

# 英語スピーキングテストにおける 英語母語話者と日本人の評価傾向に関する研究 —面接試験と英語教育等に関するアンケートの分析をもとに—

稲毛逸郎\*・ローン悦子\*\*・ローンマリー\*\*\*

(平成18年10月30日受理)

A Study of Native and Japanese Speakers of English

Grading Tendencies of Speaking Ability

—Based on the Analysis of Interview Evaluations and Background Questionnaire—

Itsuro INAGE\*, Etsuko LAWN\*\*, Murray LAWN\*\*\*

(Received, October 30, 2006)

## 1. はじめに

前稿稲毛・ローン(2005)では、スピーキングテストを実施している試験機関の多くは、評価者の評価方法の標準化に向けて様々な努力をしているにもかかわらず、検討すべき点が多いのが現状であることを指摘した。近年、実践的英語コミュニケーション能力の育成を目指した指導がより重視されるようになり、英語の授業の一環としてスピーチコンテストを行ったり、学校外で催されるコンテストへの参加を生徒に奨励したりすることも増えてきた。長崎県内でどの程度スピーキングテストを実際に学校内で行っているかについての調査(2006)<sup>1</sup>によると、スピーキングテストの実施が、中学校ではまだ十分に浸透しているとは言えないという状況が明らかになっている。しかし、それぞれの教育現場では、何らかのスピーキングテストの必要性を感じているようである。

英語のスピーチコンテストの規模は、授業中にクラスメイトの前で行うものから地域主催の大きなものまでさまざまであり、一般的にその評価方法は、生徒のパフォーマンスを複数の評価者(日本人、英語母語話者)がいくつかの観点をもとに採点し、その総合点から上位者を決めていくというものが多いためと思われる。また、英検では、受験者のパフォーマンスは一般的に1名(ただし1級のみ2名)の評価者によって採点される。試験会場により評価者が異なるのが普通で、事前に評価基準マニュアルを熟読していたとしても、多少の評価の個人差が出てくるのは当然である。前稿で触れたTOEIC LPI、ケンブリッジ

---

\* 長崎大学教育学部助教授・国際文化講座

\*\* 長崎大学大学院教育研究科教育専攻(英語教育専修)

\*\*\* 長崎純心大学人文学部英語情報学科講師

大学英語検定試験, IELTS のどの試験機関も, そのような個人差を極力縮小するための評価者のトレーニング, 資格化, 標準化に特に力を入れている。それは, スピーキングテストの評価をする場合, たとえ試験前に評価基準が明確にされていても, 誰が評価するかによって評価結果が違ってくることがあるからである。

## 2. 評価者に関する先行研究

Hadden (1991) は, 経験の違いが第2言語でのコミュニケーション能力の評価にどのように影響するかについて, 特に英語を教える ESL 教師と非教師<sup>2</sup>との間に差異が存在するかに注目し検証を行った。結果は, ESL 教師群と非教師群は, 何を重要と考えているかについて似た傾向を示すものの, 全く同一というわけではなく, 非教師群の方が ESL 教師群よりも全体的に寛容な評価をしていることが明らかになった。特に, 「言語能力」については, 非教師群の方が ESL 教師群よりも有意に高く評価しているという結果が得られた。

Nakamura (1992) は, 日本人英語学習者の口頭能力を評価する際に, 日本人英語教師 (JET) と英語を母語とする英語教師 (NET) とで評価の仕方によどのような差異が生じるかを 59 項目の質問から成るアンケートを用いて調査した。59 項目から成るアンケートの質問項目はすべて 5 段階評価で回答するようになっている。59 項目のうち, 11 項目がメインカテゴリーに関わる質問で, 残りの 48 項目がサブカテゴリーに関わる質問となっている。メインカテゴリーとは, 「文法の正確さ」, 「語彙」, 「音素」, 「イントネーション」, 「流暢さ」, 「談話」, 「内容」, 「話し手の自信」, 「社会言語的能力」, 「方略的能力」, 「発話内行為能力」を指し, サブカテゴリーとは, それぞれのメインカテゴリーの下位区分を示す。t 検定の結果, 11 の主カテゴリーの中では, 「流暢さ」と「談話能力」において日本人英語教師 (JET) と英語を母国語とする英語教師 (NET) の間で有意差があり, 日本人英語教師の方が, 英語を母国語とする教師より「流暢さ」を重視していた。また, 各項目重要度の判断においても JET と NET の間で差異が認められ, JET は「内容」「イントネーションなど」「語彙使用」の順であるのに対し, NET では「流暢さ」「談話能力」「内容」の順であった。

Katagiri (1999) では, 日本人英語教師と AET (Assistant English Teachers) の間で第2言語スピーキング評価の仕方に差異があるかを調査するために実験的試験が実施された。試験では, 受験者各1名ずつ, 試験会場で渡された日本語を英語に訳し, レポーターのように報道する作業をビデオの前で行うことが求められる。まず日本語に目を通す時間が5分間与えられ, 2分30秒以内に英訳していく。録画された回答は, 転写され, 100点の測定尺度で, 18種類の観点から評価された。結果は, 日本語教師, AET どちらも評価の際には「発話の速さ」を最重要視し, 次に「局所的誤り」<sup>3</sup>に注目する傾向があった。従ってこの実験では, 日本語教師と AET の間では, 顕著な評価の差異は明らかにならなかった。

Douglas (1994) は, 整った評価基準が設けられているスピーキングテストに関して, 全く別々の観点から同一の評価結果が下されることがありうるか, 逆に異なる評価結果でも質的に同様な内容のパフォーマンスが存在しうるかという2点から実験を行った。スピーキングテストは, 発音, 文法, 流暢さ, 理解力の4項目に渡って4段階の評価を行った。テストは後に転写され, 文法の間違い, 語彙の適切さ, 流暢さ, 内容, 修辭的な構成などについて質的な角度から分析された。結果として, 文法で高い段階評価を受けた被験

者でも、質的な分析によれば局所的な (local) 文法の間違がかえって多く認められるという例が見られた。また、図表の説明をするタスクでは、理解力で低い段階評価を受けた被験者の方が、より高い段階評価を受けた被験者よりも説明が丁寧に行われているという例が見られた。このことから、ここで行われた数値を使って表す主観的評価方法では、被験者の発話の質を必ずしも適切に評価できるとは限らないと結論づけた。

Powers et al. (1999) は、スピーキング能力をより正確に引き出すために改定された TSE テストの妥当性を検証した。そのアプローチを実証するために、主に英語を母国語とする大学生に TSE 受験者の見本試験解答を聞かせ、さまざまな反応、判断等のデータを集めた。受験者の TSE スコアは、TSE の正式な評価者によって事前に採点されており、TSE スコアの全てのレベルに及んでいる。テープに録音された受験者の TSE タスクに対する解答を、大学生がどの程度理解しているかを測定するために、Secondary Listening Test (SLT) を実施した。その目的は、TSE 受験者が伝えようとしたメッセージ (スピーキングの応答) を、リスナー (この場合、大学生の評価者) の能力で、どの程度正式なスコアを予測しうるものなのかを調査することであった。分析の結果、評価者の判断、反応、理解と TSE スコアレベルとの間に強い相関関係があることが明らかになった。

### 3. 調査の目的

本稿では、英語スピーキングテストを実施する際に、日本人英語教師 (JTE) と英語母語話者の英語教師 (NTE) の間で、また日本語教師間でも、中学英語教師、高校英語教師、大学英語教師の間で何らかの差が出てくるか、また教師経験年数などによりどのような評価の差が出てくるのかを分析することを目的とした。スピーキングテストを実施しているテスト機関が評価者の標準化に苦慮している現状をふまえ、また、中、高等学校、大学で実施されるような様々なスピーキングテストにおいて、どのような点に注意することで評価の標準化につながるかを考察することが本調査の主たる目的である。

本調査における調査課題は次の通りである。

- ① 日本人と英語母語話者の英語教師の間に評価の差が出るのか。  
もし差がある場合、どのような差なのか。差がない場合、どのような類似点が見られるのだろうか。
- ② 日本人、英語母語話者間に顕著な評価の差が出ない場合、アンケートの結果をもとに別の角度からみると、その結果どのような評価の差あるいは類似点があるのだろうか。またどの被験者の場合に評価の差が顕著に現れるのか。
- ③ 日本人の英語教師間 (中学教師、高校教師、大学教師) に評価の差が出るのか。  
もし差が出る場合、その要因は何だと考えられるのか。また、特にどのタイプの被験者に対して評価の差が顕著に現れるのか。

### 4. 方法

前節で設定した調査課題を解明するために実験的なスピーキングテストを大学生 15 名 (学部生 3～4 年生、大学院生 1 年生) に実施し、日本人英語教師 19 名 (大学教師 10 名、高校教師 6 名、中学教師 3 名) と英語母語話者の英語教師 8 名が評価者として参加した。面接官には、受験者とは面識が一切ない英語母語話者 1 名が担当した。面接官は、あるテ

スト機関の面接官の資格を有する者ではないが、面接官が面接をするにあたって注意する点を書かれた資料を事前に十分読んで試験に臨んだ。また、本実験的試験前にパイロットスタディーを実施し、その録画をもとに本試験内容の改善を図った。15名の受験者の英語力は、TOEIC400～800点程度である。試験の内容については、試験の事前に「面接試験の方法」カード(表1)を受験者に読んでもらった後、さらに面接官が試験の要領を直接受験者に説明した。試験は、受験者1名、面接官1名の1対1の形で行われた。まずウォームアップとして、挨拶と簡単な質問が2問行われ、その後、ロールプレイカード(表2)が渡され20秒間の黙読の後、その内容に従って、面接官から必要な情報を引き出し、面接官の質問に答える受験者中心の面接形式が取られた。本試験では、英検3～4級レベルにありがちな面接官による一方的質問形式はあえて避けた。試験時間は、英検準2級程度の約5～6分間にした。本試験は全て録画され、上記評価者は、ビデオに録画されているインタビューの発話をもとに、後述するように4項目を5段階に分け判定した(表3/スピーキングテスト評価基準表使用)。本試験では、SST(Standard Speaking Test)の評価基準を使用した。SSTとは、英語のスピーキング能力を判定するインタビュー形式のテストで、日本の英語教育事情を考慮した上で、米国の全米外国語教育協会(American Council on the teaching of Foreign Language; ACTFL)が中心となり開発したOral Proficiency Interview(OPI)を基にして、ACTFLとアルクが共同開発したテストである。

評価基準(表3)は大きく4つの項目から成り(①総合的タスク・機能、②テキストの型、③話題、状況、④正確さ)、特定の項目を重視するのではなく、発話を包括的に評価することが重視されている。本試験では、4つの項目を柱に置き5段階(1. Poor/劣る、2. Fair/やや劣る、3. Good/良い、4. Very good/大変良い、5. Excellent/非常に良い)で評価された(表4/スピーキングテスト評価表使用)。

また同時に、評価者は、上記のスピーキングテストの評価後、英語教育等に関するアンケートに回答した。日本人英語教師には、「英語習得、英語教育状況またはネイティブスピーカー(英語母語話者)との経験に関するアンケート」9問(表7)に対する回答を求め、また英語母語話者の英語教師には、「英語教育状況と英語を第二言語とする人との経験に関するアンケート」8問(表8)に対する回答を求めた。質問項目は、4つの選択肢から成るが、項目の中には、答えが1つとは限らないものも含まれた。

表1： 面接試験の方法

試験は初めに面接委員と簡単な自由会話を行います。次に<Role Play Card>が手渡されます(黙読時間は20秒程度)。その内容に従って、面接委員から必要な情報を引き出し、質問に答えていきます。例えばRole Playでは、面接官と受験者は、先生と生徒、店員とお客といったような役を演じながら、主に受験者がリードして会話を進めていきます。最後に<Role Play Card>を面接委員にもどして終了となります(約5分間)。

表2： Role play card

You are a foreign tourist in Auckland, New Zealand. You want to visit Rotorua, so you go to see a travel agent (your examiner is going to be the travel agent). After you have explained the situation,

- ① Ask him how to get to Rotorua.
- ② Ask about the price, traveling time, comfort etc., and ask his opinion.
- ③ Decide how you will travel and explain why.

表3：スピーキングテスト評価基準（4項目）

<p>1. 総合的タスク・機能 (Global Function) 言語を使って何ができるか &lt;上級レベル&gt; 問題解決をする。正確に時制を使う。詳細な描写をする。 &lt;中級レベル&gt; 質問し、答える。簡単なやりとりをする。自分の言葉で表現する。 &lt;初級レベル&gt; 丸覚えの表現を使う。モノの列挙をする。</p>	<p>2. テキストの型 (Text Type) どんな構文や構成を使うことができるか &lt;上級レベル&gt; 語や句、単文、複雑な文、段落 &lt;中級レベル&gt; 語や句、単文、複雑な文 &lt;初級レベル&gt; 語や句</p>
<p>3. 話題・状況 (Context / Content Area) どのような状況で、何について話すことができるか &lt;上級レベル&gt; ほとんどのインフォーマルな状況といくつかのフォーマルな状況。一般的な話題、自分自身のことと身近な事柄について話せる。 &lt;中級レベル&gt; インフォーマルな状況 自分自身のこと、身近な事柄について話せる。 &lt;初級レベル&gt; 非常に限られた日常生活の場面で 身近な事柄について断片的に話せる。</p>	<p>4. 正確さ (Accuracy) どれだけ発話をきちんと聞き、相手に伝えることができるか ① 文法の正確さ ② 語彙 ③ 発音 ④ 流暢さ ⑤ 社会言語学上の適切さ</p>

表4：スピーキングテスト評価表

評価者氏名 \_\_\_\_\_

以下の4分野別評価を目安にし、当てはまる番号を○で囲んで下さい

1. Poor (劣る)	2. Fair (やや劣る)	3. Good (良い)	4. Very good (大変良い)	5. Excellent (非常に良い)
-----------------	-------------------	-----------------	------------------------	-------------------------

1. 受験者名 \_\_\_\_\_

分野					
1. 総合的タスク・機能 (言語を使って何ができるか)	1	2	3	4	5
2. テキストタイプ (どんな構文や構成を使うことができるか)	1	2	3	4	5
3. 話題・状況 (どのような状況で、何について話すことができるか)	1	2	3	4	5
4. 正確さ (どれだけ発話をきちんと聞き、相手に伝えることができるか)	1	2	3	4	5
Total Score _____					

## 5. データ分析

まず初めに、日本人英語教師 (JTE) と英語母語話者の英語教師 (NTE) 間の評価平均値と標準偏差値を比較した (図1)。JTE の評価平均値は 10.4, NTE は、11.7 であり、NTE の方が JTE よりやや寛大な評価をする傾向を示している。また、評価値の標準偏差値は、JTE が 2.2, NTE が 2.4 であり、NTE の評価値の変動が JTE よりもやや大きいという結果が出ている (表5)。

次にやや寛大な評価値が出た NTE と JTE の評価値の差 (NTE 評価値—JTE 評価値 / 図 2) を見てみると、被験者 5 番と 13 番には 2.5 ポイントもの差があり、被験者 8 番、9 番、12 番は比較的近接した結果が出ている。被験者 5 番、13 番に関し NTE と JTE の 4 分野別評価項目の差を比較してみた。被験者 5 番の場合、項目 1 の総合的タスク・機能に関し NTE は JTE に比べより低く評価している。項目 2 と 3 は NTE, JTE どちらもほぼ同じ評価をしている。しかし項目 4 の正確さにおいて NTE は JTE より高い評価をしている。被験者 5 番、13 番の試験中の会話を転写し、会話の流れ、正確さの詳細を確認した。被験者 5 番は、“plane / train” の発音の違いが聞き取れずロールプレイ中「ロトルアまでの交通機関を選ぶ」タスクの決断まで至らなかったが、他の 14 人の被験者に比べ、全体的に丁寧な表現 (more pleasantly spoken) を使用していた印象を受ける。一方被験者 13 番の場合、全ての項目において、NTE は JTE に比べ高く評価していた。被験者 13 番の転写された会話を分析すると、文法的誤り、また発音の不正確さが目立つが、聞き取れない場面では何度も尋ね直すなど全体的に自信を持って会話を確実に進めていった印象を受ける。以上の結果から、被験者 5 番に関しては、NTE は被験者がタスクを達成できなかったことで評価項目 1 の評価が JTE より低くなっていると思われるが、項目 4 の正確さにおいては、NTE が JTE より高い評価をしたのは、表現の丁寧さが関係しているのではないかと考えられる。一方被験者 13 番の場合は、発音、文法の不正確さは目立つが、自信を持って確実に会話を進めていったことが NTE により良い印象をもたらした要因ではないかと考えられる。全体的に見て NTE は JTE に比べ内容、丁寧さ、会話の進め方を重要視する傾向にあるが、一方 JTE は文法、発音の誤りが多い被験者に対しやや厳しい評価をする傾向にあるようである。

表 5: JTE と NTE 評価平均値と評価値の標準偏差値

	評価平均値	標準偏差値
日本人英語教師 (JTE)	10.4	2.2
英語母語話者英語教師 (NTE)	10.7	2.4

本スピーキングテスト評価終了後のアンケートによる調査結果は表 7, 表 8 の通りである。このアンケートは、本スピーキングテストの評価結果と各々の英語教師の経験内容とがどのような関係があるかを調査するために実施された。日本人英語教師には、英語習得状況 (ネイティブから英語を学んだ学習状況) の質問を加えた。JTE と NTE のどちらにも第 2 言語話者 (JTE にとっての第 2 言語話者は、本調査では英語母語話者であり、NTE にとっての第 2 言語話者は、英語母語話者以外の日本人を含む話者である) との日常的な接触状況について尋ねた。結果は、表 7, 表 8 の示す通りであるが、回答者数は、日本人英語教師が 19 名 (大学教師 10 名, 高校教師 6 名, 中学教師 3 名) と英語母語話者の英語教師が 8 名であった。アンケート (5) の項目は、JTE, NET どちらも回答が複数の回答になり得る点を注記しておく。アンケート (1) の英語教育経験に関する質問については、本調査の評価者である JTE の場合は、1 年未満から 10 年以上全ての範囲の教師が含まれていたが、NTE 8 人に関しては、全員 10 年以上の英語教育経験があり、その点に関しては比較が出来なかった。

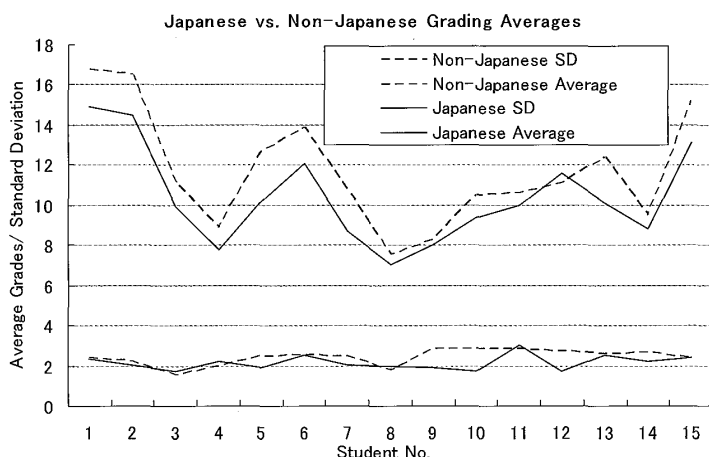


図 1：日本人英語教師 (JTE) と英語母語話者英語教師 (NTE) との評価平均値及び標準偏差値の比較

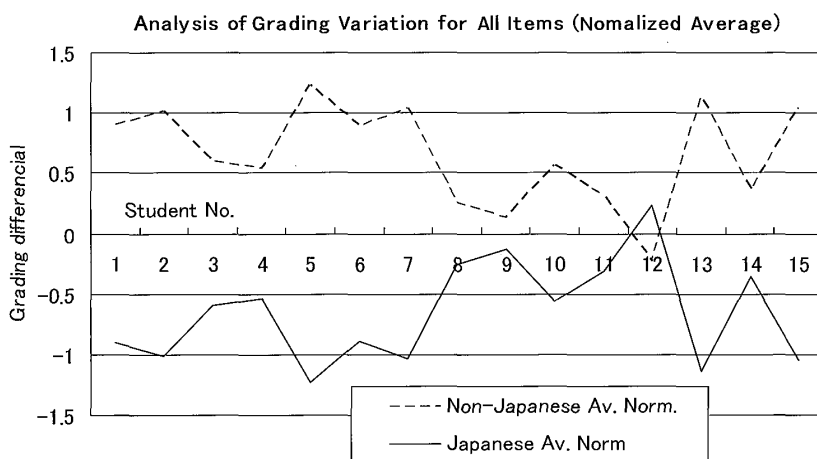


図 2：JTE と NTE の評価値の差の比較

このアンケートの目的は、NTE の評価平均値に対するそれぞれの評価者の評価がアンケートの各項目とどのような関係があるかを見出すことにある。そのために、まずアンケート各項目①～④の① = 相関係数 0, ④ = 相関係数 1 として仮に基準を置き、それに対する相関関係 (データの一貫性) を調査した。それに加え相関関係の強さを調査するために NTE 評価平均値に対する各項目レベル別 (①～④) 傾向 (標準偏差値) も出した (表 6)。NTE 評価平均値に対する各項目別傾向は非常に低く、最も高い傾向値でもわずか 16% である。

次に、JTE の英語教育に携わった年数別評価平均値<sup>4</sup>と NTE の評価平均値とを比較してみた。表 7 の日本人英語教師 (評価者) の英語習得、英語教育状況またはネイティブスピーカー (英語母話者) との経験等に関するアンケート表によると、NTE 評価平均値に対する英語教育年数別評価平均値の相関係数は、どの教育年数別評価値を取っても相関係数がほぼ 0.8 と極めて相関関係が高いことを示している。

その中でも 1 年未満の教師群<sup>5</sup>が 0.698, 1 年以上 5 年未満の教師群の相関係数が 0.798, 5 年以上 10 年未満の教師群が 0.813, 10 年以上の教師群が 0.78 であった。NTE 評価平均値に対する全体的な教育年数別評価平均値の相関係数は 0.655 であり、教育年数別の相関係数の差は非常に少ないものの、全体的に見ると教育年数が長くなればなるほど NTE の

評価平均値により近づく傾向を示している。

表6：NTE 評価平均値に対するアンケートの項目別データの一貫性と各項目のレベル別 (①～④) 傾向のJTE とNTE 間の比較

項目	JTE		NTE	
	データの一貫性 (全体的相関係数)	項目レベル別傾向 (標準偏差値)	データの一貫性 (全体的相関係数)	項目レベル別傾向 (標準偏差値)
1	66%	10%	-	-
2	-37%	6%	-	-
3	79%	6%	-85%	16%
4	-53%	6%	-	-
5	-	-	-	-
6	96%	8%	10%	14%
7	79%	10%	68%	6%
8	-100%	16%	30%	14%
9	-	-	N/A	N/A

注：データの一貫性 (相関係数から割り出し、係数値は-1～+1) と項目レベル別傾向 (標準偏差値から割り出し、値は0～-0.5) の比較値が分かりやすいように全体的相関係数 x 100, 標準偏差値 x 200 としてパーセントの値を算出した。尚、データ一貫性の欄にあるマイナス (-) 記号は、仮に立てた基準に対して反対の方向を表していることを示している。また、項目レベル別傾向のマイナス (-) 記号は、値が極わずかで表すに値しないもの。N/A(Non applicable) は、NTE のアンケートにはない質問 (項目9) を意味する。

表7：日本人英語教師 (評価者) の英語習得、英語教育状況またはネイティブスピーカー (英語母話者) との経験等に関するアンケート

経験の種類	回答者数	回答者数率	NTE 評価平均値に対する相関係数 (r)
(1) 英語教育に携わって何年になりますか？			
① 1年未満	1	5.3%	0.698
② 1年以上5年未満	6	31.6%	0.798
③ 5年以上10年未満	4	21.1%	0.813
④ 10年以上	8	42.1%	0.78
相関係数 (全体的傾向)			0.655
標準偏差値			0.051
(2) 英語圏の滞在経験はどのくらいですか？			
① ない	5	26.3%	0.76
② 1年未満	9	47.4%	0.799
③ 1年以上5年未満	3	15.8%	0.733
④ 5年以上	2	10.5%	0.757
相関係数 (全体的傾向)			-0.37
標準偏差値			0.027
(3) ネイティブスピーカー (英語母話者) と会う機会はふつう週に何回ありますか (頻度)？			
① ない	2	10.5%	0.738
② 1～2回	7	36.8%	0.799
③ 3～5回	6	31.6%	0.774
④ 5回以上	4	21.1%	0.814
相関係数 (全体的傾向)			0.793
標準偏差値			0.033
(4) ネイティブスピーカーに会った時、平均的に英語で話す時間はどのくらいですか？			
① 1分以内	3	15.8%	0.843
② 1～10分	11	57.9%	0.776
③ 10～60分	5	26.3%	0.807
④ 60分以上	0	0%	-
相関係数 (全体的傾向)			-0.53
標準偏差値			0.034



(5) そのネイティブスピーカーとはどのような関係ですか？ (ネイティブに会うと答えた教師のみ回答)			
① 友人	1	5.3%	-
② 職場の人	18	94.7%	-
③ サービス業の人 (お店)	0	0%	-
④ 生徒	0	0%	-
(6) ネイティブスピーカー (英語母語話者) に英語を学んだ期間はどのくらいですか？			
① ない	2	10.5%	0.722
② 1年未満	4	21.1%	0.779
③ 1年以上5年未満	8	42.1%	0.789
④ 5年以上	5	26.3%	0.82
相関係数 (全体的傾向)			0.96
標準偏差値			0.041
(7) 英語圏あるいは日本でネイティブスピーカーと会話をする上で不快な経験を受けたことがありますか？			
① ない	9	47.4%	0.695
② 1回ある	1	5.3%	0.790
③ 2～5回	6	31.6%	0.781
④ 5回以上	2	10.5%	0.795
相関係数 (全体的傾向)			0.789
標準偏差値			0.047
(8) 日本語を母国語としない人 (英語母語話者を含め) から、英語以外の科目を日本語で学んだことがありますか？			
① ない	14	73.7%	0.814
② 1～2回	3	15.8%	0.75
③ 3～5回	0	0%	-
④ 5回以上	2	10.5%	0.66
相関係数 (全体的傾向)			-1
標準偏差値			0.078
(9) その教師の日本語は分かり易かったですか (あると答えた人のみ回答) ?			
① 分かり易かった	6	31.6%	-
② 少し分かり辛かった	0	0%	-
③ かなり分かり辛かった	0	0%	-
④ ほとんど分からなかった	0	0%	-

表 8：英語母語話者教師の英語教育状況と第 2 言語話者との経験等に関するアンケート  
Background Questionnaire on Examiners' Teaching Experience and Understanding of Non-native speakers

Type of experience	Number of responses	Responses as a percentage	Cor. to NTEs' grading average
(1) How long have you been teaching English			
① Less than 1 year	0	0%	-
② 1-5 years	0	0%	-
③ 5-10 years	0	0%	-
④ More than 10 years	8	100%	-
(2) How long have you been living in Japan?			
① Less than 1 year	0	0%	-
② 1-5 years	0	0%	-
③ 5-10 years	1	12.5%	-
④ More than 10 years	7	87.5%	-
(3) Have you ever taught English outside of Japan? If so, for how long?			
① None	4	50%	0.878
② Less than 1 year	2	25%	0.875
③ 1 - 5 years	1	12.5%	0.729
④ More than 5 years	1	12.5%	0.754
相関係数 (全体的傾向)			-0.85

標準偏差値			0.079
(4) How much time do you speak with non-native speakers of English (in English) per week?			
① Less than a minute	0	0%	-
② 1 - 10 minutes	0	0%	-
③ 10 - 60 minutes	0	0%	-
④ More than 1 hour	8	100%	-
(5) What's your relationship with the non-native speakers referred to in the above question?			
① friends/social acquaintances	7	87.5%	-
② Colleagues/business acquaintances	5	62.5%	-
③ shop attendants	3	37.5%	-
④ students	7	87.5%	-
(6) How many courses have you learned from non-native speakers (teachers) of English (in English)?			
① None	3	37.5%	0.832
② 1-2 times	2	25%	0.879
③ 3-5 times	1	12.5%	0.729
④ More than 5 times	2	25%	0.881
相関係数 (全体的傾向)			-0.01
標準偏差値			0.071
(7) Did you ever find it difficult to understand the above mentioned non-native teachers' English?			
① Not really	2	25%	0.819
② a little difficult	1	12.5%	0.849
③ fairly difficult	1	12.5%	0.901
④ very difficult	1	12.5%	0.861
相関係数 (全体的傾向)			0.68
標準偏差値			0.034
(8) Have you ever had an unpleasant experience with a non-native speaker of English (including Japanese)?			
① None	1	12.5%	0.754
② 1-2 times	2	20%	0.913
③ 3-5 times	1	12.5%	0.824
④ More than 5 times	4	50%	0.835
相関係数 (全体的傾向)			0.304
標準偏差値			0.065

図3によると4つの全体的なグラフの傾き具合は、前半はほぼ同じであるが、被験者10番以降の評価には多少のばらつきがあることが示されている。さらに分かりやすくするために、図4を作成した。図4は、NTEの評価の平均値をゼロにし、英語教育年数別評価平均値のばらつきをNTE評価平均値に対し分析したものである。表9のNTE評価(平均値をゼロとした)平均値と5年未満の教師群の平均評価の差は、-0.9ポイント、5年以上10年未満の教師群が-2.1ポイント、10年以上の教師群は-1.5であった。この結果から言えることは、どの教師群もNTE教師群と同じような傾向の評価をするものの、被験者1～15番を通し、被験者9番と12番以外は一貫して5年未満の教師群の評価値はNTE平均値よりやや厳しい評価をする傾向にあり、10年以上の教師群が次にさらに厳しい評価をし、5年以上10年未満の教師群が3つの教師群の中で最も厳しい評価をしたと分析できる。テストの後半部分の被験者8～15番に関しては、5年以上10年未満の教師群は10年以上の教師群よりNTEの評価平均値により近い評価を示している。10年以上の教師群の場合は、5年未満の教師群と逆の傾向がありNTEとほぼ同じような寛容な評価で始まり、後半部では評価にややばらつきが出た。

次に、NTE評価平均値に対するJTEの評価平均値を学校レベル別(中学、高校、大学)に比較してみた。図5が学校レベル別の全体的な評価平均値の比較であるが、NTE

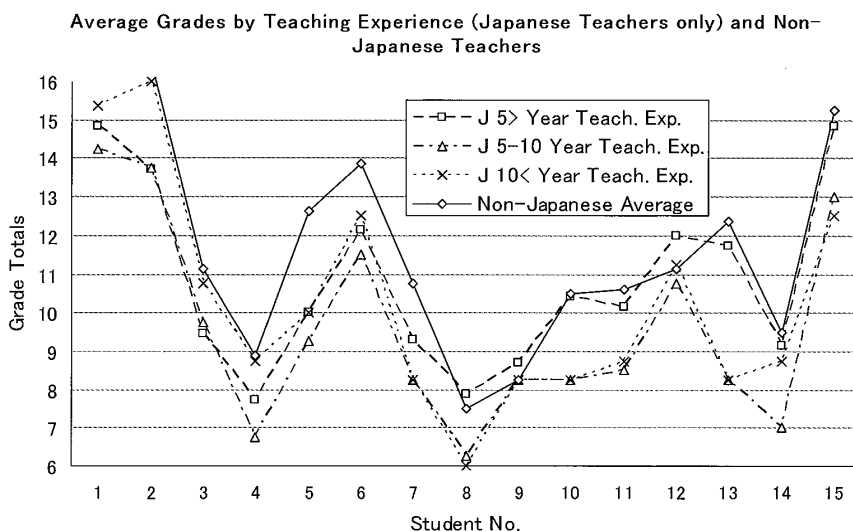


図 3：英語教育年数の異なる評価者の平均値と英語母語話者の評価値との比較

の評価は被験者 1～15 番を通し一貫して JTE の評価よりやや高めの評価をする傾向にある。しかし被験者 9 番と 12 番に関しては日本人教師群の方が NTE 教師群よりやや高い評価をしている。図 6 は、英語教育経験別の比較同様、NTE の評価平均値をゼロ（中央値 / 標準値）にし、学校レベル別に評価平均値を比較したものである。大学教師群は、図 5 の NTE の評価平均値に対する日本人英語教師学校レベル（中学、高校、大学）の評価平均値の比較によると、被験者 13 番と 15 番以外は中学教師群と高校教師群に比べ極めて NTE 評価平均値に近い評価をしている。表 10 の NTE 評価平均値と日本人英語教師学校レベル（中学、高校、大学）別評価平均値との差を見ても、大学教師群の平均値は、学校レベル別教師群の中で NTE の平均値に最も近い -1.1 ポイントである。

表 9：NTE の平均値（中央値 / ゼロ）に対する英語教育年数別評価値の差

教育年数別教師群	評価平均値
教育年数 5 年未満の教師群	-0.9
5 年以上 10 年未満の教師群	-2.1
10 年以上の教師群	-1.5

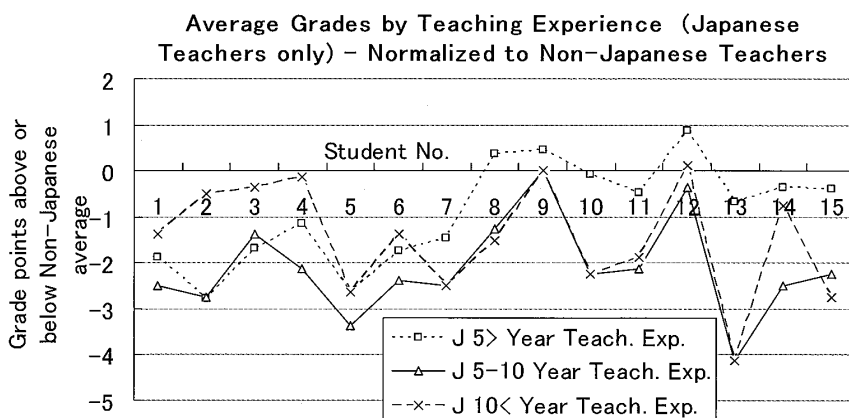


図 4：英語教育年数の異なる評価者による評価平均値と NTE の平均値（平均値をゼロに設定）との比較

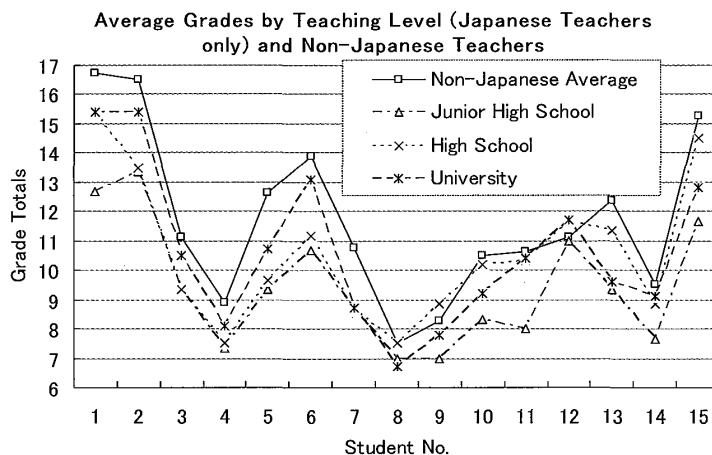


図5: NTE の評価平均値と日本人英語教師学校レベル (中学, 高校, 大学) 別の評価平均値との比較

表 11 の NTE 評価平均値に対する日本人英語教師学校レベル (中学, 高校, 大学) 別の評価平均値の相関関数の結果では, どの学校レベルの教師群も相関係数が 0.7 以上と NTE 教師群と極めて高い相関関係を示している。特に大学教師群はその中でも最も高く 0.819, 次に高校教師群の 0.768, 最後に中学教師群の 0.724 という結果が出ている。それぞれの相関係数の変動は小さく (0.1), 全体的なデータの一貫性は 0.999 と極めて一貫しており, 学校レベル別傾向は 0.048 とレベルが上がるにつれて NTE の評価平均値に近い評価をする傾向を示している。図 6 によると高校教師群の場合は, 後半部分でかなり NTE 評価平均値に近い評価を示している。一方, 中学校教師群は, 被験者 8 番と 12 番に関しては NTE と全く同じ評価をしているものの, 全体的には大学教師群, 高校教師群に比べやや厳しい評価をしている。これは, 表 10 の NTE 教師群との平均値の差が -2.29 ポイントという値にも表れている。特に, 4 つの教師群 (NTE, JTE 中学, JTE 高校, JTE 大学) の中で被験者 5 番に関しては, 全ての JTE 教師群が NTE 教師群より 2 ポイント以上低い評価をしている。また被験者 1 番, 6 番, 13 番, 15 番に関しては, 1 つあるいは 2 つの教師群が NTE の評価よりかなり低めに評価している。被験者 8 番と 12 番に関しては, NTE と JET の評価が逆転し, 1 つの JTE 教師群の平均値が NTE 教師群平均値よりやや高めに表れている。これは, NTE の教師群の中に極端に低い評価をした者がいるからである。理由として考えられるのが, 被験者 8 番の場合, 文法的な誤りは他の被験者と比較してそれほど多い方ではないが, 返答 (応答) するまでの時間が非常に長かったため一部の NTE にややネガティブな印象を与えた可能性がある。被験者 12 番の場合も文法的な誤りに関しては他の被験者と大きな違いはないが, イントネーションが平板であることが特に目立っていたので, 一部の NTE に全体的な会話がやや不自然である印象を与えた可能性があると考えられる。

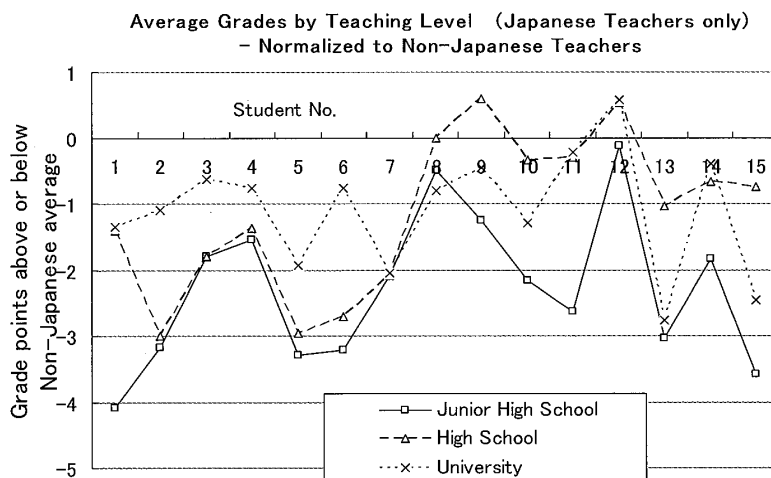


図 6：NTE の評価平均値に対する日本人英語教師学校レベル ( 中学, 高校, 大学 ) 別評価平均値との比較

表 10：NTE 評価平均値と日本人英語教師学校レベル ( 中学, 高校, 大学 ) 別評価平均値との差

学校レベル	評価平均値の差
日本人中学英語教師	-2.29
日本人高校英語教師	-1.15
日本人大学英語教師	-1.1

表 11：NTE 評価平均値に対する日本人英語教師学校レベル ( 中学, 高校, 大学 ) 別の評価平均値の相関係数<sup>6</sup>及び標準偏差値

相関関係	係数
NTE と日本人中学英語教師	0.724
NTE と日本人高校英語教師	0.768
NTE と日本人大学英語教師	0.819
上記の全体的なデータの一貫性	0.999
学校レベル別傾向 (標準偏差値)	0.048

## 6. 考察

前節で示された3つの分析の結果によると、調査課題項目①のNTEとJETの全体的な評価の比較に関しては、多少の評価のばらつきはあるものの両者が非常に高い相関関係を示していることが判る。全体的にはNTEの評価の方がJETの評価よりやや高い(寛大な)傾向があった。

他の2つの分析では、まず調査課題②に注目し英語母語話者の持つ母国語に対する絶対的な感覚を基準にして(NTEを評価の基準にし)、日本人英語教師の教育年数別(①5年未満, ②5年以上10年未満, ③10年以上)評価平均値とまた学校レベル別(中学, 高校, 大学)評価平均値とを比較した。その結果、どの教育年数別も相関係数が約0.8でNTE評価平均値と極めて高い相関関係を示していることが判った。全体的に見ると、ごくわずかではあるが教育年数が長くなるにつれNTE評価平均値との相関関係が高くなる傾向があるという結果が出た。

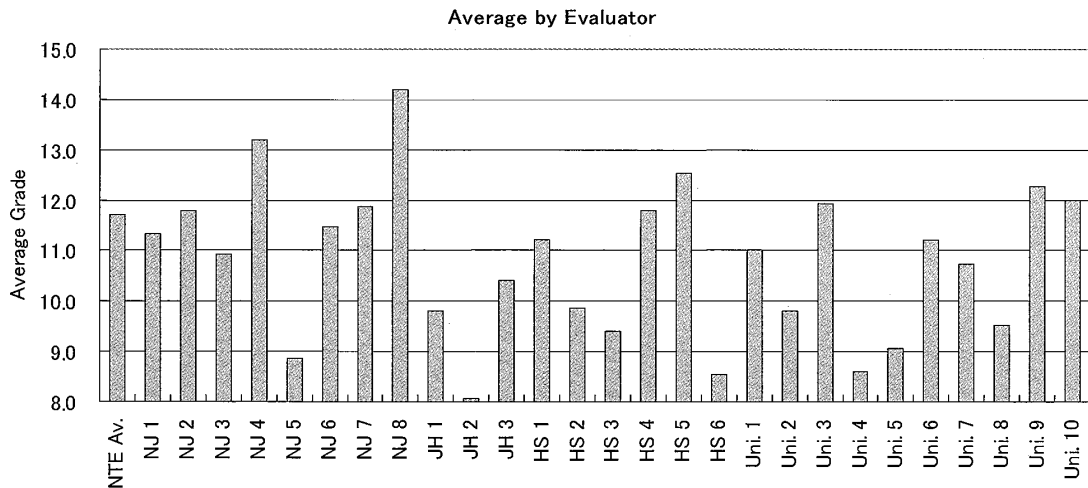


図7 評価者別評価平均値の比較 (個人差)

注：NTE Ave は英語母語話者評価平均値，NJ1 は英語母語話者1番 (Non-Japanese)，JH1 は日本人中学英語教師1番，HS1 は日本人高校英語教師1番，Uni1 は日本人大学英語教師1番を意味する。

調査評価課題③の学校レベル別評価平均値に関しても、どの学校レベルの教師群も相関係数が0.7以上でNTE評価平均値と高い相関関係を示している。NTE評価平均値との相関関係は、前述の教育年数別教師群の結果同様わずかな差ではあるが、大学教師群が最も高く0.819、次に高校教師群が0.768、最後に中学教師群が0.724の順である。それぞれの相関係数の変動は非常に小さいが学校段階が上がるに連れ、NTEの評価により近い傾向を示している。NTE評価平均値との差に関して言えば、大学レベルの教師群の差が最も小さく、次にごくわずかな差であるが高校教師群、最後に中学校教師群と続く。大学レベル教師群がNTE評価平均値により近い評価の傾向にある要因の1つに、NTE教師群のほぼ全員が同じく大学で英語を教えており、英語中学教師群、高校教師群に比べ文法中心の受験英語からよりコミュニケーション型な英語教育に携わっているからではないかと考えられる。次に高校、中学と続く理由には、今回の被験者が大学生であるため、各々の教師が現在教えている生徒のレベルを基準に評価をした可能性があるのではないかと考えられる。特に今回評価の差が中学校の教師群が最も低いのは、中学校教師群の中で極端に低い評価をした者がいたことが要因の1つにあることを注記しておきたい。

本実験の結果から、全体的に見てNETとJTE教師群の評価平均値の相関関係が非常に高いと言える。前述のように、英語教育年数が長くなればなるほどごくわずかな差ではあるがNTEの評価平均値に近づくこと、またJTEが勤務する学校段階が上昇するほどわずかな差ではあるがNTE評価平均値に近づく傾向を示している。アンケートの結果、教育年数、学校レベル別評価平均値と同様、ごくわずかな差ではあるがNTE評価平均値と相関関係がある項目がいくつか見つかっている(表6)。ここで注目すべきことは、図7を見て判るように、評価の差は主に個人差が最も大きいという点である。このことから言えるように、スピーキングテストを実施する際には、事前にいかに慎重に評価の個人差を縮小するかが、評価の客観性に大きく影響すると考えられる。そのためには、言い換えれば、評価

の観点の理解、及びそれらの観点をより客観的に評価するための十分なトレーニングがいかに必要であるかを示唆していると言えよう。英語母語話者の評価者の比較・分析については稿を改めることにする。

#### 注

1. 園田裕 (2006) は、長崎県内 70 の中学校でアンケートを取り、英語スピーキングテストがどの程度実施されているかを調査した。その結果、必要性は感じているが半分以上の中学校でテストが実施されていない現状が明らかになった。
2. 非教師とは、ここでは英語を外国人に教える ESL 教師に対し教師ではない一般職についている者を表わすものとする。
3. 「局所的誤り (local errors per 100 words)」とは単語 100 語あたりの文法的誤りを言う。
4. 年数別評価平均値は、個人の評価をより正確にする (個人差を縮小する) ために、まず教師ごとに NTE に対する評価平均値の相関係数を出し、そのトータルを教師数で割り、ここで使用している教育年数別評価平均値の相関係数 (表 7) を割り出している。
5. 図 3 は英語教育年数別 (① 5 年未満, ② 5 年以上 10 年未満, ③ 10 年以上) 評価者の平均値と NTE の評価平均値を比較したグラフである。アンケートでは、教育年数を① 1 年未満, ② 1 年以上 5 年未満, ③ 5 年以上 10 年未満, ④ 10 年以上と 4 つに区分しているが、アンケートの結果 1 年未満の教師が 1 人であるため、図 3 では①と②をまとめて① 5 年未満として比較した。
6. 表 11 の相関係数も注 4 と同様に、まず教師ごとに NTE に対する評価平均値の相関係数を出し、そのトータルを教師数で割り (平均値を出し)、学校レベル別評価平均値の相関係数を割り出している。

#### 参考文献

- 馬場哲生. 2003. 『英語スピーキング論』 河源社：東京.
- Powers, D., Schedl, M., and Wilson, S. 1999. "Validating the revised Test of Spoken English against a criterion of communicative success." *Language Testing*. 16 (4), pp. 399-425.
- Douglas, D. 1994. "Quality and quantity in speaking test performance." *Language Testing*. 11, pp. 125-144.
- Hadden, B.L. 1991. "Teacher and nonteacher perceptions of second-language communication." *Language Learning*. 41, 1, pp. 1-24.
- 金屋憲. 2003. 『英語教育評価論』 河源社：東京.
- Katagiri, K. 1999. "Evaluation of L2 Speaking JTEs versus AETs" *ARELE (Annual Review of English Language Education in Japan: 全国英語教育学会紀要)*. vol.10, pp. 93-102.
- Nakamura, Y. 1992. "Differences in native and non native teachers' evaluation of Japanese students' English speaking ability." *Cross Currents*, 19, 2, pp. 161-165.
- Underhill, N. 1994. *Testing Spoken Language*. Cambridge, England: Cambridge University Press.