

令和元年6月13日現在

機関番号：15201

研究種目：基盤研究(C) (一般)

研究期間：2014～2018

課題番号：26381206

研究課題名(和文) 英語口頭運用能力に関する英語教員の主観的評価を分析するオンラインシステムの開発

研究課題名(英文) Developing online system to help users reflect upon their rating tendency on speaking skills

研究代表者

猫田 英伸 (Nekoda, Hidenobu)

島根大学・学術研究院教育学系・准教授

研究者番号：80452598

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究では、日本人英語教員に、自らが持っている内的なスピーキング評価基準について内省を促すためのオンラインシステムを開発した。本システムでは、ユーザである英語教員は、学習者が英語を話しているビデオ映像を視聴しながら、各学習者について語彙・表現の幅、文法的正確さ、流暢さ、および発音の質という4つの尺度(から選択したもの)に基づいて評価を行う。ユーザの評価はオンライン上で多相ラッシュ・モデルを用いて分析され、その後、ユーザに全体的な評価の厳しさ/甘さの傾向とともに、評価項目ごとの評価の厳しさ/甘さが日本人英語教員の一般的な評価傾向からどの程度逸脱しているかを視覚的にフィードバックする仕様とした。

研究成果の学術的意義や社会的意義

社会的意義：2020年度から実施される大学入学共通テストや、2019年度から3年ごとに全国の中学校で実施される外国語科(英語)の全国学力調査において4技能型の試験が導入され(てい)る。これを踏まえ、評価基準などに基づくスピーキングの評価に対する英語教員の意識を高め、自身の評価の妥当性について内省を促すことを目指す。

学術的意義：先行研究からも、評定者の特性によりスピーキングに対する評価の傾向が異なる可能性が示唆されている。本研究が開発したオンライン評価システムを活用することで、多様な評価者による評価データの収集・分析が容易になることは確実であり、今後の研究の発展につながる。

研究成果の概要(英文)：This study discusses a new online system that allows Japanese teachers of English (JTEs) to reflect on their internal speaking assessment criteria. The user (an English teacher) watches three movie clips of student speaking performances and then rates them using one or more scales on vocabulary, accuracy, fluency, and pronunciation. The system then employs the Many-facet Rasch Measurement (MFRM) to immediately analyze the user ratings by accessing a database developed from JTE mass speech sample ratings data (approximately 7,500 ratings). Sixteen English teacher trainees were tentatively analyzed using both the new system and FACETS (an MFRM statistical software), and the results were compared. It was confirmed that the newly developed system was able to accurately detect any deviations in the user ratings.

研究分野：言語テスト

キーワード：スピーキング評価

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

外国語教育の分野においては、近年、ますます英語のスピーキング能力の評価に関する研究が盛んに行われている。本研究課題に着手した2014年頃、国際的な言語テスト専門誌 Language Testing に掲載された論文を見てみると、本研究に特に関連の深い「スピーキング能力についての主観的評価」を扱った研究のみを取り上げたとしても以下のものがある: Shang, Y., & Eler, C., 2011; Carey, M., D., Mannell, R., H., & Dunn, P. K., 2011; Jin, T., Mak, B., & Zhou, P., 2012; Jin, T., & Mak, B., 2013; Winke, P., Gass, S., & Myford, C., 2013。

また、日本国内において、学校英語教育に対して音声言語面(リスニング、スピーキング)の指導の充実が各方面から声高に叫ばれはじめたのもこの時期であった。このことは、「相手の意図や考えを的確に理解し、論理的に説明したり、議論の中で反論・説得したりできる能力の育成」(外国語能力の向上に関する検討委員会, 2011)、「二者間折衝・交渉/他者間折衝・交渉ができるレベルの人材の潜在的候補者層を厚くする指導」(グローバル人材育成推進会議, 2012)、「中学校における英語による授業」(教育再生実行会議, 2013)など公的な組織から示された文書の中に、産出技能、特にスピーキング能力の育成に主眼を置いた記述が散見されることから分かる。

2. 研究の目的

本研究の目的は、英語音声(スピーチ・サンプル)に対して英語教員がさまざまな評価基準(言語特徴を記述した文)に基づいてください主観的評価を継続的に蓄積し、簡便に共有、分析するためのオンラインシステムを開発することである。ユーザはweb上でシステムにアクセスし、提示される種々のスピーチ・サンプルを評価基準に基づいて評価する。評価はデータベース化されるとともに、多相ラッシュ・モデルによって自動で処理され、ユーザ(評価者)には他の評価者(一般的な日本人英語教員集団)と比較した際の自身の評価傾向を即時にフィードバックする。

本研究を行う意義として、大別して以下の2点を挙げることができる。

- (1) 社会的意義: 2020年度(2021年度入学生)から実施される大学入学共通テストや、2019年度から3年ごとに全国の中学校で実施される外国語科(英語)の全国学力調査において4技能型の試験が導入され(てい)る。そんな中、英語教員に自身のスピーキングに対する評価傾向について他の評価者との対比のもとに認識し、内省を行うための機会を提供することの意義は大きいと考える。スピーキングの指導・評価を扱ううえで避けて通ることのできない(評価基準などに基づく)主観的評価に対する英語教員の意識を高めるとともに、自身の評価の妥当性について内省を促すことを目指す。
- (2) 学術的意義: 先行研究からも、評定者の特性によりスピーキングに対する評価の傾向が異なる可能性が示唆されている。それにもかかわらず、これまで「日本人英語教員」が、例えば英語母語話者と比して、どのようなスピーチ・サンプルに対して、どのような評価基準に基づいて、どのような主観的評価をください傾向があるのかなどについて検証した研究は決して多くない。本研究課題の遂行によってシステムが開発されることで、多様な評価者による評価データの収集・分析が加速することは確実であり、今後の研究の発展につながる。

3. 研究の方法

本研究において開発を試みる評価システムは、TOEIC や TOEFL などのテストのように受験者の解答を分析対象とするものではなく、スピーチ・サンプルを聞いて評定者がください主観的評価を分析対象とする。そのため、データの形式は「受験者×テスト項目」という2相データではなく、「評定者×スピーチ・サンプル×評価基準」という3相データとなる。本研究が開発を目指すオンライン評価システムには、少なくとも以下の3つのコンポーネント・機能が必要となる。

3相データを多相ラッシュ・モデルによって分析するためのプログラム

ユーザがパソコンに向かい、スピーチ・サンプルを聞きつつ、画面に表示されている評価基準に基づいて評価を入力するインターフェースを構成するプログラム

で多相ラッシュ・モデル内のパラメータの数値を変更したり、あるいは で再生するスピーチ・サンプルを選択的したりする管理者用のインターフェースを構成するプログラム

については、研究者自身が多相ラッシュ・モデルの数理モデルに基づいて数式を組み立てた。その際、Nekoda, Nekoda and Miura (2007)を初めとする過去の自身の研究において、類似の形式で日本人英語教員から収集した評価データとその分析結果を参照した。ただし、オンラインでの分析スピードと研究期間の制約の関係から本研究で必要となる最小限のサイズのプログラムを作成、使用することとした。(ここで言う「最小限」とは、評価対象のスピーチ・サンプルの【個数】を3個に固定するなど、プログラム上の「枠」を固定したことなどを指す。

具体的にどの3個のスピーチ・サンプルが提示されるかは多数のスピーチ・サンプルの中からランダムあるいは管理者による指定で選択できるようになっており、決してごく少数のスピーチ・サンプルのみが常に提示されるような「閉じた」システムではないことを念のため申し添える。）

および については、民間のソフトウェア開発業者と相談しつつ進めた。画面遷移やサーバ管理など詳細については適宜コンサルティングを通して仕様を固めながら開発を進めた。

4. 研究成果

本研究が開発したシステムの全体的な仕様・画面推移は以下のとおりである。主な画面のみ以下に示す。

- (1) ログインとユーザ登録 (図1、図2): ユーザ登録ではユーザ(評価者)に利用条件に同意した上で自身の属性の入力を求める。これにより、新規ユーザの評価データをそのままデータベースに追加できる仕様とした。
- (2) 自己診断したい評価基準の選択 (図3): システム内でこれから視聴するスピーチ・サンプルの種類と、用いる評価基準のカテゴリを選択する。
- (3) スピーチ・サンプルに対する評価の入力 (図4): (2) でいずれのスピーチ・サンプル、評価基準のカテゴリを選択したとしても、提示されるスピーチ・サンプルの【個数】は3個、提示される評価基準の【個数】は各カテゴリ4個である。
- (4) ユーザの評価傾向の分析 [ユーザ画面には非表示: バックグラウンドで処理]
- (5) ユーザへの分析結果のフィードバック (図5): 日本人英語教員の評価データベース基準とした場合の、ユーザの評価の全体的特徴と部分的な偏り(統計的に有意なものがあれば)が報告される。
- (6) ログアウト

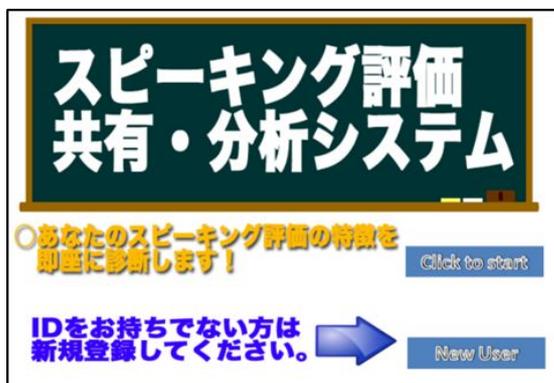


図1 ログイン画面



図2 ユーザ登録



図3 スピーチ・サンプルの種類と評価規準の選択



図4 スピーチ・サンプルの視聴と評価

1.35: やや厳しい

同じくらいの厳しさの日本人英語教員と比較して、以下の項目で評価に偏りが見られるようです。
 評価項目「ある程度の流暢さをもって、自分の考えを十分に表現することができる。」による評価が著しく低くなること
 があるようです。

評価項目「言葉に詰まること目立って多く、長い間が空くが、短いやりとりを行うことができる。」による評価が日本
 人英語教員の評価傾向と一致していません。

ラッシュモデルによる分析の結果（ロジットスケール表示）

→日本人英語教員の集団の中でのあなたの厳しさは●の位置

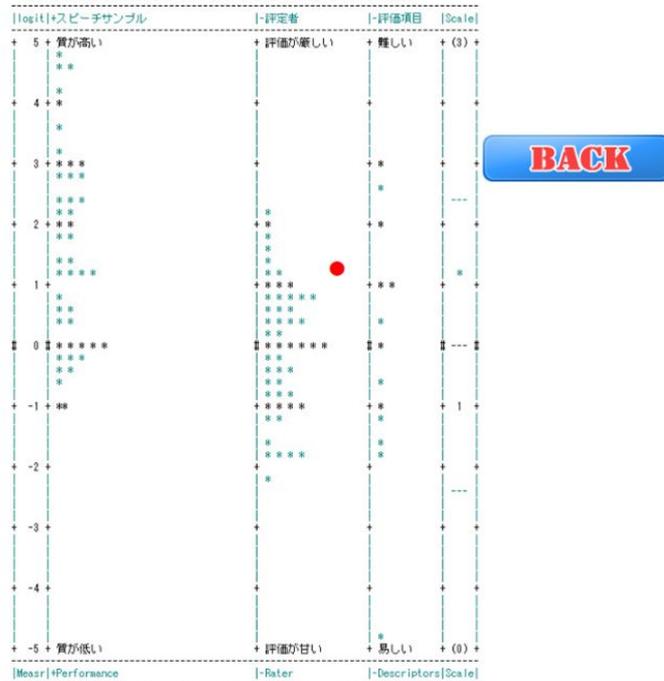


図5 フィードバック画面

本研究では、英語教員が簡便に自身の英語スピーキング評価基準について自己診断を行い、内省を深めるオンラインシステムを開発した。そして、実際に本システムを用いて教員養成課程に在籍している学生の評価基準を分析する事例研究を行うことでシステムの検証を試みたところ、本システムの出力結果は多相ラッシュ・モデルを用いた分析を行うための統計ソフトである FACETS ver.3.80.0 (Winsteps, 2017) の出力結果と一致することが確認された。このことから、今後はシステムをより大きな規模で活用していくことが期待される。

5 . 主な発表論文等

〔雑誌論文〕(計1件)

猫田英伸 (2019)「スピーキング能力の評価傾向を自己診断するためのオンラインシステムの開発」『中国地区英語教育学会研究紀要』49号, 21-31.

〔学会発表〕(計2件)

猫田英伸 (2018)「スピーキング能力の評価傾向を自己診断するためのオンラインシステムの開発」第49回 中国地区英語教育学会 山口研究大会

猫田英伸 (2014)「英語スピーチサンプルに対する主観的評価の傾向を自己診断できるオンラインシステムの開発」全国英語教育学会 第40回徳島研究大会

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

- 名称：
- 発明者：
- 権利者：
- 種類：
- 番号：
- 出願年：
- 国内外の別：

取得状況（計 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名：
ローマ字氏名：
所属研究機関名：
部局名：
職名：
研究者番号（8桁）：

(2) 研究協力者

研究協力者氏名：
ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。