

FACETS 分析による英語音声評価における評定者バイアス検証

Investigating Rater Bias in Measuring English Pronunciation: A FACETS Analysis

YOSHIDA, Hiroko

INOUE, Kumiko

IMAI, Yumiko

UEDA, Hiroko

OTSUKA, Tomomi

Osaka Jogakuin College

Abstract

This study investigates rater factors affecting English pronunciation evaluation. Yoshida (2004) found that both L1 English and L1 Japanese raters exerted significant bias against evaluation items and candidates' pronunciation ability when rating EFL Japanese pronunciation. Her study also suggested that L1 Japanese raters are well capable of evaluating English pronunciation as accurately as L1 English raters as long as they have proficient English ability and basic knowledge of phonology. However, the raters in the study were not experts, and had little experience of evaluating pronunciation. The present study, then, investigates rater bias using raters who teach courses in phonetics and regularly examine students' pronunciation at a college. The data used in this study were 76 readings that were recorded by 38 college students. The students tape-recorded a dialog, which was designed to evaluate English pronunciation before and after a one-year phonetics course. The readings were rated by five phonetics instructors using 15 analytic items that represent the features of the sound system of General American English (vowels, diphthongs, consonants, consonant clusters, aspiration, word stress, sentence stress, rhythm, intonation, weak forms, loudness, rate/tempo, smoothness, energy, clarity; Yoshida, 2005). The Many-faceted Rasch analyses were applied using FACETS software (Linacre, 1996a, 2005). The results have revealed that: (a) the ratings by five expert raters showed reasonable interrater reliability and acceptable intrarater reliability; (b) the raters significantly differed in their rating severity; and (c) each rater exerted significant unique biases both in assessment items and in students' pronunciation ability.

1. はじめに

第二言語学習者にとって対象言語の適切な音声を習得することが円滑なコミュニケーションを進める上で重要であることは、コミュニケーション能力 (communicative competence) の下位構成要素として音声の役割が認識されていることから明らかである (Bachman, 1990; Bachman & Palmer, 1996; Canale, 1983; Canale & Swain, 1980; Celce-Murcia, Dörnyei, & Thurrell, 1995)。しかし、日本の英語教育において音声指導はこれまで積極的に取り入れられることはなかった。その理由のひとつとして、英語を第二言語話者とする英語教員が音声を客観的に指導評価する手法が確立しておらず、音声評価を筆記試験による

評価で代用してきたことが考えられる。Yoshida (2004) の研究では、高い英語力を持ちかつ基本的な英語音声学の知識を持つ日本語母語英語教員は、英語を母語とする英語教員と同等に英語音声評価を行うことができることが示された。しかし、この研究では評定者は経験豊かな大学英語教員ではあったが、英語音声教育¹に従事する教員ではなかった。本研究は、Yoshida (2004) の研究を発展させ、大学レベルで英語音声教育を担当し音声評価に習熟している評定者の評価バイアス (rater bias) をパフォーマンス評価に適した項目応答理論を応用したラッシュ分析 (ファセット分析) を用いて検証した。

2. 先行研究

近年言語テストの分野ではパフォーマンス評価に対する関心が高まり、パフォーマンス評価における学習者のスコアに及ぼす様々な要因について検証が行われてきた。言語テストでは、学習者がテストで得るスコアは真のスコア (true score) に評定誤差 (error) を加算したもの (Brown, 1988, p. 99) と考えられているが、筆記試験と異なりパフォーマンス評価にはこの評定誤差に影響を与える種々の要因が存在する。McNamara (1996, p. 9) は筆記試験とパフォーマンス評価の相違点は、筆記試験では試験問題のみが受験者のスコアに影響を与えるのに対し、パフォーマンス評価では試験課題に加えて評定者と評定尺度がスコアに影響を与える要因になりうると論じている。信頼性のある評価をおこなうために評定者要因をどのように扱うかについては、ライティングやスピーキングに関して評定者が評価に及ぼす影響を中心に数多くの研究が報告されている。Weigle (1998) は第二言語話者によって書かれた英語ライティングの評価において、評価経験有無が評定者の評価の厳しさや一貫性に与える影響を評定ガイダンスの前後で検証した。その結果、評価経験無しの評定者の方が経験豊かな評定者よりも一貫性に欠けたより厳しい評価を行うことが明らかになった。また、評価前の評定ガイダンスが評定者間における評価の厳しさの差を完全に取り除くことはできなかったと報告している。Kondo-Brown (2002) は第二言語話者によって書かれた日本語ライティングの評価に関し、評定者は一貫性のある評定をおこなったが評価の厳しさには有意差があったことを示し、評価前の評定ガイダンスが評定者の評価基準の統一を図るには十分でなかったと報告している。また、被験者のライティング能力の高低並びに評価項目に対し各評定者が独特のバイアスを示し、評価の厳しさの度合いに評定者間で有意な差が生じたことを明らかにした。スピーキングの評定に関する研究においても、評定者の経験あるいは経歴がスピーキング評価に影響を与えることが報告されている (Bonk & Ockey, 2003; Brown, 1995; Chalhoub-Deville, 1995)。Lumley and McNamara (1995) はライティングと同様にスピーキングの評価でも、評定ガイダンスが評定者の評価を一致させるには至らなかったことを指摘した。また、評定者の厳しさの度合いが被験者のスピーキング能力の高低により影響をうけること (Bachman, Lynch, & Mason, 1995; Lynch & McNamara, 1998)、スピーキング課題の違いが評定者バイアスにつながる (Bachman et al., 1995; Wiggleworth, 1993) が報告されている。

これらのスピーキングやライティングのパフォーマンス評価に比べて、英語音声のパフォーマンス評価に関する研究は極めて少ない。Yoshida (2004) はラッシュ分析を用いて信頼性と妥当性のある英語音声の分析的評価を検証し、英語母語話者と日本語母語話者である6名の大学英語教員の評価特性を分析した。その結果、日本語母語話者教員も英語母語話者教員とともに評定者内一貫性を示し、高い英語力と英語音声学の基本的知識を有する日本語母語話者が、英語音声評価に関して英語母語話者と同等の評価を行うことができる

ことを示した。また、各評定者の評価に対する厳しさの度合いには有意な差が見られたが、その差は評定者の第1言語 (L1) によるものではなく、評定者の個人差であることを明らかにした。さらにライティングやスピーキングの研究と同様に音声の評定においても、評定者は被験者や課題によって独特のバイアスを示すことを確認した。しかし Yoshida (2004) の研究では評定者は英語指導経験が豊かな大学の教員であり英語音声に関する知識も有していたが、音声評価を日常的に行っていたわけではなかった。そこで本研究では、Yoshida (2004) の研究を発展させ、大学で英語音声教育に携わり音声指導と音声評価に習熟している評定者の評価特性を分析することを目的とし、次の研究質問を設定した。

1. 評定者の信頼性はどのようなものであるか。
2. 評定者は被験者の音声能力の高低や評価項目に対し、どのような評価バイアスを示すか。なお本研究の英語音声とは、アメリカ標準英語 (General American English) をモデルにした音素、超音素、準言語要素からなる英語発音と定義する。

3. 方法

3.1 データ収集

研究に使用したデータは、2003 年度に短期大学に入学した英語専攻の学生 262 名が Phonetics² のコース開始時 (2003 年 4 月) と修了時 (2004 年 1 月) に、英語音声診断を目的とした会話文 (大塚, 今井, 井上, 上田, & 米田, 2004, 付録) を音読録音したテープから無作為に抽出した 50 名分である。このうち、録音状態が不良なものを排除した結果、最終的に 38 名分のテープ (76 の会話文を録音したもの) をデータとして使用した。これらの 76 本のデータは Phonetics 学習前後の区別なく無作為順序で 1 本のテープに編集、ダビングされ各評定者に渡された。

3.2 評価項目と評価表

3.2.1 評価項目

Morley (1991, p. 493) の広義の英語音声構成要素のうち、本研究では音素 (segmentals), 超音素 (suprasegmentals), 準言語要素 (paralinguistic features) に属する次の 15 項目 (Yoshida, 2005) についてアメリカ標準英語 (General American English: GAE) を基準として評価を行った。

音素	超音素	準言語要素
母音 (Vowel)	単語ストレス	声の大きさ (Loudness)
二重母音 (Diphthong)	(Word stress)	速さ / テンポ
子音 (Consonant)	文ストレス	(Rate/Tempo)
子音連結	(Sentence stress)	スムーズ度 (Smoothness)
(Consonant cluster)	リズム (Rhythm)	エネルギーレベル
アスピレーション	イントネーション	(Energy)
(Aspiration)	(Intonation)	明瞭度 (Clarity)
	弱形 (Weak form)	

3.2.2 評価尺度

前出の 15 項目について、6 段階評定尺度 1 (Very poor), 2 (Poor), 3 (Fair), 4 (Good),

5 (Very good), 6 (Excellent) を用いて評定した。これは Yoshida (2004) で使用され、その妥当性が示されたものである。

3.2.3 評定者と評価手順

評定者は全員が TESOL を専門分野とし、大学と短期大学で Phonetics の授業を担当する 5 人の日本語母語話者である。表 1 は評定者の Phonetics 担当年数及び大学（短大含む）における英語教員歴である。これらの評定者は Phonetics の授業において課題や試験としてテープ録音した英語音声定期的に評定しているが、本研究の実施に際し統一性を保つために評定ガイダンスを行い、評定項目の擦り合わせ並びにデモテープと評価表を用い評定練習を行った。なお、評定は 2004 年 8 月に各自個別に実施した。

表 1 評定者の Phonetics 担当年数及び英語教員歴

評定者	Phonetics 担当年数	英語教員歴 (年)
1	3	5
2	13	13
3	8	10
4	11	15
5	9	10

4. 分析

本研究では FACETS ソフトウェア (Linacre, 1996a, 2005) を用いて音声パフォーマンス評価に及ぼす評定者の影響を分析した（これ以降、ファセット分析と称す）。この分析はパフォーマンステストにおける被験者の能力測定が、その評価を実施する際種々の要因（ファセット, facets）によって相互に影響をうけるという考えに基づいている (Lumley & McNamara, 1995)。ファセット分析では全てのファセットが同時に分析され、同一の「真の間隔尺度」(Henning, 1987) であるログ (log) 尺度に表示される。それぞれのファセットにおいて標準誤差 (standard error) とフィット統計 (fit statistics) が算出され、各ファセットが予測モデルに適合しているかどうか確認することができる。また、ファセット分析は信頼性指数を検出し、難易度や評定の厳しさの差が有意であるかどうか示す。さらにファセット分析はバイアス分析 (bias analyses) により、評定者の評価バイアスや評価項目と評価のスコアに見られる特別な相互関係の分析が可能である。通常、 z スコアの絶対値が 2 以上であれば有意なバイアスを示す (Linacre, 1996b; Lynch & McNamara, 1998; McNamara, 1996)。

5. 結果

主分析を行う前にファセット分析を行い各被験者のデータの外れ値 (misfitting) の有無を検証したが、すべてのデータが分析に適合していることが確認された。また、スピーマン・ブラウン方式で評定者間信頼性を算出したところ、15 項目の信頼性係数の範囲は .66 - .89 であった (表 2)。

表 2 項目別評定者間信頼性

評価項目	信頼性係数	評価項目	信頼性係数
母音	.66	イントネーション	.81
二重母音	.73	弱形	.82
子音	.78	声の大きさ	.89
子音連結	.76	速さ/テンポ	.76
アスピレーション	.80	スムーズ度	.84
単語ストレス	.66	エネルギーレベル	.83
文ストレス	.81	明瞭度	.81
リズム	.85		

図 1 はファセット分析サマリーを示し、被験者能力、項目別難易度、Phonetics 学習前後の難易度、評定者の評価の厳しさの度合いを同一尺度の log を用いて図に表している。各被験者はアストリスク (*) で示され、最も高い英語音声能力を示した被験者が一番上、最も低い英語音声能力を示した被験者が一番下に表示される。「Items」は項目別難易度を示し、最も難易度の高い項目が一番上、最も易しい項目が一番下に表示されている。「Time」は Phonetics 学習前後の難易度を表し、Time 1 (Phonetics 学習前) が Time 2 (Phonetics 学習後) より難しかったことを示す³。「Rater」は評定者の評価の厳しさの度合いを示し、上から下に評定者の厳しさの度合いがゆるやかになっている。「Scale」は各被験者の推測される評定結果を表している。例えば、被験者能力が 0 log の学生は、評価における厳しさの度合いが平均的な評定者が評価した場合、6 段階尺度で 3 ポイントのスコアを得ることを示す。

表 3 は評定者の厳しさの結果を示す。Infit 数値は各評定者の評定一貫性を表すもので、5 人の Infit 数値はすべて平均 ($M = 1.00$) より標準偏差 ($SD = 0.20$) の 2 倍値の範囲内であり ($M - 2SD < \text{Infit} < M + 2SD$)、5 人の評定者は全員評定者内一貫性があることを示した。しかしながら、評定者の評定の厳しさは同等ではなかった (Reliability index = 0.99, chi-square = 977.9, $df = 4$, $p = .00$)。表 3 の評定者は上から評価の厳しい順に並んでおり、評定者 2 が最も評価に厳しく、評定者 4 が最も評価が易しかったことを示している。

バイアス分析の結果、各評定者は被験者能力及び項目に関して独自の評価バイアスを持っていることが明らかになった。図 2 - 6 は、評定者と被験者の音声能力の相互作用を表示したものである。これらの図では線が被験者の英語音声能力 (log) を示す。「●」は有意でないバイアス ($-2 < z < +2$) を示し、「▲」は有意なバイアスを表す ($|z| \geq 2$)。z スコアの +2 以上は評定者が有意に厳しい評価を行ったことを示し、-2 以下は評定者の評価が有意に寛容であったことを表す。

Measr	+Examinees	-Items	-Time	-Rater	Scale
+ 2 +					(6)
	*				---
+ 1 +	**				
	**				4
	*	weak forms	Time 1	2	
	*	intonation		5	
	***	consonants		3	---
* 0 *	*	diphthongs	*	*	*
	*	aspiration			
	*	c. clusters			
	***	loudness			
	**	clarity			
	*	energy	Time 2	1	
	*	rate		4	3
	*	word stress			
+ -1 +	**				
	***				---
	*				
	**				

	*				
	*				
+ -2 +					2
	*				
+ -3 +					(1)
Measr	* = 1	-Items	-Time	-Rater	Scale

Note. c.clusters = consonant clusters

図1 ファセット分析サマリー

表 3 評定者の厳しさ

評定者	厳しさの度合い (log)	誤差	Infit (mean square)
2	0.73	0.04	1.30
5	0.32	0.04	1.00
3	0.10	0.04	0.70
1	-0.49	0.04	1.10
4	-0.66	0.04	1.00
<i>M</i>	0.00	0.04	1.00
<i>SD</i>	0.52	0.00	0.20

評定者 1 は、被験者 3 人に対し厳しい評価を行い、被験者 4 人に対しては寛容な評価を行ったことを示している（図 2）。評定者 1 により厳しく評価された被験者は中級あるいはそれ以下の音声能力レベルであり、寛容な評価をされたのは中級以上の音声能力をもつ被験者に対してであった。図 3 は評定者 2 の評価バイアスを示しており、9 人の被験者に対して厳しく評価し、6 人の被験者に対して寛容な評価を行ったことを示している。評定者 2 は全音声能力レベルにわたって評価が厳しい傾向であることを示している。

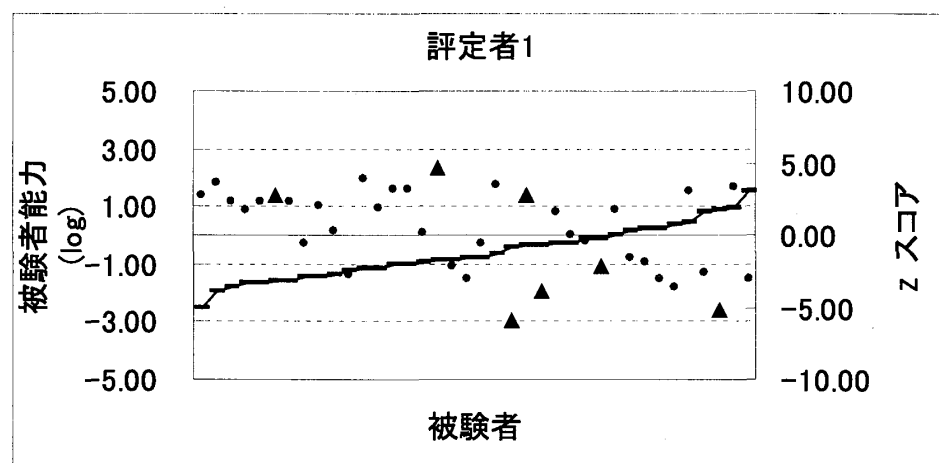


図 2 評定者 1 と被験者におけるバイアス

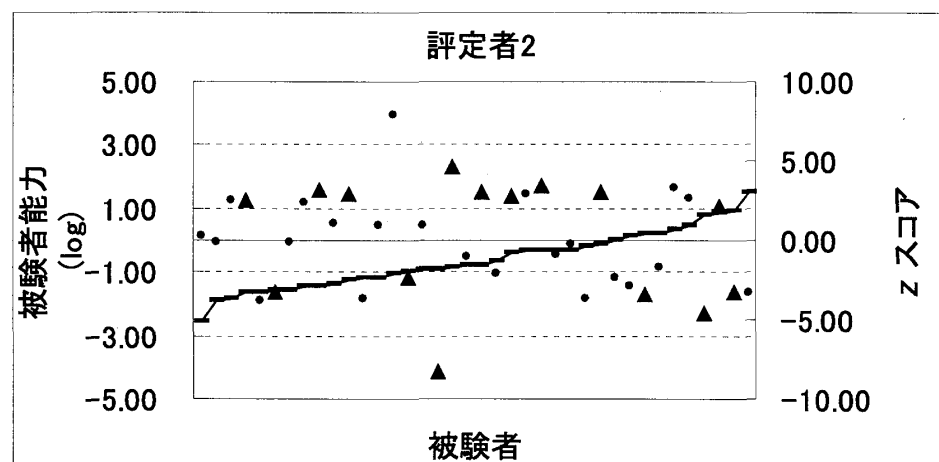


図 3 評定者 2 と被験者におけるバイアス

評定者3は、7人の被験者に対して厳しい評価を行い、寛容な評価も同様に7人に対して行った。評定者3は被験者の音声能力が高い場合に評定が厳しく、低い場合に評定が寛容であり、被験者の音声能力が中級レベルの評定においては、有意なバイアスは全くみられない（図4）。評定者4が厳しく評価したのは2人であり易しい評価は9人に対して行われ、評定が易しい評価バイアスがより多い。また、被験者の音声能力が低い場合は評価バイアスは認められない（図5）。評定者5は、6人に対して厳しい評価を行い、7人に対して評価が寛容であったが、その評価バイアスは全音声能力レベルにわたって観察された（図6）。

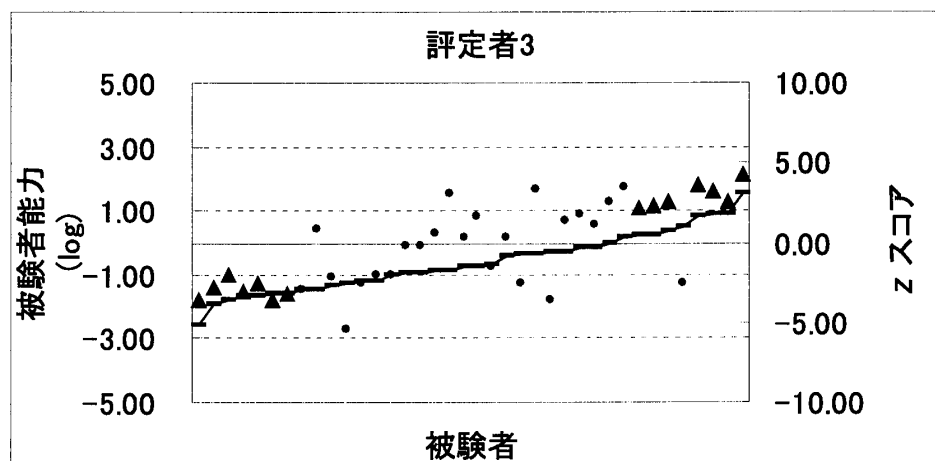


図4 評定者3と被験者におけるバイアス

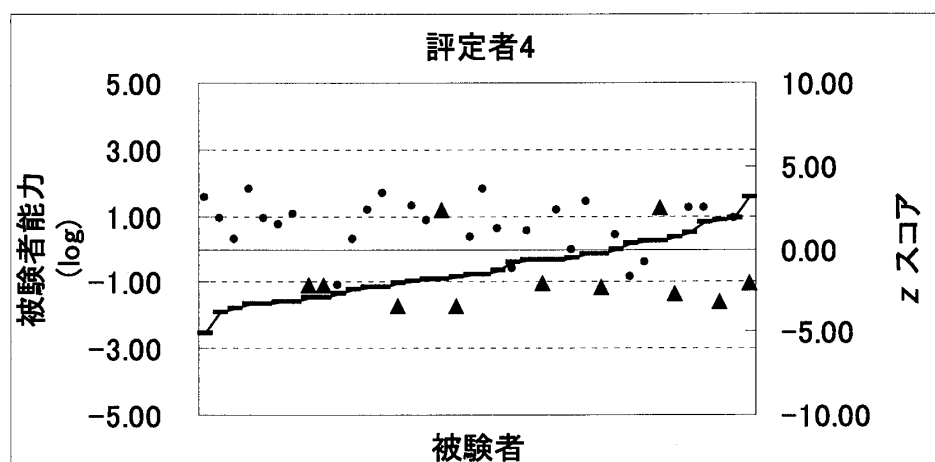


図5 評定者4と被験者におけるバイアス

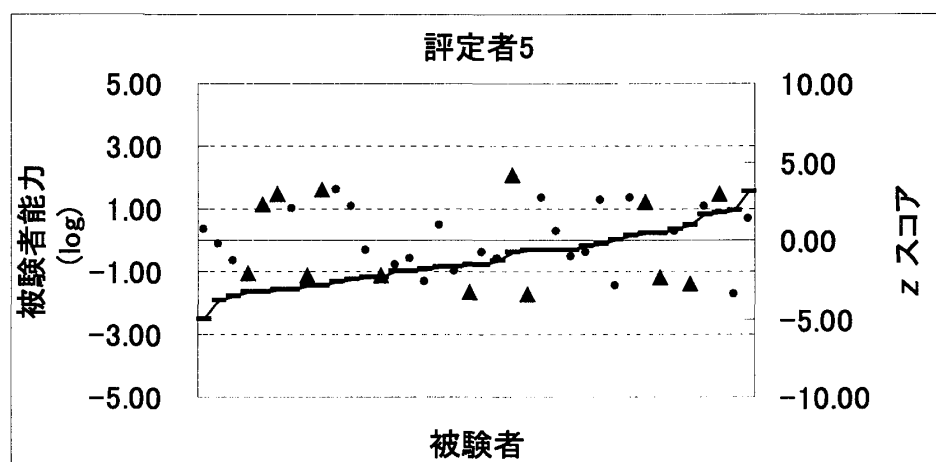


図6 評定者5と被験者におけるバイアス

図7は評定者と各評価項目におけるバイアスを示す。zスコアの+2以上は評定者が有意に厳しい評価を行ったことを示し、-2以下は評定者の評価が有意に易しいことを表す。評定者1は15評価項目のうち、アスピレーション、文ストレス、リズム、イントネーションに対して厳しく評価し、子音、子音連結に寛容な評価を行ったことを示している。評定者2は、子音連結、スムーズ度、エネルギーレベル、明瞭度に厳しい評価をおこない、単語ストレス、文ストレス、リズム、イントネーション、弱形に寛容な評価を行った。評定者3は、唯一子音においてのみ寛容な評価バイアスが観察された。評定者4は母音、子音、弱形、声の大きさに対して厳しい評価を行い、アスピレーションとエネルギーレベルには易しい評価を行った。評定者5は、子音連結に厳しい評価を行い、声の大きさには寛容な評価を行ったことを示す。

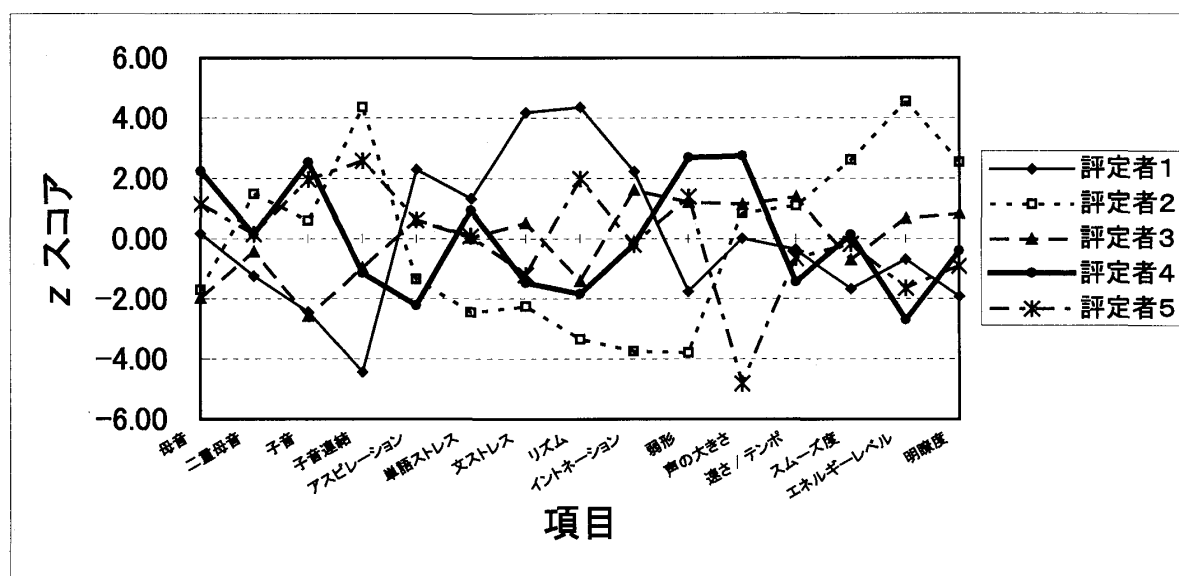


図7 評定者と各項目におけるバイアス

6. 考察と結論

6.1 評定者の信頼性について

5人の評定者による英語音声評価は評定者内一貫性を示し、評定者間信頼性も一定以上の信頼性が見られた。これはYoshida (2004)と同様に、英語音声評価に習熟している日本語母語話者が信頼性のある評価を行ったことを示す。しかし全評定者は事前に評定ガイダンスを受けたものの評定者の評価の厳しさの度合いに有意な差がみられたことは、英語音声の評価に習熟している評定者においても、ライティングやスピーキングにおける研究結果(Kondo-Brown, 2002; Lumley & McNamara, 1995; Weigle, 1998)と同様に、評定者は各自固有の評価基準(厳しさの度合い)を持っており、その厳しさの度合いは評価前評定ガイダンスによっても完全に収束することができないことを表している。このことは、従来言語テストでは一定レベル以上の評定者間信頼性(通常、信頼性係数は.70以上が望ましい(Vogt, 1999))が求められてきたが、評定者間信頼性が高い場合でも評定者が均質な評価を行っていないことを本研究結果は示唆している。

6.2 被験者の音声能力、評価項目に対する評定者バイアスについて

本研究の結果、英語音声評価に習熟しているPhonetics担当者において被験者の音声能力の高低と評価項目において評定の際にバイアスが観察された。これはYoshida (2004)で、大学英語教員がその母語にかかわらず、被験者の音声能力の高低と評価項目に対して評定者バイアスを示した結果と一致する。また、各評定者が被験者のライティングやスピーキングの能力の高低(Bachman et al., 1995; Kondo-Brown, 2002; Lynch & McNamara, 1998)や評価項目(Kondo-Brown, 2002)に対し独特のバイアスを示した研究の結果とも一致する。

では、これらの結果は信頼性のある英語音声評価並びに指導においてどのように応用することができるであろうか。言語評価はその目的から集団準拠試験(norm referenced test: NRT)と目標基準試験(criterion referenced test: CRT)の2種類に大別することができる。集団準拠試験は集団の標準からの個人の成績の乖離を測定し集団における個人の相対的な評価をおこなうことを目的とし、目標基準試験は学習目標が個人によって達成された度合いを絶対的な評価で測ることが目的である(白畑, 富田, 村野井, & 若林, 1999)。集団準拠評価においては評定者間の共通の評定基準は何より重要視されなければならない。従って評定者バイアスが評定ガイダンスの実施にもかかわらず保持されるということは、そのバイアスを相殺するために集団準拠評価においては、評定は複数の評定者によってなされるべきであることを示唆する。また、評定ガイダンスにおいて従来行われてきた評価方法や項目の説明に加えて、あらかじめ各評定者の評価バイアスを分析し、評定者が各自のバイアスを認識した上で評定の訓練を徹底することもひとつの方法である。これに対して目標基準試験は学習目標に則り評価を行うものであるから、集団準拠評価とは異なり評定者バイアスを排除するべきものとしてとらえるのではなく、評定者がクラス内で目標に準じた評価をおこなうために積極的に活用することが望ましい。例えば、コース目標が「音素習得」であれば評定者の評価は「超音素」や「準言語要素」よりも「音素」に厳しくあるべきであり、コミュニケーション能力の下位要素として音声をとらえるならば、超音素や準言語要素は当然ながら厳しく評価されるべきである。また、学習者の英語音声能力に対する評定バイアスが顕著な場合は、評定者はコース到達目標に準じた公平な評価を行うことを常に意識する必要がある。

英語学習における音声の重要性は、音声が単にコミュニケーション能力の下位構成要素

であるからだけではなく、その習得が EFL (English as a Foreign Language) 学習者に及ぼす心理的影響の存在も見逃すことはできない。学習者がどのような英語音声を目標とし習得するかは、最終的には学習者の判断に委ねられるべきであるが (Morley, 1991), 英語の音声習得がコミュニケーションを行う際の自信につながると考える EFL 学習者の意識も報告されている (Yoshida, 2003)。これらのことを鑑みればコミュニケーションな英語教育が重要視されている昨今、英語学習においてより充実した音声教育が望まれる。そして、その音声学習は信頼性のある評価を伴ってこそより実り多いものとなるであろう。本研究の結果が日本の英語教育でこれまで積極的に取り入れられることはなかった英語音声評価の発展につながり、第一歩となることを願う。

参考文献

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgments in a performance test of foreign language speaking. *Language Testing*, 12, 238–257.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion. *Language Testing*, 20, 89–110.
- Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, 12, 1–15.
- Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). New York: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6, 5–35.
- Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tests and rater groups. *Language Testing*, 12, 16–33.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, MA: Newbury House.
- Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing*, 19, 3–31.
- Linacre, J. M. (1996a). Facets (Version 3.0). Chicago: MESA.
- Linacre, J. M. (1996b). *A user's guide to Facets*. Chicago: MESA Press.
- Linacre, J. M. (2005). Facets for Windows version (Version 3.57.0). Chicago: MESA.
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for

- training. *Language Testing*, 12, 54–71.
- Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and Many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15, 158–180.
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Longman.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481–520.
- Vogt, W. P. (1999). *Dictionary of statistics and methodology (2nd edition)*. Thousand Oaks, CA: Sage.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15, 263–287.
- Wigglesworth, G. (1993). Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. *Language Testing*, 10, 305–335.
- Yoshida, H. (2003). College students' views on pronunciation: Intelligible vs. nativelike pronunciation. 大阪女学院短期大学紀要, 第32号, 147–159.
- Yoshida, H. (2004). *An analytic instrument for assessing EFL pronunciation*. Ann Arbor, MI: UMI.
- Yoshida, H. (2005). Validity of an instrument measuring English pronunciation performance. *JACET Kansai Journal*, 8, 1–14.
- 大塚朝美, 今井由美子, 井上球美子, 上田洋子, & 米田信子. (2004). *Phonetics in context*. 東京: 北星堂.
- 白畑知彦, 富田祐一, 村野井仁, & 若林茂則. (1999). *英語教育用語辞典*. 東京: 大修館書店.

註

1. 英語音声教育とはここでは英語の発音や音声学の理論を学ぶ教育をさす。
2. Phonetics は大学のカリキュラムで用いられているコース名であり, 学生は English phonetics の基本理論を学び, 発音指導を受ける。
3. Phonetics 学習前後の難易度の検証は本稿の目的ではないが, この場合 Time 1 (Phonetics 学習前) が Time 2 (Phonetics 学習後) よりも英語音声能力の評定スコアが低いことを示唆する。

謝辞

本研究は, 2005 年度第 44 回 JACET 全国大会において発表した内容に加筆したものである。本研究の実施に際して多大な御協力と御支援を賜った大阪女学院大学・短期大学の関係者の方々に心より感謝の意を表したい。

付録

Diagnosis

This dialog takes place on the plane from Japan to the U.S. A woman and a college student sitting side by side are enjoying chatting during the flight.

(W = Woman; S = Student)

W: So, were you studying or traveling?

S: I was studying Japanese language, culture, and history.

W: What is your impression of the people? Were they nice to you?

S: Yes. I miss my host family.

They were the nicest people I've ever met.

W: How long were you there?

S: About one year.

W: Time flies when you're having fun, doesn't it?

S: You're right.

My Japanese friends came to the airport to see me off.

We all cried!

W: I guess you had a good experience!

S: How about you? May I ask what you did in Japan?

W: Oh, I've been teaching English there.

I'm going back home to see my sister after ten years.

She's a college professor in Boston.

(大塚 et al., 2004)