清泉女子大学
Seisen University

英語スピーキングテストを大学入試に導入する国内外の取り組み

小泉利恵（清泉女子大学 言語教育研究所）

【挑戦科研】令和３年度第１回研究会
オンライン開催

2021年9月29日
10:15～10:45

大学入試での背景

- 2019年11月　大学共通テストの枠組みでの英語民間試験の使用の見送り
  - 2024年度より導入の可能性は残す
- 2021年6月　文部科学省の大学入試に関する有識者会議　2025年1月以降の大学入学共通テストでの記述式問題と英語民間試験の導入を断念
- 2021年7月　文部科学省　英語民間試験の活用など、入試改革に積極的に取り組む大学の補助金を増やす制度を創設する方針。各大学の個別入試改革を促進

スピーキングテスト（ST）が最も導入が困難

- 実施・採点が困難。実行可能性が低くなりがち
- テスト不安や障がいを持つ受験者へのケアが必要
- **その中でどのように導入するとするとどのような方向があるか？**
- 4技能民間試験を必須化。受験要件に追加
- 2次試験用のSTを作成、導入
- 高校でのST結果を入試にも使う

4技能民間試験を必須化
受験要件に追加

- 可能だが、共通テストに導入するときと同じ問題がある。以下は例
  - 異なる構成概念（測る力）を持つ異なる民間試験を比較することが適切か
  - 学習指導要領に合わないものもある
  - 公平性の担保が難しい。格差が起きやすく、皆同じ条件で試験実施が難しい
  - 採点の信頼性の確保が難しい
- 各大学が4技能テストを作る方法、面接でスピーキングを入れる方法等よりは容易

4技能民間試験を必須化
受験要件に追加

- 入試のいろいろなタイプや側面に導入検討できる
- 一般入試、2次試験に導入
- AO入試に一部に導入

2次試験用のSTを作成、導入

- 例：
- 京都工芸繊維大学のST
  - https://kitspeakee.wordpress.com/
- BCT-S: 大学入試用英語ST
  - https://www.britishcouncil.jp/exam/bct-s/about
  - http://www.tufs.ac.jp/NEWS/admission/180608_2.html
  - British Council TUFS- Speaking Test for Japanese Universities
  - 東京外国語大学とBritish Councilが共同開発
  - タブレット実施12分

BCT-S: 大学入試用英語ST

- タスク：4個
- 採点：総合的（holistic）
  - 評価する分野例：1) 与えられた課題に対する達成度 2)トピックに沿っているか 3) 文法の幅広さと正確性 4) 語彙の幅広さと正確性 5) 発音と流ちょうさ 6) 内容のまとまり
- Aptisが原型。日本の東京外大の大学入試用にタスクを修正（localization）

BCT-S: 大学入試用英語ST

- 2021年　東京女子大も導入
- 他大でも導入可能
  - 検討課題：予算、タスクの共有のためには、東京外大や東京女子大との同日開催が必要

BCT-S: 大学入試用英語ST

- 共通テストに導入時の難点
  - 異なる構成概念（測る力）を持つ異なる民間試験を比較することが適切か
    - → 1つのテスト。STのみ追加導入
  - 学習指導要領に合わないものもある
    - → 基づく
  - 公平性の担保
    - → 筆記試験と同じ部屋で同日開催
  - 採点の信頼性の確保が難しい
    - → 訓練された資格を持つ採点者

BCT-S: 大学入試用英語ST

- https://www.britishcouncil.jp/about/press/bct-s
  - 5日以内に結果が出る
  - 短時間の実施が可能。ある程度実施を委託可能
- プラスの証拠は提示されているものの、今後さらなる妥当性検証が必要
  - 構成概念の確認
  - 大学や将来必要になる力との関係
  - プラス・マイナスの波及効果
  - 詳細は小泉（2018a）参照

高校でのST結果を入試にも使う

- 教室内テストの活用
- School-based assessment (SBA)
- コースワーク評価（飯田, 2006）
  - この後紹介するのはすべて外国語科目としてのテスト
- 長所：生徒の力、指導内容等に柔軟に合わせて評価
- 短所：学校・教員によって評価が異なる

高校でのST結果を入試にも使う

- 使用には発想の転換が必要
  - 評価を学習のために役立つものとし，「すべての生徒が自分の持つ力を最大限に示す機会を与える」という意味での公平性を担保（Davison, 2007; Tierney, 2013）

- 以下、小泉（印刷中）より抜粋

教室内テスト：香港

- 大学入試（Hong Kong Diploma of Secondary Education Examination: HKDSE）の15%
- 他教科でも同様に使われ，日本の大学入試センターにあたるHong Kong Examinations and Assessment Authority（HKEAA）が運営母体
- ◇大学入試の内訳
- SBA 15%　　リーディング20%
- ライティング 25%
- リスニングと技能統合問題 30%
- スピーキング（面接官と会話、受験者4名で議論）10%

教室内テスト：香港

- ◇SBAの実施方法
- 高2と高3
- 授業中に，①個人やグループごとにテストを行うか，②録画してクラス一斉に行うか，③個別実施するか，また③の際には他の生徒はテストの様子を見ているか，他の活動を行っているか等，柔軟に決めることができる。
- ◇SBAの採点方法
- 授業担当教員がテスト時に採点か，録画を見て採点する。
- 事前と事後に地域、校内での研修がある。
- 学校単位で調整される。

教室内テスト：ニュージーランド

- 高1～3の間に毎年，学年末試験として実施
- 最終学年の結果は大学入試の出願要件
- 国の組織が運営
- ◇タスク形式（interact：3～5分）
- あるトピックについて，生徒同士のペアかグループでやり取りを行う。
- ・例：「学校や制服について」「ニュージーランドには猫カフェは必要か」
- ◇採点方法：授業担当教員が録音・録画で授業外に採点。テスト内容や採点等は学内で確認作業が入り，サンプルの発話記録は学外に送られ，他校と比較。校内外での研修や調整

教室内テスト：イギリス（GCSE & A level）

- 義務教育の終了する16歳でGCSE（General Certificate of Secondary Education）と呼ばれる統一試験を受け，その結果が成績付きの中等教育修了証明書となる。
- 大学進学希望者は2年間の専用コースで，普通は3科目を勉強し，18歳で大学入試にあたるA level（General Certificate of Education Advanced Level）と呼ばれる統一試験を受ける。
- 試験を作成・採点するテスト団体が複数存在
- テスト問題やタスク，配点などが異なる。
- 教師が校内で1対1で実施して録音。採点は，テスト団体が実施する。

日本での教室内ST実施・採点

- 多くある中の一例：宮城県石巻高校（小泉、印刷中）。授業内実施・授業内採点

	第 1 回 6 月	第 2 回 9 月	第 3 回 11 月	第 4 回 2 月
高 1：コミュ英 I	生徒 2 名での会話☆【状況 1】			
高 1：英語表現 I	個人 Show & Tell#【状況 2】			
高 2：コミュ英 II	Chain Talk（生徒 3 名で会話）☆【状況 3】			（生徒 4 名で）☆
高 2：英語表現 II	教員と生徒 1 名で面接形式☆	個人 Show & Tell#	個人で順番にスピーチ（教員と生徒 3 名がいて，話者以外が聞く形式）☆【状況 4】	
高 3：コミュ英 III	生徒 3 名でディスカッション☆	教員と生徒 1 名の面接☆【状況 5】	2 グループ（4 名ずつ）でディベート#【状況 6】	
高 3：英語表現 II	個人プレゼンテーション☆	生徒 4 人でディスカッション#	3～4 名でグループワーク発表#【状況 7】	

宮城県石巻高校でのST採点方法

- ①事前の打ち合わせ 10分
  - 同学年・学年間でのすり合せのために,英語科会議でテストの案を持ち寄り,改善（タスク確認）
- ②A　教員2名（授業担当者と別教員）が採点（高3）
- または
- ②B　教員1名がST時に採点。可能な範囲で録画を後で参照して再採点（高1, 2）

小泉他 (2021)：方法

- 高3　40名6クラス 225名分析
- 精密な採点者トレーニングはできなかった
  - 10分ほどの事前打ち合わせのみ
- その状態で、どの程度信頼性（評価者内、評価者間の安定性）が保てるか？
- テスト2回分　Group discussionとGroup debate
- 評価者：生徒1名を2名の教員で採点
- タスク：グループ型ディスカッション、グループ型ディベート
  - ともに、発言順番固定、役割の割り当てあり

小泉他 (2021)：分析

- Q: シンプルなルーブリックを用い、長時間の採点者トレーニングがない場合、 採点者は何名いれば信頼性が確保できるか？
- 多変量一般化可能性理論 (Grabowski & Lin, 2019; 小泉, 2018b)
  - mGENOVA（University of Iowa, 出版年不明）使用
  - テスト得点のばらつき = 受験者能力の違い+ 採点者の厳しさの違い + 誤差　　「受験者能力の違い」を測る度合いが高い = 信頼性が高い
- その他、多相ラッシュ分析 (Linacre, 2021; McNamara et al., 2019; 平井他, 2018)
  - Facets使用
  - 採点者やルーブリック観点の質も分析

小泉他 (2021)：結果

・両方とも高い信頼性。1名で十分
　グループ型、1名採点者で Φ ＝ .79 ~ .81
・グループ型でも、自由度が小さいタスクの場合。採点者全員が同じ学年、同教材で指導・評価。日頃からの教材や教育理念の共有あり

- 高校でのST結果を入試にも使う
  - そのためにはある程度の信頼性確保の手順導入が必要？（ガイドライン作成、小泉, 2021）
  - 高校でのST結果を高校卒業試験に使う方がわかりやすい

まとめ

- スピーキングテストの大学入試への導入
- 3方向
- 4技能民間試験を必須化。受験要件に追加
- 2次試験用のSTを作成、導入
- 高校でのST結果を入試にも使う
- どれが適しているかは状況によるだろう
- 今後、国内の現状把握、利点欠点を含めた議論が必要

参考文献

- Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly*, *4*(1), 37-68. https://doi.org/10.1080/15434300701348359
- Grabowski, K. C., & Lin, R. (2019). Multivariate generalizability theory in language assessment. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment* (Vol. I: Fundamental techniques; pp. 54-80). New York, NY: Routledge.
- McNamara, T., Knoch, T., & Fan, J. (2019). *Fairness, justice, and language assessment*. Oxford University Press.
- Linacre, J. M. (2021). *A user's guide to FACETS Rasch-model computer programs: Program manual 3.83.5*. http://www.winsteps.com/manuals.htm

参考文献

- Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125-144). SAGE.
- University of Iowa, College of Education.（出版年不明）. Computer programs. https://education.uiowa.edu/centers/center-advanced-studies-measurement-and-assessment/computer-programs
- 飯田直弘 (2006). 「イギリスのGCSE試験におけるコースワーク評価の研究―導入から現在までの内容と方法の変化に関して―」 『国際教育文化研究』, 6, 113-124.
- 小泉利恵 (2018a). 『英語4技能テストの選び方と使い方―妥当性の観点から―』 東京：アルク
- 小泉利恵 (2018b). 「一般化可能性理論」 平井明代（編著）. 『教育・心理・言語系研究のためのデータ分析―研究の幅を広げる統計手法』 (pp. 65-93). 東京図書

参考文献

- 小泉利恵 (2021, 9月). 「教室内英語スピーキングテストの適切な採点に向けたガイドライン」令和3年度大阪大学マルチリンガル教育センター公開講座「英語教育オンラインセミナー」
- 小泉利恵 (編) (印刷中, 2021年11月発行予定). 『実例でわかる 英語スピーキングテスト作成ガイド』大修館書店.
- 小泉利恵・初澤晋・磯部礼奈・松岡京一 (2021). 「日本の高校におけるスピーキング評価の採点者信頼性—教室内グループ型のディスカッションとディベートの場合」未刊行原稿.
- 平井明代・横内裕一郎・加藤剛史 (2018). 「項目応答理論」 平井明代 (編著). 『教育・心理・言語系研究のためのデータ分析—研究の幅を広げる統計手法』(pp. 94-137). 東京図書