

# プロフィシエンシーと日本語クラス口頭能力試験アセスメント

池田隆介（北九州市立大学）

池田富見子（久留米大学）

古賀さと子（福岡国際学院）

山下直子（鹿児島キャリアデザイン専門学校）

キーワード：アセスメント、口頭試験、ルーブリック、会話、日本語クラス

## 1. はじめに

本研究は、日本語クラスにおいて実施されている口頭能力試験（会話試験）を「プロフィシエンシー」重視の観点から改善していくためにはどのような手法が有効かを探ることを大きな狙いとする。TOEFL、TOEIC、日本語能力試験（JNPT）など不特定多数の受験者を想定した大規模試験では、困難度や識別力に問題のある項目を排除し信頼性を確保するためのアセスメントが試験実施団体により持続的に実施されている。一方、教育機関内で実施される試験についてはアセスメントが必ず実施されているとはいいがたいのが現状といえよう。たとえば、日本語教育関係の試験においては、伊東（2005）、吉川（2015）のようにプレイスメント・テストの妥当性検証を実施した報告は多数報告されている。しかし、日本語クラス内で実施される、小テスト、クイズ、期末試験などの小規模試験については、教員の経験により改善が行われるケースが大部分であり、客観的な指標を参照したアセスメントをもとに試験改善が試みられる例は少ないのではないかと思われる。そこで、本研究では、日本語クラス内の「会話試験」を対象に、ACTFL-OPI を利用したアセスメントを実施し、そのアセスメントがどの程度効果的であるかを検証する。

## 2. 調査

本研究では、日本語クラス内の代表的パフォーマンス試験である会話試験を題材に、ACTFL-OPI を使ったプロフィシエンシー重視のアセスメントを試みる。評定の妥当性を検証できるのか、試験改善のための手がかりが得られるのか、または、アセスメントにかかる手間がどの程度のものかなども含め、実行可能なアセスメントとなりうるか否かを二つの調査により明らかにしていく。一つ目は、中上級の学習者を対象にしたプレゼンテーション、二つ目は、初中級の学習者を対象にした複数種類の会話試験（ロールプレイ、ミニスピーチ）を取り上げる。

### 2. 1 中上級学習者を対象にしたプレゼンテーションのアセスメント<sup>1)</sup>

#### 2. 1. 1 対象者・概要

日本の大学の理系学部にも所属する学部 1 年次の留学生 11 名を対象に、日本語授業内でプレゼンテーション課題を実施した。同時期に実施した OPI の判定と、プレゼン課題の評価（総合得点、及び、ルーブリック各項目の得点）の相関性を検証する。

#### 2. 1. 2 結果

OPI とプレゼンの総合得点との関係を表 1、ルーブリックの項目別の得点との関係を表 2 に示す。

総合得点と OPI とは、 $r=0.88$  という高い相関関係にあった。つまり、ほぼ実力通りの結果が出ていると言える。しかし、試験が適切であったかどうかを見極めるには、プレゼンテーション課題を評定し

た際に使ったルーブリックの各項目の検討も必要となる。今回の評定に用いたルーブリックの項目は全部で 14 ある。その中でも、「フォーマルな話し方」「正確さ」「発表の態度」「発音」という項目が OPI と高い相関を示していた。一方、「司会・進行」「時間」「パワーポイント」という項目は OPI との相関がないということも明らかになった。

表 1 OPI とプレゼン試験

学習者	OPI	総合点
あ	上級-上	91.50
い	上級-中	86.00
う	上級-中	85.50
え	上級-中	82.50
お	上級-下	90.00
か	上級-下	84.50
き	中級-上	78.00
く	中級-上	77.00
け	中級-上	70.00
こ	中級-中	72.00
さ	中級-下	67.50
OPI との相関		r=0.88

表 2 OPI とプレゼン試験各項目 (抜粋)

学習者	フォーマルな話し方	正確さ	発表の態度	発音	...	司会・進行	時間	パワーポイント
あ	1.00	0.90	1.00	1.00		0.90	1.00	0.80
い	1.00	0.90	0.80	1.00		0.90	0.75	0.80
う	1.00	0.90	0.90	1.00		0.80	0.75	0.70
え	0.90	0.80	0.90	0.90		0.90	0.50	0.80
お	1.00	0.90	1.00	1.00		0.90	1.00	1.00
か	0.90	0.80	1.00	1.00		0.80	0.75	0.90
き	0.80	0.70	0.70	0.80		0.95	0.75	0.80
く	0.80	0.60	0.60	0.80		0.80	0.50	0.90
け	0.80	0.60	0.80	0.80		0.70	0.25	0.70
こ	0.60	0.60	0.60	0.60		0.80	0.50	0.80
さ	0.50	0.60	0.60	0.80		0.90	1.00	0.80
OPI との相関	r=0.92	r=0.85	r=0.75	r=0.75		r=0.28	r=0.18	r=-0.05

### 2. 1. 3 考察

ここから試験の特徴を考察すると、学習者のプロフィシエンシーがそのまま成績に反映されているという点を評価できる。まず全体と OPI との相関が高い。また、項目別に見ると、言語的な正確さを問う項目と OPI との相関が高い。このことから、学習者の日本語の実力通りの結果がでており、試験判定は十分に説得力を持つものと言える。「司会・進行」「パワーポイント」という項目は OPI との相関が低かったがその点は大きな問題にはならない。これらは事前に準備することが可能な項目であり、OPI との相関の低さを問題視する必要はなく、逆に、OPI との相関が高ければ、事前に準備できるだけの環境が整えられていないということになり、授業計画に問題があると解釈されることになる。

一方で、改善すべき点も明らかになった。それが「発表の態度」「時間」という項目だ。「発表の態度」は、発表中の身振りや視線の向きなど非言語的な振る舞いを評定する項目である。OPI との相関が高いことそのものは問題ではなく、非言語の面でも適切な振る舞いをするには一定のプロフィシエンシーが必要であるということの示唆としても受け止めることができる。だが、この試験においては、採点者が「発表の態度」を適切に観察できたかどうかという懸念が浮かび上がる。「発表の態度」を独立した項目として評定していれば問題ないが、言語的な正確さに引きずられて非言語的な要素を好意的に変更して判定しているのであれば、別項目を立てている意味がなくなる。また、「時間」は定められた時間内に過不足ないように発表できるかどうかを評定する項目だが、高プロフィシエンシーの学習者に制限時間を超過する者が多く、OPI との相関は低かった。つまり、そもそもの設定時間が短かったことが原因と考えられる。適切に実力を発揮してもらうためには、制限時間の見直しが必要と考えられる。

## 2. 2 初中級の学習者を対象にした会話試験

### 2. 2. 1 対象者・概要

日本の大学に併設された留学生別科に在籍する初中級クラスの 13 名に対し、学期末の会話試験を実施した。その期末テストの結果および同時期に評定された OPI 判定結果について、相関性を検証した。対象の 13 名は滞日 4 か月から 16 か月の学習者である。当該試験は、ロールプレイと「ニュース」をテーマとするミニスピーチという 2 種の口頭試験である。素点はロールプレイ 36 点、ミニスピーチ 19 点であり、そのまま合計した 55 点満点で成績評定を行っている。

### 2. 2. 2 結果

ロールプレイ、ミニスピーチのそれぞれに採点のためのルーブリックが準備されているが、今回は、2 種の試験の点数、および、その合計点と OPI との関係を分析する。それぞれの口頭試験、及び、合計点と OPI との相関を表 3 に示す。ロールプレイと OPI には高い相関関係がみられたが、ミニスピーチはそれほど相関は認められなかった。しかし、合計点は  $r=0.72$  と高い相関を確保している。

### 2. 2. 3 考察

合計点と OPI の相関は高く、プロフィシェンシーの現状を反映した会話試験を実行できていると考えられる。実力を反映した説得力のある成績になっており、学習者にとっても納得できる結果が出ていると推察される。

2 種の試験それぞれよりも合計点と OPI との相関が高いことから、タイプの異なる会話試験を実施することで、学習者の実力を発揮できるように配慮されていると考えられる。初中級レベルのクラスにおいても、機能・話題・場面など状況に応じたプロフィシェンシーをどの程度有しているかは、一つの会話試験だけでは測定しきれないと考えられるので、複数の会話試験を組み合わせるといふ工夫が奏功したと言える。また、複数の試験を組み合わせる際には成績評定の際の得点比を慎重に検討しなければならない。この試験においては 36:19 という比率も適切であった。

ロールプレイは授業で取り上げた内容を再現しているので到達度を測定することが狙いの課題であった。しかし、結果的に OPI との高い相関を得ることができた。到達度の測定ではあるが、受験者の点数が高止まりすることなく、出来栄を測定し分けることができるような採点基準(ルーブリック)が設けられていたことがその要因であると考えられる。実力を識別できる試験になっていると評価することができる。一方で、課題を見出すと知れば、62.04% という平均点だ。到達度の測定として

表3 OPIと2種類の会話試験

学習者	OPI	RP	(%)	ミニスピーチ	(%)	合計	(%)
		36		19		55	
A	中級一中	27	75.00	18	94.74	45	81.82
B	中級一下	29	80.56	15	78.95	44	80.00
C	中級一下	27	75.00	17	89.47	44	80.00
D	中級一下	30	83.33	10	52.63	40	72.73
E	中級一下	26	72.22	12	63.16	38	69.09
F	初級一上	28	77.78	15	78.95	43	78.18
G	初級一上	29	80.56	13	68.42	42	76.36
H	初級一上	24	66.67	17	89.47	41	74.55
I	初級一上	23	63.89	14	73.68	37	67.27
J	初級一上	24	66.67	7	36.84	31	56.36
K	初級一中	24	66.67	13	68.42	37	67.27
L	初級一中	16	44.44	12	63.16	28	50.91
M	初級一中	14	38.89	10	52.63	24	43.64
平均点		24.69	68.59	13.71	72.18	38.41	140.77
OPIとの相関		$r=0.70$		$r=0.43$		$r=0.72$	

は難易度が高すぎたかという懸念が残る。

もう一つの試験であるミニスピーチはそれほど OPI との高い相関は得られなかったが、「ニュース」をテーマとしたこの試験を行う意義はあったと考えられる。もともと学習者の実力の如何がスピーチの完成度に大きくかかわる課題なので、ロールプレイよりも OPI との相関が高くなると予測されたが、実際はそうならなかった。その要因は、初中級のレベルにおいてこの課題は、プロフィシエンシーよりも事前の準備をどの程度行ったか否かが評定を左右したという点にある。D は「中級-下」の実力を有し、ロールプレイでも最高点を取っているが、準備不足のためミニスピーチでは下から 2 番目にとどまっている。一方、「上級-上」の H はミニスピーチ課題の準備をして高得点をとることができた。しかし、この課題を取り入れたのは、このクラスの学習者が今後進学することになれば、身近なことだけでなく社会的な話題も扱っていなければならぬ状況が増える。そこで、初中級のレベルでは難しい「ニュース」という話題を敢えて取り入れることによって、今後の学習の先鞭をつけるという意義があったと言える。もちろん、そのために成績が不適切なものとなってしまうのは避けられないが、全体の会話試験のバランスを崩さない割合で組み込んでいる。説得力のある成績にしつつ、今後の学習のモチベーションも与えていくという授業設計にのっとった試験となっていると言える。ただし、学習者の準備の度合いの差をどのように縮めていくかという点への配慮は必要となるであろう。

### 3. まとめ

OPI を利用した日本語クラス口頭能力試験のアセスメントは十分な効果を期待できる。2 章では、レベルも属性もことなるクラスの口頭能力試験のアセスメントを試みたが、いずれも試験の特徴と課題がある程度確認することができた。OPI との相関の高さを評価すべき項目、OPI との相関が低くても含めておいたほうがよい項目など、さまざまなタイプの項目があるのでそれをアセスメントで精査することが重要である。OPI の情報があるのであれば、規模の小さなクラスにおけるパフォーマンス試験でもアセスメントをして改善のための手がかりを得ることはできると考えられる。

### 4. 今後の課題

今後の課題としては、さらに多様な日本語クラスにおける口頭試験においても OPI を使ったアセスメントが有効か否かを調査していく必要がある。また、少人数の学習者を対象とするルーブリックの場合、学習者の個別特性が極端に反映されてしまうケースに注意しなければならない。それから、OPI は得点段階を数値化できる基準ではないので、超級 10 点～初級-下 1 点として分析を行った単純な今回の測定の仕方が適切であったかどうかは十分に検討しなければならない。

#### 注

1) 3. 1 の調査・結果は池田 (2019) をもとに再構成している。

#### 参考文献

池田隆介 (2018) 「留学生を対象とした日本語口頭能力試験評価ルーブリックの妥当性検証—外部基準を利用したパフォーマンス評価法改善の試み—」, 2019 年 3 月 24 日, 日本語プロフィシエンシー研究学会 2018 年度第 3 回研究例会, 於: かんぼの宿柳川.

伊東祐郎 (2005) 「プレースメント・テストの妥当性確認の試み」, 『東京外国語大学留学生日本語教育センター論集』, 第 31 号, 161-174, 東京外国語大学.

吉川達 (2015) 「2 つの日本語プレースメント・テストの等質性の検証」, 『佐賀大学全学教育機構紀要』, 第 3 号, 111-124, 佐賀大学.