを考慮した評価方法 一番早く伸びる会話力を使って、潜在的な力を引き出す ラボールを築き、持っている力を思う存分發揮できるよう配慮するとともに、技能別テストが可能かどうかを判断 CF（会話の流暢度）・DLS（弁別の言語能力）・ALP（学習言語能力）の連続性と個別性 各技能別のテスト 日本語の力の6段階の「ステージ」 総合的かつ多面的に記述した「JSL評価参照枠」 「JSL評価参照枠<全体>」は、「在籍学級参加との関係」と「支援の段階」を6ステージで評価 	<ul style="list-style-type: none"> 日常的な授業実践での観察ややり取りにおける児童生徒の理解から見られる日本語の発達段階の見立て JSL バンドスケールと照らし合わせながら、確認 「日本語を第一言語としない児童生徒」(JSL児童生徒)の日本語能力を把握するための「測定基準（ものさし:scales）」として使用 複数の教員による観察と共有 定期的に使用し、児童生徒の発達段階を把握 児童生徒の「個人票」に記録し、長期にわたって使用 「ことばの力」は、児童生徒の第一言語や日本語、また日本語以外の言語と、それらの言語の知識や技能や経験が複合的に結びついているもの <p>「見立て」のプロセス</p> <ol style="list-style-type: none"> 児童生徒を観察する 観察のメモをとる 実践の中で考え、資料を集め バンドスケールのチェックリストと見比べる バンドスケールのレベルを見立てる
項目	<ul style="list-style-type: none"> (1) A-1...生活言語 (2) A-2...生活言語 (3) B-1...小学校国語教科書(低学年)から文法項目の抜粋 (4) B-2...小学校国語教科書(高学年)から文法項目の抜粋 (5) C-1...中学校国語教科書(低学年)から文法項目を抜粋 (6) C-2...中学校国語教科書(高学年)から文法項目を抜粋 	<ul style="list-style-type: none"> (1)はじめの一歩（「導入会話」と「語彙力チェック」） (2)話す (3)読む (4)書く (5)聴く 	<ul style="list-style-type: none"> (1) 小学校低学年(1,2年生) 4技能（聞く、話す、読む、書く）、1~7の見立て（レベル） (2) 小学校中高学年(3,4,5,6年生) 4技能（聞く、話す、読む、書く）、1~7の見立て（レベル） (3) 中学校・高校(中学1年生から高校3年生) 4技能（聞く、話す、読む、書く）、1~8の見立て（レベル）
特徴	短期的 集団で実施 言語の即時処理能力を測る	中期的 個別に測定 1対1のやりとり、対話	長期的 集団の中の個として観察 複数の教員の観察
実際の使用感	<ul style="list-style-type: none"> ひらがなの習得が進んでいないとやや困難 ミニマル・ペア、文字表記(特殊拍)、アルファベットによる誤答、など 性格、モチベーション、体調、周囲の環境からの影響など 指導者の評価・測定を裏付け 他者への説明の材料になる 	<ul style="list-style-type: none"> 口頭のやりとりに支障がない場合、少し評価が高くなる傾向 評価者とのラボールの形成状況による影響 1対1で行うため、場所、時間、評価者の確保が困難 事前に評価者の訓練の必要性 実施者が自分の判断に自信が持てない 	<ul style="list-style-type: none"> 口頭のやりとりに支障がない場合、少し評価が高くなる傾向 メモや資料を集め、レベルを見立てる負担 長期的かつ定期的に見立てる必要性 担任や担当の配置や周囲の環境、状況による困難さ 実施者が自分の判断に自信が持てない

各評価方法は児童生徒の発達段階や観察項目、使用目的に応じて適切に選択する必要がある。一方で、河野他（2025）では「DLA, JSL バンドスケール, SPOT-Y (A, B, C) の3種類の評価方法の結果において、それぞれ明らかな相関が見られた」ことが示されている。そのうえで、「3種類いずれかのツールだけでは困難であった評価を、それぞれの特徴を活かして相互利用することにより、様々な要因を考慮した上で、評価の結果を判断し使用することが可能」であり、「互いに補完し合う形で利用することができ、目的や対象、現場の状況などに応じて使い分けることが可能」であると述べられているように、それぞれの特性を活かし組み合わせて使用することで、単一の評価方法では見落とされる能力の側面を補うことが可能であると考えられる。

4. 3種類の評価方法の組み合わせによる評価の実践

4-1 当校の概要

当校は、幼稚園年少（3歳）から12年生（18歳）までの多様な言語文化背景を持つ児童生徒が在籍する、英語を基盤とした幼小中高の15年一貫校である。日本語の熟達度はゼロ初級者から母語話者まで学年を問わず多様である。第一言語または優位言語が英語という児童生徒が大多数であり、日本語能力は全体的に未熟達者が多い。日本にルーツや何らかの関わりを持つ児童生徒であり、他言語授業の選択が可能になる11～12年生ではごく一部日本語を履修しない者もいるものの、日本語授業に関しては基本的に全員必修である。国際バカロレア（IB : International Baccalaureate）の初等教育プログラム（PYP : Primary Years Programme）、中等教育プログラム（MYP : Middle Years Programme）、ディプロマプログラム（DP : Diploma Programme）の3つの課程の認定校であり、DPでは多くの生徒が英日のバイリンガル資格を取得して卒業している。

4-2 対象

当校の1～12年生（小1～高3）の全児童生徒。

当校で全児童生徒を対象に定期的に実施している①SPOT-Yの結果をもとに、日本語能力が当校の期待する学年相当の段階に未熟な児童生徒を選定し主な対象とした。さらに、普段の授業での様子などから、教師がさらなる日本語能力の評価が必要だと判断した児童生徒についても検証の対象とした。

4-3 手順

- (1) 全員に、①SPOT-Yを実施、結果をExcelに入力
- (2) 教師が、組み合わせ評価を実施する対象の児童生徒を選抜
- (3) 対象の児童生徒に、②DLA、③バンドスケールを実施
- (4) ②DLAの評価参照枠におけるステージ、③バンドスケールのレベルを、教師が判定
- (5) ②DLA、③バンドスケールの結果をExcelに入力

- (6) ①SPOT-Y の結果を基準にヒートマップを作成、3種類の評価方法の関係性を可視化
- (7) 外れ値の児童生徒について個別に解釈

当校における継続的な実施として、①SPOT-Y は YA1, YA2, YB1, YB2, YC1, YC2 の全 6 種類で測定を行った。(ただし、YC1, YC2 は対象が中学生以上そのため、小学生には実施せず。)

②DLA については、DLA 実践ガイドを基に、当校のカリキュラムとの兼ね合いから、教師の判断によりレベル相当と想定される教材などを使用し、授業内での 1 対 1 のやり取りをもとに DLA 評価参考枠を使用した評価を行った。

当校はインターナショナルスクールであり、児童生徒は日常の日本語の使用頻度が全体的に少ないことから、これまでの SPOT-Y の結果において、一条校で学習する児童生徒と比べると日本語の熟達度は 2 年程度遅れていることが証明されている。そのため、③バンドスケールについては、実際の在籍学年でのバンドスケールを主として使用し、それに加えて 2 学年下のバンドスケールを使用した評価も補足的に行い参考にした。

3 種類の評価方法は、それぞれ目的や方法、評価基準が異なるため、一概に比較分析することは不自然であり不可能である。しかしながら、本研究の目的はそれぞれの評価方法の特徴を活かし、状況に応じた適切な組み合わせによる使用のあり方を検討することであるため、3 種類の関係性を見る手段として、分析には各評価方法の以下の①から③のデータを使用した。

- ①SPOT-Y : YA1, YA2, YB1, YB2, YC1, YC2 のスコアの総計
- ②DLA : 話す、読む、書く、聞くの 4 技能のステージの数値の総計
- ③バンドスケール : 聞く、話す、読む、書くの 4 技能のレベルの数値の総計

その上で、①SPOT-Y, ②DLA, ③バンドスケールの結果について、それぞれの関係性を分析した。

5. 結果と分析

まず、3 種類の日本語能力の評価方法について、図 1 のような相関関係が見られた。評価の方法は異なるが、各評価結果が整合していることから、それぞれの評価方法が適切に評価しており、信頼性があることがわかった。

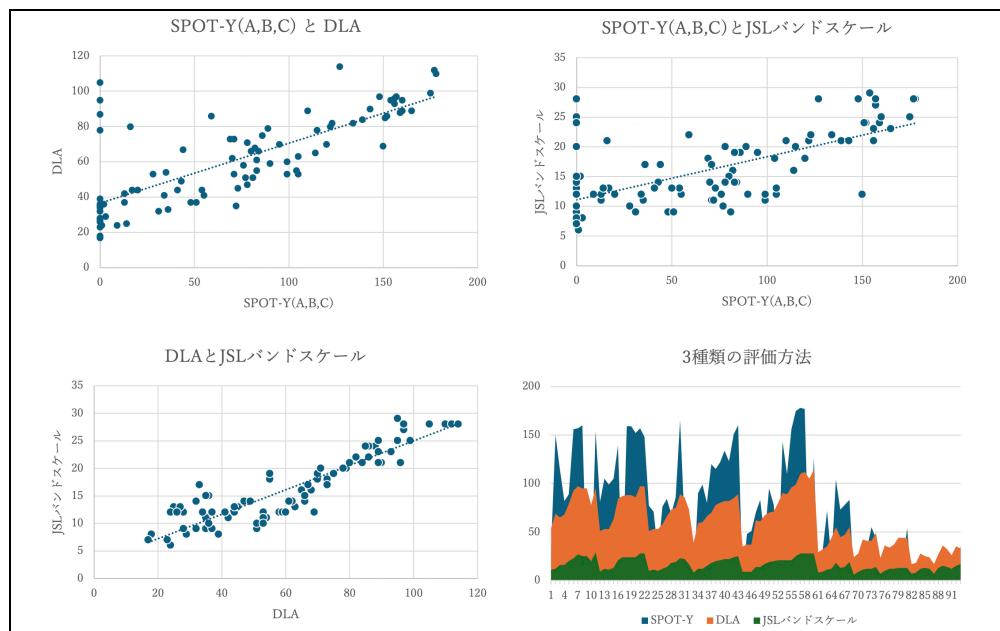


図 1：3 種類の評価方法の関係性

次に、3 種類の日本語能力の評価結果の関係性を視覚的に把握するため、ヒートマップを作成した。

図 2 に示すヒートマップでは、①SPOT-Y, ②DLA, ③バンドスケールの各評価項目間の得点を色の濃淡で表現している。

濃い赤：得点が高い（日本語能力が高い）

薄い赤：やや高い得点

薄い青：やや低い得点

濃い青：得点が低い（日本語能力が低い）

このヒートマップにおいて、3 つの評価方法の色が一致して濃淡が揃っている場合、3 つの評価方法が同じ傾向を捉えていることを示す。すなわち、色の濃淡が 3 つの評価方法で一致しているということは、各評価結果が整合しており信頼性が高いことを示す。具体的には、日本語能力が高い児童生徒は 3 つの評価で高得点を示し、濃い赤で揃い、日本語能力が低い児童生徒は 3 つの評価で低得点を示し、濃い青で揃う。

一方、外れ値のように色が一致しない場合も存在する。この場合は、児童生徒の学習背景や状況を考慮し、得点の意味を個別に解釈する必要がある。例えば、3 つの日本語能力の評価方法の特性によって着目する観点が異なるため、黄色で示した児童生徒のように、色の不一致が生じることがある。

また、3 種類の評価方法を実施した時の児童生徒の身体的・精神的状況が影響する場合もある。例えば、評価の当日に体調不良の場合や睡眠不足で集中力散漫な場合、直前の授業や休み時間に友達と喧嘩して感情が不安定になっている場合、昼食前で空腹状態の場合など、小さなことであっても影響する可能性がある。

子どもの日本語教育研究会
子どもの日本語教育研究 第8号

	児童生徒	SPOT-Y	DLA	バンドスケール
1		1.779377238	1.983246935	1.957320518
2		1.761733538	2.02286094	1.957320518
3		1.726446139	1.548531562	1.46939248
4		1.550009144	1.153335769	1.144107122
5		1.461790646	1.390453245	1.46939248
6		1.461790646	1.153335769	1.46939248
7		1.444146947	1.11381619	1.306749801
8		1.444146947	1.11381619	1.306749801
9		1.408859548	1.469492404	1.794677839
10		1.408859548	1.469492404	1.957320518
11		1.391215848	1.311414086	1.144107122
12		1.391215848	1.429972824	0.818821764
13		1.355928449	1.390453245	2.119963197
14		1.32064105	1.034777031	1.306749801
15		1.30299735	0.995257452	1.306749801
16	a	1.285353651	0.362944183	-0.644962348
17		1.250066252	1.469492404	1.957320518
18		1.161847754	1.192855348	0.818821764
19		1.091272956	0.955737872	0.818821764
20		1.003054458	0.204865685	0.005608368
21		0.879548567	2.141325252	1.957320518
22		0.808973764	0.876698714	0.981464443
23		0.791330064	0.976595555	0.818821764
24		0.756042665	0.402463762	0.330893726
25		0.667824167	0.718620396	0.656179085
26		0.650180468	0.204865685	0.005608368
27		0.579605657	1.153335769	0.818821764
28	b	0.491387172	-0.269369087	-0.644962348
29	c	0.491387172	0.125826707	-0.482319669
30	d	0.473743472	-0.190329928	0.330893726
31	e	0.385524975	-0.269369087	-0.807605027
32	f	0.385524975	0.007267969	-0.644962348
33		0.314950177	0.402463762	0.493536406
34	g	0.226731679	-0.032251611	-0.644962348
35		0.209087979	0.758139976	0.656179085
36		0.156156881	0.600061658	0.493536406
37		0.120869482	0.244385445	-0.31967699
38		0.103225782	0.046787548	-0.31967699
39	h	0.103225782	-0.190329928	0.493536406
40		0.085582083	0.323424603	0.005608368
41	i	0.067938383	-0.348408245	-1.132890386
42		0.050294684	0.244385445	-0.157034311
43		0.015007285	0.441983341	0.656179085
44		0.015007285	-0.506486563	-0.31967699
45		0.002636415	-0.348408245	-0.970247706
46		-0.020280114	-0.07177119	-0.644962348
47		-0.073211213	-0.585525721	-0.482319669
48		-0.090854913	0.980721515	-0.807605027
49		-0.108498612	-0.269369087	-0.807605027
50	j	-0.108498612	0.5210225	0.168251047
51	k	-0.126142312	0.086307127	-0.31967699
52		-0.143786011	0.5210225	0.330893726
53	m	-0.320223006	1.034777031	0.981464443
54		-0.390797805	-0.743604039	-0.644962348
55		-0.408441504	-0.625045301	-0.482319669
56		-0.461372603	-0.901682356	-1.132890386
57		-0.514303701	-0.901682356	-1.132890386
58	n	-0.584878499	0.283905024	0.168251047
59		-0.602522199	-0.427447404	-0.31967699
60		-0.637809598	-0.625045301	-0.482319669
61	o	-0.726028096	-1.059760673	0.168251047
62		-0.743671795	-0.229849507	-0.807605027
63		-0.761315495	-0.743604039	-0.644962348
64		-0.814246593	-1.099280252	-1.132890386
65		-0.867177692	-0.269369087	-0.970247706
66		-1.008327288	-0.625045301	-0.644962348
67		-1.061258387	-0.625045301	-0.482319669
68		-1.114189485	-1.375917308	-0.482319669
69		-1.131833185	-0.704084459	-0.807605027
70		-1.131833185	-0.901682356	-0.644962348
71		-1.202407983	-1.415436887	-0.644962348
72		-1.308270118	-1.21783899	-1.295533065
73	p	-1.32591388	-0.941201935	-0.157034311
74		-1.343557579	-1.415436887	-1.620818423
75		-1.361201279	-0.822643197	-1.295533065
76		-1.361201279	-0.980721515	-1.132890386
77		-1.361201279	-1.25735857	-1.132890386
78		-1.361201279	-1.454956466	-1.458175744
79		-1.361201279	-0.941201935	-0.970247706
80		-1.361201279	-1.020241094	-0.644962348
81		-1.361201279	-1.692073942	-1.458175744
82		-1.361201279	-1.652554363	-1.295533065
83		-1.361201279	-1.25735857	-0.644962348
84		-1.361201279	-1.692073942	-1.458175744
85		-1.361201279	-1.296878149	-0.482319669
86		-1.361201279	-1.099280252	-0.31967699
87		-1.361201279	-1.336397728	-0.644962348
88	q	-1.361201279	-0.980721515	-0.157034311

図2：3種類の日本語能力の評価結果のヒートマップ

さらに、評価の実施環境や実施者が影響する場合もある。例えば、SPOT-Y やバンドスケールでは、集団の中での評価の実施となるため、実施環境により、周囲の友人の目が気になったり、真面目であるが故に評価されていることに敏感になったりすることがある。個別に評価を実施する DLA について、文部科学省（2014）によると「実施者の質問や応答の仕方によって、また、実施するための環境要因によって、子どもの反応が異なる場合があつたり、実施者の採点や評価の仕方が主観的になる傾向もあつたりして、結果における信頼性の確保はむずかしい」と述べられている。

加えて、文部科学省（2014）では、DLA の実施において「『JSL 評価参照枠』のように、子どもの性格も結果に影響する一つの要因となり得る」とされており、酒井・河野・小林（2015）では、SPOT-Y の実施において「慎重な性格だったり、全体的にのんびりとした性格で字を丁寧に書くために速度に追いつけないケース、間違うことを恥ずかしく思い誤答よりは空欄にすることを選ぶケース」などがあることが明らかにされているように、もともと児童生徒が持っている性格が影響する場合も否めない。

外れ値に関する個別の解釈の必要性も含め、ヒートマップは評価方法間の整合性や一致度を直感的に把握するための有効な手段であり、評価結果の解釈や指導計画立案に活用できる。

表2は、外れ値となった児童生徒について結果の傾向と日本語学習背景についてまとめたものである。

表2：外れ値の児童生徒の評価結果の類似グループと日本語学習

グループ	SPOT-Y, DLA, バンドスケール	対象児童生徒の学年	特徴
A	赤, 青, 青	b (中学生) e (中学生) g (中学生) i (中学生)	日本語非母語話者で、家庭での日本語使用もなし、当校には比較的上級学年になってから在籍
B	赤, 赤, 青	a (高校生) c (中学生) f (中学生)	家庭内に日本語熟達者がおり、家庭での日本語使用場面がある
C	赤, 青, 赤	d (小学校高学年) h (小学校高学年)	日本語母語話者だが、海外生活が長いなど日本語から離れて生活
D	青, 赤, 赤	j (中学生) l (中学生) m (中学生) n (小学校高学年) o (小学校低学年)	日本語非母語話者だが、乳幼児期から日本語に触れている／滞日年数が長く、小学校低学年から当校に在籍
E	青, 赤, 白	k (小学校高学年)	滞日年数および日本語学習歴は短いが、学習に対するモチベーションがある
F	青, 青, 赤	p (小学校低学年) q (小学校低学年)	日本語の背景があり口頭でのやり取りはよくできるが、低学年のためひらがなが未定着

当校に比較的上級学年⁽⁴⁾から編入、在籍し、家庭での日本語使用場面も少ない日本語非母語話（グループA）は、SPOT-Yでは高得点でも、DLAやバンドスケールでは評価が低い。基礎的な日本語能力は習得していても、実践の使用場面が少ないことが予想され、対話ややり取りにおける日本語能力とその使用はまだ十分ではないことがわかる。

一方で、当校に比較的上級学年から編入、在籍しているが、家庭内に日本語熟達者がいたり、日本語使用場面が予想されたりする日本語非母語話（グループB）は、SPOT-YとDLAでは高い評価を得られるが、バンドスケールでは評価が低く、授業時などにおける学習日本語の使用には困難さが残ることが予想される。

日本語非母語話者であっても、当校に幼児期から在籍し日本語授業を履修していたり、家庭内に日本語母語話者がいて乳幼児期から常に日本語に触れていたりする日本語非母語話者（グループD）は、SPOT-Yでは評価が低いものの、DLAやバンドスケールでは評価が高い。日本語能力はまだ十分ではないが、日常的に頻繁に日本語使用環境、場面に身を置いていることから、様々なストラテジーを駆使し日本語での対話ややり取りは可能である。

上級学年では SPOT-Y で高得点を取得しやすいが、下級学年では SPOT-Y で得点を取得しづらい。例えば小学校低学年（グループ F）では、音声は聞き取って理解することができていても、ひらがなが未定着のため解答が困難な場合がある。また、文字の習得の度合い以外に、発達段階の未熟さに伴う言語処理能力の有無や、取り組むためのストラテジーの有無、集中力や性格なども影響している。

さらに、日本語母語話者であっても、家庭内で 2 言語以上の使用が予想される日本と外国の両親を持つダブルの児童生徒や海外での生育期間が長い児童生徒は、授業時のやり取りを中心とした観察ベースのバンドスケールの評価は高いが、対話型で日本語能力を診断する DLA は評価が低い（グループ C）。また、これらグループ C の児童生徒は、数年前にはグループ F の児童生徒と同様の結果であったことから、成長に伴い、今後グループ F の児童生徒はグループ C の様相に移行していくことが予測される。

また、滞日年数が短く、SPOT-Y の評価は年齢相応の期待値からは低いが、言語背景、また口頭でのやりとりに積極的に取り組む本人の性格や日本語を学ぼうという姿勢から DLA の評価は高く、授業時には教師や友人からある程度の支援を受けければ自身で活動に取り組めるためバンドスケールの評価は高くも低くもないという児童生徒もいた（グループ E）。

6. 評価方法の組み合わせの意義

3 種類の評価方法を組み合わせた使用により、個別指導の精度が向上し、学習支援の的確性が高まる。評価方法を状況に応じて組み合わせることで、より正確かつ包括的な日本語能力の把握が可能になる。

具体的には、以下のような活用方法が考えられる。

a. 段階的評価

SPOT-Y の一斉評価で日本語習得に困難がある児童生徒を選抜、さらに DLA で 1 対 1 の対話型評価を行い実態を確認、その後バンドスケールで授業時の活動における取り組みやクラスメイトとのやり取りを観察することで、総合的な支援方法を検討する。この段階的評価により、日本語能力の習熟度および各発達段階に応じた適切な教材や指導が選択できる。また、はじめに SPOT-Y を一斉に実施することで、例えば見逃されがちな日本語母語話者の児童生徒の中から日本語指導が必要な者を拾い上げることが可能になる。

b. 補完的評価

3 種類の評価方法はそれぞれ特徴と強みを持つため、単一使用では把握が困難だったり見落とされたりする日本語能力の側面を補完し合うことが可能である。複数の評価方法を併用すると、例えば、SPOT-Y で日本語能力に問題が見られない場合でも、DLA では文脈に応じた発話の適切性や会話理解の精度に課題がある場合がある。また、バンドスケールでは授業活動の遂行に困難が生じていることもある。この場合、各評価の結果をもと

に個別の指導計画を策定することができる。また、評価者や関係者が既存の評価について保証が必要だと感じる場合、別の評価方法を用いることでその評価の適切性を担保することも可能であろう。

c. 状況依存型評価

一見、口頭では日本語使用に問題がないように見えるが、実は日本語能力以外の能力を駆使しているだけで、日本語能力の習得には困難が生じている場合などもある。具体的には、周囲を観察し模倣することで困難な状況をなんとか乗り越えている場合や、試験では形式への慣れから日常生活での評価以上の結果を残している場合などが考えられる。例えば、SPOT-Y の結果では目立った点がない場合でも、教師の経験上、授業や日常生活の中で何かしらの困難を抱えているのではないかと感じられる場合、DLA やバンドスケールを使った評価を行うことで、児童生徒が必要とする支援が明確になる可能性がある。評価対象の児童生徒の文化的言語的特徴や性格なども考慮に入れながら、各評価方法で得られた結果を読み解いていくことが、それぞれの児童生徒にとって適切な指導計画を作成するために肝要である。

このように、状況に応じて適切な評価方法を選択することで、教育効果の最大化が図れる。組み合わせ評価は、それぞれ場面に即した意義があるが、評価結果の解釈には注意が求められる。

児童生徒の日本語能力の評価結果について、評価者や日本語指導担当者から、日本語指導が専門ではない教員や学級担任、他教科の担当教員、保護者などと共有する場合、例えば SPOT-Y の実施結果などは具体的かつ視覚に理解しやすい客観的データとして受け入れられやすいかもしれない。一方で、日本語教育に精通していない者や結果の見方に不慣れな者が単なる数値として受け取ってしまい、何らかの誤解を生む可能性もある。そのため児童生徒の日本語能力の評価については、専門的かつ多角的に評価することが必要である。

評価結果はあくまである一時点の特定の方向から見たものにすぎない。児童生徒が各々にとって適切な支援を受けることができるようになるためには、データの数値のみで判断するのではなく、児童生徒の文化的言語的背景や性格、評価実施時の状況なども念頭に置くべきである。また、児童生徒の優劣を判定したり、できないことに注目したりするのではなく、児童生徒ができる事を知り、どのような支援が必要かつ適切かを、保護者を含めた関係者と確認することが大切である。

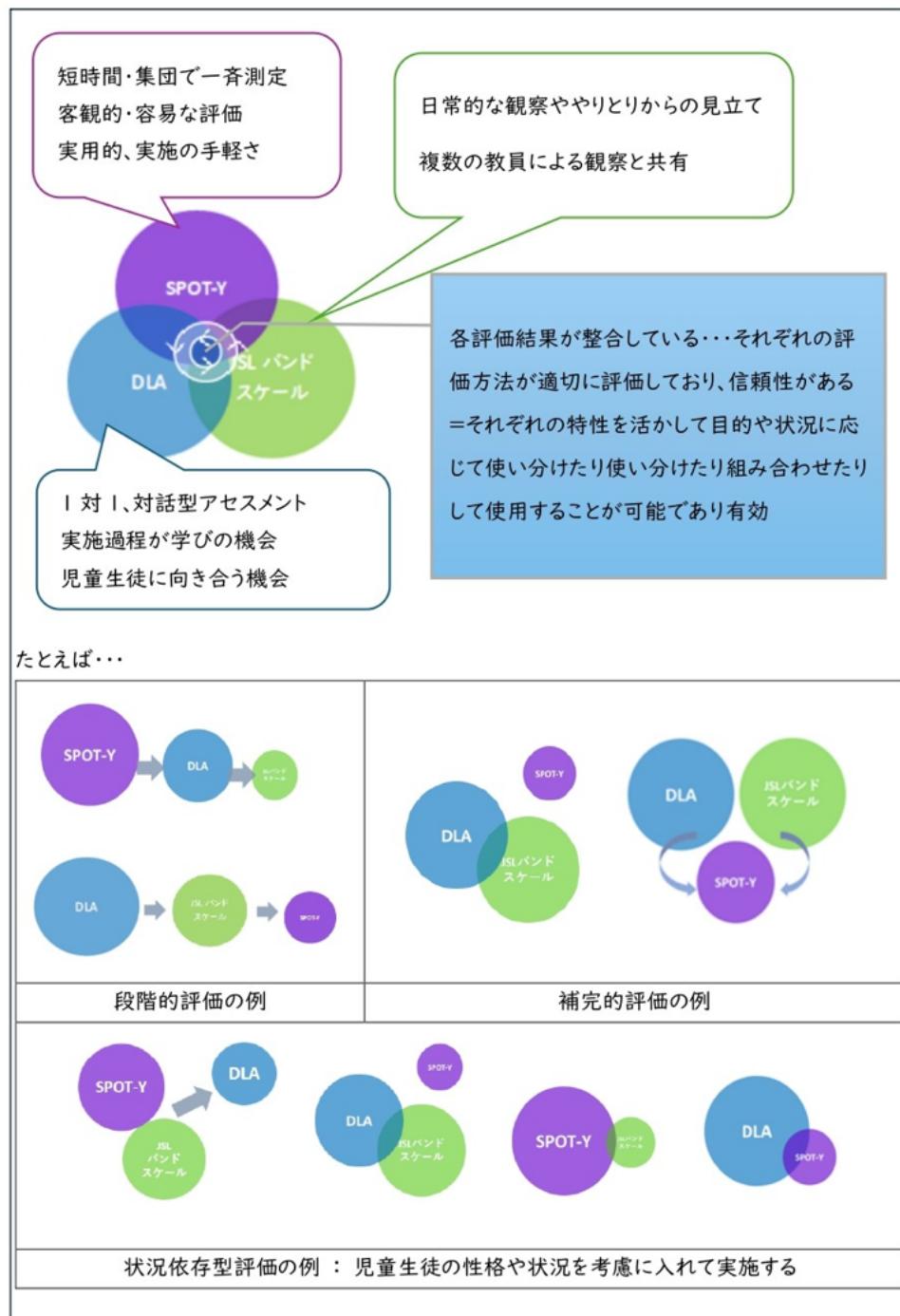


図 3：3種類の評価方法の関連性と組み合わせの可能性

また、3種類の評価方法の相関関係から、指導者および実施者の評価を相互に裏付けることが可能である。例えば、評価や指導経験の少ない指導者にとって、DLA やバンドスケールによる評価は評価者の観察力や判断力に依存するため、なかなか自信が持てない場合があるが、他の評価方法と組み合わせることで、自身の評価の妥当性が裏付けられる。あるいは

は、単一の評価方法だけでは難しい個別の長所や短所について、他の評価方法を参考にしながら、気付くことができる。

以上の通り、評価の目的を明確にし、どの評価方法が適切かを見極めることが大切である。例えば、発達段階ごとの習得状況を比較するため、児童生徒の日本語習得の習熟度の伸びを確認するため、個別の指導計画を立案するため、教材を選定するため、児童生徒へのフィードバックを行うため、保護者の理解を得るため、他の教師との情報交換やチームとしての支援計画作成のためなど様々考えられる。これにより、単一の評価方法では見えなかった課題が明確となり、個別指導の質を向上させることが可能となる。

また、組み合わせた評価結果を活用する意義は多岐にわたる。個別の指導計画の作成、授業や教材の改善、授業の進め方や活動内容の工夫、保護者との連携による家庭学習の方向性や支援方法の改善、他の教師との情報共有と連携、日本語教育プログラムの改善、ひいては学校全体の教育方針への反映などの教育的意義や効果が考えられる。

一方で、組み合わせ評価の実施の際には負荷についても考える必要がある。複数の評価方法を併用すると、評価に時間を要したり、児童生徒にも評価実施者にも負荷がかかったりする場合がある。そのため、適切なスケジュール調整と効率的な記録方法の工夫が求められる。

7. 結論

本研究で示したように、SPOT-Y, DLA バンドスケールはそれぞれ異なる強みを持ち、單一では把握できない側面を補完し合うことができる。目的や状況に応じて評価方法を選択し組み合わせることで、児童生徒一人ひとりの日本語能力をより正確に把握し、適切な支援や指導に結びつけることが可能である。

今後もさらなる評価事例の蓄積により、効率的で信頼性の高い日本語能力の評価方法が確立されることが期待される。また、2025年4月には、DLAの改訂版および、文化的言語的に多様な背景を持つ外国人児童生徒等のための「ことばの発達と習得のものさし」(以下、「ことばの力のものさし」)が公開されたため、それらの活用も含めて、今後もさらに複数の評価方法を組み合わせた効果的な児童生徒の日本語能力の評価について検討していくたい。

【注】

- (1) 「本調査における『日本語指導が必要な児童生徒』とは、日本語で日常会話が十分にできない児童生徒、もしくは、日常会話ができても学年相当の学習言語が不足し、学習活動への参加に支障が生じている児童生徒を指す」としている。
- (2) 2025年に改訂版「文化的言語的に多様な背景を持つ外国人児童生徒等のための対話型アセスメント DLA」が公開された（文部科学省、2025b）。

- (3) 本研究では、以下の各項目内の語を同義として扱うこととする。①「評価」「測定」「アセスメント」②「日本語能力」「日本語力」「日本語運用能力」③「評価者」「実施者」「日本語教師」「日本語指導担当者」④「児童生徒」「子ども」
- (4) 児童生徒の特定を避けるため詳細な学年の記載はしないこととする。また、本研究では中高生を「上級学年」、小学生を「下級学年」とする。

【引用文献】

- (1) 川上郁雄 (2020a) 『JSL バンドスケール【小学校編】—子どもの日本語の発達段階を把握し、ことばの実践を考えるために』明石書店
- (2) 川上郁雄 (2020b) 『JSL バンドスケール【中学・高校編】—子どもの日本語の発達段階を把握し、ことばの実践を考えるために』明石書店
- (3) 河野あかね (2017a) 「多言語背景の児童生徒を対象とした年少者用 SPOT の実施—インターナショナルスクールにおける測定結果と分析—」『日本語教育方法研究会誌』, Vol.24, No.1, pp.106-107, 日本語教育方法研究会
- (4) 河野あかね (2017b) 「年少者用 SPOT の縦断的試行と結果—インターナショナルスクールにおける 4 年間の経過—」『CASTEL/J 2017 予稿集』, 127-134
- (5) 河野あかね・白田千晶・三木杏子・秋本瞳 (2025) 「子どもの日本語能力の評価・測定ツールの関係性—DLA, JSL バンドスケール, SPOT-Y の 3 種のツールの相関分析—」『日本語教育方法研究会誌』Vol.31, No.2, 92-93, 日本語教育方法研究会
- (6) 酒井たか子・河野あかね・小林典子 (2015) 「年少者用 SPOT の開発：問題作成とインターナショナルスクールにおける試行」『筑波大学留学生センター日本語教育論集』30 号, 21-33, 筑波大学
- (7) 酒井たか子 (2017) 「多言語背景の児童を対象とした日本語診断テストの開発研究」『日本語教育学会春季大会予稿集』, 332-337, 日本語教育学会
- (8) 酒井たか子 (2018) 科学研究費助成事業 「多言語背景の児童を対象とした多層分岐適応型日本語力診断オンラインテストの開発」研究成果報告書
- (9) 文部科学省 (2014) 「外国人児童生徒のための JSL 対話型アセスメント DLA」URL: https://www.mext.go.jp/a_menu/shotou/clarinet/003/1345413.htm (2025 年 1 月 15 日閲覧)
- (10) 文部科学省 (2024a) 「報道発表資料」URL: https://www.mext.go.jp/content/20240808-mxt_kyokoku-000037366_01.pdf (2025 年 9 月 30 日閲覧)
- (11) 文部科学省 (2024b) 「日本語指導が必要な児童生徒の受入状況等に関する調査（令和 5 年度）」の結果について」URL: https://www.mext.go.jp/b_menu/houdou/31/09/1421569_00006.htm (2025 年 9 月 30 日閲覧)
- (12) 文部科学省 (2025a) 「文化的言語的に多様な背景を持つ外国人児童生徒等のための

ことばの発達と習得のものさし（略称「ことばの力のものさし」）」 URL:
https://www.mext.go.jp/a_menu/shotou/clarinet/003/1345413_00002.html （2025 年
9 月 30 日閲覧）

- (13) 文部科学省 (2025b) 「文化的言語的に多様な背景を持つ外国人児童生徒等のための
対話型アセスメント DLA, Dialogic Language Assessment for Culturally and
Linguistically Diverse Students.」 URL:
https://www.mext.go.jp/a_menu/shotou/clarinet/003/1345413_00003.html （2025 年
9 月 30 日閲覧）

(河野あかね；つくばインターナショナルスクール/
白田千晶；元つくばインターナショナルスクール/
三木杏子；つくばインターナショナルスクール/
秋本瞳；つくばインターナショナルスクール)