

令和 6 年 5 月 13 日現在

機関番号：11101

研究種目：若手研究

研究期間：2020～2023

課題番号：20K13101

研究課題名（和文）スピーキング要約課題のための具体的な発話内容の評価基準設定

研究課題名（英文）Concrete standard setting for contents of utterance in oral summary tasks

研究代表者

横内 裕一郎（Yokouchi, Yuichiro）

弘前大学・教育推進機構・助教

研究者番号：40782800

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、口頭要約課題のパフォーマンス評価を行う際により効果的なルーブリックの開発を目標としていくつかのルーブリックが作成された。その中で、Performance Decision Tree (PDT; Fulcher, et al., 2011) と呼ばれるルーブリックの形式は作成に手間がかかる一方で、極めて多様なフィードバックを学習者に返すことができるため、有効であると判断した。PDTは項目数（記述子）の数を n とすると 2^n の乗までフィードバックの種類を増やすことができるため、指導上有意義であると判断した。

研究成果の学術的意義や社会的意義

本研究の背景として、現在外国語教育のみならず、他の教科教育、人事評価、医療現場における診断など、様々な領域において評価が行われている現状においてより効果的である可能性の高いルーブリックの形式（PDT）を示すことができたことに学術的・社会的意義があるといえるだろう。

現状、公開済みの研究成果、及び現在査読中の論文では部分的なものではあるが、PDTの有用性を示すことができた。この形式のルーブリックの有用性をさらなる研究を通じ示すと同時に、今後この形式のルーブリックを応用した教材・教科書を刊行し、学習者・指導者に使用してもらえるように努めたい。

研究成果の概要（英文）：In this study, two rubrics were developed with the purpose of developing more effective rubrics for assessing performance on oral summary tasks. The Performance Decision Tree (PDT; Fulcher, et al., 2011) rubric format, while complicated to create, was found to be effective because it provides an extremely wide variety of feedback to students. Since the PDT can increase the number of feedback types up to the n th power of 2 when the number of items (descriptors) is n , it was determined that the performance of the PDT can be fine-tuned to the actual performance of the learner. However, at present, the validity, reliability, practicality, and washback have not been fully examined, and only partial examination of these factors has been completed. It is recognized that additional research is needed in the future to collect comments from learners, mainly to examine the study of washback effects.

研究分野：言語テスト

キーワード：PDT 妥当性 信頼性 実用性 波及効果 ルーブリック評価 パフォーマンス評価

1. 研究開始当初の背景

本研究で着目している口頭要約とは、文章を読んだり聞いたりしたあとに、記憶した内容を再生したり、理解した内容をオリジナルの言葉に置き換えたりして文章の内容を整理して話す活動である。この活動はテストタスクとして様々なテストで使用されており、実生活で似た状況が発生しやすいため真正性が高く、回答者全員が同じ情報を得て回答することになるため、与えられた文章の内容を理解することができる場合、公平性も高くなるため優れたタスクであると言える。スピーキングパフォーマンスを評価するためのタスクでは、一般的に発話の流暢さや表現・文法・発音などの正確さ、そして発話の内容（内容の一貫性や収束性）、あるいはタスクの完遂度を評価することが多い。しかし、発話内容自体を扱ったルーブリックは、研究開始当時にはほとんどみられず、特に日本人英語学習者を対象とした要約課題における発話内容の評価基準で適切に妥当性の検証が行われた研究は無かった。

要約課題に似た課題である再話 (retelling) のルーブリックは Hirai and Koizumi (2008, 2013) など、いくつか開発されている。これらの研究では日本人英語学習者を対象としたルーブリックの開発がなされており、妥当性の検証を適切に行なっている。その後、Hirai and Yokouchi (2019) で発表されたチェックリスト形式のルーブリックでは、発話内容の評価基準を設定しているが、『6点: 大枠と詳細が十分あり、時間内によくまとまっている』、『4点: 詳細も少しある』、『2点: 大枠だけで詳細がない』、『0点: 内容が違う』、となっており、具体性に乏しい。また、タスクの特性は要約と再話で類似しているとは言え、再話よりも要約のほうが発話の自由度がより高くなる傾向があるため、そのまま上記のルーブリックを援用することはできない。

口頭要約課題のための発話内容を評価するためのルーブリックを構築する上で最大の難関は、刺激文の内容や文章の形態に応じてルーブリック内の記述子を作成してしまうと、全く異なるテーマを扱った要約課題に応用できないことにある。かといって一般化を図るために、『Thesis statement に言及しており、重要なポイントを半分以上カバーしている』のような表記にしまうと「重要なポイント」が評価者間でずれてしまう可能性があるため、ある程度具体的な表記も必要となる。そこで本研究では、一般化可能な口頭要約課題のための発話内容を評価するためのルーブリックの構築を試みた。

2. 研究の目的

本研究の目的は、口頭要約課題における発話内容の評価をより適切に行うためのルーブリックを構築することにあった。そのうえで、様々な標準テストで使用されることも多い要約課題であるため、各種テストのルーブリックを精査するとともに、ルーブリックバンク（世界各国の教育者が個人で作成したルーブリックを共有しているウェブサイト）に登録されている要約課題に関するルーブリックを精査し、どのような内容が評価基準に組み込まれているかを検証することを第一の目的に掲げた。その情報をもとに、独自に日本人大学生を対象とした口頭要約課題の発話データを収集し、そのデータをもとに一般化可能な口頭要約課題のルーブリックを作成することを第二の目的とした。そして最後に、そのルーブリックの妥当性と信頼性を中心に質の検証を行うことを最終目的とした。

3. 研究の方法

まず、標準テストで口頭要約を使用しているルーブリック（例. TOEFL iBT）やルーブリックバンクに公開されているものを網羅的に検証した。対象としたルーブリックバンクは Rubister (Advanced Learning Technologies in Education Consortium at the University of Kansas, n.d.) とし、これに英語で記載されているルーブリックのうち、要約課題の内容を評価しているものを対象とした。ここでタスクの対象をスピーキングに絞らず、ライティングも含めることとしたが、これはパフォーマンスの内容を評価しているルーブリックの数が限定的であったためである。ここで、公開されているルーブリックの形式が一樣に Grid/ Matrix 形式であることに気づき、後の研究で作成するルーブリックについては他の形式（例. Empirically derived, Binary-choice, Boundary-definition (EBB), Performance Decision Tree (PDT)）の形式を参考にすることを考慮に入れた。

続いて、実際のルーブリックの作成については、報告者が担当する英語科目の定期試験において口頭要約課題のパフォーマンスを収録し、そのデータをもとに新たなルーブリックの作成に取り掛かった。データ収集は、協力者に 300 words 程度の文章を読んでもらい、その後 2 分間の準備時間を経て回答してもらう形式で実施した。その後、EBB scale (Turner & Upshur, 2002; Upshur & Turner, 1995) 作成の手順に則って、大学で英語指導経験のある 3 名が発話の内容の良し悪しの観点でパフォーマンスの分割作業を行った。その結果を受けて報告者がルーブリック

クを作成した。なお、ルーブリックの形式については、Fulcher, et. Al. (2011)による Performance Decision Tree (PDT)を採用することとした。これは、EBB や Grid/ Matrix 形式のルーブリックに比べて多様なフィードバックを返すことが可能となると考えたためである。口頭要約課題における発話内容の評価のしやすさという観点から PDT の有効性を示すため、Grid/ Matrix 形式のルーブリックを別に作成した。PDT と Grid/ Matrix 形式のルーブリックのどちらがより効果的な評価につながっているかを検証するためである。

その後、作成したルーブリックを用い、5名の評価者(うち2名は前述のデータ分割作業に携わった教員、残り3名は教員免許状取得を目指す大学4年生)が前述の口頭要約課題のパフォーマンスを評価した。データ分割作業から実際の評価まで約3ヶ月期間を空けているため、データ分割作業に携わった2名の練習効果はないものと判断した。また、評価の実用性を示すため、評価にかかった時間を計測し、どちらのルーブリックのほうがより素早く評価できるかを検証した。

ルーブリックの妥当性と信頼性については、多相ラッシュモデルを用いた分析と一般化可能性理論を使用した分析で検証した。ただし、Knoch and Chappelle (2017)の妥当性検証の枠組み全てを網羅することはできていない。したがって、上記の枠組みの限定的な部分において妥当性の検証を実施している。

評価後、各評価者に評価のしやすさの観点からどちらのルーブリックのスタイルがより効果的であるのかを評価者にインタビューを行った。その結果を実用性や波及効果の観点からの裏付けとすることができた。

4. 研究成果

上記の研究計画に基づいて分析を行った結果、妥当性の観点では、多相ラッシュモデルによる分析の結果から、部分的に妥当性が認められた。また、一般化可能性理論による結果では、Grid/ Matrix 形式のルーブリックと PDT 形式のルーブリックで信頼性の差はそれほど小さくなく、いずれも一定程度の信頼性があることが認められた。上述した評価にかかった時間という観点では、全評価者が PDT を使用したケースにおいて Grid/ Matrix を用いて評価した場合に比べて大幅に時間を短縮することが可能であることが明らかとなった。

さらに、すべての評価者が PDT のほうが評価しやすかったと回答しており、Grid/ Matrix 形式のルーブリックの場合は各レベルの中間に該当するようなパフォーマンスにどう対応してよいか分かりづらいという反応が得られた一方、PDT の場合、すべての記述子に対して Yes/ No で回答していくという結果から自然楽に評価できたという反応を得ることができた。

本研究では、口頭要約課題における発話内容の評価という観点でより効果的なルーブリックを作成することを目的としていたが、PDT 形式のルーブリックのほうがより効果的であることが明らかになった。例えば、Grid/ Matrix 形式のルーブリックでは、受験者に与えられるフィードバックが1種類に限られる。EBB の場合、途中にある記述子にある情報からもフィードバックを得られることが推定されるが、経験上多くの学習者は最終的に到達した記述子に記載されたフィードバックの内容に注目する傾向がある。一方、PDT の場合、分岐点が複数あるため、受験者はフィードバックをそれぞれの記述子に注目しながら結果を確認できれば学習上極めて有効であると判断した。実際、それぞれの記述子が Yes/ No で分岐することになるが、これはつまり、フィードバックの種類を2のn乗(n = 項目数)まで増やすことができることが可能となる。したがって、他のルーブリックの形式よりも圧倒的に多く学習者にフィードバックを返すことができると判断した。実際、PDT の作成には、EBB 同様専門家によるパフォーマンスの分割作業が必要であり、手間はかかるものの、発話内容の評価という観点で言えば極めて有効であると判断した。

本研究の成果は、最終目的を中心に口頭発表をしているものの、その結果をうけて執筆した論文が現在査読中である。詳細な報告については、別途改めて行うものとする。

引用文献

- Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5-29.
- Hirai, A., & Koizumi, R. (2008). Validation of the EBB scale: A case of the story retelling speaking test. *JLTA Journal*, 11, 1-20. https://doi.org/10.20622/jltaj.11.0_1
- Hirai, A., & Koizumi, R. (2013). Validation of empirically derived rating scales for a story retelling speaking test. *Language Assessment Quarterly*, 10(4), 398-422. <https://doi.org/10.1080/15434303.2013.824973>
- Hirai, A., & Yokouchi, Y. (2019). An investigation of EFL learners' diagnostic assessment capabilities for a classroom-based speaking test. *ARELE: Annual Review*

of English Language Education in Japan, 30, 209-224.

- Knoch, U., & Chapelle, C. A. (2018). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4), 477-499. <https://doi.org/10.1177/0265532217710049>
- Turner, C. E., & Upshur, J. A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Quarterly*, 36(1), 49-70. <https://doi.org/10.2307/3588360>
- Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3-12. <https://doi.org/10.1093/elt/49.1.3>

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 5件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Yuichiro YOKOUCHI	4. 巻 5
2. 論文標題 Correlation Between EBB Scale Scores and CAF Indices : Evidence from Speakers' Actual Performances	5. 発行年 2020年
3. 雑誌名 Learner Corpus Studies in Asia and the World	6. 最初と最後の頁 67-78
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 横内裕一郎	4. 巻 5
2. 論文標題 教養教育英語科目受講生のスピーキング能力の推定	5. 発行年 2021年
3. 雑誌名 弘前大学教養教育開発実践ジャーナル	6. 最初と最後の頁 49-56
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横内裕一郎	4. 巻 41
2. 論文標題 合成音声技術を活用した小学生向け英語リスニングテストの開発	5. 発行年 2021年
3. 雑誌名 東北英語教育学会研究紀要	6. 最初と最後の頁 75-88
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 高木修一・久保田恵佑・横内裕一郎	4. 巻 41
2. 論文標題 東北英語教育学会研究紀要22-40号掲載論文におけるテーマおよび研究デザインの系統的レビュー	5. 発行年 2021年
3. 雑誌名 東北英語教育学会研究紀要	6. 最初と最後の頁 33-42
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 立田夏子, 横内裕一郎, パードセール・ブライアン・ジョン, ソロモン・ジョシュア・リー	4. 巻 8
2. 論文標題 教養教育英語科目1年次科目のカリキュラム改革と検証	5. 発行年 2024年
3. 雑誌名 弘前大学教養教育開発実践ジャーナル	6. 最初と最後の頁 27-35
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横内裕一郎, ソロモン・ジョシュア・リー, パードセール・ブライアン・ジョン, 立田夏子	4. 巻 8
2. 論文標題 教養教育英語科目高年次科目のカリキュラム検証：学生アンケートを中心とした考察	5. 発行年 2024年
3. 雑誌名 弘前大学教養教育開発実践ジャーナル	6. 最初と最後の頁 37-48
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件 (うち招待講演 3件 / うち国際学会 2件)

1. 発表者名 立田夏子・横内裕一郎
2. 発表標題 大学教養教育英語科目のカリキュラム改革 ICT教育を組み込んだ1年次科目
3. 学会等名 大学英語教育学会第61回国際大会 (国際学会)
4. 発表年 2022年

1. 発表者名 横内裕一郎
2. 発表標題 VELC Test Onlineを使用したオンラインプレイスメントテストの試み
3. 学会等名 全国英語教育学会第46回長野研究大会
4. 発表年 2021年

1. 発表者名 横内裕一郎
2. 発表標題 CAF指標を用いたルーブリックの 構成概念妥当性（外的側面）の検証
3. 学会等名 第5回学習者コーパス国際シンポジム（招待講演）
4. 発表年 2020年

1. 発表者名 横内裕一郎
2. 発表標題 合成音声を用いた小学生向けリスニングテストの作成: Google Cloud Text-To-Speechを用いて
3. 学会等名 第20回小学校英語教育学会
4. 発表年 2020年

1. 発表者名 平井明代・横内裕一郎
2. 発表標題 教室内技能統合型スピーキングテストにおけるルーブリックと採点
3. 学会等名 第52回日本言語テスト学会研究例会（小泉利恵科研プロジェクト第2回例会）（招待講演）
4. 発表年 2021年

1. 発表者名 小泉利恵・深澤真・横内裕一郎・井上千尋
2. 発表標題 高校の英語授業内スピーキング評価における「信頼性確保のための採点指針」の作成に向けて
3. 学会等名 第52回日本言語テスト学会研究例会（小泉利恵科研プロジェクト第2回例会）（招待講演）
4. 発表年 2021年

1. 発表者名 横内裕一郎
2. 発表標題 ChatGPTを使ったテスト作成
3. 学会等名 東北英語教育学会第41回福島研究大会
4. 発表年 2023年

1. 発表者名 横内裕一郎
2. 発表標題 口頭要約課題における「発話内容」評価のためのルーブリック作成
3. 学会等名 全国英語教育学会第48回香川研究大会
4. 発表年 2023年

1. 発表者名 Tatsuta Natsuko, Yokouchi Yuichiro, Solomon Joshua Lee
2. 発表標題 大学教養教育英語科目のカリキュラム検証
3. 学会等名 大学英語教育学会第62回国際大会（国際学会）
4. 発表年 2023年

1. 発表者名 横内裕一郎
2. 発表標題 一般化多相ラッシュモデルの実行例の紹介：Facetsによる分析との差別化
3. 学会等名 外国語教育メディア学会関西支部 メソドロジー研究部会2023 年度第 3 回研究会
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------