

〔実践報告〕

オーラルテストにおける目標の可視化を目指した 形成的評価の一試案

— Can-do Statements によるチェックリスト型評価シートの開発 —

近藤 行人・田中 典子・藤森 秀美・古本 裕子

名古屋学院大学留学生別科

要 旨

初級の日本語クラスで課ごとに行うオーラルテストにおいて、効果的に形成的評価を行うことを目指して Can-do Statements チェックリスト型評価シートを開発した。この評価シートは、その課のタスク達成に関わる「内容」とコミュニケーション言語能力に関わる「言語」に分かれ、項目をチェックすることで、タスクで要求される項目ができたかどうかを確認される。この評価シートは、目標の可視化、教師と学習者の評価基準の共有、パフォーマンスの振り返りの促進などに有用であった。本稿ではこの評価シートの特徴と開発経緯について報告する。

キーワード：形成的評価, Can-do Statements, タスク型オーラルテスト, コミュニケーション言語能力

A proposal for the formative assessment designed to visualise
learning targets on oral performance tests:

Development of checklist of “Can-do Statements”

Yukihito KONDO, Noriko TANAKA, Hidemi FUJIMORI, Yuko FURUMOTO

Institute for Japanese studies
Nagoya Gakuin University

1. はじめに

本稿では、交換留学生を主な対象とした日本語教育機関において行われた形成的評価に関する実践を取り上げる。実践が行われた機関では、初級から上級までレベル別のクラスが開講されている。このコースは学生のレベルや個別の学習状況に合わせて柔軟に運営され、日本語教育に対する高い満足度を得てきた。しかし、初級クラスは、修正を加えながらではあるが、15年以上前に設定された目標やそれに合わせた教材を使用したコースであったため、近年の外国語教育の流れや現在の学生の実態に即したコース運営の必要性が生じた。そこで、使用する教科書を変更し、初級のコース運営の見直しを行った（國澤・近藤 2016）¹⁾。本稿は、この初級コースにおける取り組みの一つであるオーラルテストとその評価シートについて報告するものである。

本稿で報告する評価シートを使用する対象となる初級レベルのクラスは、1学期間週8コマ（1コマ90分）開講されている。このクラスの学習者は、北米からの交換留学生が多く、学習歴はあっても、日本語によるコミュニケーション能力が十分ではないことも多い。2015年に実施されたニーズに関するアンケートでは、4技能のうち、「話す」能力を伸ばしたいと回答した学習者が最も多かった。このため、コース見直しの過程においては、とりわけ口頭能力の向上を実現することを目的とし、1課ごとに口頭能力を評価するためのオーラルテスト²⁾を取り入れることにした。このオーラルテストは形成的評価を目的としており、課ごとの行動目標を実現するようなタスクを設定し、実施している。評価にあたっては、形成的評価とフィードバックを円滑に行うための Can-do Statements によるチェックリスト型評価シートを開発し、利用している。本稿では、このオーラルテストの評価に用いるための評価シートの開発について報告する。

2. 口頭能力評価のための理論的枠組み

このコースで実施するオーラルテストを評価するためには、どのような枠組みを用いるべきであろうか。近年の言語教育では、ある言語がどの程度できるかを示す物差しとして、ヨーロッパ共通言語参照枠（Common European Framework of Reference, 以下、CEFR）が広がっている。CEFRの特徴の一つは行動中心主義的理念に基づく能力観であり、単に言語に関する知識（語彙や文法など）を問うのではなく、学習者が当該言語を用いて「何が」できるのかを問う。行動が「どの程度うまく」遂行できるのかに基づいており、全て Can-do Statements で記述されている点特徴的である。

また、CEFRでは、コミュニケーション意図を実現させるために必要な能力があると考えられており、このうち言語に関連の深いコミュニケーション能力をコミュニケーション言語能力と呼んでいる。このコミュニケーション言語能力は、言語構造的な能力、社会言語能力、語用能力によって構成される（Trim, North, Coste 2014, p.116, 国際交流基金 2017, pp.7-8）³⁾。「言語構造的な能力」とは、語彙、文法、発音、文字、表記などに関する能力であり、「社会言語能力」は、相手との関係や場面に応じて適切に言語を使う能力である。「語用能力」は、ディスコース（談話）を組

み立てたり、コントロールしたりする能力であるディスコース能力と、コミュニケーションの中での言語使用の役割や目的（例：事実を報告する、説得するなど）を理解したうえで適切に使用できる能力である機能的能力の2つで捉えられるものである。

コースで用いるオーラルテストにおける口頭能力評価では、テストで設定されたタスクにおける活動の成否を評価すること、そして、そのタスクが実現される言語活動の質を測るための言語コミュニケーション能力を評価することが必要であると考えた。このため、本稿では、CEFRやCEFRを日本語教育へ適用することを試みているJFスタンダードの考え方を基に、「行動目標の達成の成否」、及び「コミュニケーション言語能力」に関するCan-do Statementsを用いた評価シートの開発と使用について報告する。

3. Can-doチェックリスト型評価シートの開発

3.1 口頭能力を測るためのパフォーマンステスト

対象となる初級クラスでは、先にも述べたコース運営の見直しに伴い、初級教科書『できる日本語』⁴⁾を採用した。この教科書では、課ごとに行動目標が示されているが、オーラルテストでは、学習内容に合わせたタスクを課し、各課で学習した内容を用いて、各課の行動目標が達成できたかどうかを見る。準備した会話の暗記・再生を測るのではなく、タスクの目的にかなったやりとりができるか、まとまりのある話を聞き手に向けて話すことができるかを評価する⁵⁾。このため、

表1 『できる日本語初中級』の第3課オーラルテスト学生配布用カード

テスト1 (学生Aと学生Bの会話)

ロールカードA

{春・夏・冬} 休みの予定について、クラスメートにたくさん質問してください。

例：「休みになったら、何(を)する？」

「休み、どこか行く？」「そこはどんなところ？」

「いつ・どうして・どうやって…」

ロールカードB

あなたは{春・夏・冬} 休みに何かしようと思っています。

自分の休みの予定について、クラスメートからの質問に答えてください。

できるだけ、詳しく話してください。

テスト2 発表

スピーチ

大学を卒業したら、どんなことをしたいと思っていますか。あなたの将来の夢について発表してください。将来の夢は仕事でも他のことでもいいです。

例：・その夢のために、今、どんなことをしていますか。

・その夢のために、これから、どんなことをするつもりですか。

・その夢のために、何をしなければなりませんか。

オーラルテストは、基本的に、各課の行動目標を達成するようなロールプレイなどによるやり取り形式のタスクと、発表やスピーチなどによるモノログ形式のタスクを組み合わせで作成し、そのパフォーマンスを評価している。テストは課ごとに実施し、各課の到達目標に合わせて2～4つのタスクを設けている。

例えば、「目標や計画を話したり進路の参考のために周りの人から話を聞いたりすることができる」ことを目標とする『できる日本語初中級』第3課では、「春休みの計画について話したり、相手に質問したりする」というやりとり形式の会話タスクと、「自分の将来の夢や目標について話す」というモノログ形式の発表タスクを設定している。

表1は、このテストで実際に使用した学生配布用資料である。このテストではテスト1で、学生同士がロールプレイを行う、テスト2で将来の夢について簡単なスピーチを行うというタスクを課している。後述の評価シートだけでなく、このテストで何を話すのか、どんなことを行うのかについては、事前に学生用カードとして配布し、説明を行っている。学生はこの学生配布用カードと、評価シートを基にオーラルテストの準備を行うことになる。

3.2 評価シートの開発

3.2.1 評価シート開発の方針

先にも述べたように、オーラルテストが評価するものは、各課に設定された行動目標の達成がどの程度できたかである。このため、各課のオーラルテストで行われるタスクに対しての目標が達成できたかどうか、そして、そのタスクにおける会話や独話の質的な側面を見られるかどうか、を評価できるような評価シートの開発を目指した。実際の評価シート開発にあたっては、1) 具体的な到達目標が目に見える形で示される、2) 評価基準が学習者・教師間で共有できる、3) テスト実施直後にフィードバックができるという3つの方針で開発を行った。

そこで、これらの方針を実現するため、チェックリスト型評価シートの作成以前に、ループリック形式の評価シートを作成した。次項からは、ループリック形式の評価シートの開発について述べ、そこで生じた問題意識を基に行われたチェックリスト型の評価シートの開発について述べてい。

3.2.2 初期の評価シートの開発

コースの見直しに伴い、当初はCan-do Statementsを用いたループリック形式の評価シートを作成した。ループリックとは、「成功の度合いを示す数値的な尺度あるいは評語と、それぞれの数値や評語に見られる認識や行為の質的特徴を示した記述語からなる評価基準表」である(田中2010, p.48)。このループリックは、「ある課題について、できるようになってもらいたい特定の事柄を配置するための道具」であり(スティーブンス他2014, p.2)、評価に用いられるとともに学習目標を示すことができるとされる。当初作成した評価シートでは、これらの特徴を生かすように、「内容」と「言語」に関するループリックを作成した。「内容」では、各課のタスクの達

成に必要な項目をループリックに書き込んだ。初中級第3課のテスト2では、表2のようなループリックを作成した⁶⁾。

最も高い評価である **excellent** の記述として、「長い休みの予定について、友達に、日程、場所、理由、手段など」を「色々と上手に話すことができます」、**good** では、「上手に」を削り、「色々話すことができます」のみとし、**poor** では、「あまり話すことができません」、**bad** では、「ほとんど話すことができません」とした。

この評価シートを用いることによって、学生に対し、テスト前に評価基準を開示したうえでオーラルテストを実施することが可能になり、学習者に対するタスクや評価の可視化をもたらすことができた。また、テスト実施前にクラス担当でループリックを作ることで、あるタスクができるというのはどのようなことなのかという点についての共通理解を形成することができた。そしてタスク達成について、学習者がどの到達度にあるのかを示すことができるようになった。しかし、運用を続けていくうちに、次第に次のような問題点が生じてきた。一つは、学習者の読む負担が大きいう点である。学習者は、テストタスクごとに書かれた詳細で複雑な文を4段階に分けて読まなければならない、また、読み込まなければテストタスクで要求していることが明確に伝わらないという評価シートとなっていた。また、基本的に条件や話すべき内容を取り上げている部分が重なってしまい、全て読む必要がないものも見られた。特に初級の学習者には負担が大きく、媒介語による翻訳をつけたとしても、評価まで厳密に確認してテストに臨む学習者は多くなかった。教師からも「ループリックは一見シンプルに見えるが、どの部分が違うのか、その差の境目がわかりづらかった」とのコメントが出ていた。

また、テストタスクごとに異なる内容のループリックを作成しており、休みの予定について話すタスクに対しては、「休みの日程、場所、理由、手段が話せる」といったように、ループリックの記述には特定の場面やタスクを想定した具体的な内容を記述した。これは開発方針に即したものであるが、「何ができるか」という観点から差異をつけるのは難しく、4段階とするために、「詳しい」「上手」「わかりやすい」「あまりわかりやすすくない」といったような条件をつけていた。これ以外に話題や活動といった要素では区別することが難しかったため、何段階ものループリックを記述する必要性はなく、ある基準に対して数値をつけるような形式の評価基準との差異化を図るのが難しかった。

表2 初期の評価シートの記述

excellent	長い休みの予定について、友達に日程、場所、理由、手段などについて、色々上手に話すことができます。
good	長い休みの予定について、友達に、日程、場所、理由、手段など、色々話すことができます。
poor	長い休みの予定について、友達に、日程、場所、理由、手段など、あまり話すことができません。
bad	長い休みの予定について、友達に、日程、場所、理由、手段など、ほとんど話すことができません。

また、評価の際に混乱が起こってしまうということがあった。このテストで求めたルーブリックの Can-do Statements には、「日程、場所、理由、手段」という4つの特定の話題が含まれていた。実際のパフォーマンスでは、ある学習者は日程（夏休み）、場所（どこへ行くか）、は上手に話しているが、「理由」や「手段」は話せなかった。また、別の学習者は「理由」は話せたが、「場所」については詳しく話せなかった。このような場合に評価の差異化をどこで測ればよいかについて、教員間で混乱が見られるようになったのである。これは、評価基準作成時の想定による4段階の差異と評価担当者の解釈がずれてしまっていることや、4段階の差異を示している「詳しい」「上手に」などの条件のみで画一的に決められるものでなかったことが要因として挙げられる。

3.2.3 チェックリスト型評価シート

このような問題点を解消するため、学習者の読む負担を減らすこと、課ごとに個別に設定した様々な話題を別々に評価すること、フィードバックできることを目指す評価体制の構築を課題とした評価シートの開発を行うことにした。目標の可視化を目指した評価体制では、教師と学習者がタスクで目指すことを共有し、そして学習者はそのパフォーマンスに対する教師や仲間、あるいは自分自身による振り返りで得られるフィードバックを受け入れることが重要である。ただし、学習者が参加し、フィードバックが機能するには、受け入れる側にも条件が必要であり、自発性の訓練が必要であることも指摘されている。すなわち、学習者を訓練したり、意識を高めたりすることが必要であり、チェックリストはこの訓練に有用な手段だと言う (Trim, North, Coste 2014, p.208)⁷⁾。

一般的にチェックリストとは、「照合を行う表」のことであり、あるタスクの遂行にあたり、その諸内容、諸基準といった様々な項目を Yes か No の2段階で記入するものであり、ある作業ができたかできなかったかというように、作業手順の確認や点検に用いられることが多い。このため、あるタスクを達成するために必要な項目をリスト化して提示し、全ての項目ができれば高いパフォーマンスを得られるようリストを提示すれば、学習者にとっても教師が何を求めているかが明確にわかり、準備を行う波及効果が期待できると考えた。また、テスト後にチェックリストを用いてパフォーマンスを振り返ることで、学習者自身による振り返りや教師によるフィードバックが伝わりやすくなるのではないかと考えた。

新たに開発する評価シートでは、学習者がタスクの準備をしたり、パフォーマンスを振り返ることを支援したりするため、一つの項目に詳細に必要な事項を書き込むのではなく、できるだけ少ない要素の組み合わせで構成した Can-do Statements によるチェックリストを作成することにした。CEFR で書かれる Can-do Statements はそれ自体がチェックリストとして示されうる (p.203) ものではあるが、CEFR の Can-do Statements は「個人的な情報を求めたり、与えたりすることができる」のように抽象的な文言であることも多い。しかし、タスクに応じて「自己紹介できる」「どこに住んでいるか言うことができる」のように、内包されている構成要素に分解できる (p.202) ことも指摘されている。今回新たに作成を目指した評価シートでは、テストタスクに応じて具体的な能力記述文を作り、自己評価、もしくは教師による評価に用いることを目指した。評価シ

トにおける Can-do Statements の具体的な作成方針については、3.2.3.2 で示す。

評価シートは課ごとに設定されたテストのタスクで要求される言語行動の Can-do Statements がリスト化され、これができるか、できなかったかを照合する表となっている。学習者の読む負担を減らし、それぞれのタスクが求める言語活動を明確にし、判断しやすくできることを目指したのである。

3.2.3.1 「内容」と「言語」

評価シートは大きく、「内容」と「言語」に分けられる。「内容」は、テストタスクで目標としている項目の行動目標とそれを達成するために必要な言語活動を記述したものである。その課に特有のタスク内容を示す項目であり、課ごとのタスクに応じて全て異なる Can-do Statements を作成している。

一方「言語」はコミュニケーション言語能力の質的側面⁸⁾に関わる項目であり、テストタスクに関わらず、どの課でもある程度共通する項目である。このため、「言語」に関しては、全てのテストで共通して使用できるリストの作成を目指した⁹⁾。次項より、内容と言語の順に述べていく。

3.2.3.2 「内容」

「内容」は、行動目標を達成するために必要な言語活動の Can-do Statements である。一般的に Can-do Statements は、条件、話題・場面、対象、行動といった要素を組み合わせることによって、その言語活動において、どの程度うまくできるかが記述される（国際交流基金2017）。

一方で、本稿の評価シートでは、採用したチェックリスト型の利点を生かすため、一つの Can-do Statement に多くの要素を加えるのではなく、テストタスクで目標としている項目の行動目標とそれを達成するために必要な言語活動を細分化して記述するような Can-do Statements の作成を目指した。

例えば、前述の3課のタスクのルーブリックに用いた Can-do Statement については、以下のような要素が組み合わせて用いられている。

(教室で) 友達に 長い休みの計画について, 日程, 場所, 理由, 手段 を話すことができる。
場面 対象 大きな話題 a b c 行動

すでに、開発の経緯で述べたように、このようなチェックリストの場合、「日程」「場所」「理由」「手段」の4つの要素が一つの Can-do Statement の中に含まれている。これらの要素を、学習者に「日程は話せた」、「行き方は話せなかった」のように示すことは、フィードバックとしても、あるいはテスト前に準備するための項目としても重要であるが、一つの Can-do Statement の中に含まれていると、評価として示すのが難しいだけでなく、それを学習者が読み解くのも難しくなる。このため、3課では、以下のような Can-do Statements とした。

長い休みの計画について、

- a) 日程やどこに行くかを、話すことができます。(=日程, 場所)
- b) どうやって行くかなど、話すことができます。(=手段)

c) どうしてその計画にしたのか、理由を話すことができます。(=理由)

「長い休みの計画」という話題を構成する要素を分解し、一つのCan-do Statementでチェックするものをできるだけ少なくしたのである。これにより、評価者の負担が減るだけでなく、学習者にとっても一つ一つ確認しながら何を話せばいいのか準備をすることが可能になる。また、振り返りも一つ一つ確認することが容易になると言える。以前に作成していたループリック形式の評価シートでは、詳細な文言の中から、該当の場所を見つけていく必要があったが、チェックリスト型にすることによって情報量は減っていないものの、確認作業は容易になり、学習者及び教師の負担は減ったと考えられる。

3.2.3.3 「言語」

すでに述べたように、「言語」は、話し言葉の質を見るものである。CEFRでは、話し言葉の

表3 チェックリスト型評価シートの「内容」部分

ないよう 内容	<p>なが やす けいかく 長い休みの計画について、クラスメートと話すことができます。</p> <p><input type="checkbox"/> なが やす けいかく 長い休みの計画について、クラスメートに質問することができます。</p> <p><input type="checkbox"/> なが やす けいかく についで い 長い休みの計画について、日程やどこに行くか、どうやって行くかなど、話すことができます。</p> <p><input type="checkbox"/> なが やす けいかく けいかく り ゆう はな 長い休みの計画について、どうしてその計画にしたのか、理由を話すことができます。</p> <p><input type="checkbox"/> あいて はなし かんそう い しつもん てきせつ はなし つづ 相手の話について、感想を言ったり質問したりして、適切に話を続けることができます。</p> <p><input type="checkbox"/> かいわ てきせつ お クラスメートとの会話を適切に終わることができます。</p> <p><input type="checkbox"/> あいて かんけいせい おう はな かた はな 相手との関係性に応じた話し方で話すことができます。</p>	<p>outstanding</p> <p>excellent</p> <p>good</p> <p>fair</p> <p>poor</p>
	<p>じぶん きょうみ けいかく しょうらい ゆめ はっぴょう 自分の興味があることや計画など、将来の夢について発表することができます。</p> <p><input type="checkbox"/> じぶん しょうらい ゆめ なに はな 自分の将来の夢が何か、話すことができます。</p> <p><input type="checkbox"/> じぶん しょうらい ゆめ いま はな 自分の将来の夢のために、今していることを話すことができます。</p> <p><input type="checkbox"/> じぶん しょうらい ゆめ はな 自分の将来の夢のために、これからすることやしなければならないことを話すことができます。</p> <p><input type="checkbox"/> じぶん しょうらい ゆめ り ゆう はな 自分の将来の夢について、どうしてそれが夢なのか、理由を話すことができます。</p> <p><input type="checkbox"/> こみせい はっぴょう まとまりのある構成で発表することができます。</p> <p><input type="checkbox"/> はっぴょう てき ていねい ひょうげん はな かた はな 発表に適した丁寧な表現や話し方で話すことができます。</p> <p><input type="checkbox"/> はっぴょう てき たいど はっぴょう 発表に適した態度で発表することができます。</p> <p><input type="checkbox"/> あいて いけん き てきせつ しつもん 相手の意見を聞いて、適切に質問したり、コメントしたりすることができます。</p>	<p>outstanding</p> <p>excellent</p> <p>good</p> <p>fair</p> <p>poor</p>

質的側面を記述するために14の観点が表示されている。しかし、全ての項目を評価することはできず、学習者の必要性に合致したもの、実施される評価課題が求めるもの、教授環境の形に応じて、評価者が取り扱える数を採用する必要があると指摘されている(Trim, North, Coste 2014, p.215)。そこで、新たに開発する評価シートでは、コミュニケーション言語能力の質的側面を評価する「言語」について、「文法・語彙・表現」「流暢性・ディスコース」「話し方」の3つのカテゴリーを設定した。以下に3つのカテゴリーについて述べていく。

3.2.3.3.1 文法・語彙・表現

「文法・語彙・表現」は、「言語構造的な能力」に関し、テストでのパフォーマンスを測るカテゴリーである。ここでは文法・語彙、表現についての正確さ、及び、適切さを見る。

この正確さとは、「正確な文法・語彙・表現の使用」を見るものであり、以下のような観点から評価を行う。

- ・文法、語彙や表現についての形式面の正確さ（ex 助詞の欠落等の文法規則の誤り）
- ・産出されたものの意味が聞き手に理解できるか。

適切さは、「適切な形式、レジスター（表現・語彙、文末スタイル）」の使用を見るものであり、具体的には、以下のような観点から評価を行う。

- ・効果的な文型の使用
- ・「話し言葉」と「書き言葉」の区別
- ・「普通体」と「丁寧体」の区別
- ・稚拙な表現の使用の有無

3.2.3.3.2 流暢性・ディスコース

このカテゴリーは、話し言葉の滑らかさを測る「流暢さ」と、ディスコース（談話）を組み立てたり、コントロールしたりする能力を測る「ディスコース」からなる。

流暢性として、フィラー、言い淀み、相手とのやり取りのスムーズさ、あいづち、聞き返しを項目として取り上げた。以下の観点から評価する。

- ・無音の沈黙がなく、話が継続していることを示すようにフィラーを使用できているかどうか。
- ・いいよどみがない、すなわち話を停止することなく続けているかどうか。
- ・相手の反応を見ながら、話し方やスピードを変えて話せているかどうか。

また、ディスコースについては、単語単位、文単位、談話レベルの談話構成能力、談話自体の一貫性を取り上げた。やり取りを伴うタスクでは、相互行為の中で保たれる談話の一貫性に関する発言権や相手の反応を見ながら発言ができるかといった項目を取り上げた。以下の観点から評価する

- ・単語単位でぶつぶつ切れていないかどうか。
- ・簡単な接続表現を用いて語句と語句の間につながりを作って話せているかどうか。
- ・ポイントを挙げて、まとまりのある談話を構成しているかどうか。

- ・談話が一貫性を保っているかどうか。
- ・会話の話題や流れの一貫性を保った会話ができているかどうか。
- ・その場の状況や聞き手に応じて、内容や話し方を調節し、談話の一貫性を保持しているかどうか。
- ・ふさわしいときに発言権を得たり、保持したりすることによってスムーズな談話を実現しているかどうか。

3.2.3.3 話し方

話し方は、タスク達成に伴う「関心・意欲・態度」といった情意面の要素に関し、テストでのパフォーマンスを測る。以下の観点から評価する。

- ・話す意欲を持ち、自信を持った聞きやすい話し方であるかどうか。
- ・最後まであきらめずに話せたかどうか。
- ・聞き手を見ながら話せたかどうか。

3.2.4 採点基準

チェックリスト型の評価シートの採用によって大きく変更したのが採点方法である。ルーブリック評価シートでは、個々の教師がタスクのパフォーマンスについて「どの程度できているか」を基に、評価を行っていた。しかし、チェックリスト型の場合、できたCan-do Statementにチェックを入れるという形式をとっているため、チェックが入ったCan-do Statementsの割合を評価に反映させることにした。また、4段階から、outstanding, excellent, good, fair, poorの5段階評価へ変更した。4段階では、少し気になるが、不合格にするほどでもないといったレベルのパフォーマンスであっても、good (75%) の評価となり、不合格の基準の点数となるpoorは実質的につけにくいという傾向が見られたためである。目安としてチェックが入った割合がoutstandingは100%, excellentは70-90%, goodは40-60%, fairは10-30%, poorは0%の5段階評価とした。例えば、前述のテスト1の「内容」では、チェック項目が8つあるため、outstandingはチェックが8つに入った場合、excellentは6-7つ、goodは4-5つ、fairは1-3つ、poorは0の場合とした。

3.2.5 フィードバック

オーラルテスト終了後に、すぐにパフォーマンスを録画した映像を見返すなどすることで、学習者は自分のパフォーマンスを振り返りながら、自己評価や相互評価を行う。教師や学習者同士で、評価シートで挙げられた項目等について、気が付いた点やうまくできた点、効果的な表現などについて話し合う。また、先にも述べたように教師から学習者へのフィードバックとしても、このシートを活用している。Can-do Statementsの中で、うまくできている部分には、○をつけ、できていなかった部分には下線を引き、それ以外に学生に伝えるべき点は、内容の右側にコメント欄を設け、そこに書き込んで返却している。このコメント欄には、誤りの事例や修正案、タス

表4 チェックリスト型評価シートの「言語」部分

文法 Grammar 語彙 Vocabulary 表現 Expression	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ぶんぽう ひょうげん</small> つか 習った文法や表現を使うことができます。	outstanding excellent good fair poor
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ぶんぽう ひょうげん</small> つか 習った文法や表現について、意味の分からなくなるような誤りなく、使うことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ぶんぽう ひょうげん</small> つか 習った文法や表現をほとんどの場合適切に使うことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ぶんぽう ひょうげん</small> つか 習った文法や表現をほとんどの場合適切に使うことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ご</small> つか 習った語を使うことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ご</small> つか 習った語をたくさん使うことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	なら <small>ご</small> つか 習った語を、ほとんどの場合、適切な使い方で使うことができます。	
流暢性 Fluency ディスコース Discourse	<input type="checkbox"/> <input type="checkbox"/>	し <small>ぜん</small> まじ <small>はな</small> 自然なフィラーやポーズを交えて話すことができます。	outstanding excellent good fair poor
	<input type="checkbox"/> <input type="checkbox"/>	い <small>はな</small> 言いよどみがほとんどなく話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	たん <small>ご</small> たん <small>い</small> き 単語単位でぶつぶつ切れずに、話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	かん <small>たん</small> せ <small>つ</small> ぞ <small>く</small> ひょう <small>げん</small> も <small>ち</small> <small>ご</small> く <small>ご</small> <small>あい</small> だ <small>つ</small> 簡単な接続表現を用いて語句と語句の間につながりを作ることができます。	
	<input type="checkbox"/> <input type="checkbox"/>	あ <small>はな</small> ポイントを挙げて、まとまりのある話をするすることができます。	
	<input type="checkbox"/> <input type="checkbox"/>	あ <small>い</small> て 相手とのやりとりがスムーズにできます。	
	<input type="checkbox"/> <input type="checkbox"/>	ひ <small>つ</small> ぶ <small>やう</small> <small>ば</small> <small>あい</small> <small>し</small> <small>ぜん</small> <small>つ</small> か (必要な場合) 自然にあいづちを打って話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	ひ <small>つ</small> ぶ <small>やう</small> <small>ば</small> <small>あい</small> <small>ほ</small> <small>つ</small> げ <small>ん</small> け <small>ん</small> え <small>ほ</small> <small>じ</small> (必要な場合) 発言権を得たり、保持したりすることができます。	
	<input type="checkbox"/> <input type="checkbox"/>	い <small>っ</small> かん <small>せい</small> た <small>ち</small> <small>かい</small> <small>わ</small> <small>だ</small> <small>ん</small> <small>わ</small> <small>い</small> <small>じ</small> 一貫性を保って会話・談話を維持できます。	
<input type="checkbox"/> <input type="checkbox"/>	あ <small>い</small> て <small>は</small> <small>つ</small> わ <small>き</small> <small>と</small> 相手の発話が聞き取れなかったとき、聞き返すことができます。		
<input type="checkbox"/> <input type="checkbox"/>	あ <small>い</small> て <small>は</small> <small>ん</small> の <small>う</small> <small>み</small> <small>は</small> <small>つ</small> げ <small>ん</small> 相手の反応を見ながら発言することができます		
話し方 Attitude	<input type="checkbox"/> <input type="checkbox"/>	き <small>はな</small> <small>か</small> <small>た</small> <small>はな</small> 聞きやすい話し方で話すことができます。	outstanding excellent good fair poor
	<input type="checkbox"/> <input type="checkbox"/>	じ <small>し</small> <small>ん</small> <small>はな</small> 自信をもって話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	あ <small>き</small> <small>ら</small> <small>め</small> <small>ず</small> <small>さい</small> <small>ご</small> <small>はな</small> あきらめずに最後まで話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	き <small>て</small> <small>ち</small> ょう <small>し</small> ゅう <small>み</small> <small>はな</small> 聞き手/徴収を見て話すことができます。	
	<input type="checkbox"/> <input type="checkbox"/>	つ <small>た</small> <small>き</small> <small>も</small> <small>い</small> <small>よ</small> く <small>かん</small> <small>はな</small> <small>か</small> <small>た</small> 伝える気持ちや意欲を感じるような話し方ができます。	

クのうまくいった点等を書き込むことで、振り返りにつなげている。

3.2.6 まとめ

以上に述べてきたように、本稿ではループリック形式の評価シートで生じた学習者の読む負担

の軽減、タスクで要求されることの可視化、評価判断の主観性によって生じる教員間の認識のずれの軽減を試みるためのチェックリスト型評価シートについて述べてきた。

チェックリスト型評価シートを用いることで、学習者は短いCan-do Statementsの文を見て、テストの準備を行ったり、テスト時の自身のパフォーマンスを振り返ったりすることができるようになった。自己評価と教師評価のすり合わせなどの使用にもつながった。また、タスクの評価項目を分解することで、タスクで要求される言語活動を再分析できた。また、「上手に」等、程度を表す文言が何を指すのかについての教員判断をできるだけ排し、タスクの成否を照合する「内容」と、言語コミュニケーション能力を基に、「文法・語彙・表現」、「流暢性・ディスコース」、「話し方」を設けた「言語」のみで評価を行えるようになったことから、主観による揺れを少なくすることができたのではないだろうか。

4. 評価シートに対する反応

4.1 学習者の反応

2017年度の冬学期に、新たな形式の評価シートを使用した学習者11名に、自由記述式のアンケートへの記入を依頼した。依頼に際しては、口頭で、参加は自由であり、回答によって不利益を被ることがないこと、辞退をした場合にはいつでも辞退が可能であること、得られた回答は回答者がわからないように匿名化し、公表する際には個人が特定されないように配慮することを説明している。回答を依頼した質問は、下記に示す(1)から(4)の4問であり、依頼した11名の学生全員から回答があった。しかし、アンケート用紙に不備があったため、11名中2名の学習者からは、(3)の回答がなかった。以下に、質問順に結果を報告する。

(1) オーラルテストの準備のために、評価シートを使いましたか。どのように使いましたか。

11名中10名の学習者が使用したと答えた。特にタスクで要求されることができるかどうかを自身でチェックしながら、タスクの予行演習などに使用していた、目標に合わせて準備ができた、などの回答を得られた。また、準備に評価シートを使用せず、学生配布用タスクの指示を用いて準備していたと回答した学習者も1名いた。

(2) 評価シートをもらった後、自分のできた部分とできなかった部分がわかりましたか。自分の評価と比べて、違いがありましたか。

全ての学習者がわかったと答えた。また、自己評価との違いについては、自信がない、ナーバスになったなどの理由で、教師の評価よりも低く自己評価していたと答える学習者が3名見られた。

(3) この評価シートを使ってみて、どんな点がいいと思いましたか。感想を聞かせてください。

(4) この評価シートで改良する点と悪い点は何ですか。

(3)と(4)の質問については、重複する回答も多かったため、合わせて報告する。これらの評価シートについては、11名中10名が「役に立つ」と答えた。テストタスクで何を求めているかがわかった、評価シートを用いたフィードバックで自分の誤りを理解できた、等の肯定的な意

見が見られた。一方で、評価シートの記述が明確ではない、英語と日本語の文の意味がずれているなどの改善を促す声もあった。

4.2 教員の反応

新たな形式でテストの実施を担当した教員7名に対しても、自由記述式のアンケート記入を依頼した。一部の教員には、口頭で回答を得た。教員に対しても、参加は自由であり、回答によって不利益を被ることがないこと、得られた回答を公表する際には個人が特定されないように配慮することを説明している。教員に対しては4つの質問¹⁰⁾への回答を求めたが、内容は大きく分けて、よかった点と改良すべき点について、評価シートの使用方法についてであったため、この3点に分けて報告する。

(1) よかった点

7名中4名が判断のしやすさを挙げた。チェックリスト型になることで、項目を一つ一つ見ていけばよいと、判断への迷いが少なくなったことや、評価基準で要求される記述の理解が容易になったなどの理由が挙げられた。また、学習者に対してタスクの説明が容易になった(1名)、フィードバックの際に何ができていたかを示しやすい、学習者と評価のポイントを共有することができていることがよい(3名)との回答が見られた。

(2) 改善すべき点

改善すべき点は、特に内容について、どの程度できたかを評価することが難しいという回答が7名中4名から得られた。具体的行動についてはチェックが容易であるが、質の高い議論やスピーチの面白さといったような抽象的な項目については、できたかできないかのチェックだけでは評価しにくく、特に初級後半のタスクでは、程度を見られるような形式が望ましいとの回答も見られた。また、項目数を絞るべきであるとの回答が、2名の教師から得られた。これは、多すぎると学習者の負担が増える、理解する意欲をなくす、煩雑であるなどの理由である。

また、言語の部分の評価項目についても、項目数が多い(1名)という意見があった。言語の項目では、「正確に」や「適切に」といったような抽象的な文言も見られ、これらの基準のわかりにくさや評価のしにくさについて2名から言及があった。また、言語は全てのタスクで同じものを使用しているため、タスクによっては当てはまらないものもある(1名)や、やり取りには必要であるが、スピーチには不要な項目などは、タスクに応じて削除すべきである(2名)との声もあった。

4.3 まとめ

使用の反応を見ると、改善を目指した学習者の読む負担の軽減、教員間の評価判断の揺れについては、一定の改善が見られたという実感があるようである。また、タスクで要求されることの可視化については、以前の評価シートでもある程度実現していたが、要素に分けることで学習

者、教員にとってポイントを把握しやすくなったことから、目標や評価基準がより共有されやすくなったようである。ただし、その一方で、項目数の多さや評価判断が揺れる可能性がある記述についても言及があり、改善が必要である。

5. まとめと今後の課題

本稿では、Can-do Statementsによるチェックリスト型評価シートの試案とその開発経緯について述べてきた。このチェックリスト型の評価シートを用いることで、学習者はテストの準備を行ったり、テスト時の自身のパフォーマンスを振り返ったりすることができる。振り返り際には短いCan-do Statementsで書かれた評価シートを教師も学習者もともに使用するため、教師評価と自己評価の比較も容易になった。タスクの評価項目を分解したことで、タスクで要求される言語活動を再分析できた。また、「上手に」といった条件についての主観的判断をすることが少なくなり、タスクの成否のみで評価を行っていることから、教師の評価時の迷いを少なくすることができたとも言える。このように、チェックリスト型の評価シートは、目標の可視化、学習者との評価基準の共有、パフォーマンスの振り返りの促進などに有用であった。しかし、教員や学習者から挙げたように、項目数の多さや不明確な記述など、まだまだ改善点は多い。タスク達成の程度をどのように評価するかという点も残された大きな課題であると言える。

チェックリスト型評価シートを用いることによって、様々な課題を解決することができたものの、決してこれが最善ではなく、継続して改善を図っていく必要がある。本稿における評価シートはあくまでも開発途中にある試案であり、タスクが変わったり、評価者が変わったり、あるいは、学習者が変わったりする中で、新たな評価シートの形が必要となることもあるだろう。今後は本評価シートの試みにおける改善点とともに課題の存在を十分に考慮し、「目標の可視化」を実現する評価体制の構築が課題であることを述べ、本稿を締めくくりたい。

謝辞

本稿は実際にコースで行われた教育実践を基にした論文であり、実践の企画・立案・実施にあたり、様々な方の協力を得ている。特に初期の評価シート作成を含む評価体制の構築は、國澤里美講師（当時）との協働による。本稿の執筆の契機となった日本語教育学会中部研究集会交流広場での発表では、梶原彩子講師にも協力をいただいた。また、実際のテストタスクや評価シート作成は、当該機関の非常勤講師諸氏が担当している。ここに記し、謝辞とさせていただきます。

注

- 1) 中級以降のコースについても、見直しを始めている。
- 2) これは、國澤・近藤（2016）で、「話すテスト」として記述したテストと同一のテストを指している。

- 3) 国際交流基金によって作成されている。JFスタンダードは、CEFRを日本語教育へ応用する試みであり、CEFRのCan-do Statementsに加え、現場で用いられるタスクに対応した独自のCan-do Statementsを加えた構成となっている。
- 4) 『できる日本語初級』と『できる日本語中級』を初級クラスの教材として採用した。
- 5) ただし、テストタスク及び評価基準は事前に予告されるため、テストタスクに対する準備は可能である。
- 6) この評価は実際に使用した際の文言とは厳密には異なっている。3-2-3でのチェックリスト型評価シートの記述との対比をしてわかりやすくするために若干の修正を行ったが、基本的な文言や記述、形式は変更していない。
- 7) 意識訓練のためには、印象（チェックリストの中でできると答えた項目）と現実（チェックリストの中でできると答えた項目を実際に聞いてみて、それが理解できるかどうか確かめること）を比べることという例が挙げられている。（Trim, North, Coste 2014, p.208）
- 8) 國澤・近藤（2016）では、ルーブリック形式の評価シートのコミュニケーション能力をCanal and Swain（1987）、Canal（1983）に求めたが、チェックリスト型の評価シート開発にあたり、CEFRにおける言語コミュニケーション能力の構成概念やレベル記述を参考に全面的に見直した。CEFRでは口頭能力に関連のある14の質的カテゴリーとして、「発話権取りの方略」、「協調の方略」、「説明を求めること」、「流暢さ」、「柔軟性」、「一貫性」、「話題の展開」、「正確さ」、「社会言語能力」、「総合的な使用言語範囲」、「使用語彙領域」、「文法的正確さ」、「語彙の使いこなし」、「音韻面の把握」が挙げられている（Trim, North, Coste 2014, p.215）。
- 9) CEFRには、「上のレベルにならないと出てこない能力記述文がある」ことも明記されている。全てのレベルに同じ項目が使用できるかどうかについて考慮する必要があるが、テストは初級・中級のみを対象としていることもあり、全て同じチェックリストを使用することにした。
- 10) 質問は、以下の4問である。
 - (1) この評価シートを使ってみて、（前のルーブリックの形式と比べて）どんな点がいいと思えましたか。
 - (2) あるいは、使いにくいと感じた点はありますか。／以前のルーブリック形式のほうが優れていると思う点はありますか。
 - (3) 以下の点についてのご意見をお聞かせください。（改良すべき点、その理由など）
 - ・この評価シートで改良すると思う点は何か。理由も教えてください。
 - ・項目の内容／項目数／成績のつけ方
 - (4) フィードバックの際、この評価シートをどのように使用していますか。

参考文献

- Canale, M. (1983) "From Communicative Competence to Communicative Language Pedagogy" In J. C. Richards and R. W. Schmidt eds. *Language and Communication*, London: Longman, pp. 2-27
- Canale, M. and M. Swain. (1980) "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing," *Applied Linguistics*. 1(1), pp. 3-47
- Trim, J., North, B., Coste, D. (2014) 吉島茂・大橋理枝（他）訳『外国語教育 2 外国語の学習、教授、評価のためのヨーロッパ共通参照枠 第三版』朝日出版社；追補版
- 国際交流基金（2017）『JF日本語教育スタンダード【新版】利用者ガイドブック』（https://jfstandard.jp/pdf/web_toc.pdf（2018.11.21最終アクセス）

名古屋学院大学論集

國澤里美・近藤行人（2016）「交換留学生を主体とした初級日本語クラスのコースデザイン」『名古屋学院大学論集 言語・文化篇』27(2), 123-137

できる日本語教材開発プロジェクト（2012）『できる日本語』アルク

ダネル・スティーブンス, アントニア・レビ（佐藤浩章監訳）（2014）『大学教員のためのルーブリック評価入門』玉川大学出版部

田中耕治（2010）「ルーブリック」『よくわかる教育評価』ミネルヴァ書房；第2版