



Title	発話能力テストの現状と問題点
Author(s)	清池, 恵美子; 松本, 隆; 西口, 光一
Citation	日本語教育. 1986, 58, p. 127-145
Version Type	VoR
URL	https://hdl.handle.net/11094/25292
rights	©日本語教育学会
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

発話能力テストの現状と問題点

清地恵美子, 西口光一, 松本 隆

(1985. 10. 18受)

要 旨

近年, コミュニケーションのための外国語教育の必要性が強く叫ばれるようになった。本論は, それに対応する発話能力の評価という問題について, 筆者らによる面接テストの試行をふまえて論じるものである。はじめに, 現在実施されている面接テストの中で, 実際の生活場面を反映した形で実施されているものを二つ紹介する。つぎに, 面接方式のテストとしての妥当性及び信頼性が検討された。面接方式は一般に, ひじょうに妥当性が高く, 客観性も十分あるものと認められた。面接の広範な採用を阻んでいる要因についても検討された。そして, 実施容易性という観点からすると, 面接は対人的なやりとりを行う能力にしばって評価するものとして活用するのが望ましいと提案する。発話能力の他の面の評価は, 反復テスト, クローズテスト他の方法により補うことができる。

【キーワード】 コミュニケーションのための外国語教育, 発話能力,
面接テスト, 評価, 談話能力

1. はじめに

1984年度より日本語能力試験が実施され, 日本語能力測定への関心は高まっている。読む力や聞く力だけでなく, 話す力を測定してもらいたいという学習者も少なくないようである。また, 日本の大学や企業にはよりいっそう多くの外国人の参加が期待されており, 受け入れ側にとって日本語でどの程度話せるかがわかること, 外国人にとっては自分がどの程度日本語が話せるかを証明できることは, 早晚, 日常的に需要が出てくるものと思われる。しかし, 話す力のテストは一定した評価をすることが難しいとされており, 特に精度を要求されたり, 多くの受験者を処理する必要のある場合, 日本語能力認定試行試験のように実施を見合わせるか⁽¹⁾, あるいは全員に同じ質問を

し、予想される答え方のひとつひとつへの配点を細かく決めておくといった現実のコミュニケーションのあり方とは異なる柔軟性のない形で面接テストを行って、その結果をもって受験者の話す能力とするかのどちらかになりがちである。

本稿の目的は、現実の「話す」活動に近い形で行われるいくつかの外国語の面接テストを紹介し、筆者らの行った面接テストの試みをふまえて、それらの問題点について検討したのち、話す力の現実的な測定方法を探ることである。

2. 外国語テストについての考え方

外国語の評価法には、二つのアプローチがある。一つは項目テスト (discretepoint test)、もう一つは全体テスト (global/holistic test) である。

前者は、特定の言語項目を出題し、それに正解することによって得点が積み上げられていくタイプのテストで、一般に行われているテストにはこのタイプのものが多い。たとえば、初級段階を修了したかどうかを認定するテストで、ある特定の動詞を使って受け身形の作り方を出题した場合、これは項目テストである。項目テストは、個別の言語項目 (文法・語彙・文字・音声) の習得状況を確認するものであり、学習者の弱点を発見する診断テストとして有用である。

それに対して、後者は言語能力を全体として評価しようとするものである。面接・作文などは、この方法で実施・採点されることが多い。これは一般に採点が主観的で信頼性がないとして、公的な性格をもつテストの場合、回避されがちなようである。

次に、テスト目的による分類では、先にあげた診断テストのほかにここでとりあげようとしている能力テスト (proficiency test) がある。

能力テストの目的は、たとえば日本語能力試験1級の場合、受験者が日本語で社会生活を営み、また大学での学習・研究を始めることができるかどうかを認定することである。このような能力の認定は単に学習者の励みとする

ためだけではなく、大学の入学許可担当者や企業の人事担当者にとって、ある人物がその外国語（ここでは日本語）を使ってどの程度のことのできるかを知る目安として役立つ。この種のテストとしては、典型的な例ではないが、広く知られているものとして実用英語検定試験があげられる。

項目テストで高い得点をとることのできる学習者が必ずしもその外国語を使いこなすことができるとはいえないことは、経験的に知られているとおりである。言語に関する個別的な知識の量の多寡だけでは学習者の言語運用能力を間違いなく予測することは難しいため、言語運用能力そのものを測定する方法がとられなくてはならない。次節では外国語教育における口頭による言語運用能力（発話能力）測定の試みをいくつか紹介する。

3. 外国語発話テストの紹介

3.1. 日本語教育における発話能力の測定

外国語を話す能力を測定するのは難しいということは、しばしば指摘されているが、日本語教育の現場で、言語運用能力としての発話能力を測定しようとしている例を2つ紹介したい。

この2つのテストは、どちらも様々なテクニックを使って受験者から発話を引き出し、いくつかの尺度で評価して、その合計点を最終的な評価とするものである。植松（1978）の報告によると、国際学友会でたてている尺度は、理解・文法・内容・発音・印象の5つである。各尺度は5段階に評価され、2名の教師が採点にあたり、満点はその合計の50点となる。

アメリカ・カナダ11大学連合日本研究センター（以下「日本研究センター」と呼ぶ）で採用している尺度はやはり5つであるが、①聞いて理解する力 ②正確に表現する力 ③表現力 ④話し方のなめらかさ ⑤発音 と、国際学友会の場合と異なっている。ここでの評価は、面接担当者以外のすべての教師（約10名）が参加して行われる。その中で、他の採点者の評価と最もへだたりのある採点者の評価を除外したうえで、その他全員の評価点の平均が出されて最終的な評価を決めている⁽²⁾。

植松 (1978) は、このような評価法での問題点として、各尺度において設定した5つのレベルがそれぞれどのような能力を標準とするかを記述する必要があると述べている。単にある尺度をいくつかにわけただけでは、採点者によってその解釈が異なり、その結果、採点者間での評価の変動を生み出してしまふからである。植松 (1978) にはレベルごとの標準能力記述の具体例が若干示されている。日本研究センターでは評価に際してFSIテスト(次項参照)の標準能力の記述を参考にして評価基準を決めている。

また、植松 (1978) では、テスト実施上の問題としては受験者の日本語力の水準にあった課題が与えられなければならないことが指摘されている。

三枝 (1985) はゲーテ・インスティテュートの例を紹介し、テストの問題が場面性に富んでいることに注目して、日本語教育においてもコミュニケーション場面での言語運用能力を測定する必要があるのではないかということを示唆している。これは植松 (1978) でみられる、現実のコミュニケーション場面での話す能力を測定しようとする志向と一致する。

英語をはじめとする他の外国語教育でもこの必要性は指摘されており、多くの試みがなされてきた。ゲーテ・インスティテュートのテストのように、実用化されて実際に多数の受験者の評価に使われているものも多い。次に、それらの例をみることにする。

3.2. 日本語以外の言語での外国語発話能力テスト

The Communicative Use of English as a Foreign Language (以下Communicative Use) は、上に述べたような外国語教育でのコミュニケーション能力を測定する必要性から開発されたものである。既に多くの外国語の面接テストが紹介されているが⁽³⁾、これはあまり知られていないので少し詳しく紹介したい⁽⁴⁾。

Communicative Use テストは、近年の外国語教育の動向に対応するためにイギリスのRoyal Society of Arts Examination Board (以下RSAEB) で開発が進められ、1981年から実施されている。読む・聞く・書く・会話する(Oral Interaction)の4能力のそれぞれを初級・中級・上級の3レベル

に分け、別々にテストして合否を決める。このテストで測ろうとしているのは、受験者がある事柄を英語を使ってできるかどうかで、実生活で言語使用者が会うであろう現実的な課題が出題される。モデルとする英語はイギリス国内で使われている英語で、native および non-native の英語話者との交際、公的機関でのやりとり、買い物やサービス利用、関心のある場所や娯楽のための場所の訪問、旅行と旅行の手配、情報収集と娯楽のためのメディア使用、治療と健康、学問的・職業的・社会的目的での勉強にかかわる内容である。会話テストでの評価基準は①言語形式の正確さ(音声・文法・語彙)②場面に応じた適切な言語使用③使えることばの range④新しい話題・機能・談話の種類への適応力⑤会話への参加度の5つである。それぞれの尺度でのレベルごとの標準能力は資料1にあげておく。テストは次の3部に分かれ、すべて対面で行われる。ただし電話での会話が出題されることはある。所要時間はそれぞれ約5分ずつである。

第1部 学校またはテストセンターを代表する人物との会話。面接者は通常、受験者と初対面ではない。

第2部 受験者同士の会話。2人1組が原則だが、人数の関係で3人のこともある。

第3部 室外にいた人物に、第2部で行った会話の報告。

Communicative Use テストは一般的・日常的なコミュニケーション能力を扱うものであるが、Test of English for International Communication (以下 TOEIC) は「企業人のビジネス・コミュニケーション能力」を評価しようとするものである⁽⁵⁾。

TOEIC は、アメリカの民間テスト機関 ETS (Educational Testing Service) が経団連などの後援を受けて日本人のために開発したテストである。読解・聴解の筆記テストを全員に課し、5段階に評価した上で、上位2レベルの受験者のうちの希望者に面接による発話能力テストを行う。これは、ある程度までの発話能力は聴解能力と構文力・語彙力から予測可能であるという見解に基づいている。なお、TOEIC の場合、この予測の信頼性を

保証するために、別途に受験者の5%を抽出して20分程度の面接テストを実施し、筆記テストの得点との関係を調整している。

このテストの開発にあたって留意されたのは得点の解釈容易性 (score interpretability) である。

得点の解釈容易性とは、テストの評価が単なる点数でしかないのではなく、その受験者の能力が具体的にどのようなものであるかが、テスト専門家以外の一般の人間にも把握できるかどうかをいう。TOEIC ではこの点を重視し、各レベルの標準能力の記述を公表している。

TOEIC の標準能力の記述およびテスト方法は米国国務省外務研修所 (FSI) の開発した発話能力テスト (以下「FSI テスト」) と流れを等しくする。そのテスト方法は面接者と受験者が実際に自然なコミュニケーションを行いながら評価するというもので、日本語教育においても参考となるところが大きいと思われる。そこで、次項では TOEIC 面接テストの方法について述べることにする。

3.3. TOEIC の面接方法

TOEIC の面接は受験者の個性・興味に合った自然な会話の形で行われるので、事前に個々の面接での会話内容や順番を予想することはできない。原則として4段階に分かれるが、各段階間の移行ははっきりとした区切りをつけず、自然な話の流れの中でなされる。下に述べる段階(2)・(3)の順は固定的なものではなく、熟練した面接者は、この二つを適宜組み合わせながら受験者のレベルを見定める。また、全体テストであるので、誤用や不適切さは、それが一貫した傾向としてみなせる程度に繰り返されてはじめて、レベル決定の資料となる。

〔面接の段階〕

(1) The Warm-up 受験者をリラックスさせ、その外国語で話すことに慣れさせる。面接者はここで受験者のだいたいのレベルの見当をつけるが、はじめのうちの発話は、まだ本来の能力が出し切れていないので、レベル決定の資料とはしない。

(2) The Level Check 受験者が一貫して維持することのできる最も高い発話レベルを見る。様々な話題や機能がとりあげられなければならない。発話内容の事実関係の正誤は不問とし、言語能力のみが評価の対象となる。

(3) The Probes 前段階で判定したレベルより上のレベルを受験者が発話できないことを確かめる。急激に誤用が増えたり、話すスピードが落ちたり、言葉を探して考え込んだりという現象が何度かおこれば、そのレベルの発話ができないと判断できる。

(4) The Wind-Down 受験者は Probes で能力の限界まで試されている。そのまま面接を終えると挫折感を残してしまうので、ここでもう一度受験者に発話できる最も高いレベルに戻って面接を終える。

以上見てきた方法で行われる面接テストは、現実のコミュニケーションに近い形で、実際的なコミュニケーション能力を測定しようとしており、テストが教育に及ぼす影響を考えると、日本語教育のどの段階においても、試みられるべきであると考えられる。しかし、実際にとりいれることがためらわれがちなのは、いくつかの問題点が予想されるからである。次節からは、面接法による発話能力テストの問題点をあげ、検討していく。

4. 面接法による発話能力テストの問題点

一般に面接テストの問題点としてよく指摘されるのは、実施の容易性と評価の信頼性である。米国国務省日本語研修所の卒業試験は受験者のその後の仕事と待遇に大きな影響を持つものであるが、このときは口頭試験だけで一人あたり合計70～80分を費やす。学期末試験のときには一人あたり30分ということである⁽⁶⁾。TOEICでは、一人あたり15～30分が望ましいとされる⁽²⁾。両方で期待されているような形、つまり、あらかじめ質問項目を決めておくのではなく、受験者の反応を利用しながら、「大人同士の一つの自然な話の流れの中で」⁽⁷⁾ 受験者のレベルを見定める、という形でテストする場合、30分間で評価のための十分な発話のサンプルを得るのは熟練を要する。日本語能力試験のように、将来大量の受験者が予想されるテストでは、受験

者全員に熟練した面接官と30分間1対1になる時間を与えることが實際上可能かどうかはかなり疑問である。また、第一回日本語能力認定試行試験でも問題となったように⁽⁸⁾一人の採点者が、自分の受け持つどの受験者に対しても一貫した尺度で評価が下せるか、複数の採点者間で評価に食い違いがおきかないか、ということが問題になる。

〔面接法の妥当性と信頼性〕

米国国務省日本語研修所で使われているのはFSIテストである。これは外国語テスト研究で発話能力テストとしてしばしばとりあげられている。まず、このテストの妥当性について Bachman と Palmer の研究を見ておきたい。

Bachman & Palmer (1981) は FSI テストを形式と内容の二つの観点から検討した。75人の被験者の読む能力と話す能力をそれぞれ面接法・翻訳法・自己評価法の3つの方法でテストして分析したところ、(1)話すテストにおいて、面接法はテスト方法の成績への影響が最も少ない。(2)FSIテストの成績に最も大きい影響を持つのは話す能力である、という結果が出た。

複数の採点者間での評価の相違の有無については Mullen の実験が示唆に富む⁽⁹⁾。ここで使った評価基準は、発音・流暢さ・聴解・構文力・全体評価の5つであり、これらの観点から被験者を5つのレベルとその中間、計9段階に評価した。基本となる5レベルに関しては観点ごとにそのレベルの標準的な発話の特徴を記述したもの(2-2参照)が採点者に渡された。採点者は、外国人への英語教育とこの種の面接テストの経験のある言語学専攻の大学院生である。実験に際しては、事前に採点基準・面接手順・実施上の注意事項についての説明・指示を受けているが、それ以上の訓練や打ち合わせ等を行われていないようである。二人1組で98人の被験者を採点したところ、二人の採点が大きく異なった例が数例あったが、6組の採点者のうち4組の採点結果には満足できる信頼度があった。5つの観点のうち、全体評価はどの組の評価も高い一貫性が認められた。

採点者間での評価の信頼性は採点者が事前に受ける研修の内容によって増す

ことができるようである。British Councilが行っている面接テストでは採点者は本国から送られてくるテープをもとに討論して調整する⁽¹⁰⁾。FSIでは各言語の公認採点者は1名しか置かず、日本語研修所では他の採点者と実際のテストについて討論することで評価基準の一定化をはかっている⁽¹¹⁾。ある研究でFSIテストが使われたときに採点者間の信頼性がチェックされたが⁽¹²⁾、このときは15例のインタビューを収めた訓練用のテープを使って採点者の研修が行われた。その結果、発音を除く他の5つの観点すべてにMullenの実験より高い信頼性が認められた。ここでも最も信頼性の高かったのは全体評価においてであった。

筆者らは1985年にTOEIC方式に基づいて低いレベルの発話能力の識別力を増す方向で評価基準を改定し、19例の面接テストを試行した。被験者は英語・中国語・トルコ語を母語とする成人で、発話能力は資料3に付した評価基準でレベル1-1からレベル4+にわたった。採点にあたった者は計7名である。この試行テストを通して、評価の信頼性に関して二つの示唆を得た。一つは上述のように事前研修において採点者同士が討論の機会をもつことの重要性である。これはBritish Council, FSIの例を見てもわかるように、基準を決めるための討論というよりは、既に決められている基準をよりよく理解するための討論であるといえよう。事前研修のたびに討論して基準を決めることは、そのつど基準にずれを生じさせることになり、評価の一貫性を損なうもととなる。

もう1つの重要な点は、評価基準をできるだけ簡潔にすることである。面接中にマニュアルを参照することは難しいし、受験者に与える心理的影響のうえからも好ましくない。また、細部に気を取られることで全体としての印象が損なわれるおそれがある。その結果受験者の発話能力を正しく評価することができなくなり、評価の信頼性が保てなくなる。

簡潔な評価基準とは、記述の短さだけではなく、基準になる観点を整理し一本化していくことも意味する。筆者たちのテストで採点基準統一上の問題として残ったものの中には、acceptabilityとnativenessの問題があった。こ

それは、そのテストでどのような日本語を最上級のモデルとするかということであるが、これは外国語教育上の重要な問題を含んでいる。ほとんどの外国人話者は母国語話者と同じ量の日本語知識・技術を持たないが、限られた知識・技術を活用して必要な行動目的を遂行する。このときに用いる言語的・非言語的な工夫のことをストラテジーというが、communicativeな話者は様々なストラテジーを駆使して、簡単な文型だけでかなりの言語活動ができる。筆者らのテストでは、目的遂行の成否という観点からは資料2の評価基準で高いレベルに位置づけた言語活動を限られた文型だけでこなせることがわかった。

TOEICではレベル3までの評価は実現できる言語機能を基準とし、それ以上のレベルでは正確さ・自然さを基準とすることになっている。しかし、日本語による描写力・自己表現力の点では非常に劣りながらも対人的な自然さの点で非常に高いレベルの発話によるやりとりのできる受験者がいる一方で、書き言葉と変わらない単純な文を使って比較的高度な内容が表現できるのにもかかわらず、単に流暢さの点だけでなく、対人的な会話力の点で非常に劣る受験者などの例が出てきて、採点者間の評価の食い違いを招いた。このような面接例を通して、口頭による描写力・心的態度の表現力・やりとりにかかわる能力は原則としてそれぞれ独立に発達する別々の要素ではないかという印象を得たのだが、これらの受験者はTOEIC方式の評価法では結果的に同じレベルとして評価されることがある。仮に採点者間の採点結果を統一できたとしても、このように特徴の異なるものを同一と認定することは能力テストに必要な要件—(1)他の受験者との能力の区別(2)本人の能力についての情報提供を満たさないと考える。

TOEICの場合は、受験者は日本に住む日本人であり、ほぼ単一の教育背景と言語環境が期待できる。そのため、この種の例外的なことといえよう。しかし、日本にある日本語教育機関でのテストや日本語能力試験のように、日本在住経験も含めて幅広い教育背景・言語環境をもつ受験者を対象とする場合には、TOEICのもつ前提をそのままに踏襲することはできない。発話

能力を構成する下位能力を特定する観点から、発話能力と他の3能力（聞く・読む・書く）の関係を明らかにし、それに基づいて評価基準そのものが検討しなおされなければならない。

5. 実施容易性の高い発話能力テストの形態についての研究と考察

1対1の面接を行わないで発話能力を測ることは、たとえ可能であったとしても、現実の発話によるコミュニケーションのあり方から遊離してしまい、教育に及ぼす影響と受験者の心理的満足感の観点から問題がある。ここでは、1対1の面接を行うことを前提とし、その所要時間を極力短縮する方向で、発話能力テストの実施容易性を高める方法を考察したい。

1対1の面接時間の短縮は、1対1の面接でなくても十分効果的に測定できる項目はほかのテストで代行することで達成できる。

従来、言語の4技能の別が強調されてきたが、それは話す技能を支える知識技術の集合と他の技能、たとえば書く技能を支える知識・技術の集合の間に全く重なりがないということを意味しない。むしろ、1970年代以降は4技能の相関性や誤用の研究などを通して、どの言語技能にも共通する中核的な言語能力を想定する傾向がある。そういった中核的な言語能力については、面接テストから除外できる。Oller & Perkins eds. (1980)はその中のいくつかの論文で、この中核的な言語能力の存在を言語テスト研究の立場から実証しようとした⁽¹³⁾。

Oller たちは、南イリノイ大学とテヘラン大学の学生を対象に合計27種の言語テストを行って、その成績を比較したところ、何らかの共通した言語能力の存在を示唆する結果が出たと述べている。ただし口頭面接テストに関しては言語能力の共通部分が他のテストほど顕著ではなかったようである。この事實は、面接テストの重要性を示していると考えてよいかもしれない。

発話能力が他の3技能と違う点は話し相手との間で常に feedback がなされながら進められる点であろう。FSIが、近年、従来の5尺度（発音・文法・語彙・流暢さ・理解）のうち、発音の観点を廃して談話能力の観点を加え

た⁽¹⁴⁾のはこの点を重視したためであると思われる。Munby (1978) は話すことに特徴的な技能として次のようなものをあげている。

(1) 会話を始める

- ① 口火のきり方
- ② 新しい観点の持ち出し方
- ③ 新しい話題の出し方

(2) 会話を維持する

- ① 反応のし方
- ② 続け方
- ③ 相手の反応へ合わせての話の進め方 (特に発話の途中での feedback)
- ④ turn-take のし方
- ⑤ 間のとり方 (いいよども・「息つく暇」をとるための決まり文句など)

(3) 会話の終わらせ方

- ① 会話の区切りの示し方
- ② 会話からの脱け方
- ③ 話題のしめくり方

また、語彙・文法・構文などにおいて、書き言葉と対比した意味での話し言葉に特徴的な言語使用があるだろう。これらの点に関しては面接以外の形でテストすることは難しいと思われる。このような点を評価対象としてのレベル設定、それに基づいた具体的なテストの構成は今後の課題である。

では、面接テストの役割を、実施容易性の観点から上記のようなポイントにしぼった場合、残った部分を代行するものとして、どんなテスト方法があるだろうか。

Hendricks et al (1980) は、FSI 方式の面接テストを数種の実施容易性の高いテストと同時に70人の学生に実施し、それらのテスト間の相関について調べている。

実験に使われたテストはFSI方式面接テスト・反復テスト・口頭クローズド・テスト・音読テストの4種である。面接テストは、発音・文法・語彙・流

暢さ・理解・総合評価の6つの観点から採点された。

その結果、反復テストに面接テストの総合評価との最も高い相関があることがわかった。クローズ・テストも高い相関を持つが、音読テストは流暢さと理解の得点と非常に高い相関を持ちながらも、総合評価との相関は他の2つのテストと比べて比較的低かった。

反復テストは、ひとまとまりの文を一度聞かせたあと、いくつか区切って聞かせながら受験者に口頭で繰り返させるものである。この実験では、70語前後の長さの文が3つ出題され、解答はLL設備を使って録音されて、本文中の語と、本文中にはないが許容できる語の両方に得点を与える方法で採点された。

このテストでは、受験者は自分の持っている当該外国語能力を反映した一貫した誤りの傾向を示すと考えられる。ただし、記憶力のテストにならないよう、聞き取った言葉を正確に繰り返すことより、聞き取った意味を正確に表現することに集中するように受験者に指示を与えておくべきかと思われる。このテストが先に論じた発話能力のどの部分を測定するか、また、他にどのようなテストが面接テストの補助となりうるかは、今後の研究に待たなければならない。

5. おわりに

以上、現実のコミュニケーションに近い形で外国語の発話能力を測定しようとする試みを紹介し、その妥当性を確認し、実施上の障害となる信頼性の確保、実施容易性の観点からの改良の方向について考察した。

日本語能力試験はいうまでもなく、個々の日本語教育機関で行われるものについても、テストが教育のありかたに及ぼす影響は非常に大きい。TOEICのように一定レベル以上の希望者のみに面接テストを行うのはひとつの現実的なやりかたと言えるかもしれない。しかし、上のような視点からは、むしろ限られた文法力・語彙力しか持たない低いレベルからの日本語によるコミュニケーション能力が測定できるような視点で1対1の面接テストを行う方

が好ましいと思われる。また多人数に対して面接テストを実施するためには、1対1の面接テストでは話し言葉での対人的な談話構成能力にポイントをしぼって採点することが必要ではないかと考える。その際に不足する部分については、本稿であげた反復テストをはじめとして、クローズテストその他の筆記テストなどで補っていく方法があるのではないだろうか。このような観点からの評価基準の設定と具体的なテストの実施が望まれる。

この論文は、財団法人国際教育振興会主催の日本語教師養成講座担当者研究会での討論をふまえたものである。米国国務省日本語研修所主任教官の高見沢孟氏、ブリティッシュ・カウンシルのコリン・スローリー氏、ケビン・ドビン氏には貴重な時間をさいてご協力いただいた。ここにお礼を申し上げたい。また被験者の確保や試験場の提供など、日本研究センター・インターカルト日本語学校・ELEC英語研修所のご協力がいただけた。ここにあわせて感謝の意を表したい。

注

(1) 日本語教育学会 1985 に次のような1節がある。「アンケート調査の結果、話す能力、作文の能力も測定してもらいたいという要望を持つ受験者があることが判明した。話す能力については、第1回の試行で面接試験を実施した。しかし、採点の基準が定まらず、採点者間の評点に大差が出て、その時点で採点は不可能という結論に達し、現在もその対策は出ていない。採点者を常時雇用し、採点者間の採点基準が常に一定であるように訓練し続けるといった制度的な問題も含めて今後の検討に委ねなければならない。」

(2) 谷1978による。

(3) 主な面接テストについて、名称と参考文献をあげておく。

[アメリカ]

・ Interagency Language Roundtable Oral Interview, Madsen 1983.

・ Ilyin Oral Interview, Ilyin 1976, Oller 1979, 垣田 1984.

・ Upshur Oral Communication Test, Upshur 1972, 1974, Oller 1979, 垣田 1984.

[イギリス]

・ Oral Examinations (Association of Recognized English Language Schools Examination Trust), Madsen 1983.

・ Cambridge First Certificate, Jones 1984a.

・ Cambridge Proficiency, Jones 1984b.

[フランス]

・ French Speaking Proficiency Test, Pimsleur 1961, 1967, 垣田1984.

- (4) Communicative Use テストについては SAEB の1983年発行のブックレットを参考にした。
- (5) TOEIC に関する記述は Woodford 1978, TOEIC 運営委員会1981に基づく。
- (6) (7) 米国国務省日本語研修所主任教官高見沢孟氏との1985年9月6日の面談による。
- (8) (注) 1. 参照
- (9) Mullen 1980.
- (10) The British Council の英語教育担当官 Colin Sloley 氏, 英語教育試験担当官 Kevin Dobbyn 氏との1985年7月8日の面談による。
- (11) 高見沢氏との面談による。
- (12) Hendricks et al, 1980.
- (13) Irvine et al. 1974もこの問題を扱っている。
- (14) 高見沢氏との面談による。

[参考文献]

- (1) 安藤淑子, 1979「口頭表現能力の評価(初級日本語)」講座日本語教育, 第15分冊
- (2) 植松 清, 1978「会話テストの改善をめざして」国際学友会日本語学校紀要, 第2号
- (3) 大友賢二, 1966「言語テストの考え方(1), (3)」現代英語教育, 7・10月号, 研究社
- (4) 垣田直巳監修, 岡 秀夫編, 1984「英語のスピーキング」大修館
- (5) 加藤俊一, 1968「評価のしかた」講座日本語教育, 第9分冊
- (6) 三枝令子, 1985「Goethe-Instituteの統一試験の構成と評価法」日本語教育55号
- (7) 斎藤修一, 1979「外国語教育における試験」日本語と日本語教育, 第8号
- (8) 秦 喜美恵, 1982「日本語熟達度測定法としてのクローズ法に関する実験的研究」国際基督教大学大学院教育学研究科提出修士論文
- (9) 高見沢 孟, 1977「米国務省日本語研修所の場合」日本語教育32号
- (10) 谷 すみえ, 1978「各種の試験」アメリカ・カナダ11大学連合日本研究センター紀要, 第1号
- (11) TOEIC 運営委員会編, 小林薫監修, 1981「ビジネスマンの国際英語力テスト——第1回 TOEIC 問題と解答」朝日イブニングニュース社

- (12) TOEIC 運営委員会編, 三枝幸夫解説, 1982「第3回 TOEIC 問題集——解答と解説」朝日イブニングニュース社
- (13) 日本語教育学会, 1985「外国人のための日本語能力認定試験に関する調査・研究の経過報告V」
- (14) 長谷川恒雄, 1970「留学生に対する日本語能力テスト——その結果と分析(上)」日本研究, 第1号
 ——, 1974「留学生に対する日本語能力テスト(下)——表現力テストについての分析」日本語と日本語教育, 第3号
- (15) ラドー, ロバート, 1971「言語テスト」大修館
- (16) Bachman, L.F. and A.S. Palmer, 1981, The Construct Validation of the FSI Oral Interview, *Language Learning*. Vol. 31 No. 1 67—86.
- (17) ——, 1982 The Construct Validation of Some Components of Communicative Proficiency. *TESOL Quarterly* Vol. 16 No. 4 449—464.
- (18) Campbell, R. and R. Wales, 1970, The Study of Language Acquisition, in *New Horizons in Linguistics*, J. Lyons (ed) 1970, Penguin Books.
- (19) Canale, M. and M. Swain, 1980, Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing, *Applied Linguistics*, Vol. 1 No. 1 1—47.
- (20) Canal, M. 1983, From Communicative Competence to Communicative Language Pedagogy. in *Language and Communication*, Richards, J. C. and Schmidt, R.W. eds. 1983, Longman.
- (21) Carrol, J.B. 1961, Fundamental Consideration in Testing for English Proficiency of Foreign Students, in *Testing*, Allen, H.B. (ed) 1965.
- (22) Chomsky, N. 1965, *Aspect of the theory of syntax*, MIT Press.
- (23) Gronlund, N.E. 1981, *Measurement and Evaluation in Teaching*, Macmillan.
- (24) Heaton, J.B., 1975, *Writing English Language Tests*. Longman Handbook for Language Teachers, Longman.
- (25) Hendricks, D., G. Scholz, R. Spurling, M. Johnson and L. Vandenburg, 1980, Oral Proficiency testing in an intensive English language program, Oller and Perkins (eds), 1980.
- (26) Hymes, D. H., On Communicative Competence, In *Sociolinguistics*, J. B. Pride and J. Holmes. (eds), 1972, Penguin Books.
- (27) Ilyin, D., 1976, *Ilyin Oral Interview*, Newbury House.
- (28) Irvine, P., P. Atai and J.W. Oller Jr. 1974, Cloze, Dictation, and the Test of English as a Foreign Language, *Language Learning*, Vol. 24 245—252.
- (29) Jones, L. (ed), 1984a, *Cambridge First Certificate: Examination Practice 1* :

- Teacher's book, Cambridge University Press.
- , 1984b, Cambridge Proficiency: Examination Practice 1: Teacher's Book, Cambridge University Press.
- ⑩ Lado, R., 1961, Language Testing, Longman.
- ⑪ Madson, H.S., 1983, Techniques in Testing, Oxford University Press.
- ⑫ Mullen, K.A., 1980, Rater reliability and oral proficiency evaluations, Oller and Perkins (eds).
- ⑬ Munby, J. 1978 Communicative Syllabus Design Cambridge University Press.
- ⑭ Oller, J.W. Jr., 1979, Language Tests at School: A Pragmatic Approach, Longman.
- ⑮ Oller, J. W. Jr. and K. Perkins. 1980 "Research in Language Testing" Newbury House.
- ⑯ Pimsleur, P., 1961, A French Speaking Proficiency Test, The French Review, Vol. 345, 470—479.
- ⑰ —, 1966, Speaking Skill, in Trends in Language Teaching, Valdman, A. (ed), 1966.
- ⑱ Royal Society of Arts Examination Board, 1983, The Communicative Use of English as a Foreign Language and English as a Foreign Language.
- ⑲ Savignon, S.J., 1983, Communicative Competence: Theory and Classroom Practice, Addison-Wesley.
- ⑳ Upshur, J.A., 1972, Objective Evaluation of Oral Proficiency. English Teaching Forum, Vol. 10—2, 15—19.
- ㉑ —, 1974, Productive Communication Testing: Progress Report, in Focus on the Learner, Oller, J.W. and Richards, J.C.(eds), 1973, Newbury House.
- ㉒ Woodford, P. 1978, Test of English for International Communication, A handout given at Keidanren, ToKyo, Japan Dec. 19, 1978.

資料1 イギリス Royal Society of Arts Examination Board が実施している
 “The Communicative Use of English” の評価基準

	初級	中級	上級
正確さ	発音は母語の干渉が強 く、なまりがあるが、 普通聞き分けられる。 文法・語彙の誤りが混 乱をひきおこすが自己 回復できる。	発音は明らかに母語の 干渉があるが明瞭に聞 き分けられる。文法・ 語彙の正確さは一般に 高い。誤用は現れる が、容認できる程度で、 コミュニケーションを 阻害しない。	発音は正確である。若 干のなまりは認められ る。文法・語彙の正確 さは非常に高い。
適切さ	求められる機能に対し て概ね適切な言葉を使 用できるが、微妙な使 い分けはできない。話 し手の意図は多少努力 すれば理解できる。	求められる機能に対し て一般に適切な言葉を 使用できる。話し手の 全般的な意図は常には っきりわかる。	コンテキスト・機能・意 図について完全に適切 な言語使用ができる。
広がり	極めて限られた範囲の 表現で合格とされる。 しばしば表現方法を模 索しなければならない。	かなり広い範囲の言語 を使用できる。あから さまに単語につまずく ことなく自己表現でき る。	非常に広い範囲の言語 が使用でき、困ること はほとんどない。回避 ストラテジーのあから さまな使用もほとんど ない。
柔軟性	会話で主導権を握るこ とは普通ない。話題の 変化への対応に時間か かかることがある。話 し相手が会話を支える ためにかなり話し手に 合わせる必要がある。	会話で主導権が握れ、 新しい話題や話の方向 ・変化に適応できる。 (それが常に表面に現 れるわけではない。)	会話に十分貢献し、も ちろん主導権がとれ る。話し相手に緊張を 与えることはほとんど ない。
発話の 長さ	ひとことかふたことの 簡単な発話でも合格と される。	ほとんどの発話は短い がより複雑な発話を産 出しより長い談話に発 展できることを示す証 拠が必要。	長い応答など、会話を 発展させることができ なければならない。

資料2 試行テストに使用した評価基準

レベル0 日本語によるコミュニケーションはできない。

0-0 日本語でまったく何も言えない。

0-1 こんにちは・さようなら・ありがとう・すみません・はい・いいえ・わかりません程度のことと言える。

0-2 0-1以外の、コミュニケーションには役立つ断片的な暗記した語フレーズの再生ができる。若干のラベリングができるだけで、コミュニケーション能力はない。

0-3 コンテキストの助けがなければコミュニケーションできない。好都合な条件のもとで、否定・能力否定・禁止・依頼・意志・質問のうちのいくつかができる。文法能力はない。

レベル1 限られた日常生活上の社会的コミュニケーションができる。

1-1 自己紹介・一日のあいさつ・値段聞きを含むサバイバル・レベルの日本語がなかばできる。

1-2 サバイバル・レベルの日本語ができる。やさしい言葉で言いかえたり、相手の助けを借りたりして、言いたいことはなんとか伝達できる。コミュニケーションへの参加はたいてい受動的。

レベル2 レベル1以外の諸種の日常生活上の社会的コミュニケーションができる。コミュニケーションに積極的に参加することができる。対面でなくてもコミュニケーションできる。会話の展開をコントロールすることができる。

レベル3 フォーマル・インフォーマル両場面で社会的コミュニケーションがおおむねできる。複雑な内容を表現し、伝達することができる。まとまった考え・意見を表明し、ディスカッションできる。得意分野において専門的な話ができ、またディスカッションできる。実質的に専門的でないあらゆる会話に参加できる。

レベル4 特殊なものを除いて、教養ある母国語話者が行なうほとんどすべての言語活動をこなすことができる。ニュアンスの違いが使い分けられる。

レベル5 教養ある母国語話者が行なうすべての言語活動が適切にできる。文法・語彙・発音・談話の構成・待遇表現・レジスターの使い分け・語彙量・イデオロム・口語表現・格言・挿話など、どの側面から見ても母国語話者と同じ言語活動ができる。

(ELEC 英語研修所, アメリカ・カナダ11大学連合日本研究センター,
東京学園高校)