目　次

第1章 言語テストを捉える

李 在鎬

1 言語教育と言語評価

応用言語学の分野で「評価」は2つの意味で用いられる。1つ目は,「アセスメント（assessment)」の意味として，2つ目は,「エバリュエーション（evaluation)」の意味として用いられる。前者は，学習の成果を確認する目的で行われ，典型例としては，クラス活動の一環として行われる小テストや学期末テスト,さらには本書で紹介する「日本語能力試験（Japanese Language Proficiency Test；JLPT)」《➡第2章参照》や,「BJTビジネス日本語能力テスト（Business Japanese Proficiency Test；BJT)」《➡第3章参照》などが挙げられる。一方，後者は，教育現場を取り巻くさまざまな実態調査とプランニングを目的に行われ，典型例としては，学期末に行う授業評価や卒業生へのアンケート調査などが挙げられる。一般的な言語テストとしての「評価」は,「アセスメント」としての活動であり,「教師が教室内で用いる評価など，学習者のパフォーマンスや達成度を評価する様々な方法である」(Gipps 1999）と定義できる。こうした「アセスメント」としての評価は教育に関わるあらゆる活動において日常的に行われている。一方,「エバリュエーション」は,「アセスメント」よりも広域の概念であり，教育の質や価値の維持，もしくは質的向上の必要な評価のすべてがその対象になる（近藤 2012)。なお，文献によっては「アセスメント」のことを「評定」,「エバリュエーション」のことを「評価」とし，両者を区別する場合もあるが，本書では,「アセスメント」の意味で「評価」という用語を使う。

さて，アセスメントとしての評価は，規模の違いはあるにせよ，学習者

第2章 日本語能力試験

大隅 敦子・谷内 美智子

1 背景と目的

日本語能力試験（Japanese Language Proficiency Test；JLPT）は，日本語を母語としない者の日本語能力を測定し，認定する試験である。日本語能力試験は 5 年にわたる試行試験を経て 1984 年に開始され，2013 年で満 30 年を迎えた。試験開始当初の受験者数は 7,019 人だったが，2012 年には 64 カ国で実施され，受験者数も 572,169 人となった。

日本語能力試験が開始された 1980 年代は，国内外において日本語学習者が急激に増加した時期である。この時期は文部科学省の依頼を受け，関係有識者によって「21 世紀への留学生政策に関する提言」（1983 年），「21 世紀への留学生政策の展開について」（1984 年）の 2 つの提言がなされた。これらを踏まえ，政府は 21 世紀初頭までに 10 万人の留学生を受け入れる計画を策定した。2003 年に達成されたこの計画は，日本国内の学習者数の増加を力強く後押しした。また 1980 年代は日本経済がおおむね好景気であったことも学習者数の増加に影響し，1980 年には 127,167 人であった海外の日本語学習者数が，1990 年には 981,407 人と，10 年間で約 8 倍に増加した。日本語学習者数，日本語能力試験受験者数と実施国・地域数，日本経済成長率を一覧にしたのが表 1 である。

第3章 BJT ビジネス日本語能力テスト

小野塚 若菜

1 背景と目的

BJT ビジネス日本語能力テストは，ジェトロビジネス日本語能力テストとして，日本貿易振興会（現・日本貿易振興機構；JETRO）によって創設されたテストである。JETRO は，1980 年代に貿易摩擦や市場開放問題が注目されるようになる中で，「関税以外に貿易の障壁となるもののひとつが『日本語』である」との指摘が外国政府からなされたことから，ビジネス日本語支援を 1 つの事業として位置づけて取り組みを始めた。

JETRO のビジネス日本語支援事業は，在日大使館員や海外の公的機関の職員を対象とした日本語スクールの運営，研修用教材の作成，インターンシップ支援など多岐にわたった。中でも，日本人とのビジネスコミュニケーションの手段として日本語を使用し学習する外国人の増加，また，日本企業で働く外国人や就職を希望する外国人の増加に従い，日本企業および外国人ビジネス関係者双方から外国人のビジネス日本語能力を測定するテストの実現が望まれるようになった。JETRO が 1993 年に行った日系海外進出企業に対するアンケート（計 260 社）の結果でも，60% がこの種のテストの実現を望んでいた。そしてそのニーズに応えるために開発されたのが，ジェトロビジネス日本語能力テストである。

ジェトロビジネス日本語能力テストは，日本企業の国際化・海外進出日系企業の活性化に資することを目的として，外国人を雇用する企業，日本企業への就職を希望する外国人に役立つように，ビジネス日本語のコミュニケーション能力を客観的に示すテストとして，企業の協力を得て開発された。1993 年に日本語教育の専門家，ビジネスマン，外国人から成る検

第4章 J-CAT（Japanese Computerized Adaptive Test）

今井 新悟

1 背景と目的

J-CAT は検定試験ではなく，学習者が自身の日本語能力の把握，特に継続して学習する人たちが，自分で日本語能力の伸長を確認できるテストである。このテストは，科学研究費補助金基盤(A)「インターネットによる日本語のコンピュータ適応型テストの開発と検証」などの助成を受けて開発された（今井・赤木・中園 2012）。複数の大学の日本語教育およびテスト理論を専門とする研究者および多数の協力者によって現在の形になり，インターネット上で無料で公開している。J-CAT の開発は 2004 年から，筆者の前任校であった山口大学において始まった。当時，留学生のための日本語の授業へのプレースメントテストとして，紙による試験を年 2 回実施していた。しかし，その実施には困難と課題があった。

第一に，テストの実施の負担が大きすぎる。テストの作成，実施，採点にかかる労力は相当なもので，また，プレースメントテストであるから，短期間でクラス分けを行わなくてはならず，それが負担を増す。

第二に，テストの点数に信頼性がない。テストが変われば，点数の意味も変わってしまう。テストが違えば，テストの難易度が異なっている可能性が高く，同じ 80 点でもその持つ意味が異なってしまう。ある年の学生の平均点が上下しても，それは学生の能力によるものなのか，テストの難易度が変化したものなのかわからない。よって，プレースメントテストで，点数帯ごとにレベルを区切ったとしても，その点数がレベルを正しく反映していると保証できない。毎回同じテストを使えば点数の意味は変わらないが，紙を使ったテストでは問題用紙の持ち出しが懸念され，毎回テ

第5章 TTBJ（Tsukuba Test-Battery of Japanese）

酒井 たか子・加納 千恵子・小林 典子

1 背景と目的

筑波大学留学生センターでは，1980 年代から日本語授業に適切な日本語力を持つ学生をクラスに配置するために，プレースメントテスト開発に力を注いできた（酒井 1989；1991 など）。受験者数が少ないときには，時間をかけて対面でのインタビューや，作文などのパフォーマンスを見るテストや，個々の受験者のレベルに合わせた問題の選択ができたが，受験者数が多くなると個別対応が難しくなり集団テストを行わざるを得なくなった。具体的には受験者数が 100 人を超える頃から，短時間に効率よく行うために，マークシートによる解答方式を取り入れ，そこにさまざまな工夫を加えた。たとえば，事前に受験者に日本語レベル自己申告やコンピュータによる簡易レベルチェック診断を行い，初級者と中上級者にそれぞれ適した異なる文法テストを受験させていた。しかし，音声のテストでは同時に異なる問題が実施できないことや，採点処理の複雑さなどが問題となっていた。

SPOT《➡第 6 章参照》は，プレースメントテストの一部として用紙版で長く実施してきたが，コンピュータに載せることにより，ランダムに問題の提示ができることや提示時間のコントロールなどが容易になるなど多くの利点が考えられ，また共同研究を進めてきたアメリカの大学において，コンピュータを使わないものは認めないという方針が出された[1]こともあり，2004 年より WEB 版 SPOT の開発を始めた（小林 2005；2008 など）。

1　米国カリフォルニア大学サンディエゴ校およびコロラド大学と共同研究を進めてきた。

第6章 SPOT（Simple Performance-Oriented Test）

小林 典子

1 背景と目的

本章では日本語の運用力が得点に反映していると考えられる客観テスト（機械的に採点可能）として SPOT（Simple Performance-Oriented Test）を紹介する。SPOT がテストとして利用されるようになったのは，1995 年頃からである。小林・フォード（1992）は学習者の音声聴取能力が聴覚能力だけによるのではなく，言語知識の有無によって左右されるものであるという仮説を立て，それを証明する目的で，SPOT の原型であるひらがな 1 文字分の音声を書き取るテスト，「聞きテスト」を作成し，その結果を報告している。調査用テストとしては，無限にある語彙ではなく，項目数に限りがあり，言語使用において，より必然性の高い文法を「聞きテスト」の問題項目とした。この「聞きテスト」の得点をほかの文法知識テスト得点と比較することで，聴取能力と言語知識の関係を見ることにしたのである。その結果，「聞きテスト」は当時の筑波大学留学生センターの日本語学習者の文法知識テスト得点との高い相関を示し，「知識」と「聴取」の関係が深いことを実証できた（小林・フォード 1992，小林・フォード丹羽・山元 1996）。また，総合得点との相関はさらに高く，この「聞きテスト」は日本語能力の測定の道具となり得ると気付き，評価を目的とするテストとして問題を検討し作り直した（SPOT ver. 2 と呼ばれているもの）。以後，このテスト形式は，「日本語能力簡易テスト」などと呼ぶ時代を経て，その特徴から，SPOT と呼ぶようになったのである。現在では，日本語教育現場における学習者のプレースメントテストとして，また，習得研究における能力別グループ分けのテストとして，広く知られている。

第7章 OPI（Oral Proficiency Interview）

鎌田 修

1 背景と目的[1]

外国語を学ぶ最大の動機は何といってもその言葉を使ってコミュニケーションしてみたいという思いではなかろうか。とりわけ，昨今の交通手段の発達，さらに，インターネット通信の普及は，「空飛ぶ鳥のように」とは言えなくも，人の行き来を物理的にも，また，インターネット上でも大変容易なものにしている。そのように「外国語を使ってコミュニケーションを行う」という考えは，今や広く外国語教育の指針に影響を与え，欧州の *Common European Framework of Reference for Languages: Learning, teaching, assesment*（CEFR『ヨーロッパ言語共通参照枠』），米国の *National Standards for Foreign Language Learning: Preparing for the 21st Century*（『21 世紀の外国語学習スタンダーズ』）だけでなく，日本における『JF 日本語教育スタンダード』の基本的な骨子にもなっている。また，外国語能力の測定・評価においても，作成当初からこのような考えをもとに書かれた *ACTFL Proficiency Guidelines*（ACTFL 言語運用能力基準）は言うまでもなく，かつては学習時間が能力評価の軸であった「日本語能力試験」（国際交流基金，国際日本語教育支援協会）を，日本語を使って何ができるかという“Can-do Statements”の考えを取り入れた新しい評価に変えるところにまで至っている。

このように広く行き渡った外国語能力観は，しかしながら，その最も基本的な部分であろう「話す能力」の測定においては，ここで詳細す

1 本稿は「2014 年度南山大学パッヘ研究奨励金 I-A-2（特定研究助成・一般）」を受けて完成した。

第8章 「生活者としての外国人」のための日本語能力判定

村上 京子

1 背景と目的

「生活者としての外国人」とは，「だれもが持っている「生活」という側面に着目して，我が国において日常的な生活を営むすべての外国人を指すもの」と文化庁の報告書では定義されている[1]。1990年の入管法の改正に伴って，多くの日系外国人が日本の工場などで働くために来日し，日本各地で生活している。そこには，長い間日本で暮らしていても，簡単な挨拶程度しか日本語を話せない人も多くいることが指摘されている[2]。

名古屋近郊の豊田市でも自動車関係の工場で多数の外国人が働いているが，工場の中では通訳を介して仕事の指示がされ，団地と工場の間は工場のバスが送迎するため一般の日本人と接触する機会は限られている。スーパーやレストランなどでは，住民の多くを占めるブラジル人のためにその母語であるポルトガル語で買い物や食事ができ，同じ言葉を話す人々とだけ交流をして暮らしているうちに，日本語の読み書きはもちろん，簡単なやりとりもできないまま，5年，10年と経ってしまった人も多い。このような外国人集住都市などでは，日本語を習得しなくても生活できる環境があることや，工場のシフト制や残業で就業時間が不規則なため日本語を学習する時間がなかなか取れないことが，そこに住む外国人の日本語未習得の主な原因である。

1　文化庁（2010）『「生活者としての外国人」に対する日本語教育の標準的なカリキュラム案について』文化審議会国語分科会，p. 2.

2　国立国語研究所日本語教育基盤情報センター学習項目グループ・評価基準グループ（2009）『「生活のための日本語―全国調査―」結果報告〈速報版〉』

第9章 日本語語彙認知診断テスト

孫 媛・島田 めぐみ・谷部 弘子

1 背景と目的

言語教育における診断テストは，学習者の知識や運用面での習得状況，特に弱点に焦点を当てて把握するためのテストである（Alderson 2005）。学習を支援するという観点から，診断テスト開発の意義は大きい。学習者が現在どのような知識状況にあるかを把握することができれば，教師は適切な診断情報を学習者に与えることが可能となり，それに基づいて指導を改善することができるからである。しかし，当然ながら，適切な診断情報を得るためには，診断をしようとする領域の学習に必要な知識やその習得過程を知らなければならない。中高生の英語診断テストの開発にあたった金谷・英語診断テスト開発グループ（2006）は，「現在のところ，英語学習（習得）についての詳細で的確な診断情報を得ることは容易なことではない。英語学習（習得）過程が十分明らかになっていないからである」（p. 10）という状況を踏まえた上で，生徒の学習をモニターしてサポートする役割が相対的に高まってきている今，教師に診断という発想を定着させることが重要であり，そこに，不完全であっても診断テストを開発しなければならない理由がある，と述べている。日本語教育においても，英語教育と同様，教師に求められる役割は変わってきており，今後ますます診断テスト開発への期待は高まるものと思われる。

広い意味での診断目的では，近年，言語能力の簡便なレベル判定を目的として作成されたテストが Web サイト上で受験できるように公開されているが，これらはいずれも正答率から総合的に判定するもので，どの部分が習得できていてどの部分ができていないかなど分析的に提示できるシス

第10章 コンピュータによる日本語口頭能力テスト

安高 紀子

1 背景と目的

口頭能力を測定するテストには，試験官と受験者，または，受験者同士が対面して話す対面方式や，電話やコンピュータからの音声に回答する非対面方式のものがある。対面方式で行われるテストとしては，ACTFL-OPI《➡第７章参照》が挙げられる。OPI は試験官と受験者が対面し，インタビュー方式で行われるテストで，実際の会話場面に即した真正性の高いテストだと考えられる。しかし，大規模テストにおいては，試験官と受験者が 1 対 1 の対面で行う方式は，試験官の養成や人員確保という人的資源や，多数の受験者に対して一斉にテストを行うことができないため，時間的な効率の点で問題となる面もある。

TOEFL や TOEIC などの英語の大規模テストにおいて，口頭能力を測定するスピーキングテストには，コンピュータを用いた方式がすでに実用化され，運用されている。コンピュータを用いたテストの形式上，対話者のいない非対面方式で行われ，受験者はコンピュータからの音声に従って回答し，受験者が回答した発話は録音される。このような対話者とのやりとりがない準直接テスト[1]について，マクナマラ（2004）は，全受験者に同一の指示を与えることができ，対話者がいる場合に生じる変数が排除できるため，より公平なテストとなる可能性を持つと述べている。言語テストにおいては，テストの実施方法，受験者自身，評定者といった要因によって測定誤差が生じ，テストの得点に影響を与える（ブラウン 1999）とさ

1　テープなどに録音された音声によって指示が行われるテストのこと。半直接テストともいう。一方，対面する対話者からの直接の音声によるテストは直接テストと呼ばれる。

第11章 大規模言語テストの世界的動向：CEFRを中心として

野口 裕之

1 大規模言語テスト

本章では，これまで述べてきた日本語のテストが置かれた状況および今後の方向性を示唆するものとして，言語テストの世界的な動向について述べる。とりわけ，広範囲な内容を持つ言語教育評価の中でも全世界規模で実施され，多数の受験者を持ち，結果が受験者の処遇に影響を与える可能性の高い high-stakes test である大規模言語テストを取り上げる。

大規模言語テストは具体的には，日本語の場合は日本語能力試験（国際交流基金日本語試験センター）《➡第2章参照》，英語の場合はケンブリッジ英検や IELTS（Cambridge English Language Assessment），TOEFL（Educational Testing Service），フランス語の場合は TCF や DELF/DALF（CIEP），ドイツ語の場合はゲーテドイツ語検定（Goethe-Institut），中国語の場合は漢語水平考試（北京語言大学 HSK センター，2010 年以降は新 HSK として中国国家漢語国際推進事務室（中国漢弁））などをはじめ，言語・測定目的に応じて多数のものがある。

また，このような大規模言語テストの開発機関が相互に協力して情報を共有し，言語テストのテストとしての質を維持し，向上させることを目的として，欧州域内では ALTE（The Association of Language Testers in Europe）が 1989 年に結成され，欧州域内にある 33 のテスト開発機関が現在メンバーとして加盟している。そして，毎年各種のワークショップや 3 年ごとに国際研究大会が行われている。

欧州では現在の Cambridge English Language Assessment が 1913 年に最初の言語テストを実施して以来 100 年の歴史があるが，わが国の場合，日