

『外国語教育メディア学会(LET)  
関西支部2015年度春季研究大会』5/23  
於:大阪電気通信大学寝屋川キャンパス

# ペア型会話テスト:授業内での実 施と採点を目指して

シンポジウム 「英語教育における音声指導と評価:ス  
ピーキングと発音」

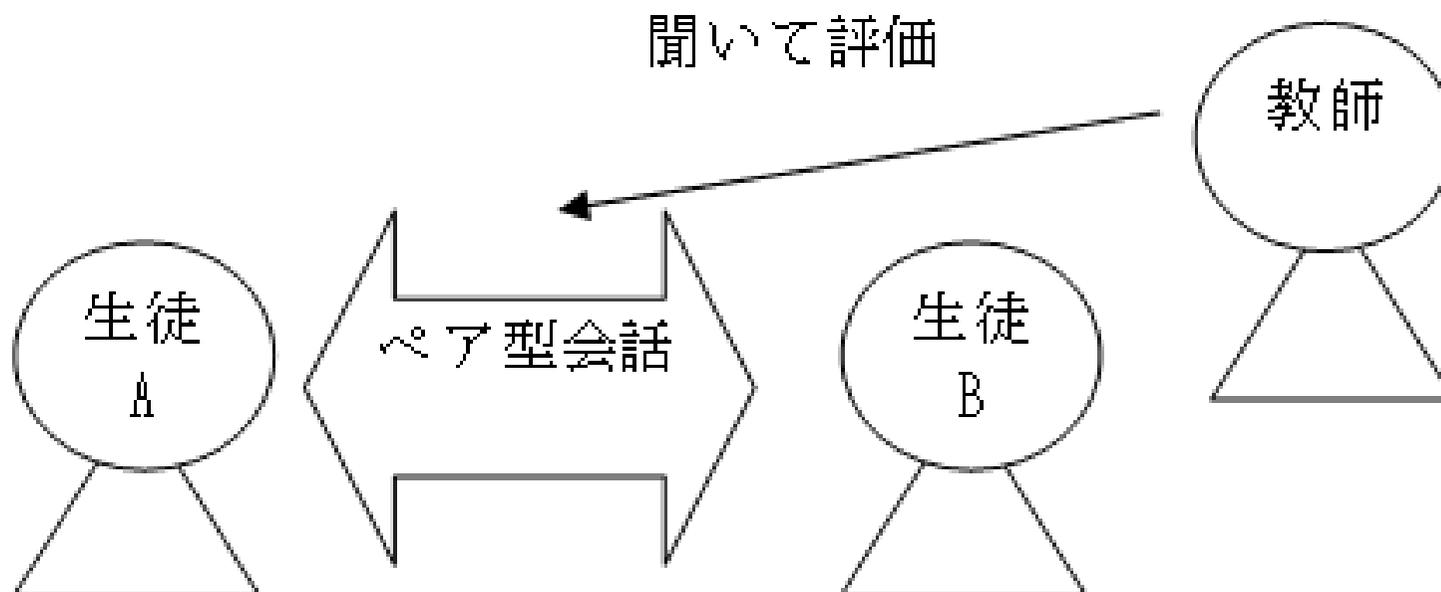
小泉 利恵 (Rie KOIZUMI) 順天堂大学

rie-koizumi@mwa.biglobe.ne.jp

資料Webに掲載

<http://www7b.biglobe.ne.jp/~koizumi/KoizumiHP.html> <sup>1</sup>

# ペア型会話テスト:どんなテスト?



- Paired orals (相互評価・自己評価も可)
- 授業でよく行うペア活動の評価

Role play: A: 店長、B: バイト希望者  
A←B 質問 (UCLES, 2010)

## Discussion: ABで同じ役割

- 共通の友達のKenが明日彼女 (girlfriend) と初めてのデートをします。2人に以下の質問をしてきました。意見を出し合い、Kenにしてあげるアドバイスを1つ決めてください。
- 質問: Kenがデート (date) に誘ったが、費用 (cost) を全部出すべきか。全部出さないとする、どのくらいは出すべきか (考える時間なし、3分話す)

## ペア型会話の評価：なぜ研究？

- ペア活動を授業では実施
  - 上手く評価に入れられないか
  - モノローグ型で測りにくい力も測りたい
- 言語テストニング研究では、グループ型会話 (group orals) とともに、研究が進んでいる (e.g., Taylor, 2011; Taylor & Wigglesworth, 2009)。日本でも必要

# スピーキング評価の現状 (小泉, 2014)

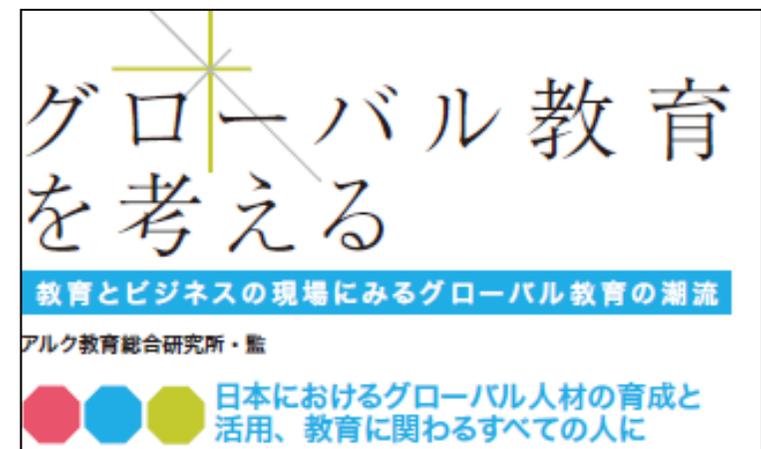
- 学習者が話す形式の評価
- 普通科「英語 I」では限定的 (文部科学省, 2011)
- 理由: 実施・採点に時間がかかる。  
タスク作成や採点の方法が知られていない
- ペア型会話は、スピーキング形式の重要な一つ

# スピーキングテストでのタスク

- モノローグ型
  - 音読、文の並び替え等、絵・地図・グラフの描写、まとまった内容の説明・意見陳述 (技能独立型・多技能統合型)
- 面接者との会話型
  - ロールプレイ、インタビュー
- 生徒間での会話型 (ペア型会話・グループ型会話)
  - チャット、ロールプレイ、ディスカッション 7

# 仕事での英語のスピーキング (小泉, 2015a)

- 仕事で英語を使う社会人825人
  - あなたが、直近1年以内で、仕事で英語を使った(話す、書く)シーンをすべてお選びください。
- 電話で話す (61.2%)
- 会議で発言 (48.4%)
- 交渉 (33.3%)
- プレゼン (31.3%)



## モノローグ < 会話型

- TOEICスコア別
- 465点以下の人 ( $n = 122$ )、470~595点の人 ( $n = 174$ )
- 電話で話す (35.2%, 48.5%)
- 会議で発言 (32.7%, 38.2%)
- 交渉 (23.0%, 38.2%)
- プレゼン (14.5%, 22.4%)

## 授業でいつ、どのように実施？

- 指導の直後に、録音なしで評価
  - 教師がその場で聞いて評価
  - 練習後、一部または全員の生徒が会話発表、評価 (個別 vs. 皆の前)
  - 自己評価・相互評価
- 指導の直後に、録音して後で評価
- 定期テストで、録音なしで評価
  - 別室で、教師評価
- 定期テストで、録音して後で評価

## ペアの組み合わせはどう決める？

- ・生徒間で決める vs. 教師が指定

## 評価ルーブリック(規準+基準)は？

- ・分析的尺度 vs. 全体的尺度

— 学習に役立つ vs. 詳細すぎずに継続可能

— 1・2分の会話だと、細かく評価できない

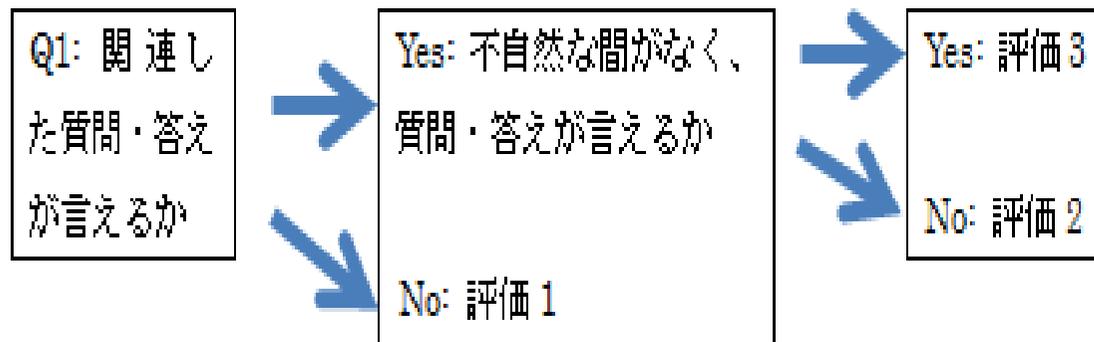
# 分析的尺度の例 1/3 (Ockey et al., 2014の改変)

- 1) 会話スキル
  - 参加度、会話のスムーズさ (発話交代、他者の発言に反応、質問、会話の開始、言い換え、間接的に述べる)
- 2) 言語表現、タスクの到達度
  - 流暢さ、語彙・文法、発音
- 研究では5～7レベルでの評価が多いが、授業の評価では3レベル程度では？

## 分析的尺度の例 2/3

- 1) 会話スキル: 参加度、スムーズさ
- 3: 以下が無理なく行える。①会話の開始。②他者の発言に適切に反応。③意味が不明なときに確認。④言い換え、明確化など、他者を助ける発言ができる。
- 2: ①～④ができるが、容易ではない
- 1: ①～④の中で、できないことがある

- 別例:  
(Ducasse,  
2010の改変)



## 分析的尺度の例 3/3

- 2) 言語表現、タスクの到達度  
– 流暢さ、語彙・文法、発音
- 3: ①発話に詰まり、誤りは少しあるが、意図は無理なく理解できる。かつ、②指示されたタスク (例: 3 best methodsを決める) ができる。
- 2: ①か②のどちらかに問題がある。
- 1: ①と②両方に問題がある。

# 全体的尺度の例 1/2 (Koizumi, In'nami, & Fukazawa, 2014; Nakatsuhara, 2013 に基づき作成)

- 観点: タスク実施に必要な英語表現を使用し、効果的にやりとりをしながら、与えられた役割を適切に果たすことができる
- 3: タスク・ポイントを適切にこなしている。  
。適切に交互に話しながら、効果的に英語でコミュニケーションができる。会話の進行がスムーズに行く程度に流暢に話している(これらをほとんど満たす)
- 「適切に」で発音も含めて見ている

## 全体的尺度の例 2/2

- 2: タスク・ポイントを一部適切にこなしている。ほとんどの場合にかろうじてコミュニケーションができるが、答えたり、意見を言ったりする際に受け身のことが多い(または、自分一人で話すことが多く、会話を独占してしまっている)。流暢さの点で会話の進行が妨げられることがあるが、英語で会話を続ける努力を行っている
- 1: 上をほとんど満たしていない

# タスク・ポイント:できるだけ明確に

- 質問指定型:

- 質問者は、指定された質問を全部適切にできる。関連する会話を適切に続けられる。
- 質問に答える者は全部の質問に適切に答えられる。関連する会話を適切に続けられる。

- 役割指定型:

- 与えられた役割や指定された発言を適切にできる。関連する会話を適切に続けられる。

- テーマ提示型・議論型:

- 関連したテーマで会話を適切に続けられる。結論まで到達しなくてもよい。

## 評価ルーブリック

- 会話スキル（コミュニケーションが効果的か） + 言語面など
- 3レベルの2が「合格基準をほぼ満たしている」、3が「合格基準を大きく満たしている」
- 1回聞いて評価できるように、評価ポイントを絞る。継続して評価しても負担にならないようなルーブリックを使う

# フィードバック返却によるよい影響 重視の場合：スコアレポートの例

生徒へのスコア・レポート

名前 ( 関西 花子 )

学習目標：タスク完成に必要な英語表現を使える。効果的にやりとりができる。与えられた役割を適切に果たすことができる。

基準：3(十分満足)、2(おおむね満足)、1(努力を要する)(詳細は別紙)

Task 1	Task 2	Task 3	Total
3/3	2/3	3/3	8/9
あなたの強みと弱み			
◎：特に良いところ    無印：普通    △：改善が必要			
・必要な英語表現が使える			
・スムーズに話すことができる			◎
・効果的にやりとりができる			△
・会話が続くように努力している			
・与えられた役割を適切に果たすことができる			◎

## どう評価する？ (深澤, 2015; 小泉, 2015b)

- ・基準の明確化
- ・採点者で、基準を合わせる練習
- ・2人以上の採点が望ましい
  - －教師・相互・自己評価の利用
  - －難なら、一部だけでも確認 (例:20%)
  - －特に、重みづけ多く、重要なテストでは実施
- ・評価規準を事前に生徒に示す

# なぜペア型会話テストが必要？ 1/2

- 形式によって引き出せる力が違う
  - Wang (2014): スピーキング研究のメタ分析。ペア型テストと他の形式の違いで大きな違い ( $d = 2.29$ , 95%CI: 1.93-2.65,  $k = 9$ )
  - Ockey et al. (2014): 即興Group oralとモノローグ型タスクとの相関は中程度 ( $r = .67-.76$ )
  - In'nami & Koizumi (in press): 過去のG-studyスピーキング研究 (28件) の統合。モノローグ型の中でのタスクの違いで順位が変わるのは 9.8~13.1% > 評価者1.9~6.4%

## なぜペア型会話テストが必要？ 2/2

- 熟達度によって、タスク難易度、力の出しやすさが変わる
  - Negishi (2015): 全体:(難易度高) 生徒間でペア型会話 > モノローグ型 > グループ型会話
  - 熟達度別: 上位者、下位者: paired > monologue、中位者: paired < monologue
- いろいろな形式を使わないと、有利不利が出てくる。不公平な評価になる
- 様々な形式・タスクを使うべき。1回ではできないため、実施回数も増やす必要あり

## ペア型会話テストの利点は？

- ・ 指導と一体化しやすい
- ・ 対話力が見られる（＞モノローグ）
- ・ 共に作り出す会話（co-construction）や、対等な会話（教師の援助がなく行う普通の会話）が引き出せる（＞面接者との会話型）
- ・ テストで見られる発話機能が豊富  
（＞モノローグ・面接者との会話型）
- ・ 教師は主に評価に専念できる。2人同時実施で時間短縮（＞面接型）

## 発話機能 (O' Sullivan et al., 2002)

- ・ 情報に関する機能
  - 意見・根拠を述べる      詳細に述べる
- ・ やりとりに関する機能
  - 修正する      意見を求める
  - 説得する      会話を修復する
- ・ やりとりを維持する機能
  - 会話を始める      トピックを替える
  - 決定する      会話を止める

## ペア型会話テストの難点は？ 1/3

- ・ 生徒だけでは会話が続かないこともある。指導が必要
- ・ 奇数人数の場合、3名で行う場合もある（または教師とペア）。3名だと参加しない生徒も出る
- ・ 誰とペアになるかで結果が変わるかも：性格、親しさ、英語力、会話の独占度 (e.g., Ockey, 2009)

## ペア型会話テストの難点は？2/3

- 授業直後に数名を評価すると、授業間で異なるタスクを使った場合には、タスクが違ってしまいう
  - 継続的に、繰り返しやることでブレは修正
  - 指導重視の形成的評価 (formative assessment)、学習のための評価の視点 (assessment for learning [AfL]; learning oriented language assessment [LOLA]) の一環として、学習への良い波及効果、指導・学習に役立つ情報の取得を重視 (vs. 総括的評価: summative assessment)

## ペア型会話テストの難点は？ 3/3

- 2人同時に採点は難。練習必要
- 2人で作り上げた会話を評価。1人の能力として捉えてにくい (McNamara, 1997)
- スピーキング力に加えて、リスニング力も問うテスト
  - ・ 対話力を測るなら問題ない。主に測る力はスピーキング力。必要なリスニング力は基本的なもの
- 難点 < 利点 (Galaczi & ffrench, 2011; O' Sullivan & Green 2011)

# 実施上の注意点は？ 1/6

- ☆タスクの設定：難易度に関わる要素
- どのような状況設定か
  - よくある場面、交渉が必要な困難な場面にするか、会話時間、意見の一致を求めるか
- どの程度指定、補助するか
  - 話す流れ・機能、使用表現 (Zero vs. hints vs. phrases)
- どのくらい準備時間をとるか
  - 即興～5分(短い方が自然な会話; Nitta & Nakatsuhara, 2014)。課題事前提示？

## 実施上の注意点は？2/6

- ☆録音だけで2人の声を聞き分けるのが難しいことあり
  - タスクごとに、名前を言うような設定
  - 話し始める人を指定（話し始める力は見られなくなる）
- ☆隣のペアの声で生徒の声がよく聞こえないことあり
  - 離れて座るように指示
  - 別室実施

## 実施上の注意点は？ 3/6

- ☆一斉実施だと、テストタスクの統制が難しい（見ながらやってしまう等）
  - 見ないように指示
  - 教師の直接の評価を少しは入れる
- ☆録音を後で採点する時間が取れない
  - 原則、その場で評価。Backupとして録音

## 実施上の注意点は？4/6

- ☆授業中の採点中に、他の生徒が話しかけてくる
  - 定期テストでは別教室で行う
- ☆先にやるペアが準備時間が少なくて不利になる
  - 別なペアで再挑戦できる設定
  - テストごとに違うペア、異なる順番で行う
  - 1/3をとった生徒には、指導後再テスト<sub>31</sub>

## 実施上の注意点は？5/6

- ☆授業と同じタスクで評価でいいか
  - 学んだ表現を同じ状況で使えるかを見るなら同じでもよい。即興で、学んだ表現を使えるか、似た状況で使えるかを見たいならば、タスクを若干修正、または似た応用タスクを使う
- ☆会話が途中で止まってしまおう
  - ペア同士、または教師に、質問OKとする(減点など方法を統一しておく)
  - ヒント提示。方法を決めておく。Dynamic assessmentを行う可能性につながる

## 実施上の注意点は？6/6

- ☆十分に安定した (信頼性の高い, .70以上) 結果を得るには、評価者何人、タスクはいくつ必要か (Koizumi et al., 2014)
  - 全体的尺度の場合：評価者1名時、タスク3つ、2名ではタスク1つ
- ☆タスクごとに評価を出すか
  - 2個以上のタスクを合わせて、1つのスコアを出すのでもOK。1分以内のタスクだと、複数のタスクで1つでないは無理

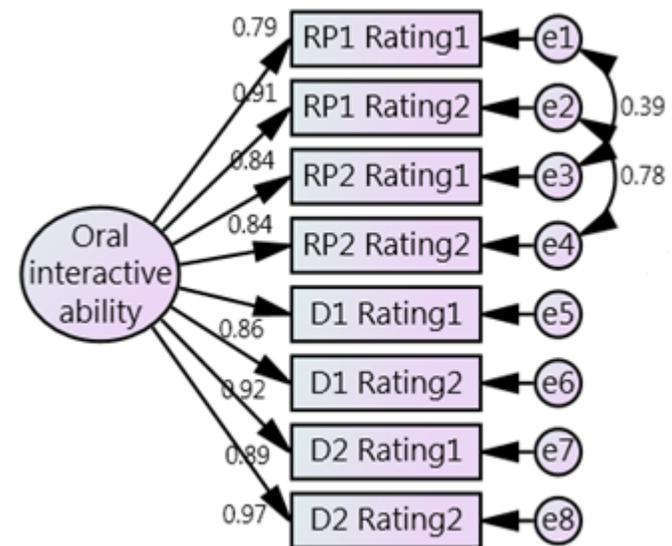
# 大学入試改革で活用が検討されている、英語4技能試験との関係は？

- ・ ケンブリッジ英検 (KET [Key English Test]~CPE [Certificate of Proficiency in English] 全て) にペア型テスト形式あり (Taylor, 2011)
- ・ TEAP (Test of English for Academic Purposes) に試験官に対してインタビューする形式あり
  - 相手は試験官だが、使う英語は、ペア型テスト形式に近い面あり (Nakatsuhara, 2014)
- ・ それ以外のテストでも、ペア型会話の指導と評価で伸ばしたスピーキング能力は生きる

# ペア型会話テストの妥当性は？

(Koizumi et al., 2014)

- $N = 163$  日本人大学生。5 観点から検討。さらに継続
- タスク・評価者・ルーブリックはRaschモデルに適合
- TOEFL ITPとの相関 .38。異なる能力測定
- ロールプレイとディスカッションのタスクは似た能力を測定。一次元構造



# 今後の研究

- ・ 日本人英語学習者におけるペア型会話の研究は少ない(例外: Koizumi et al., 2014; Negishi, 2015)。今後の研究・実践が必要
- ・ 適切なタスク、タスクプールの開発
- ・ ペアの組み合わせの影響
- ・ ペア型会話が何を測るか(構成概念)
  - 教師主導型のロールプレイとの違い
- ・ 他のテスト形式と比べ、発話や波及効果がどのように異なるか
- ・ 複数回実施し、どのように力が伸びるか

興味を持たれた方へ

- 理論編
  - スピーキングの評価全般
- 実践編
  - 様々な形式のテスト

# 英語4技能評価の理論と実践

—CAN-DO・観点別評価から技能統合的活動の評価まで

望月 昭彦・深澤 真・印南 洋・小泉 利恵 編著



## ペア型会話テストのまとめ

- 難点もあるが、利点の方が上回る
- 利点：
  - 継続的に実行可能。学習に良い波及効果ありの可能性
  - 対話力が測れる
  - 教師は主に評価に専念できる
- 最初は無理せず、授業内でできる形を考える
- 有力な形式だが、これだけでも偏る。他のテスト形式とともに使う

# 謝辞・引用文献

- 科研の共同研究者(印南洋先生、深澤真先生)
- 平成26, 27, 28, 29, 30年度科学研究費補助金, 基盤研究(C), 日本学術振興会. No. 26370737
- Ducasse, A. M. (2010). *Interaction in paired oral proficiency assessment in Spanish*. Frankfurt am Main, Germany: Peter Lang.
- 深澤真. (2015). 「Q&A より良いテストの作り方・使い方 スピーキングの評価・2: テストの採点方法」. 『英語教育』, 63(11, 1月号), 62–63.
- Galaczi, E., & French, A. (2011). Context validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 112-170). Cambridge, UK: Cambridge University Press.
- In'nami, Y., & Koizumi, R., (in press). Task and rater effects in L2 speaking and writing: A synthesis of generalizability studies. *Language Testing*.
- 小泉利恵 (2014). 「スピーキング評価の実際」『全国英語教育学会第40回研究大会記念特別誌 英語教育学の今—理論と実践の統合—』 (pp. 82-85). 東京: 全国英語教育学会

- 小泉利恵 (2015a). 「グローバル社会で通用する英語」アルク教育総合研究所編『グローバル教育を考える』(pp. 141–192). 東京:株式会社アルク
- 小泉利恵 (2015b). 「スピーキングの評価」望月昭彦・深澤真・印南洋・小泉利恵 (編)『英語4技能評価の理論と実践—CAN-DO・観点別評価から技能統合的活動の評価まで』(pp. 43-57). 東京:大修館書店
- Koizumi, R., In'nami, Y., & Fukazawa, M. (2014, September). *Rating scale for paired oral assessment in Japanese classrooms*. Paper presented at the 2nd British Council Symposium on Supporting Reform in Education New Directions English: Role of English assessment in Internationalization, Meiji Kinenkan, Tokyo, Japan.
- McNamara, T. F. (1997). "Interaction" in second language performance assessment: Whose performance? *Applied Linguistics*, 18, 446-466. doi: 10.1093/applin/18.4.446
- 文部科学省 (2011). 「平成22年度公立高等学校における教育課程の編成・実施状況調査(B票)の結果について」. Retrieved from [http://www.mext.go.jp/a\\_menu/shotou/new-cs/1301650.htm](http://www.mext.go.jp/a_menu/shotou/new-cs/1301650.htm)
- Nakatsuhara, F. (2013). *The co-construction of conversation in group oral tests*. Frankfurt am Main, Germany: Peter Lang.

- Nakatsuhara, F. (2014). *A research report on the development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese university entrants—Study 1 & Study 2*. Tokyo: Eiken Foundation of Japan. Retrieved from <http://www.eiken.or.jp/teap/group/report.html>
- Negishi, J. (2015). Effects of test types and interlocutors' proficiency on oral performance assessment. *Annual Review of English Language Education in Japan*, 26, 333-348.
- Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing*, 31, 147-175. doi:10.1177/0265532213514401
- Ockey, G. J., Koyama, D., Setoguchi, E., & Sun, A. (2014). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing*, 32, 39–62. doi:10.1177/0265532214538014
- Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing*, 26, 161–186. doi: 10.1177/0265532208101005

- O’Sullivan, B., & Green, A. (2011). Test taker characteristics. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 36–64). Cambridge, UK: Cambridge University Press.
- O’Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing*, 19, 33–56. doi: 10.1191/0265532202lt219oa
- Taylor, L. (Ed.). (2011). *Examining speaking: Research and practice in assessing second language speaking*. Cambridge, UK: Cambridge University Press.
- Taylor, L., & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 26, 325–339. doi:10.1177/0265532209104665
- University of Cambridge ESOL Examinations (UCLES) (2010). Speaking test preparation pack for Key English Test. Cambridge, UK: Author.
- Wang, L. (2014). A meta-analysis of peer-peer interaction in L2 English speaking assessment. *English Teaching & Learning*, 38(3), 103–137. doi:10.6330/ETL.2014.38.3.04
-