

論文

英語スピーチコンテストのための 分析的評価に代わる全体的評価の実用の可能性

福田 昇¹

¹ 一般教育科—英語 (Liberal Arts-English, National Institute of Technology, Nagaoka College)

The Possibility of Implementing a Holistic Evaluation Instead of an Analytic Evaluation for the English Speech Contest

Noboru FUKUDA¹

Abstract

This research explored the possibility of implementing a holistic evaluation instead of an analytic evaluation for the English speech contest. Ten speakers were evaluated by their scores based on the sub-rank method. The participants of the two different evaluation systems were 52 incumbent English teachers grouped into three kinds of English abilities: 17 Japanese English teachers who were second level for the Society for Testing English Proficiency(STEP), 20 Japanese English teachers who were more than pre-first level for STEP, and 15 native English speakers who were Assistant Language Teachers(ALTs). This research clarified three results: (a)there were no significant differences between the holistic evaluation and the analytic evaluation relating to the score of ten speakers. However, it revealed a significant difference concerning the score of two speakers between the 2nd level teachers for STEP and the ALTs; (b)the holistic evaluation significantly reduced the evaluating time compared to the analytic evaluation; (c)it was also apparent that the participants never thought that holistic evaluation was easier than analytic evaluation.

Key Words : *holistic evaluation, analytic evaluation, speech contest, sub-rank method*

1. はじめに

スピーチの発表者は時間をかけて原稿を準備し、繰り返して練習をして聴衆の前に立ってスピーチを伝える。そこに至るまでのプロセスにはさまざまな苦労があることを現場で教育に携わる英語教師は知っている。しかしながら、スピーチコンテストの審査方法となると大会の規模が大きければそれなりの審査員が選出されて審査の評価を行うが、校内大会や市大会レベルでは得点の集計方法や評価方法も明

確でなく審査が行われているところもある。また、審査を日本人英語教師が一切関与せずに、英語母語話者(ALT)に一任している学校もある。あるいは、逆に日本人英語教師だけで校内大会をしている学校もある。英語スピーチコンテストは文科省が述べている、実用性のある英語力を生徒に身につけさせるための大きな力となるものである。審査はそもそも主観的であるが、審査を出来る限り公的評価であろうとすれば、その主観的審査にも基準が必要である。また評価方法や評価集計方法の相違が実際にど

のような相違があるのかを中学校の現場で用いられているデータの集計方法に基づいて検証することも必要である。

本研究では、評価方法として、大きな大会で採用されている「分析的評価」と、校内大会や市大会で比較的用いている「全体的評価」の相違を検証する。また、評価集計方法と評価集計方法による相違を調査分析することで、スピーチコンテストのための、より公正な審査方法を検証し、実用性のある評価方法と集計方法を大会運営で用いていくことが可能となると考える。

2. 先行研究

2. 1 全体的評価と分析的評価の相違

「分析的評価」とは、口頭能力がいくつかの要素から構成されていると考え、能力を要素ごとに分析して評価する方法である。それに対して、「全体的評価」では、能力を全体として1つの総合体としてみなす立場をとる。評価対象を複数の項目に分割することなく、全体に対して1つの「印象点」を与えるという方式をとる。スピーチコンテストの審査においては、前者の分析的評価が一般的である。つまり、英語の発音やリズム、内容、デリバリーなどの複数の評価項目の得点をし、それらの総和をスピーチの最終評価とする方式である。野村¹⁾は、スピーチコンテストでこの方式が広く採用されているのは、評価項目が細かく分かれているため、コンテストの審査結果に説得力を持たせることができ、個々のスピーチの良い点や悪い点を具体的に指摘できるため、判断の材料を提供できることであると述べている。しかしながら、大友²⁾が言うように分析的評価にはこういったメリットがあるが分析的評価は複雑であり、スピーチコンテストの評価では、評価の信頼性と評価方法が「測定しようとしていることを本当に測定しているかどうかという度合い」を示す妥当性を得ることが難しい。

実際、校内大会や市大会レベルの審査では現実的にはそのような時間を十分に設定することは不可能に近い状況にある。このため、スピーチコンテストの評価方法の開発は、英語の口頭能力を育成するためにも意義のあることと考える。なお、今回の研究は、スピーチコンテストに限定している。その背景として、暗唱コンテストは、参考となるテキストとデモテープがあり、発音、アクセント、リズムなどの判断基準が明確であり、同じ内容を発表することで、その目的基準に対しての相違点が際立ってくる。

それに対して、スピーチコンテストの評価は、論理の展開も、個人的な体験も異なる複数のスピーチを比較してその優劣を比較して判定しなければならないという側面を持っている。その意味で、両者を一括して論じることは無理である。

2. 2 英語を母語とする教師 (NET) と日本人英語教師 (JET) による評価の相違

Nakamura³⁾は日本人英語学習者の口頭能力評価を76名のNETとJETの大学英語教師に行った。32名はJETであり、44名はNETであった。アンケート項目は59項目からなり、全ての項目は5段階評価で回答するものであった。重要度の順位でJETは「内容」「イントネーション」「語彙」であったが、NETは「流暢さ」「談話能力」「内容」の順であった。Anderson-Hsieh et al.⁴⁾は、母語話者が非母語話者の発音に対する評価研究をした。彼らは発音を「分節」「音節構造」「韻律」の3つの構成要素に分け、母語話者の発音評価を調査した。その結果、発音評価と最も相関関係が深かったのは韻律であった。韻律は評価者の主観的判断が高く評価基準の一貫性に欠ける課題があるとしている。高梨⁵⁾は、スピーチ評価には、母語話者同士、非母語話者同士で類似性が見られるが、母語話者と非母語話者との間ではあまり類似性はないとしている。以上のことを本研究に当てはめると、日本人英語教師はALTよりも内容を重視し、ALTは流暢さや談話能力に評価を与えると考えられる。

2. 3 第二言語としての英語教師 (ESL) と非英語教師による評価の相違

第二言語としての英語教育で、Hadden⁶⁾は経験の相違がコミュニケーション能力の評価与える影響をアメリカの大学院に入学した上級ESLクラス8名の中国人で調査した。25名のESL教師と32名の非英語教師に、彼らが原稿なしに最大3分30秒のスピーチをした録画を見てもらい、24項目のアンケート調査を行った。その結果、ESL教師は「理解度」を最も重視し、「社会的受容度」「言語能力」「人柄」「ボディ・ラングエージ」の順となった。一方、非英語教師では、「理解度」「言語能力」はほぼ同程度に重視され、「社会的受容度」「ボディ・ラングエージ」「人柄」の順となった。また、非英語教師の方がESL教師よりも全体的に寛容な評価をしていることや「言語能力」を有意に高く評価していることが示された。以上のことより、ALTは理解度を重視し、日本人英語教師では、英語能力をALTよりも

有意に高く評価する傾向があると考えられる。

3. 仮説

本研究では、日本人英語教師とALTのスピーチコンテストにおける評価方法に対する調査結果から得られる、以下の3つの仮説を掲げる。

- 1) スピーチの順位づけを行う場合、分析的評価と全体的評価との審査結果に相違はない。
- 2) 分析的評価よりも全体的評価の方が、評価に用いる時間は短くなる。
- 3) 分析的評価よりも全体的評価の方が、より評価が容易である。

4. 研究の方法

4. 1 被験者

本研究はスピーチ審査員として日本人英語教師とALTを想定した。また、日本人英語教師の英語能力差（英検準1級と英検2級レベル）がALTの審査結果との相違に影響を与えるのかも比較したいと考えた。このため準1級レベル、2級レベルの英語教師とALTを審査員として考えた。しかしながら、5.2の実験材料にもあるように本調査にはかなりの時間を要するため、実際に協力してくれる被験者の確保は困難を極めた。協力の内諾を頂いた方々に、実験材料を手渡したり、郵送等で送付したりした方々は158名を超えた。しかしながら、最終的に全ての実験に参加した被験者数は、準1級レベルの英語教師20人、2級レベルの英語教師17人、ALT15人であった。

4. 2 スピーチ審査方法

スピーチ評価の集計方法と評価項目については以下のとおりである。

(1) スピーチ評価の集計方法

スピーチコンテストの集計方法は野村（2009）によれば大きく4つの方法がある。それぞれメリットとデメリットがある（表-1）。本実験では得点の影響差を最小に抑えるため、サブランク法を用いた。まず、評価者ごとの得点をもとに、上位1位～3位までを決めた。それをもとに1位には10点、2位には9点、3位には8点と割り当てた。また、評価者の得点と同じであった場合は、その順位から0.5点を引いて割り当てた。たとえば、2位が2人であった場合は、それぞれに $9 - 0.5 = 8.5$ 点を与えた。その後、評価者ごとの順位から与えられたスピーチ者の得点総計を

求めた。これを元に1位～10位までの順位を決定した。

表-1 審査集計方法の比較

	メリット	デメリット
得点集計法	審査員全体の合計点の数字が細かく得点差が出やすいため、順位がわかりやすい。	ある発表者に対して極端に高い得点や低い得点をつける審査員がいた場合、他の複数の審査員の評価を上回る影響をもつことがある。
順位集計法	計算が簡単であることから結果の集計に時間がかからない。	合計点の数字の少ないものが上位となるが、集計の結果、合計点が同一となることが起こりやすい。
サブランク法	得点集計法のデメリットである極端な得点差の影響を最小に抑えることができる。	審査員個々に集計を行った後、上位の発表者から重み付け得点を加え、その作業を全審査員の得点に対して行うため、集計時間を多く要する。
Toastmasters法	サブランク法の一つであるが、審査員は上位3名の名前だけを報告するため、集計係が細かな得点の集計作業を必要とせず、迅速に結果が出る。	主催者あるいは実行委員会側が各審査員の与えた実際の得点や分析評価を把握したり資料として残していくことができない。

(2) スピーチの評価項目

評価項目については、「全体的評価」では10人のそれぞれの全体的な印象を100点満点でつける形にした。一方、「分析的評価」ではRobson⁷⁾を参考にした。彼はスピーチコンテストの審査員のレポートをもとに、分析的評価の項目の難易度を調査した（表-2）。その結果、最も審査が難しいのは「話題の内容への興味」であった。話題の内容に興味を引くかどうかは、評価項目として最も大切であるが評価基準を設定するのは困難であるとされた。一方、最も審査が容易であるのは「音声」であった。また、ジェスチャーとアイコンタクトの評価については審査員により評価項目の対象とすべきか意見が二分している。この理由から、デリバリーの中からジェスチャーとアイコンタクトとの2つを排除した。そのため、実験材料としてのスピーチ発表者は音声のみのCD録音とした。「分析的評価」の評価項目のその他の項目については決まったものはないので、できるだけ多くの有名な伝統あるコンテストの分析的評価表実例を参考（トースト・マスターズのスピーチ評価表、高円宮杯スピーチ評価表等）にして作成した。本研究では、項目別スピーチ評価表を作るにあたって、最も協議時間をかけた点は、本研究のための分析的評価の評価項目をどうするか、またその

評価点をどのくらいにするかであった。この結果、分析的評価では文の内容に関わる3つの項目に40点、英語の発音や表現に関わる2つの項目を60点、総計100点満点とした。

表-2 分析評価項目の難易度

Criteria	Severity	
Interest	0.51	harder to be scored high
Body Language	0.36	↑
Eye contact	0.03	
Organization	-0.07	
Pronunciation	-0.16	
Voice	-0.67	easy to be scored high
Mean	0.00	
SD	0.38	

4. 3 実験材料

実際の中学生スピーチコンテストで平成21年～23年までの3年間で県大会最終決定に残ったレベルあるいはそれと同等レベルの10人のスピーチをCD録音したものを用意した。スピーチ指導は全て同じ指導者の日本人英語教師であり、補助として指導に関わったALTもカナダのブリティッシュコロンビア地域の出身者である。スピーチの長さは全て4:00～4:33以内に収まるように編集してある。男女による性差の影響を極力避けるために男女のバランスを考慮したが、入賞者レベルにしたため、録音した生徒は男4名、女6名となった。また、CD録音した順番は男女を極力交互に収録してある。収録したスピーチ10人分の全てを単純に聞くだけで43分37秒かかる。それに評価に要する時間がこれに加わる。全体的評価と分析的評価を別々の日に2回に分けて10人分のCDを聞くことを考えれば、これだけで合計2時間以上を要する。さらに、最後にアンケートを実施するため、それに要する時間を10分程度見ておく必要がある。また、実験に関わる手順を理解する時間を含めるとさらに時間が必要である。時期をずらして同一の被験者に2度CDを聞いてもらわなければならないため、被験者の負担は大きいですが、スピーチ数を減らすと今度はスピーチ実験の調査実態から離れてしまうので、やむを得ないと判断した。

4. 4 実験の手順

本実験は評価者の得点の平均値をもとに、①発表者の結果順位と②評価者の2つの異なった評価方法による評価をするまでにかかった時間の相違を分散分析で検証した。①は評価者（3：ALT、準1級、2級）×発表者（10：Ms A～Ms J）×評価方法（2：

全体的評価、分析的評価）の3要因混合計画析（AsBC）で、②は評価者×評価方法による2要因混合計画（AsB）で分散分析を行った。混合にしたのは、被験者の数を少なくすることができ、また等質化の手続きも不要となる利点があったからである。ただし、被験者内計画では1回目の実験が2回目の実験の練習になってしまう「練習効果」が生じる可能性がある。このため、分析的評価と全体的評価の間に日にちを一週間程度以上空け、CDの録音順、カウンターバランスをした。

4. 5 アンケート調査

2回目の実験後に2つのスピーチ評価方法に対するアンケートを五件法で実施した。

5. 研究結果

5. 1 仮説1の検証

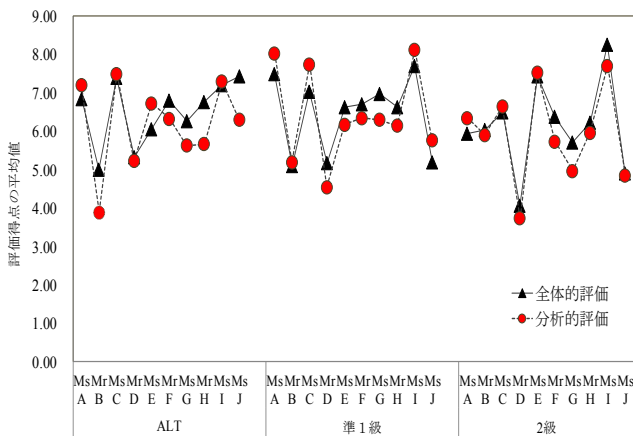
「スピーチの順位づけを行う場合、分析的評価と全体的評価との審査結果に相違はない。」統計的有意差を検証するため、評価得点の平均値をもとに、評価者（3：ALT、準1級、2級）×発表者（10：Ms A～Ms J）×評価方法（2：全体的評価、分析的評価）の分散分析（AsBC）を行った（表-3と表-4、図-1）が、交互作用は有意ではなかった。

表-3と表-4 分析的評価と全体的評価のサブランク法による評価者の評価得点の平均値

発表者	全体的評価											
	ALT (15人)			準1級 (20人)			2級 (17人)			全体 (52人)		
	Mean	SD	順位	Mean	SD	順位	Mean	SD	順位	Mean	順位	
Ms A	6.83	1.90	4	7.50	2.09	2	5.94	1.95	7	6.76	3	
Mr B	5.00	1.66	10	5.10	1.93	9	6.03	2.02	6	5.38	9	
Ms C	7.40	1.78	2	7.03	1.50	3	6.50	2.24	3	6.98	2	
Mr D	5.33	1.96	9	5.18	1.48	8	4.06	1.66	10	4.86	10	
Ms E	6.07	1.91	8	6.63	2.01	6	7.44	2.19	2	6.71	4	
Mr F	6.80	1.58	5	6.70	1.97	5	6.38	2.21	4	6.63	5	
Ms G	6.27	2.77	7	6.98	1.95	4	5.71	2.14	8	6.32	7	
Mr H	6.77	1.96	6	6.63	1.89	6	6.24	2.59	5	6.54	6	
Ms I	7.20	2.00	3	7.70	1.50	1	8.26	1.90	1	7.72	1	
Ms J	7.43	2.32	1	5.20	1.83	7	4.91	2.26	9	5.85	8	

発表者	分析的評価											
	ALT (15人)			準1級 (20人)			2級 (17人)			全体 (52人)		
	Mean	SD	順位	Mean	SD	順位	Mean	SD	順位	Mean	順位	
Ms A	7.20	2.01	3	8.03	1.98	2	6.35	2.31	4	7.19	3	
Mr B	3.90	1.94	10	5.20	1.86	9	5.91	2.28	6	5.00	9	
Ms C	7.50	1.48	1	7.75	1.70	3	6.65	2.15	3	7.30	2	
Mr D	5.23	2.30	9	4.55	1.74	10	3.74	2.04	10	4.51	10	
Ms E	6.73	2.29	4	6.18	2.14	6	7.53	1.61	2	6.81	4	
Mr F	6.33	1.81	5	6.35	2.17	4	5.74	3.00	7	6.14	5	
Ms G	5.63	2.15	8	6.30	2.63	5	4.97	1.97	8	5.63	8	
Mr H	5.67	2.67	7	6.15	2.27	7	5.97	2.46	5	5.93	6	
Ms I	7.30	2.20	2	8.13	1.31	1	7.71	2.14	1	7.71	1	
Ms J	6.30	2.94	6	5.78	2.18	8	4.85	2.51	9	5.64	7	

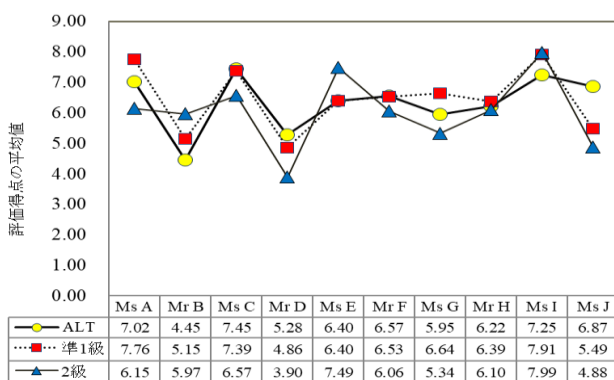
図-1 全体的評価と分析的評価の平均値



評価者×発表者の交互作用のみ有意であった ($F(18,441)=1.95, p<.05$) ので、評価者別に発表者ごとの単純主効果を検定した (図-2)。

その結果、発表者 Mr D, Ms J の2名の評価得点で有意に評価の差があることがわかった (それぞれ $F(2,49)=3.49, 4.40, p<.05$)。次に評価者ごとの水準別誤差項を用いて群の単純主効果を検定した。その結果、それぞれの評価者群は発表者に有意に評価得点で差をつけていることが示された (それぞれ $p<.01$ で $F(9,441)=4.13; 5.32; 6.76$)。LSD法を用いた多重比較によれば Mr D および Ms J に対して ALT は2級の評価者よりも有意に高く (それぞれ $p<.05$ で $MSe= 4.95; MSe= 7.09$)、それ以外の発表者に対しては評価者群の間に有意差はなかった。

図-2 サブランク法の評価得点による評価平均



分析的評価と全体的評価で、それぞれ審査評価者から出された得点の平均値をもとに1~10位の順位に相関があるかどうかを検定するためスピアマンの順位相関係数 (the Spearman Rank-order Correlation) を用いた。結果は表-5の通りであった。分析的評価と全体的評価の間には各評価者に対して、それぞれ

正の相関が1%水準で有意に認められた。以上の結果、分析的評価と全体的評価とでは発表者の評価得点に関して有意差はないことが示された。また、分析的評価と全体的評価のALT, 準1級, 2級といった評価者内、あるいは評価者全体での発表者の順位づけに関しては極めて高い順位相関があることが示された。尚、ALTと2級の評価得点では特定の発表者 (10人中2名) の評価得点で有意に評価差があることが示された。

表-5 審査の評価に関わる変数の相関分析

評価者	分析的評価	全体的評価
準1 評価者	分析的評価	1 .855**
	全体的評価	.855** 1
2級 評価者	分析的評価	1 .857**
	全体的評価	.857** 1
ALT 評価者	分析的評価	1 .795**
	全体的評価	.795** 1
全部 の 評価者	分析的評価	1 .827**
	全体的評価	.827** 1

** 相関係数は $p<.01$ 水準で有意 (両側)

5. 2 仮説2の検証

「分析的評価よりも全体的評価の方が、評価に用いる時間は短くなる。」

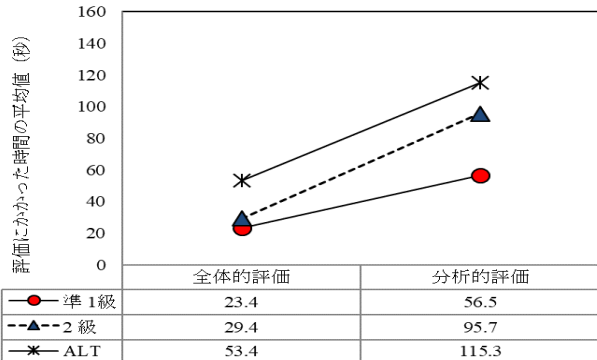
これを証明するために、評価者がそれぞれスピーチの評価に関わる時間を記録しておいたものをもとに、その平均値の差を比較し、全体的評価にかかる時間の方が統計的に有意に短いかどうかを調査した。まず、分析的評価と全体的評価の、それぞれの評価方式ごとに、すべてのスピーチの評価に費やした合計時間 (秒) の平均値とその標準偏差を求めた。それらの数値をもとに、2要因混合計画 (AsB) の分散分析を行った。評価者がスピーチの評価を行ったときに記録されている時間はスピーチの評価に関わった時間と聞いている間の時間が合計されていたため、「評価者がスピーチごとにかかった時間」から「スピーチを聞いている時間」を差し引いたものをデータとした。評価者別に見た評価方法と評価時間の平均値の結果は以下のとおりであった (表-6, 図-3)。評価者 (3) × 評価方法 (2) の交互作用は有意ではなかった ($F(2,49)=1.38, ns$)。評価者と評価方法の単純主効果のみが有意であった (それぞれ $F(2,49)=6.95, 38.01, p<.01$)。LSD法による多重比較を行った結果、準1級評価者は2級とALTの評価者よりも有意に評価時間が短いことが示されたが、2級とALTの評価者間には有意差はなかった。以上のことから、分析的評価は全体的評価よりも有意に時間がかかることが示された。また、評価に関しては、ALTと2級の評価者は準1級よりも有意に評価時間が

かかることが示された。

表-6 評価者別評価方法と評価時間

評価者	n	全体的評価		分析的評価	
		Mean	SD	Mean	SD
準1級	20	23.41	12.06	56.49	28.54
2級	17	29.44	19.66	95.69	68.54
ALT	15	53.40	43.00	115.31	79.08

図-3 評価方法と評価時間の比較



5. 3 仮説3の検証

「分析的評価よりも全体的評価の方が、より評価が容易である。」

五件法を用いたアンケートの質問事項のなかで、「全体的評価は分析的評価よりも評価が簡単に行えると思う」という項目を、 χ^2 検定を用いて検証した。アンケート項目は、ABCDEの5段階で評点した。評価基準は「全くそう思わない」、「ややそう思わない」、「どちらでもない」、「ややそう思う」、「全くそう思う」であった(表-7)。 χ^2 検定の結果、人数の偏りは有意ではなかった(表-8)。

表-7 「全体的評価は評価しやすい」項目

		全くそう 思わない	ややそう 思わない	どちらで もない	そう思う	全くそう 思う	合計	
評価者	ALT	度数	1	1	8	3	2	15
		期待度数	1.4	2.3	5.2	3.5	2.6	15.0
		標準残差	-0.4	-0.9	1.2	-0.2	-0.4	
	準1	度数	1	5	7	4	3	20
		期待度数	1.9	3.1	6.9	4.6	3.5	20.0
		標準残差	-0.7	1.1	0.0	-0.3	-0.2	
2級	度数	3	2	3	5	4	17	
	期待度数	1.6	2.6	5.9	3.9	2.9	17.0	
	標準残差	1.1	-0.4	-1.2	0.5	0.6		
合計	度数	5	8	18	12	9	52	
	期待度数	5.0	8.0	18.0	12.0	9.0	52.0	

表-8 「表-7」の χ^2 検定の結果

	値	自由度	漸近有意確率(両側)
Pearson χ^2	7.758 ^a	8	.457
尤度比	7.718	8	.462
線型と線型による連関	.006	1	.937
有効なケースの数	52		

a. 12セル(80.0%)は期待度数5未満 最小期待度数 1.44

5.4. 全体的評価と分析的評価の評価方法に対するアンケート調査

表-9 「評価方法で結果に相違がある」項目

		全くそう 思わない	ややそう 思わない	どちらで もない	そう思う	全くそう 思う	合計	
評価者	ALT	度数	1	1	3	5	5	15
		期待度数	0.3	0.9	2	6.9	4.9	15.0
		標準残差	1.3	0.1	0.7	-0.7	0.0	
	準1	度数	0	1	3	12	4	20
		期待度数	0.4	1.2	2.7	9.2	6.5	20.0
		標準残差	-0.6	-0.1	0.2	0.9	-1	
2級	度数	0	1	1	7	8	17	
	期待度数	0.3	1	2.3	7.8	5.6	17.0	
	標準残差	-0.6	0.0	-0.9	-0.3	1.0		
合計	度数	1	3	7	24	17	52	
	期待度数	1.0	3.0	7.0	24.0	17.0	52.0	

表-10 「表-9」の χ^2 検定の結果

	値	自由度	漸近有意確率(両側)
Pearson χ^2	7.262 ^a	8	.509
尤度比	7.433	8	.491
線型と線型による連関	2.251	1	.134
有効なケースの数	52		

a. 10セル(66.7%)は期待度数5未満 最小期待度数 .29

2つの評価方法実施後のアンケート調査は、有意差はないにしても、評価方法に相違があると思うという回答が多かった(表9と表10)。また、評価方法に対してさまざまな意見が記述されていた。

6. まとめ

スピーチコンテストの審査評価をサブランク法で得点化し、「全体的評価」が「分析的評価」に代わる評価方法として可能か検証した。参加者は準1級レベル英語教師20人、2級レベル英語教師17人、ALT15名であった。実験は、「a. スピーチの順位づけを行う場合、分析的評価と全体的評価との審査結果に相違はないか。 b. 分析的評価よりも全体的評価の方が、評価時間は短くなるか。 c. 分析的評価よりも全体的評価は評価が容易であるか。」の3つを調査した。結果は、a. 分析的評価と全体的評価とでは発表者の評価得点に関して有意差はないことが示された。しかしながら、ALTと2級の評価者得点では分析的評価と全体的評価で特定の発表者(10人中2名)の評価得点で有意に評価差があることが示された。 b. 全体的評価は分析的評価よりも時間的に有意に評価者の負担を軽減した。評価者別に見た場合、ALTと2級レベルの評価者は分析的評価時間は全体的評価よりも有意に時間が長くかかったが、準1級では有意差はなかった。 c. ただし、参加者は全体的

評価は分析的評価よりも評価が容易であるとは思っていないことが示された。

7. 今後の課題

今回の調査結果から、全体的評価方法は分析的評価方法と同等の評価結果が期待される。しかしながら、全体的評価方法は評価基準が明確でないため、評価者にとっては評価時間が少なくなる半面、評価結果に対して常に不安を抱えて評価するといった不安感を高める結果が生じることも示されている。このことが、評価の妥当性や信頼性を発表者に提示する具体的な資料として分析的評価方法が適切であると評価者が考える理由となっている。分析的評価に比較して全体的評価で評価得点差が少なくなるのは、全体的評価は評価項目がないため、発表者のスピーチを聞きながら評価観点を評価者が独自につくっていく必要があるため、無意識のうちに評価得点差が少なくなるのかもしれない。

今回の調査には限界もある。まず、評価者が52名の結果だけである。また、CD録音の音声評価のみであり、映像を含めた総合的な評価ではない。しかしながら、この結果を前提として述べるなら、予備選考では、全体的評価で審査を行い、決勝大会では時間をかけた分析的評価を行うことが合理的であろう。

さらに、評価者の評価方法のあり方のコメントを借りるならば、決勝大会では出場者ひとりひとりに大会結果の審査評価表をもとに評価結果のコメントを述べる時間を確保すべきであろう。そして、評価点の説明をし、生徒へ次なる大会出場へ向けての動機づけを与える機会とするといった評価方法の使い分けが必要であることを今回の調査は示唆している。

いずれにしても本研究が今後のスピーチ評価への審査を再考するきっかけとなれば幸いである。

参考文献

- 1) 野村和宏, 「英語スピーチコンテストの課題 —より教育的な活動とするために—」, 『神戸外大論叢』, 第60巻第5号, pp.1-26, 2009.
- 2) 大友賢二, 「言語テストと第二言語」 『第二言語習得研究に基づく最新の英語教育』, 大修館書店, 1994.
- 3) Nakamura, Y.: Differences in native and non-native teachers' evaluation of Japanese students' English speaking ability, *Cross Currents*, 19(2), pp.161-165, 1992.
- 4) Anderson-Hsieh, J., Johnson, R., & Koehler, K.: The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and

syllable structure, *Language Learning*, 42(4), pp.529-555, 1992.

- 5) 高梨庸雄, 「身近なコミュニケーション14」 『現代英語教育』 1月号, pp.42-43, 研究社, 1996.
- 6) Hadden, B. L.: Teacher and nonteacher perceptions of second-language communication, *Language Learning*, 41(1), pp.1-24, 1991.
- 7) Robson, G. G.: Applying Rasch Measurement to Judged Ratings from a Speech Contest at a Japanese University, *JACET Journal*, 47, pp.51-66, 2008.

(2018. 9. 27 受付)