

令和元年6月12日現在

機関番号：12102

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K13256

研究課題名（和文）技能統合型言語テストの妥当化：タスク遂行に必要な構成概念の測定に向けて

研究課題名（英文）Validation of Integrated Language Tests Toward the Measurement of Task-Induced Performance Skills

研究代表者

卯城 祐司 (USHIRO, Yuji)

筑波大学・人文社会系・教授

研究者番号：60271722

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：本研究では日本人英語学習者の読解タスクパフォーマンスを評価するためのテストおよび採点ルーブリックの妥当化を行った。情報転移課題のパフォーマンスを3名の評価者が採点し、ルーブリックがテスト開発者の意図通りに機能したのかを多相ラッシュ分析、および一般化可能性理論に基づき論証した。加えて、構造方程式モデリングによりテスト得点が学習者の語彙力・文法力・意味理解力と密接に関連することを明らかにした。

研究成果の学術的意義や社会的意義

本研究によって技能統合型テストの実効性が示されたことで、単語や文法、訳読などに焦点を当てたエクササイズ中心の読解指導から、タスク中心教授法に基づくコミュニケーション重視の読解指導への移行を促すことが期待される。テストによって英語教授法の質を向上させるためには、テスト自体が学習の一助となるように妥当性と実用性を兼ね備えたテストを実施しなければならない。本研究は指導の成果や学習者の達成度を適切に測定できるテストを提供することができる。

研究成果の概要（英文）：This study validated a rating rubric to assess the levels of task-induced performance of EFL reading. An argument-based approach was employed to examine how well the task performance reflected constructs of reading. In a survey, six information transfer tasks were presented to Japanese EFL learners to elicit their communicative reading protocols. Three raters scored the task performance; the rating appropriateness was argued using many-facet Rasch measurement and generalizability theory. Structural equation modeling was run to reveal the relationship between task-based reading performance and vocabulary, grammar, and comprehension skills. The results showed that the rating rubric functioned as intended, the six tasks with three raters were sufficient for reliable scoring, and the scores were reflectiveness of unidimensional construct of reading. These validity arguments depict the integration between the task-based reading assessment and argument-based approach to validation.

研究分野：英語教育学

キーワード：英語教育
リーディング テスト タスク中心教授法 妥当性 信頼性 項目応答理論 構造方程式モデリング

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

1. 研究開始当初の背景

現在広く使用されている英文読解テストは、システム準拠テスト、すなわち受験者の読解力を推定するために、読解力の構成概念とされる知識や技能（語彙・文法知識、推論能力など）を個別に測定するテストである (Robinson & Ross, 1996)。しかし、このようなテストでは学習者の認知的側面やタスクに内在する文脈的要素を考慮することができないため、実生活での言語運用能力を推定することは難しい (McNamara, 1996)。一方、近年注目を集めているパフォーマンス準拠テストは、例えば受験者に「旅行会社のパンフレットを読み、旅行のスケジュールについて友人にメールを書く」というような具体的なシチュエーションを提示し、その中で実際に英語を使わせるテストである (Bachman, 2002)。単語や文法知識、英文の意味を問うのみのテストとは異なり、パフォーマンステストの結果は受験者が将来、実際に言語を使用する場面に立った時のパフォーマンスレベルの予測に使用することができる。このようなテストには読み書きなどの異なる技能が介在するため、技能統合型テストとも呼ばれる。

日本の英語教育において技能統合型の英文読解テストはほとんど浸透しておらず、また実証研究も非常に限られていた。なぜなら、技能統合型テストの得点には様々な要素が反映されており、語彙・文法の知識 (e.g., 旅行会社のパンフレットにある語彙の理解)、明示的内容の理解力 (e.g., パンフレットに書かれている内容の理解)、暗示的内容の理解力 (e.g., 相手に伝える際に補うべきテキスト外の情報の推論) などを受験者のパフォーマンスから数値化できるような採点基準は開発されていなかったためである。そのため技能統合型テストに関する先行研究は「コミュニケーションの目的を達成できたか」を合否や文言等の形で評価するに留まっていた (Ushiro et al., 2015)。

そこで本研究では、技能統合型のコミュニケーションという観点を反映した英文読解テストが受験者の言語知識・技能をどれだけ正確に反映できるかを検証することに取り組んだ。これにより、受験者が「コミュニケーションの目的を達成できなかった」と判断された場合にも、学習プロセスの改善に向けた具体的なフィードバックができる、教育的意義の高いテストを開発することを目指した。

2. 研究の目的

本研究は3年の期間内で、以下に述べる2点の研究課題に取り組むことを目的とした。

- 技能統合型テストにおける受験者のパフォーマンスから、受験者の英文読解力に関わる構成概念（語彙力・文法力・明示的理解・暗示的理解）を数値化する採点基準を開発する。
- 技能統合型テストが「タスクを完遂するためにテキスト内容を適切に理解できる」言語運用能力を妥当に測定できているかを複数の統計的解析により明らかにする。

3. 研究の方法

本研究では、主に教室で使用するためのタスク型読解テストおよび採点ルーブリックを開発し、テスト得点の解釈がどの程度適切であるかを論証に基づく妥当化の枠組みで検証した。論証に基づくテストの妥当化は、テスト得点の解釈の適切さに対する様々な主張を、それに対応する推論や裏付けによって論証していくアプローチである。具体的には以下に示す3つの主張を定義し、それに対応する推論・裏付けをテストの妥当性を示す証拠として収集した。

- (1) 得点化の推論：受験者のパフォーマンスはテスト開発者が意図した観測得点が得られる手順によって得点化されている。
 - ・ 採点ルーブリックが適切に開発されていることを証明する。
 - ・ 採点ルーブリックによる採点が適切に行われていることを証明する。
- (2) 一般化の推論：観測得点は採点者間で一貫した値を示す期待得点の推定値である。
 - ・ タスク数が信頼性のある得点を得るのに十分な数であることを証明する。
 - ・ 評価者の数が信頼性のある得点を得るのに十分な数であることを証明する。
- (3) 説明の推論：期待得点はテスト開発者が定義したテストの構成概念に起因している。
 - ・ テストの難易度はタスクの難しさと連関していることを証明する。
 - ・ テスト得点は一貫して1つの構成概念を反映していることを証明する。
 - ・ テスト得点は英文読解に必要な知識・技能と連関していることを証明する。

日本人大学生 122 名を対象に、6 種類のタスクをテストした。タスクに書かれている内容を全員で確認し、タスクで求められていることを周知してテストに取り組みさせた (e.g., 友人に伝えるためのメールを作らなければならない)。その後、ペアでタスクの完成度を評価し、英文内容の理解度確認をクラス全体で行った。読解パフォーマンスの採点基準として、Ellis (2003)、Hudson (2005) や Norris et al. (2002) に従い、5 件法 (タスクの完成度に応じて 1. inadequate ~ 5. adept) の採点ルーブリックを作成した。各評価規準にはタスクの完成度を評価するための判断基準として、1~2「受験者はタスクで求められたコミュニケーションの目的を達成できていない」、3~5「完成度に差はあるものの受験者はタスクで求められたコミュニケーションの目的を達成できた」といった内容の記述を含めた。

4. 研究成果

(1) 得点化の推論について

採点ルーブリックが適切に開発されていたかを調べる方法として、評価の段階性が適切に区別されたかを検証できる多相ラッシュ分析を行った。5段階それぞれの間にある閾値は単調増加しており、閾値の間隔は1.4ロジット以上5.0ロジット未満であること、明確な頂点があることを確認した。さらに、本研究が開発した採点ルーブリックは受験者を3レベルの熟達度群に明確に識別できることが分かった。

採点者は採点ルーブリックを適切に使うことができているかについて、多相ラッシュ分析を行った結果、全体的に一貫しない採点をした採点者はおらず、信頼性係数も十分な値を得ることができた。ただし、バイアス分析をさらに行ったところ、採点者が特定の受験者に対して妥当でない採点を行うケースが0.3%含まれていたことも明らかとなった。特定のタスクで一貫した採点ができなくなるケースも4.8%あったことから、パフォーマンステストの採点の際には評価者トレーニングを事前に行う必要があること、およびラッシュモデル等による調整済み得点を使用した方がよいことを確認した。

(2) 一般化の推論について

テスト得点に含まれる測定誤差がどの程度採点プロセスのノイズに起因したのかについて、一般化可能性理論による検証を行った結果、6タスクを使用する場合は2名の採点で十分であることを明らかにした。採点者間における採点の厳しさは依然として有意であったため、ルーブリックを用いたパフォーマンス評価を行う際には複数名による採点を行うこと、採点に齟齬がある場合は採点者間での議論を行う必要性があることを確認した。

(3) 説明の推論について

テスト得点はテキストやタスクの複雑さと強く相関していた。6種類のタスクが一貫して1つの構成概念(すなわち英文読解力)を測定しているのかを因子分析で検証したところ、結果として、6種類のテスト得点は1つの潜在因子に集約されることが分かり、タスク型読解テストは一貫して単独の構成概念を測定していることを示した。テスト得点が反映しているものが英文読解力であることを証明するため、多特性・多方法行列、および構造方程式モデリングによる検証を行った。多特性・多方法行列ではタスク型読解テストの得点が多肢選択式の読解テスト得点と相関すること、一方で同一形式のリスニングテストとの相関はそれよりも低くなることを示した。さらにタスク型読解テストの得点が語彙知識・文法知識・意味理解とどの程度連関しているのかを検証したところ、テスト得点の分散は各知識・技能の程度によって十分に説明されることを明らかにした。これらの結果は、本研究のタスク型読解テストはリスニング等の他技能ではなく、リーディングの知識・技能を測定できていることを示している。

本研究は、技能統合型の英文読解テストの妥当化に取り組むことで、言語運用能力のテスト分野におけるパフォーマンステスト開発の一例を示すことができた。具体的には、理解したテキスト内容を基にコミュニケーションを行うタスクから得られた言語データが、学習者のどのような能力を測定し得るのかについて検証を進めたことで、単なる言語パフォーマンスの質だけでなく、英語リーディングの構成概念に関わる言語知識や技能を適切に評価するためのテスト作成に貢献できる可能性が示された。テスト得点の信頼性と妥当性を兼ね備えたパフォーマンステスト開発のためのガイドラインを今後提案することで、言語テスト研究における当該分野のさらなる発展が期待できる。

5. 主な発表論文等

[雑誌論文](計5件)

1. Ushiro, Y., Ogiso, T., Hosoda, M., Nahatame, S., Kamimura, K., Sasaki, Y., Kessoku, M., & Sekine, T. (2019). How EFL readers understand the protagonist, causal, and intentional links of narratives: An eye-tracking study. *ARELE: annual review of English language education in Japan*, 30, 161-176, 査読在.
2. Ushiro, Y., Hamada, A., Mori, Y., Hosoda, M., Tada, G., Kamimura, K., & Okawara, N. (2018). Goal-oriented L2 reading processes in maintaining the coherence of narrative comprehension. *JACET Journal*, 62, 109-128, 査読在.
3. Ushiro, Y., Hosoda, M., Nahatame, S., Mori, Y., Suzuki, K., Tada, G., Ogiso, T., Kamimura, K., Sasaki, Y., Mandokoro, R. (2018). Understanding protagonist, causal, and intentional links during EFL narrative reading. *ARELE: annual review of English language education in Japan*, 29, 81-96, 査読在.
4. 卯城祐司 (2017)「状況モデルから見た日本人 EFL 学習者が抱える読解の困難性 眼球運動測定法による検証」*LET Kyushu-Okinawa BULLETIN*, 17, 1-17, 査読在.
5. Ushiro, Y., Hosoda, M., Mori, Y., Tada, G., Kimura, Y., Tanaka, N., & Amagai, Y. (2017). Supporting the maintenance of global coherence with situational instruction: Evidence from eye movements during EFL reading. *JACET Journal*, 61,

〔学会発表〕(計 10 件)

1. 卯城祐司・神村幸蔵・小木曾智子・佐々木大和・名畑目真吾・細田雅也・尾島巧・川島葉月・結束萌香・関根貴則 (2018)「物語文の因果・意図・登場人物に関するつながりの理解 視線計測による検証」第 44 回全国英語教育学会京都研究大会にて.
2. Ushiro, Y., Ojima, T., Kawashima, H., Suzuki, S., Mandokoro, R., & Kamimura, K. (2018). *How do L2 readers monitor intentional coherence during narrative reading? Evidence from an eye-tracking study*. Poster session presented at the 25th Annual Meeting Society for the Scientific Study of Reading, England.
3. 卯城祐司 (2017)「インフォメーション・トランスファーに基づく能動的な英語リーディング指導の可能性: 旧態依然なのは大学入試か高校か?」第 43 回全国英語教育学会島根研究大会にて.
4. 卯城祐司・細田雅也・名畑目真吾・小木曾智子・森好伸・鈴木健太郎・神村幸蔵・佐々木大和・政所里佳・多田豪 (2017)「物語文における登場人物・時間・空間情報の一貫した理解 読解時間の分析から」第 43 回全国英語教育学会島根研究大会にて.
5. Ushiro, Y., Hosoda, M., Mori, Y., Suzuki, K., Tada, G., Sasaki, Y., Ogiso, T., & Mandokoro, R. (2017). *Monitoring the coherence of three situational dimensions of comprehension during L2 narrative reading*. Poster session presented at the 24th Annual Meeting Society for the Scientific Study of Reading, Canada.
6. 卯城祐司・森好伸・細田雅也・多田豪・神村幸蔵・大河原にし香 (2016)「英文理解の一貫性の破綻に対する気づきと修復 目的志向の読解時における思考発話から」第 42 回全国英語教育学会埼玉研究大会にて.
7. Ushiro, Y., Hosoda, M., Mori, Y., & Tanaka, N. (2016). *Effects of the situational instruction on the detection of global inconsistencies among second-language readers: Evidence from eye-tracking and recall*. Poster session presented at the 23rd Annual Meeting Society for the Scientific Study of Reading, Porto Portugal.
8. 卯城祐司 (2016)「なぜ読めない、『読めたつもり』に終わるのか インフォメーション・トランスファーから探る英語リーディングの謎」立教大学英語教育研究所主催講演会にて.
9. 卯城祐司 (2016)「クリティカル・リーディングで迫る、深い英文の理解」関西英語教育学会 (KELES) 第 21 回研究大会招待講演にて.
10. 卯城祐司 (2016)「なぜ読めない、『読めたつもり』に終わるのか 英語リーディング研究からの示唆」外国語教育メディア学会 九州・沖縄支部 第 45 回支部研究大会招待講演にて.

〔図書〕(計 1 件)

1. 卯城祐司 (編著) (2018)『初等外国語教育』ミネルヴァ書房, 192 頁.

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://www.u.tsukuba.ac.jp/~ushiro.yuji.gn/>

6. 研究組織

研究協力者

研究協力者氏名: 濱田 彰

ローマ字氏名: (HAMADA, Akira)

研究協力者氏名: 小木曾智子

ローマ字氏名: (OGISO, Tomoko)

研究協力者氏名: 佐々木大和

ローマ字氏名: (SASAKI, Yamato)

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。