

# 英語音読評価項目と熟達度の関係

(パイロットスタディー)

## Relationships Between Reading Aloud Measurements and Learners' Proficiency

鈴木 政 浩

Masahiro SUZUKI

### 要旨

学習者の英語音読音声を聞けば、英語熟達度がある程度わかるという知見がある。評価項目に関する提案もあるが、何をどのように評価するかについて必ずしも定説があるとは言えない。また、熟達度との関係を実証したも数が限られる上に、測定方法に課題を残すものがある。本研究は熟達度との関係から、音読の評価基準として何が採用できるのかを探索的に検証する。

対象者は埼玉県内の大学生32名であった。対象者が音読した音声を実験者4名が評価した。評価項目は単語の発音の正確さ（語認識精度）、プロソディーを構成する項目（ポーズ、ストレス、イントネーション）を6件法により評価し得点化した。さらに音読速度（wpm）と筆記テストの得点を記録し、プロソディーに関係する項目を投入したモデルの適合度を確認した。その結果、プロソディーを構成する評価項目のうちイントネーション得点を投入したモデルには十分な適合度が確認できなかった。これらの結果から、熟達度との関係では、音読速度・ポーズ・ストレスの3つが評価項目として妥当なのではないかという提案を行った。また、6件法という印象による評価では熟達度との関係を一部確認することができなかったため、音読評価に関しては熟達度との関係で明確に数値化できる指標が必要であると考え、具体的な提案を行った。

### Abstract

Although many researchers and teachers suggested the possibilities of relationships between the competence of reading aloud and learners' proficiency of English, few experimental studies have been conducted in this area. The aim of this study is to examine the relationships between the reading aloud measurements and learners' proficiency of English. Four American raters assessed the voice data in measuring word recognition, stress, pause, intonation, and rate of reading aloud by structural equation modeling. The subjects of this study were thirty two Japanese university students (twenty six males and six females). The first model consisted of the fluency factors indicated that stress and pause was the main factors, which explained intonation, and pause mainly affected the rate of reading aloud. The model composed of the measures of reading aloud and learners' proficiency

showed that word recognition and rate of reading aloud affect the prosody differently, and they affect learners' proficiency only through prosody. The author discussed that further scrutiny of the measurements and their validity should be examined when learners' reading aloud performances were assessed.

[キーワード]

英語音読評価 プロソディー 語認識精度 音読速度 熟達度

Keywords : prosody, word recognition accuracy, wpm, proficiency

## 1. 問題の所在

英語音読能力（以下音読能力）が学習者の英語学力（以下、熟達度）を反映するという見解は、国内の文献に散見できる（石井編, 1970 : 94-98 ; 財団法人語学教育研究所, 1988 : 26-31 ; 田垣, 1990 : 150 ; 伊藤他, 1995 : 81-85 ; 鈴木, 1998）。しかし、これらはすべて経験的知見である。そのため、前述の石井編（1970 : 98）は、「1クラス分のある程度の長さの英文をすべて評価するのは、ただでさえ主観的になりがちなもの、いっそう主観的にしてしまい、信頼性も妥当性も満たされない」と指摘している。教育現場では複数の教員が生徒1人1人の音声データを聞き、信頼性を確保して評価することや、定期テストの得点との関係を分析して妥当性を確保することは、時間的にも労力の面でもきわめて難しい。それほどの手間隙をかけるのであれば、音読により熟達度を予測するよりも、むしろ熟達度を測定するテストを短時間で実施する方がより効率的であるということになってしまう。しかし、常に高い得点を取る学習者は別として、テストによる熟達度の測定は、熟達度の低い学習者や苦手意識の強い学習者にとって学習動機を下げることになりかねない。そういう点では、英文の音読を通じて学習者のおおよその熟達度を把握し、教材選択や指導法

決定の材料とすることは有意義であろう。

そのため、音読能力と熟達度との関係について実証的に検証した研究が求められるわけだが、こうした研究は数が限られる（京堂, 1989 ; 宮迫, 2002）。宮迫（2002）は回帰分析により高校生の音読能力と熟達度の関係を検証した。イントネーションと熟達度との関係については、評価項目同士の相関が高いために生じる多重共線性が認められたため、分析対象からはずしている。

音読評価に関しては、コンピュータソフトによる測定方法（Pendergast, 1985 ; Coniam, 1999 ; Beck & Mostow, 2004 ; 鈴木・阿久津・飯野, 2009）と、人間が音声を聞いて評価する方法に分けることができる。そのうち評価者が評価する方法に関する先行研究や文献の中には、評価基準を示したものが散見できる。小川編（1982 : 421-423）は音読の際の態度や姿勢に関するものを含め、16に及ぶ評価項目を提起している。これらすべてについて評価をするのは教育現場においては現実的ではなく、実際の授業では、音読指導の目的に沿って個々の教員が項目を抽出して評価することになる。その他評価項目を提示しているものに、京堂（1989）、Coniam（1990）、宮迫（2002）がある。これら

の研究で示された評価項目を表1にまとめる。その他評価方法について提案した鈴木（2009）もあるが、測定の精度を確保するため個々の評

表1 音読の評価項目

小川編 (1982)	1. 読める語彙の習得度 5. 感情をこめているか 7. 同じところを繰り返していないか 9. 逐語読みにっていないか 11. 句読点を無視していないか 13. 緊張で声が不自然でないか 15. 指さしながら読んでいないか	2. 発音	3. 句切り方 6. 適度な声か 8. 頭を不自然に動かしていないか 10. 余計な語句の挿入、省略はないか 12. あて推量でないか 14. 姿勢が悪くないか 16. 読む箇所を見失っていないか	4. 小さな語句の読み方
京堂 (1989)	1. 1分間に読むスピード (6件法) 3. 発音	2. 流暢さ (自然ななめらかさ) 4. 総合評価		
Coniam (1990)	1. 音読速度 4. phrasing	2. accuracy 5. intonation	3. pronunciation	
宮迫 (2002)	1. 発音 4. 内容の伝達	2. イントネーション 5. 内容理解	3. ポーズ 6. 音読速度 (wpm)	

価項目に関してさらに具体的な基準を示すことが求められる。海外のリーディングにおける Fluency (流暢さ) や Repeated reading (繰り返し読み) 指導の領域には、基準を示した研究が散見できる。単語の発音 (語認識精度) に関しては Rasinski (2003: 159-160) が次のような基準を提示している。

1. 明らかな発音の間違い、他の単語への置き換えや発音の反転、読み飛ばしは読み間違いとする
2. 3秒以上経っても発音をしない場合は読み間違いとする
3. 同一単語の繰り返しおよび、自分で読み間違いを修正した場合は、読み間違いとしない

こうした研究は、児童生徒に対する母語のリメディアル教育や読みの苦手な児童生徒を対象とする、いわゆるボトム・アップの視点からアプローチしたものであり、これらの研究も学習者の音読能力がリーディング能力を予測することが可能であるという立場を取っている。Kuhn & Stahl (2003)、Klauda & Guthrie (2008) が過去の研究を概観し、評価項目を概観しているが、その中で国内の研究と共通していると思われる評価項目を整理すると次のよう

になろう。

1. 語認識精度 (word recognition accuracy)
2. 音読速度 (fluency, speed)
3. プロソディーに関わる要素 (Appropriate expression: pitch, pause, phrasing, the tonal and rhythmic characteristics)

これらの項目の習得には順次性があり、まず個々の単語がすばやくかつ正しく発音できるようになり (語認識精度の向上)、単語が流暢に読めるようになるにつれて音読速度も向上し、その後さらに理解、解釈や表現に関わるプロソディー (ポーズ・ストレス・イントネーション) の能力が向上する。熟達度の向上がこうした評価項目に反映されると考えられる。プロソディーに関する項目は、speed と fluency、expression と prosodic features、pause と phrasing など異なる用語を使用しているが、本研究では国内で一般的となっている用語 (ストレス・ポーズ・イントネーション) を採用する。

本研究は、上記の音読能力の発展過程にもとづき図1のようなモデルを想定しプロソディーに関わる音読評価項目が熟達度にどのような影響を与えているのかを分析することを目的とする。

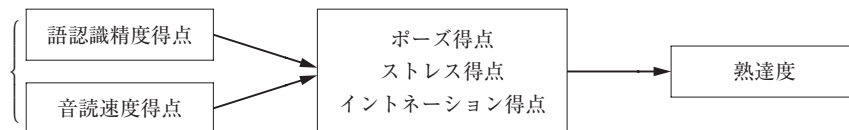


図1 音読評価項目と熟達度からなるモデル

## 2. 方法と手順

対象者は埼玉県内の私立大学生32名（男子26名、女子6名）で2008年11月から2009年2月に実施した。

事前テスト（英検3級2006年度第3回）を実施し、対象者の熟達度を測定した。その翌週初見のテキスト（英検準2級二次試験問題62 wordsを一部改変）を対象者に音読してもらい、その音声を録音した。音声データをアメリカ人評価者（以下評価者）4名が評価し、評価者間信頼係数を産出した。国籍が異なる場合、音読評価の信頼性が損なわれる可能性がある。同じ国籍の母語話者間でも居住する地域により発音の判定に差があることが考えられるが、可能な限り評価者の評価の信頼性を確保するために評価者をアメリカ人のみに限定した（資料1）。

評価項目は次のように設定した。

1. 音読精度：Rasinski（2003：159-160）による読み間違いの判定基準を採用した。この評価基準にもとづきアメリカ人評価者が評価し、正しく読めた単語の語数が、テキストに占める割合を算出し、100点満点で語認識精度得点とした。
2. 音読速度：テキストを読み終わるまでにかかった時間を計測し、wpmに換算した数値を音読速度得点とした。
3. ストレス
4. ポーズ
5. イントネーション

3～5は6段階評価とし、評価者の評価を、イントネーション得点・ポーズ得点・ストレス得点とした（資料2）。音声データは聞きながら読み間違いと判定した単語に下線を引くようにした。その後ストレス・ポーズ・イントネーションを6件法により評価しそれぞれストレス得点、ポーズ得点、イントネーション得点とした。評価者の評価については、評価者間信頼係数（Chronbachの $\alpha$ ）を産出し、音読評価の信頼性を確認した。

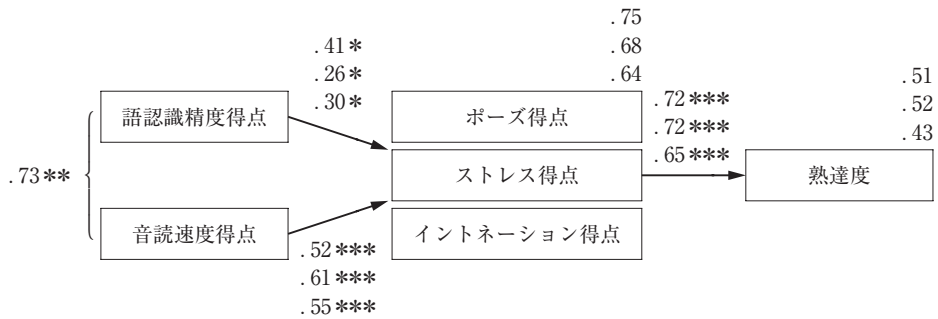
## 3. 結果

評価者間信頼係数は極めて高かった（ $\alpha = .84$ ）。評価項目の得点の中で、多重共線性を示す指標数値であるVIFが、イントネーション得点で16.17、ポーズ得点で16.76と高かった。プロソディーに関わる評価項目を順次投入したのが図2である。各々のモデルの適合度を表2に示す。

語認識精度得点からプロソディーに関する評価項目へのパス係数が若干低く、かつ有意確率も低かった。パス係数や重相関係数の平方（ $R^2$ ）はイントネーション得点を投入したモデルが低く、また適合度も確保できなかった。

## 4. 考察

ポーズは統語能力を、ストレスとイントネーションは内容理解度を反映すると考えられる。これらは別々に熟達度に影響を与えていると考えるのが妥当であるが、 $R^2$ はポーズ得点のみ、ストレス得点のみで分散の50%以上を説明する



\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$   
 パス係数および  $R^2$  は上から、ポーズ得点・ストレス得点・イントネーション得点

図2 音読精度得点・音読速度得点・プロソディー得点・熟達度からなる音読モデル比較

表2 プロソディー評価項目別モデル適合度

	GFI	AGFI	RMSEA	AIC
ポーズ得点を投入したモデル	.990	.951	.000	16.62
ストレス得点を投入したモデル	.981	.904	.000	17.23
イントネーション得点を投入したモデル	.955	.777	.129	19.04

結果となっている。これは多重共線性によるものであると考えられる。6件法の評価ではポーズ、ストレス、イントネーション評価に関しては明確な違いを生じさせることができない可能性を示唆している。イントネーション得点を投入したモデルに関しては適合度が低いことを考えると、イントネーション評価はアメリカ人母語話者であっても明確な基準を設定した上で評価する必要があることを示していると思われる。

イントネーション得点を投入したモデルの適合度が十分でなかったことから、ポーズ得点とストレス得点を合算して図2のモデルに投入してみた(図3)。その結果、適合度指標は  $GFI = .994$   $AGFI = .971$   $RMSEA = .000$  であった。しかし、音読精度得点からポーズ得点+ストレス得点へのパスが有意ではなかった。

以下図3をもとに熟達度と音読評価および評価項目に関する考察を進める。

1) 図3のモデル適合度が良好であったことから、ポーズとストレスは6件法でも評価が

可能かもしれない。しかし、より精緻な評価をするためには、具体的な指針が必要となろう。ポーズに関しては、Rasinski & Padak (2008: 259) は、一語ごと、2から3語、節ごとの区切りと長いフレーズで区切るほど高い評価をするという提案をしている。センスグループごとに正確に区切れているかというポーズ精度と、ポーズ間に含まれる語数の平均値等を評価項目に含め数値化すればよいだろう。ストレスに関しては評価者があらかじめストレスを置くべきと考えた語を決めておき、対象者の音読音声は何割程度ストレスを置いて音読しているかを評価することで数値化することが可能となろう。

2) イントネーションに関しては6件法という印象では評価が困難であると考えられる。イントネーションは理解の先にある、解釈・表現の領域である。熟達度を予測することを目的に音読評価するのであれば評価項目からはずすことを考えてもよいかもしれな

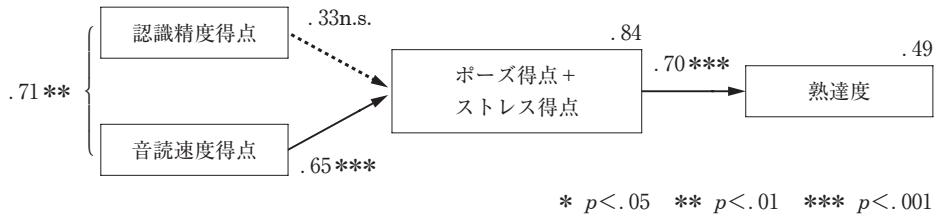


図3 ポーズ得点とストレス得点を合算し投入したモデル

い。イントネーションに関しては、英語母語話者間でも違いが生じることが少なくないため、一般疑問文における上昇イントネーション等違いが生じる余地のないものに限定することになる。しかし、これは句読法などにより予測が可能であるため、熟達度予測にはあまり役立たないと考えられるためである。近江 (1997: 24-26) はイントネーションを表現や表出に関わるところが大きいとしている。イントネーションの評価は音読の最終段階として別途評価することも考えてもよいだろう。

- 3) 実際図3のモデルに示した熟達度はあくまでペーパーテストの得点であり、スピーキングやライティングなどの能力は測定していない。あくまで理解中心のスコアであり、その点で言ってもポーズ得点とストレス得点を合算した数値が、熟達度のおよそ49%を説明しており、パス係数も.70と比較的高いことは納得できる。
- 4) 語認識精度得点からポーズ得点+ストレス得点へのパスが有意でなかった点について。単語を見ればすぐ発音できる能力は、音読速度との相関が高く、音読速度向上には役立つことが考えられるが、理解に直接寄与する可能性が低いかもしれない。単語の発音がスムーズであれば音読速度が向上し、音声面での自動化が促進される。その分、脳のリソースをポーズやストレスに回すことができるため理解が促進される。Rasinski (2010: 182) にある通り、テキスト中

音読できる単語 (語認識精度) が92%を下回る場合、指導されても理解が困難であるとすれば、音読評価の際語認識精度が92%を超えるものを選び、語認識精度は評価項目からははずすことも考える。

## 5. 本研究の限界と課題

本研究では音読の評価項目に関してある程度の方向性を示すことができた。しかし、音読評価は時間と労力を要するため、本研究では限られたサンプル数で分析を進めざるを得なかった。そのため適合度が十分であったとは言え、その結果に関しては信憑性を十分担保できず、再調査が必要である。今後、考察でまとめた視点から評価項目と指針を作成し、十分なサンプル数で追実験を行う必要がある。今回採用した変数のうち、熟達度に関しては前述の通り理解力を測定する問題で得られた得点であった。音読はスピーキングへの橋渡し (金谷・谷口, 1995: 4-9) であるとするならば、熟達度にはスピーキングやライティングの能力を加味したのもも変数として設定する必要があるだろう。

## 謝辞

本研究は、大学英語教育学会 (JACET) 関東支部大会 (2009)、全国英語教育学会全国大会 (2009)、関東甲信越英語教育学会大会 (2009) における口頭発表に加筆・修正を加えたものである。飛田ルミ (足利工業大学)、阿久津仁史

(東京都文京区立茗台中学校)の両氏には、データ分析等で多大なご協力をいただいたことに感謝申し上げたい。

#### 引用文献

- Beck, J.E., Jia, P., & Mostow, J. (2004). Automatically Assessing Oral Reading Fluency in a Computer Tutor that Listens. *Technology, Instruction, Cognition and Learning*, 1, 61-81.
- Coniam, D. (1990). Reading Aloud Speed as a Factor in Oral Fluency and General Language Proficiency? *Hong Kong Papers in Linguistics and Language Teaching*, 14, 47-69.
- Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System*, 27(1), 49-64.
- 石井正之助編. (1970)『講座・英語教授法第5巻 読む領域の指導』研究社：東京
- 伊藤健三他. (1995)『英語の新しい学習指導』リーベル出版：東京
- 金谷憲・谷口幸夫 (1995)『英語授業のアイデア集 英語教師の四十八手 スピーキングの指導』研究社：東京
- Klauda, S.L., & Guthrie, T. (2008). Relationships of Three Components of Reading Fluency to Reading Comprehension. *Journal of Educational Psychology*. 100(2), 310-321.
- Kuhn, M. R. & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3-21.
- 京堂政美 (1989). 「Reading Aloudと言語能力の相関について—言語能力の測定としての Reading Aloud」*STEP Bulletin*, 1, 117-129. 財団法人日本英語検定協会.
- 宮迫靖澗 (2002). 「高校生の音読と英語力は関係があるか?」*STEP Bulletin*, 14, 14-25. 財団法人日本英語検定協会
- 小川芳男編 (1982). 『英語教授法辞典新版』三省堂：東京
- 近江誠 (1997). 『英語コミュニケーションの理論と実際』研究社：東京
- Pendergast, T.M. (1985). OLAF 73 - Computerized Oral Language Analyzer and Feedback System. In Lee, Y.P., Folk, A., Lord, R., & Low, G. (eds.), *New Directions in Language Testing*. OUP.
- Rasinski, T.V. (2003). *The Fluent Reader*. Scholastic Inc.
- Rasinski, T.V. (2010). *The Fluent Reader. Oral & Silent Reading Strategies for Building Fluency, Word Recognition & Comprehension, 2nd ed.* Scholastic Inc.
- Rasinski, T.V. & Padak, N.D. (2008) *From Phonics to Fluency Effective Teaching of Decoding and Reading Fluency in the Elementary School*. Pearson Education, Inc.
- 鈴木寿一 (1998). 「音読指導再評価—音読指導の効果に関する実証的研究」『LLA (語学ラボラトリー学会) 関西支部研究集録』7, 13-28.
- 鈴木政浩 (2009). 「音読の評価をどうするか」『英語教育』11月号, 33-35.
- 鈴木政浩・阿久津仁史・飛田ルミ (2009). 英語音読評価に影響する要因—日米評価者の評価比較— A Comparison of Evaluation of Reading Aloud Performances Between Japanese and American Informants. 『大学英語教育学会 (JACET) 第3回関東支部大会 (青山学院大学) 要綱』, 27-28.
- 鈴木政浩・阿久津仁史・飯野厚 (2009) 「音読20回は妥当か? 音読練習ソフトを活用した反復音読中の音読スコアの推移分析 An Analysis of Computer Software Measurement and Its Gains of Repeated Reading Performance. The Efficacy of Twenty Times of Repeated Reading Aloud」*Language Education & Technology*, 46, 61-78.
- 田垣正義 (1990). 『落ちこぼれのない英語教育—誤答分析の視点から』南雲堂：東京
- 財団法人語学教育研究所編著 (1988). 『英語指導技術再検討』大修館書店：東京

### 資料1 評価者のプロフィール

性別・年齢 (日本滞在年数)	職業
男性・28歳 (1.5年)	大学教員
男性・25歳 (2年)	ALT
男性・29歳 (1年)	大学教員
男性・31歳 (6年)	大学教員

※年齢、職業等は実験実施時のもの

### 資料2 音読評価用紙

CD Track \_\_\_\_\_ ← Please write the track number of voice data CD

#### Transcript

A girl asked her father if she could go on a home stay program in Australia. She told him that she wanted to practice her English and meet interesting people. At first he didn't want her to go, because he didn't want her to travel alone. However, when her friend agreed to go with her, he let her go on the trip.

**Assessment 1** Please underline the words which you judged as having incorrect pronunciation while listening to the voice data. After evaluating, count the number of the underlined words then write the number. Errors include mispronunciations, substitutions, reversals, or omissions. If the text says one thing and the reader says another, or if the reader skips a word, it's counted as an error. If the reader hesitates or pauses for three seconds or more, it is also counted as an error. Repetitions are not counted as errors, though, nor are mispronunciations, substitutions, reversals, or omissions that the reader corrects him/herself.

**Assessment 2** Please evaluate the following three points with regard to enunciation and articulation. Please draw only one circle around the number you evaluated.

Intonation : (Inappropriate) 1 · 2 · 3 · 4 · 5 · 6 (Appropriate)

Pause : (Inappropriate) 1 · 2 · 3 · 4 · 5 · 6 (Appropriate)

Stress : (Inappropriate) 1 · 2 · 3 · 4 · 5 · 6 (Appropriate)