

平成 30 年 6 月 22 日現在

機関番号：34426

研究種目：挑戦的萌芽研究

研究期間：2014～2017

課題番号：26580113

研究課題名(和文) グローバル人材の会話能力を評価する1万人のグループオーラルテストシステムの構築

研究課題名(英文) Group oral test systems for 10,000 students in order to evaluate global citizens' speaking skills

研究代表者

藤田 智子 (Fujita, Tomoko)

桃山学院大学・経営学部・教授

研究者番号：80329002

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：「グローバル人材の会話能力」を測るための対話型スピーキングテストは、3人同時受験、人と機械採点で客観的、効率良く短時間でできるテストとして開発された。しかし、音声認識システムの技術的問題により文字変換が正確性を欠き、現時点での実用化には複数の問題点が残された。限られた条件のもとで、人による全体的評価、機械による発語数と会話バランスの3評価基準で予備テストを試みた。少ないサンプル数ではあったが、他の評価基準に比べ、会話バランスが全体評価との低い相関を示した。また、受験者へのアンケート結果は、高得点を得るため間違いを恐れず発語数を増やすよう練習する等のテスト波及効果があることが明らかにされた。

研究成果の概要(英文)：The group oral test for measuring “global citizens’ conversation skills” was developed as an objective test which is able to evaluate three examinees simultaneously in a short time by a human and machine raters. However, because the current speech recognition systems still have technical problems and their conversion from voice data to character data lacks accuracy, several problems remained for their practical use. Under the limited condition, the pilot tests were conducted by three evaluation standards, a human rater with holistic scale, the number of words spoken by each participant (TNW) and conversation balance. In spite of insufficient data, the results indicated only a low coefficient correlation between conversation balance and total score. In addition, as ripple effects of the test, the students’ questionnaire results revealed that examinees tended to practice more in order not to be afraid of mistakes and to increase TNW for higher scores.

研究分野：教育測定学 英語教育学

キーワード：スピーキングテスト 対話型スピーキングテスト グループオーラルテスト インタラクション能力  
評価基準 機械採点 発語数 音声認識システム

## 1. 研究開始当初の背景

### (1) テスト開発の動機:

私の前任校 T 大学では約 1 万人の学生が受講する必須英語科目においてスピーキング能力評価は、各クラスの担当教員に委ねられていた。その評価は対話型スピーキングテストを中間と期末の 2 回、それぞれ 90 分授業 2 回分を費やして実施するが評価者が 1 人であるためか、信頼性は非常に低く、同じ受験者を 3 人の教員が評価した別のスピーキングテスト結果との相関関係は  $r = 0.046\sim 0.475$  であった (Fujita 他, 2013)。

### (2) 対話型スピーキングテストの特徴:

ペアやグループでのスピーキングテスト(対話型)は単独の話者では生み出せない話の流れがある (McCarthy, 2010)。受験者 1 人にインタビューするモノログ型に比べ、対話型スピーキングテストには多くの相違点がある。スピーキングの機能の中でも、対話型は情報伝達機能が少ないが、対話機能は 2 倍近いことが知られている (Taylor, 2001)。そして、インタラクション能力をより正確に測ることができる (Ducasse & Brown, 2009)。テストという特別な状況より、自然な会話のなかで実際に行われているやり取りに近い状況で能力を測ることができるからである (Johnson & Tyler, 1988)。

グローバル化が進む我が国で世界の人々と自由に言葉のやり取りができるタイプの人間が求められているなか、「グローバル人材の会話能力」を測るには、この対話型スピーキングテストが適切である。しかし、対話型では他の受験者からの影響を避けることはできないという問題点もある。例えば、ペアやグループの相手の性格が内向的か外向的に左右されることがある (Berry, 2004; Nakatsuhara, 2011)。他にも、他の受験者の性別、相手との親密さ、英語能力の違い、グループ内の輪がどのように形成されるかなどによって大きく影響を受けると言われている (例: Brown, 2003; Bonk & Van Moere, 2004; O'Sullivan, 2006)。

このような問題点もあるものの、グローバル人材養成をめざす 2020 年の大学入試改革も目前に控え、大量受験者に対応でき、インタラクション能力も測定可能な対話型スピーキングテストの登場が切に待ち望まれている。ここでは、まず今まで信頼性が低い評価を受けてきた可能性が高い T 大学の 1 万人の対話型スピーキングテストの改善が本研究の原点となった。

## 2. 研究の目的

従来の英語スピーキングテストでは重視していなかった、インタラクション能力、対人対応力、認知能力や集団形成力、これらを総称して「グローバル人材の会話能力」として能力を測定し、かつテストの波及効果を得るための対話型テストの開発をめざす。

まず、1 万人規模の必須英語科目の対話型スピーキングテストを改良するため、評価時間短縮のために 3 人同時受験と機械採点(発語数、話者間の発語数バランス)を導入する。さらに、

主観的評価とならないように機械と人による複合評価を実施する。本テストが実際にどのくらいの規模で実現可能か挑戦的萌芽研究として始め、将来はさらなる普及をめざしたい。大規模に実施することで、日本の学習者たちが「グローバル人材の会話能力」を伸ばし、シャイで発話しない傾向を打開して発語数を増やすというテスト波及効果をめざす。

## 3. 研究の方法

### (1) 被験者:

日本の大学必須英語科目を履修する 1 年生と 2 年生合計 40 人の学生が本研究に参加した。彼らは、習熟度別クラスの上級、中上級、中初級、初級の 4 レベルに属する学生たちで、各レベルからアトランダムに 5 組(10 人)ずつ被験者を抽出した。

### (2) 方法:

全員同じ課題についてペアの受験者たちが 5 分間英語で会話する。それぞれの受験者はヘッドセットをつけて発話は録画されると同時にそれぞれのパソコンに音声認識ソフト(ドラゴンスピーチ)によって文字入力される。評価は人による全体的評価と音声認識ソフトにより文字入力された各受験者の発語数と会話バランスである。テスト終了後、全受験者に本テストに対してのアンケート調査を実施した。

### (3) 人による評価:

従来の対話型スピーキングテストでは、分析的評価として発音、語彙、文法、流暢さ、会話方略などの小分けされた評価基準を用いて、受験者にフィードバックを返すのが一般的であった。Fujita 他(2013)では、各評価基準を総合評価点と比較してみた。相関係数は発音のみやや低く  $r = 0.825$  であったが、残り 4 つの評価基準と総合評価点は、 $0.938 < r < 0.965$  という高い相関係数を示した。この結果からみると全体的(holistic)評価として一括に評価しても結果が大きく変化しないことを示唆している。また、本研究においては、採点時間を短縮しテスト結果を迅速に示すことを優先するため、人による採点は全体的評価を実施することとなった。

### (4) 音声認識ソフトを利用した機械採点

機械採点では、受験者の発語数と会話バランス(2 人の受験者の発語数を 100%とし、ペアの 1 人ずつが 50%からどのくらいかけ離れているか数値化したもの)を 2 つの評価基準とした。

### (5) 研究の変更

音声認識ソフトが正確に機能しなかったせいもあり、発語数は人間が数えることになった。そのため、得られたデータの 1 部分しか利用できなくなった。

## 4. 研究成果

### (1) ペアか 3 人組か

初年度に早速 3 人のグループオーラルテストの予備実験を実施した。Nakatsuhara(2009)によるとペアより 3 人の方が対話相手からの影響は小さく、また 3 人の方が 4 人より個々の受験者の主体性が損なわれ、「Me too」のような発言が

増えると言われている。上級レベルの3人のグループオーラル5組のうち3組は、比較的バランスよく話すことができたが、初級レベルの5組中4組は受験者間で極端に発語数の違いがあり、極端に全員の発語数が少なかった。3人組の対話型テストはペアのオーラルテストに比べ特に初級の受験者間で、バランス良いやりとりをしづらい傾向を印象づけた。

また、2015年時点で3人の対話型テストを諦めざるを得なかった最大の理由は、音声認識ソフトドラゴンスピーチの問題である。本テストは音声認識ソフトによって受験者の発話を文字化し、テスト終了と同時にそれぞれの受験者が5分間で作り出した発語数を評価者に示すことを前提としていた。ところが、音声認識ソフトは3人のグループオーラルの会話を1人ずつの会話に分けることができない。従って、このソフトをインストールしたパソコン3台を、1台ずつ受験者に割り当て音声文字化する方法を試みた。しかし、音声認識ソフトの正確性は非常に低く、50~60%は、人が手入力で直す必要があった。その上、システムが途中で止まり文字変換をしなくなる事態が頻発し、1人の受験者に1人のサポートがついてシステムを絶えず動かすようにしたが、会話を中断せざる得ないことも多くあった。この為、実際は膨大な時間をかけて人が再入力して受験者の発語数を数えることになった。

以上の点から、3人組のグループオーラルは諦め、2015年より対象をペアワークに変更した。

### (2) グローバル人材の会話能力の測定

人による全体的評価、機械採点の発語数と会話バランス、これら3評価基準と受験者の総合テスト結果を比較した。人による全体的評価( $r = 0.84$ )や発語数( $r = 0.92$ )と相反して、会話バランスだけ低い相関係数を示した( $r = 0.25$ )。インタラクション能力、対人適応力、認知能力や集団形成力を測るのに、会話バランスを評価基準に含めたことが短絡的であったと考えられる。上級者の場合、発語数も多いなかで会話バランスも良いやりとりとなるが、初級者のペアでは、発語数が少なく、たとえ会話バランス良くても、拙くたどたどしいやりとりとなっている場合がある。今後、会話バランスは評価基準から外すか、発語数との兼ね合いで得点が変わるなどの工夫が求められる。

### (3) テスト波及効果

本テストの受験者にアンケートと一部インタビューを実施した。アンケートに最も多く書かれていたのは、「スピーキングテスト受験は初めてで最初はとても緊張した。」というものであった。確かに、ほぼ全員の受験者がパソコンを使用したスピーキングテストは初めてであった。それが、ヘッドセットを装着させられていきなりパソコンに繋がれ、同時にビデオ録画、ICレコーダーで録音など、驚き緊張するのも無理はない。しかし、中には「途中から面白くなってもう一度体験したい。」という声も複数聞かれた。

また、テスト終了後に「本テストは発語数が多いほど良い点数がとれる可能性が高くなる。」

と説明すると、「もっと積極的に発言すれば良かった。」とか、「間違ってもいいから、もっと話せば良かったという声が聞かれた。」ほぼ全員に近い受験者が、次に本テストを受けるときまでに、もっとたくさん話せるように練習したい。」と書いていた。

### (6) 今後の展望

現状で本テスト開発の最難関となっているのは、各受験者の発語数を個別に提示するシステムが開発できていない技術上の問題である。音声認識して文字化する技術はこの数年で急速に普及した。しかし、その技術はテストに使用できるほどではなく、複数の受験者の声を別々に音声認識してそれを正確に文字化するレベルには至っていない。特に初心者の英語学習者の発音を正確に音声認識して英語で文字化することは、現時点ではどの音声認識ソフトでも不可能だと思われる。

この事実を確認するために、Cambridge AssessmentのIELTSスピーキングテストの開発を行っている部門を2016年9月に訪れて、音声認識システムについて質問した。商業的テストを最前線で開発している団体であっても、今の技術では、3人の声を聞き分けて別々に発語数を数えることはできないとの回答であった。

本テストは、テスト終了とほぼ同時に受験者それぞれの発語数を提示できなければ、短時間で大量の受験者の評価に結びつかない。従って本テストの開発には、この音声認識システムの技術革新が重要な鍵を握っている。

### <引用文献>

- Berry, V. (2004). A study of the interaction between individual personality differences and oral performance test facets. Unpublished PhD thesis, King's College University of London.
- Bonk, W. J., & Van Moere, A. (2004). L2 group oral testing: The influence of shyness/outgoingness, much of interlocutor's proficiency level, and gender on individual scores. Oral presentation at The Language Testing Research Colloquium, California, USA.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25.
- Ducasse, A., & Brown, A. (2009). "Assessing paired orals: Raters' orientation to interaction". *Language Testing*, 26(4), 423-445.
- Fujita, T., Durand, J., & Tsuchiya, K. (2012). Fluency and word production in paired oral tests. *The Bulletin of Foreign Language Center Tokai University*. 33, 29-42.
- Johnson, M., & Tyler, A. (1998). Re-analyzing the OPI: How much does it look like natural conversation? In R. Young & A. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency*. (pp. 27-51). Amsterdam: John Benjamins.
- McCarthy, M. (2010). "Spoken fluency revisited." *English Profile Journal* 1: 1-15.
- Nakatsuhara, F. (2009). Conversational styles in group oral tests: How is the conversation co-constructed? Unpublished PhD thesis, the University of Essex.
- Nakatsuhara, F. (2011). Effects of test-taker characteristics and the number of participants in group oral tests. *Language Testing* 28(4), 483-508.

O' Sullivan, B. (2000). Towards a model of performance in oral language testing. Unpublished PhD thesis, the University of Reading.

Taylor, L. (2001). "The paired speaking test format: Recent studies". *UCLES Research Notes*, 6, 16-18. retrieved from [http://www.cambridgeesol.org/rs\\_notes/rs\\_nts6.pdf](http://www.cambridgeesol.org/rs_notes/rs_nts6.pdf)

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

藤田智子(2015)『大学英語教育プログラムの客観的評価手法の開発に関する研究』東京工業大学社会理工学研究科人間行動システム、博士論文、pp. 1-140.

藤田智子(2016)スピーキングの対話の評価『日本言語テスト学会誌：JLTA Journal』査読無、20周年記念特別号 第19号 pp.112-116.

〔学会発表〕(計1件)

藤田智子(2015)第1回言語教育プログラム評価アンケート結果の考察 日本言語テスト学会全国大会(JLTA 2015)

〔図書〕(計0件)

## 6. 研究組織

### (1)研究代表者

藤田智子(FUJITA, Tomoko)  
桃山学院大学・経営学部・教授  
研究者番号：80329002

### (2)研究分担者

( )

研究者番号：

### (3)連携研究者

( )

研究者番号：

### (4)研究協力者

中津原文代(University of Bedfordshire,  
Centre for Research in English Language  
Learning and Assessment)

### (5)研究協力者

仲谷佳恵(東京大学大学院)

### (6)研究協力者

濱田弘史(東海大学大学院)