

日本語テスト学会 (JLTA)
第27回 (2024年度) 全国研究大会発表要綱 (簡易版)

Handbook
of
the 27th Annual Conference
of
the Japan Language Testing Association

Conference Theme:
Next Generation of Language Assessment

日時：2024年10月5日 (土) 8:55-17:50 ~ 10月6日 (日) 9:00-12:00

会場：上智大学 四谷キャンパス
〒102-8554 東京都千代田区紀尾井町7-1
https://adm.sophia.ac.jp/jpn/campus_kengaku_ad/kengaku_yotsuya/

主催：日本語テスト学会 事務局
〒960-8516 福島県福島市栄町10-6
久保田恵佑研究室
TEL: 024-581-5533 (直通)
<http://jlta.ac/>



The Japan Language Testing Association
Keisuke Kubota (Fukushima Medical University)

Address: Integrated Center for Science and Humanities /
School of Health Sciences,
Fukushima Medical University,
10-6 Sakae-machi, Fukushima City, Fukushima, 960-8516,
JAPAN

TEL: +81-024-581-5533 (main switch number)
e-mail: kkubota86@gmail.com /
kkubota@fmu.ac.jp

URL: <http://jlta.ac/>

Presentation Overview

Time	Part	Room 2-408	Room 2-409	Room 2-410	Room 2-415	Room 2-401
9:30— 10:00	I	Exploring the Fluency Corpus of Academic English Lectures (FCAEL): The profiles of repetitions 西澤 倫	学習者に社会的コンテキストを注目させるためのライティング・タスク—開発のための予備調査— 久留 友紀子・山西 博之・野田 三貴†	Performance Decision Tree の有用性の再考察 横内 裕一郎	--	--
10:05— 10:35	II	A case study of novice-rater engagement with an online, CA-informed, L2 oral pragmatics assessment F. Scott WALTERS	日本人中学生のスコアレポートの理解と発話改善—学習支援に向けた AI 採点型スピーキングテスト活用の場合— 長嶺 洋平・小泉 利恵	More communicative, more accurate, better test construct validity - introducing next generation AI scored language testing Stuart Connor	CEFR-JSL (日本手話) の開発試案—「手話言語コミュニケーション能力」を中心に— 中野 聡子・浅田 裕子・藤田 元†	--
10:35— 10:50	Break					
10:50— 12:10		--	--	--	--	KEYNOTE
12:10— 13:40	Lunch Break					
13:40— 14:10	III	日本の大学における Linguaskill と TOEIC スコアの比較 青山 智恵	「思考・判断・表現」を評価する英語問題の特徴—公立高等学校英語入試問題の観点別分類の試み— 渡邊 大志・大鋸 雄介・佐藤 扶・鵜籠 恵一・斉田 智里	口頭要約のための Performance Decision Tree の作成事例: 構成概念の齟齬とその解消 久保田 恵佑・前田 啓貴・高木 修一・横内 裕一郎†	Classroom practice, assessment implementation, and Artificial Intelligence Myles GROGAN†	--
14:15— 14:45	IV	全国学力・学習状況調査中学校英語の IRT 分析からわかること、わからないこと—調査問題と CEFR との対応づけ— 斉田 智里・境奈 津希・渡邊 大志	An Investigation Into the Washback of a National Standardized Examination for University Admissions in Japan: Through Analyses of Semi-structured Interviews Yusuke OGA	Assessment of functional adequacy of monologic speaking among Japanese senior high school students: Focusing on rater reliability Rie KOIZUMI・Masakazu UENO・Maki IMAZAWA・Mariko ABE	--	--
14:50— 15:20	V	小学校英語における児童の英語力と学習者要因との関係 白鳥 金吾・志村 昭暢	Yes No 語彙テストにおける採点方法の検討: バイズモデル選択とテスト情報関数に基づいて 小室 竜也 黒川 皐月	Developing and evaluating a group discussion test: Focusing on language functions, conversation style, and test-taker perception Yujia Zhou・Masashi Negishi・Asako Yoshitomi	--	--
15:20— 15:35	Break					
15:35— 17:05		--	--	--	--	Symposium

†Assessment practice presentation

To Presenters

1. Presenters will have 20 minutes to present their paper, followed by 10 minutes for discussion.
2. There will be no chairperson in the presentation room. A timekeeper will show you the time left.
3. Please register at the registration desk first. Please go to the designated room 5 minutes prior to the starting time of the presentation.
4. Presenters are expected to bring a PC. There will not be an audio terminal connector (for PC connection through a stereo mini plug). If necessary, please prepare an adaptor. Mac users should bring their own Mini DisplayPort to VGA Adapter. Third-party adapters do not work properly sometimes.
5. Please bring handouts in case your PC or the projector does not work.
6. If you need a letter of invitation/participation, contact Keisuke Kubota (JLTA Secretary General) at kkubota86@gmail.com
7. Name tags including participation certificates will be provided on the day of the conference.

発表者へのご案内

1. 20分の発表と10分の質疑応答の時間があります。
2. 司会者はありません。タイムキーパーが残り時間を提示します。
3. 受付を済ませ、発表開始5分前には、発表会場にお越しください。
4. ご発表にはご自分のコンピューターを持参してください。音声ケーブル(ステレオミニプラグによるPC接続用)は発表会場にはございません。必要に応じて、変換用アダプター、並びにMac用のケーブルはご自身でご準備ください。持ち込んだアダプターは機能しないことが時折ございます。
5. 予測できない不具合に備え、ハンドアウトのご持参をお勧めします。
6. 出張依頼状などが必要な方は、kkubota86@gmail.com (久保田恵佑, JLTA 事務局長) までご連絡ください。
7. 参加証明を含むネームタグは、大会当日に配布します。

Abstracts (発表要旨)

Paper Session

Room2-408 Part I (9:30-10:00)

研究発表

Exploring the Fluency Corpus of Academic English Lectures (FCAEL): The profiles of repetitions

西澤 倫 (麗澤大学)

Corpus-based studies have offered a domain definition inference for language tests. For instance, Biber's (2004) T2K-SWAL has been a valuable resource of lexical information in academic settings. Yet, currently, there are limited resources for temporal fluency measures such as articulation rate, filled pauses (e.g., uh), and repetitions, especially in the context of academic lecture settings. To address this issue, Nishizawa (2024) created the Fluency Corpus of Academic English Lectures (FCAEL), which is a free, publicly available source of 100 lecture videos. Nishizawa (2024) compared the FCAEL to lecture passages in TOEFL iBT and IELTS to assess their domain definition inference. The study identified some similarities and many dissimilarities across the three. One notable finding was more frequent repetitions in FCAEL than in TOEFL iBT and IELTS. Critically, there were zero observations of repetitions in IELTS. One limitation of Nishizawa (2024) was the lack of investigations on other aspects of repetitions. Thus, the present study examines what words are repeated, what types of words are repeated, and where repetitions occur. I used descriptive statistics and visualizations to compare the FCAEL to TOEFL iBT.

Data was taken from Nishizawa (2024). The FCAEL is a collection of 6-minute excerpts of 100 videos from publicly available resources (e.g., YouTube; 10 hours of data). The TOEFL iBT data was taken from official test preparation books ($k = 54$) to ensure the data was representative.

Preliminary analysis found some similarities between the two, while some dissimilarities were also identified. Similarities included a higher proportion of function-word repetitions than content-word repetitions, a higher proportion of single-word repetitions than multi-word repetitions, and a higher proportion of clause-within repetitions than clause-initial repetitions. Dissimilarities included more frequent multi-word repetitions in FCAEL, which comprised of 18.12%, than in TOEFL iBT (4.86%). Moreover, while the difference in frequency between clause-initial and clause-within repetitions was small in TOEFL iBT (clause-initial: mean = 0.20, SD = 0.30; clause-within: mean = 0.27, SD = 0.36), the difference was large in FCAEL (clause-initial: mean = 0.53, SD = 0.61; clause-within: mean = 1.10, SD = 1.04). This suggests that TOEFL iBT needs more repetitions, especially at the clause-within position to increase authenticity. I conclude the talk by suggesting thresholds of authenticity, which would be useful for making and revising lecture passages in many high-stakes English tests.

"

Keywords

listening, authenticity, fluency

A case study of novice-rater engagement with an online, CA-informed, L2 oral pragmatics assessment

F. Scott WALTERS

(Seoul National University of Science and Technology)

Since conversation analysis focuses attention on rich, complex, co-constructed language sequences that may seem irrelevant or opaque to teachers trained to view an L2 in purely individualistic terms (Huth, 2021), and since, moreover, the emic perspective required by CA may take time to master with repeated viewings of recordings (Mori, 2021) and is not (yet) a common skillset among raters of tertiary-level L2 oral tests, it was of interest to determine, if possible, a “lower-bound” of skill with CA by university-level test-raters that would enable valid interpretation of test scores (Messick, 1989; Kane, 2006).

Thus, this study investigated the progress of an oral-test rater with novice-level exposure to CA when rating responses from a test of L2 pragmatics, administered to EFL-using, university-level adults and older adults of Korean L1 (n = 35). A second rater with CA expertise was also employed. Tester-examinee interactions were video-recorded on Zoom, the recordings being independently scored by the two raters. The novice transcribed targeted interactions to facilitate acquisition of analytical skill and to determine whether the rater could achieve “sustained critical reflection” on his/her rating (Huth, Betz, and Taleghani-Nikazm, 2019, p. 99). Rating criteria were drawn from empirical CA studies: assessment responses (Pomerantz, 1984), compliment responses (Pomerantz, 1978), pre-sequence responses (Schegloff, 2007), and other-repair (Wong and Waring, 2010).

Standard quantitative test results were analyzed to help determine rater/test usefulness (Bachman and Palmer, 2010). Qualitative analyses included triangulation of raters’ CA-style transcriptions of the tester-examinee interactions with authoritative transcriptions done by the PI, and examination of notes from post hoc inter-rater discussions and rater questionnaire-responses. Results suggested that for L2 classroom settings, use of novice raters was acceptable vis-à-vis practicality, reliability, and valid test-use, though issues with the rating scale emerged. Implications for research into CA-informed L2-test rater training are discussed. Finally, implications for CA-informed testing online are also considered.

Keywords

conversation analysis, L2 oral assessment, rater training, online assessment

日本の大学における Linguaskill と TOEIC スコアの比較

青山 智恵 (Cambridge University Press & Assessment)

本研究では、日本の大学生を対象として、Linguaskill と TOEIC® Institutional Program (IP) テスト (以下、TOEIC) のリスニングとリーディングのスコアの換算表を作成し、2つのテストのスコアがどの程度類似した数値情報を与えるかを調査し、2つのテストの異なる構成要素間の相関関係を算出することを目的とした。Linguaskill のスコアはCEFR と Cambridge English スケールで報告される。TOEIC と CEFR、そして Linguaskill と CEFR の換算表はそれぞれの機関によって公表されているが、本研究で作成された換算表は、CEFR のレベル間の境界線だけでなく、スコアの全範囲を扱っており、大規模かつ最近の日本の大学生の経験的データに基づいているため、日本の大学にとって、より正確で有用と考える。換算表は、リスニングとリーディングのそれぞれについて、広く用いられている「等パーセンタイル対応づけ (equipercentile linking)」と「プリ・スムージング(pre-smoothing)」という方法を用いて作成した。また、本研究では、Linguaskill を受けた学生がライティングとスピーキング・モジュールも受けたことから、さらなる分析が可能となり、Linguaskill の4技能のスコア間の相関関係を調査し、どの程度の相関性を示すのかについても明らかにした。(本研究は弊機関 Cambridge University Press & Assessment 英国本部の研究チームが担当、発表は日本オフィスが行うものである。) Cambridge University Press & Assessment English Research: Edmund Jones, Safiatú Lopes, Angeliki Salamoura, Jing Xu Tomoe Aoyama (Japan Office)

Keywords

換算表, conversion table, Linguaskill, TOEIC

全国学力・学習状況調査中学校英語の IRT 分析からわかること、わからないこと
—調査問題と CEFR との対応づけ—

齊田智里 (横浜国立大学)

境奈津希 (東京学芸大学大学院連合学校教育学研究科)

渡邊大志 (横浜国立大学大学院教育学研究科)

IRT (Item Response Theory) を用いてテストデータを分析すると、問題の困難度と受験者の能力値を同じ尺度上で推定することができる。等化計画の下で複数のテストを作成・実施すれば、共通尺度上で各テスト問題の困難度と能力値を推定することが可能となり、項目プールの作成や学力の経年変化の可視化ができる。こうした利点から、PISA などの国際的な学力調査や TOEFL などの英語能力試験には IRT が導入されている。日本の全国学力・学習状況調査 (悉皆調査) にも令和 7 年度から CBT と IRT が導入される。調査結果は、IRT スコアとして学校や自治体に返却される予定である。しかし IRT は万能ではない。令和 5 年度全国学力・学習状況調査中学校英語 (以後、R5 英語) では、(1)「話すこと」の平均正答率が 12.4%と非常に低かったこととや、(2) R5 英語の平均正答率 46.1%が前回調査 (H31) の 56.5%より低かったことが注目され、その真意が問われた。R5 英語は、選択式、短答式、記述式の問題で構成されており、Graded Response IRT Model を用いて困難度と能力値を推定した。その結果、(1)については、「話すこと」の問題に 50%以上正答するには、能力値 3 (偏差値で 80) 以上が必要で、現在の中学 3 年生には難しすぎる問題であることがわかった。一方、中学 3 年生の英語力の水準がどの程度なのかは IRT 分析からはわからない。学習指導要領外国語科の目標と内容は CEFR を参考に作成されている。そこで、CEFR の能力記述文を参照し、問題と CEFR レベルを一対一で対応付けて、困難度との関係を検討した。その結果、問題の多くは A1 から A2 レベルであり、日本人中学生の話すことの力は CEFR に照らして低いことが伺えた。(2)については、R5 と H31 の英語は等化されていないため、それぞれ IRT 分析して結果を比較しても、中学 3 年生の英語力が変化したかどうかはわからない。注：本研究は、「令和 5 年度全国学力・学習状況調査の英語の結果を活用した専門的な分析」(令和 5 年度文部科学省委託研究) の成果の一部である。

Keywords

全国学力・学習状況調査, IRT 分析, CEFR

小学校英語における児童の英語力と学習者要因との関係

白鳥金吾 (北星学園大学短期大学部)

志村昭暢 (北海道教育大学)

2020年度から小学校における新学習指導要領が全面的に実施され、外国語活動の開始が3年生に引き下げられ、5,6年生においては新たに「外国語(英語)」として教科化された。教科としての小学校英語教育は黎明期にあることから、学習指導要領に示された目標達成状況を客観的に把握することなど、小学校英語教育の成果や課題について実証的な検証が不足している。そこで、本研究では、群馬県にあるイマージョン教育を行う小学校5年生と6年生、43名を対象とし、2名の外国人教師が主導する外国語科授業において、Cambridge English A2 Key for Schoolsを用いて児童の英語力を測定した。また、児童の英語学習に対する質問紙及び自由記述を実施した。質問紙では、英語学習の動機づけ・コミュニケーションへの積極性・外国への興味・Can-Do・英語に対する自信や不安、他者との関係性等の学習者要因について分析し、それらが英語力に与える影響について考察した。児童の英語力と情意面の違いを明らかにするために、児童の英語力を上位・中位・下位の3群に分け、学習者要因の中から、WTC・英語への関心・動機づけ・外国への興味・外向性・不安・技能別自信・英語への自信の8項目について、2元配置分散分析を用いて比較した。結果は主効果(英語力)と主効果(情意面)において、統計的に有意な差がみられ、効果量もLargeであった。交互作用も統計的に有意な差が見られ、効果量はMediumであった($F(2,40) = 2.40, p = .03, \eta^2 = .11$)。単純主効果の検定を行ったところ、不安・技能別自信・英語力の自信において有意差が見られ、多重比較(Bonferroni法)を行ったところ、不安において英語力の下位群よりも中程度の児童が低く、技能別自信と英語力の自信については、英語力の下位群よりも中位群・上位群の児童の方が高いことが示された。以上の結果から、不安・技能別自信・英語力の自信が児童の英語力の差に影響を与えている可能性があることが示唆された。

Keywords

小学校英語、英語力、学習者要因、CEFR、ケンブリッジ英語検定

学習者に社会的コンテキストを注目させるためのライティング・タスク
—開発のための予備調査—

久留 友紀子(中央大学)
山西 博之(中央大学)
野田 三貴(大阪公立大学)

第二言語ライティングにおけるジャンルアプローチは、ライティングに関する社会的コンテキスト（読み手や目的、目的達成のための方略など）に注目する。本研究は、大学での英語ライティング指導の文脈に、ジャンルアプローチの考えを適用したライティング・タスク群を提供することを目指す、タスク開発研究である。このタスク群により、教室の中であっても学習者は、多様な社会的コンテキストに注目しながら、それとの関係性の中でテキストを作成する練習ができる。今回は予備調査として、タスク試案ならびに指導試案の使用、改善、有効性の確認、今後の方向性の議論を行った。

タスク群の開発にあたり、タスクはライティングに不可欠な「書き手」「読み手」「トピック」のどれを志向するかによって整理した (Scholes, Comley, & Peritz, 1994)。今回は試案として、不服申し立て (appeal) のような「読み手」志向のタスクを、実社会のコンテキストをシミュレーション的に反映するよう作成した。

予備調査として、タスク試案を、2022～24年に私立大学理工学部1・2年生（計270名、CEFR B1 レベル）を対象とした授業の中で使用した。指導試案には、社会的コンテキストに対して学習者の気づきを促す一連の問いかけを含めた。タスク・指導試案の有効性を確認するため、プロダクトであるテキスト、タスク遂行のプロセスに関する質問への回答等を収集した。それらのデータから、タスク・指導試案の改善が可能となったのと同時に、学習者はタスク遂行にあたり社会的コンテキストに注目しながらテキストを作成していたことが確認され、一定の有効性が認められた。

今後は、社会的コンテキストをシミュレーション可能なライティング・タスク群の開発と提供、タスクのプロセス・プロダクトを評価するためのツール（例：ループリック）の開発を予定している。

Keywords

第二言語ライティング、タスク開発、社会的コンテキスト、ジャンルアプローチ

**日本人中学生のスコアレポートの理解と発話改善
—学習支援に向けた AI 採点型スピーキングテスト活用の場合—**

長嶺洋平(福島県いわき市立内郷第三中学校)
小泉利恵(筑波大学)

教室内評価において、生徒が評価の結果を適切に解釈し、学習目標を立て、実行に移すことが重要である(国立教育政策研究所, 2023)。しかし、スピーキング評価ではテスト実施や採点に時間がかかり、テスト結果の効果的な活用が難しかった。人工知能(AI)の技術発展により、テスト実施や採点が容易なスピーキングテストが利用可能になってきた。もちろん、テスト使用の際には、教員がその適切さについて確認することも必要である。その上で採点結果やそれが記載されたスコアレポートの情報を活用して指導や学習支援に活かせる環境も身近になってきたと考えられる。しかし、適切な学習支援を行うためには、スコアレポートの情報に対して、生徒がどのように反応し、学習し、それがスピーキング力の改善につながられるかについて知る必要があるが、そのような研究は限られている。

本研究の目的は、日本人中学生が AI 採点型スピーキングテストを2回受け、スコアレポートを受け取った際の反応と目標設定を調べ、スコアの変化との関連を調べることである。参加者は公立中学校3年生5名で、生徒は各自のタブレットで AISATS (ポテンシャルプラス株式会社, 2024) を2回受験した。生徒の発話は AI によって直後に採点され、スコアに加えて書き起こしや音声等がスコアレポートとして提示された。生徒は、教員の読み取り方の説明の後に、スコアレポートの内容を読み取り、振り返りシートを記入した。その後各自で話す練習を行った。本研究では、2回のテスト結果と振り返りシートの記載内容を比較・分析した。その結果、生徒はスコアレポートの情報を的確に捉えて目標設定を行ったことが分かった。2回目のスコアでは、目標に沿ったスコア上昇も一部見られたが、改善しようとしたスコアが下がった場合もあり、生徒の自律的な学習活動を支える教員の的確な助言の必要性が示唆された。発表では詳細な結果と考察、教育への示唆を述べる。

Keywords

スピーキングテスト、自動採点、自動書き起こし、振り返り、タスクの繰り返し

「思考・判断・表現」を評価する英語問題の特徴
—公立高等学校英語入試問題の観点別分類の試み—

渡邊大志 (横浜国立大学大学院教育学研究科)
大鋸雄介 (東京学芸大学大学院連合学校教育学研究科)
佐藤扶 (東京学芸大学大学院連合学校教育学研究科)
鵜籠恵一 (横浜国立大学大学院教育学研究科)
斉田智里 (横浜国立大学)

H29/30 学習指導要領改訂を受けて、どの教科でも3つの観点から目標に準拠した評価を実施することとなり、高等学校でも新たに指導要録に観点別学習状況の記載欄が設けられた。観点別にペーパーテストの問題を作成する際、各問題がどの観点を評価するかについて、教員間での認識のずれがみられることが多く、教師の言語評価リテラシーの向上が求められている。一方で、外国語科では、「知識・技能」と「思考・判断・表現」の区分が曖昧であることが指摘されている(村越・江原、2015; 今井、2022; 馬場、2023など)。そこで、本研究では、公立高等学校の英語入試問題を対象とし、各問題を観点別に分類する作業を通して、「知識・技能」と「思考・判断・表現」を判別する基準と、「思考・判断・表現」を測る問題の特徴を明らかにすることを目的とした。令和6年度神奈川県公立高等学校入学者選抜学力検査問題の英語全27項目を対象とした。まず共同発表者5名が各自で、各問題を「知識・技能」か「思考・判断・表現」かに分類した。その際、国立教育政策研究所の『「指導と評価の一体化」のための学習評価に関する参考資料』の中学校外国語と高等学校外国語を参照した。その結果、5名の判別が一致したのは14項目で、52%の一致率である。一致した問題は全て「知識・技能」で、「思考・判断・表現」の判別は意見が分かれた。判別が一致しなかった13項目について、各々の判断根拠の共有を行い、統一見解を得るための議論を行った。その中で、指示文の内容(例、概要を把握する)で判断しているか、実際の解答プロセス(例、選択肢の内容や解答のために参照する情報の質や数)まで鑑みて判断しているかで、判別が異なっていたことがわかった。『参考資料』では、何を問うているか、という指示文の内容で「思考・判断・表現」を定義しているが、それだけではなく実際に必要となる解答のプロセスまで含めて考えることが、より精度の高い「思考・判断・表現」の問題を作成するために必要であることが示唆された。

Keywords

観点別学習状況評価、知識・技能、思考・判断・表現

An Investigation Into the Washback of a National Standardized Examination for University Admissions in Japan: Through Analyses of Semi-structured Interviews

Yusuke OGA (Graduate School, Tokyo Gakugei University)

The introduction or revision of a test has the potential to affect stakeholders such as teachers and students. Such effects are called washback and play an important role in English education in Japan. In 2021, the National Center Test for University Admissions (the Center Test) was revised for the first time in 30 years and renamed the Common Test for University Admissions (the Common Test). The purpose of this study is to investigate the washback of the introduction of the Common Test on classrooms; more specifically, on teachers' teaching methods and their assessment.

Semi-structured interviews were conducted with eight high school English teachers who taught at Japanese senior high schools. The interviewees were selected through purposive sampling, which is widely used in qualitative research. A pilot interview was conducted with one high school English teacher, who was not the interviewee of this study, to refine the interview questions. The interview was conducted using the Zoom video call function due to the COVID-19 pandemic. The interviews were audio recorded using the recording function of Zoom and transcribed by the author. The transcribed interview data were coded and analyzed thematically to find frequent patterns that show the washback of the Common Test on the teachers.

The findings reveal that the Common Test influences the teachers on how to teach and/or assess reading, listening, and pronunciation although the effects seem to be limited. The main possible factors for the impacts would be an increase in the number of words used in the Common Test, and in the score allotted to the listening component in the Common Test. On the other hand, the washback on grammar instruction was not observed although the test items that measure the knowledge of grammar were removed from the Common Test.

This study suggests that just revising the standardized test is not sufficient to produce the intended washback, and professional development and support should be provided for teachers to make educational innovations more successful. However, I investigated only teachers' perspectives in this study, which is just a small part of the washback. Furthermore, since the Common Test was implemented in 2021 for the first time, it is anticipated that some of the washback has not appeared in classrooms yet. Thus, it would be worth continuing to investigate the washback of the Common Test on not only teachers' perspectives but also other perspectives, which are all still under-researched.

Keywords

Washback, The Common Test, The National Standardized Test, Semi-structured Interview

Yes No 語彙テストにおける採点方法の検討：ベイズモデル選択とテスト情報関数に基づいて

小室竜也 (東北大学大学院 国際文化研究科)

黒川皐月 (東北大学大学院 国際文化研究科)

Yes No 語彙テストでは、実在語に Yes と回答する Hit、実在語に No と回答する Miss、非単語に Yes と回答する False Alarm、非単語に No と回答する Correct Rejection の 4 つのパターンがある。4 つのパターンを区別し、YesNo 語彙テストにおけるランダムな当て推量を検討した Zhang et al. (2020) では、実在語および非単語に対して個別にラッシュモデルを実施している。そのため YesNo 語彙テストで用いられている計算式への重みづけへの具体的な数値は示されていない。そこで、本研究では採点方法を 1 つのモデルに含め、YesNo 語彙テストにおける項目レベルの採点方法について、1 パラメタロジスティックモデル、多次元項目応答理論、部分得点モデルの 3 つの分析で比較を試みた。Nicklin and Vitta (2022) で採用された Yes No 語彙テストに、日本語を母語とする大学生 99 名 (CEFRA2 レベル) が回答した。この Yes No 語彙テストは Academic Word List から抽出された 180 項目 (うち 108 項目が実在語、72 項目が非単語：受験者に割合は知らされない) に対して、知っていれば Yes、知らなければ No と回答する Y ものである。採点 1 (Hit に 1 点、Correct Rejection・Miss・False Alarm に 0 点)、採点 2 (Hit に 1 点、Correct Rejection に 1 点、Miss と False Alarm に 0 点)、採点 3 (Hit に 2 点、Correct Rejection に 1 点、False Alarm・Miss に 0 点)、採点 4 (Hit に 3 点、Correct Rejection に 2 点、Miss に 1 点、False Alarm に 0 点)、採点 5 (Hit に 3 点、Correct Rejection に 2 点、Miss に 1 点、False Alarm に 0 点) の 5 つの採点結果を比較した。Bürkner (2021) に基づき、受験者と項目の事前分布に平均 0、分散 3 の正規分布を用いた 2 値ロジスティック回帰および隣接ロジスティック回帰によるベイズ推定を実施した。モデル選択のアプローチからは、採点 2 よりも採点 1 が望ましかったが、テスト情報曲線は Correct Rejection を Hit と同様に扱う採点方法の方が広い受験者能力を弁別できた。実際に多次元項目応答理論の結果は実在語と非単語への応答が同一構成概念である可能性を示していた。本研究で得られた結果は心理言語学で採用されてきた Lexical Decision Task に対し、非単語の分析も考慮に入れる必要性を示している。

Keywords

語彙テスト、非単語、項目応答理論

Performance Decision Tree の有用性の再考察

横内裕一郎 (福島大学)

Performance Decision Tree (PDT: Fulcher, et al., 2010) は Empirically derived, Binary choice, Boundary definition scale (EBB: Upshur & Turner, 1995) の拡張系のルーブリックで、複雑な構成概念の要素を評価をすることもできる。EBB は 2 択式で最終的に記述子の数を n としたときに $n+1$ のフィードバックパターンが発生するが、PDT は n^2 のパターンが生じるため、フィードバックが非常に多岐にわたり、学習者はフィードバックの情報をより多く得ることが可能である。厳密には両ルーブリックの作成手順は異なるのだが、簡易的に EBB の作成方法に即した形で PDT も構築が可能であると考えた。そこで、口頭要約課題の発話内容を評価するための PDT と一般的な grid 形式のルーブリックを作成し、それらを妥当性・信頼性・実用性の観点でどちらが優れているかを検証した。Knoch and Chapelle (2018) の妥当性検証の枠組みの一部を使い、PDT の有効性について検証を行った。6 名の評価者 (うち 2 名が教員、4 名が学生) が口頭要約課題のパフォーマンスを評価を行い、多相ラッシュモデルと一般化可能性理論による分析を行った。その結果、1 名の評価者が Misfit であったことを除いて、PDT も Grid も概ね有効に機能していることがわかった。一般化可能性理論を用いた分析では、Grid のほうがやや E_p 係数が高かったものの、概ね両者とも十分に安定した結果であったと言える。また、評価者へのインタビューの結果からは、PDT のほうが択一式で記述子を概観することになるため、評価がしやすいという回答が得られており、概算での計測にはなるが、PDT を使った評価のほうが Grid を使った評価より、評価にかかった時間が短い傾向にあることが明らかとなった。この点については、評価のカウンターバランスを取っていないため、偶然の範囲内であるとみなされる可能性もあるが、フィードバックのパターンが幅広く、かつ時間的資源を削減して評価できるうえに一定の安定性があるというのであれば、PDT は Grid 形式のルーブリックと同等程度異常に機能すると判断した。

Keywords

Performance Decision Tree, MFRM, Generalizability theory, validation

**More communicative, more accurate, better test construct validity
- introducing next generation AI scored language testing**

Stuart Connor (Pearson)

In 1996 two Stanford professors devised a way of using big data, machine learning and natural language processing to accurately measure speaking skills without the need of an examiner. That technology became known as AI scoring, and the test became known as the Versant English Test. Over 350 million Versant tests have been delivered since it was first launched nearly 30 years ago. Earlier this year Pearson released a new version of the test - the Versant English Speaking and Listening Test - that brings the benefits of updated technology, research and design principles to organizations and institutions that rely on accurate measurement of speaking skills in both learning and recruitment contexts. Two significant changes include the addition of a listening skills metric to get a better view of receptive and productive skills in the act of speaking, and more emphasis on longer spoken turns on the part of the test taker, to get a better view of how well they use language when speaking in their own words. This session will introduce this next generation Versant test and share the latest comparative data to evaluate how the test is performing. We will also share the latest innovations in Pearson's Global Scale of English (GSE) which is embedded in the Versant scoring logic, making it easier to set cut scores for job roles and create personalised learning pathways that support the career goals of students in Japan.

Keywords

AI, genAI, big data, fast, communication skills

口頭要約のための Performance Decision Tree の作成事例: 構成概念の齟齬とその解消

久保田 恵佑 (福島県立医科大学)
前田 啓貴 (松山大学)
高木 修一 (福島大学)
横内 裕一郎 (福島大学)

本発表では、教室内での使用を想定した口頭要約課題のための Performance Decision Tree (PDT) ルーブリック (Fulcher et al., 2011) の作成事例を報告する。PDT とは、学習者パフォーマンスの分析から明らかになった特徴に基づき、パフォーマンスについて Yes/No の 2 値で判断を行う項目が段階的に構成されているルーブリックである (Fulcher et al., 2011)。本実践では言語テストを専門とする 4 名の大学教員 (平均教育経験年数 8.83 年, SD = 3.57) が口頭要約課題の PDT を作成した。評価対象とした口頭要約課題は、5 分で指定の刺激文を読んだ後に 2 分以内でその内容を要約し、関連する意見を述べるというものであった。PDT 作成には 2 つの異なる刺激文に対する日本人大学生 63 名の発話データ (N=126) が使用された。作成手順は Turner and Upshur (1996) によるパフォーマンスの全体的な弁別手順を参考にした。加えて、口頭要約課題の構成概念をより広く反映する PDT 作成のために、パフォーマンスのレベルを弁別可能な特徴を抽出した。具体的には以下の 3 段階で PDT 作成に取り組んだ: (1) 個人で全てのデータを確認し、全体的なレベルの弁別と特徴の抽出を行う、(2) 2 グループに分かれて議論し、刺激文ごとに PDT を作成する、(3) 各グループで作成した PDT をもとに作成者全員で議論し、1 つの統合した PDT を作成する。本実践で最終的に作成された PDT は 5 つの評価観点 (i.e., タスク完遂度、パフォーマンスの質、発音、流暢さ、文法的正確さ) と 15 個の項目となった。統合プロセスでは作成者らの経験や知識による構成概念の齟齬が生じたが、統合の議論を経て、評価の目的、発話データの実態、理論的側面などを反映した PDT を作成することができた。また、先行研究 (e.g., Turner & Upshur, 2002) で指摘される作成者の異なる考えや観点の影響の可能性もあることも確認され、ルーブリック作成における議論の重要性が再認識された。例えば、作成グループ間において「発音」の捉え方に不一致が確認されたが、統合の議論により評価観点としてより包括的なものにまとめることができた。当日の発表では、本実践で作成したルーブリックの特徴 (e.g., 構成概念、記述子) とそのプロセスの詳細を報告する。加えて、議論の内容や作成後の振り返りから PDT における今後の研究課題と示唆について議論する。

Keywords

ルーブリック作成、口頭要約課題、Performance Decision Tree

**Assessment of functional adequacy of monologic speaking among Japanese senior high school students:
Focusing on rater reliability**

Rie KOIZUMI (University of Tsukuba)
**Masakazu UENO (Graduate school, University of Tsukuba;
Kitakyushu City Einomaru Junior High School, Fukuoka)**
Maki IMAZAWA (Graduate school, University of Tsukuba)
Mariko ABE (Okayama University)

Functional adequacy (FA) has been conceptualized as consisting of four aspects: Content, Task requirements, Comprehensibility, and Coherence and cohesion (Kuiken & Vedder, 2018). Although FA is an essential component of speaking proficiency, along with fluency, accuracy, and complexity, assessment and examination of FA has only recently begun (Kuiken & Vedder, 2022). More studies are required on the FA of various types of learners and the characteristics of FA's measurement. Therefore, our study examines the FA of monologues of Japanese learners of English at a senior high school level, using an analytic scale comprising four criteria. Specifically, we investigated rater reliability to ensure consistent measurement of FA.

This study is part of a large-scale project to investigate the longitudinal development of speaking proficiency among 122 Japanese high school students, using the Longitudinal Corpus of Spoken English (LOCSE; Abe & Kondo, 2019). The students were from one public school and were assessed eight times over 23 months using the Telephone Standard Speaking Test (ALC Education, 2024). The test elicited 45-second speeches using 10 tasks and produced human-scored holistic ratings across 1–9. The range of students' scores was 2–7. We aimed to develop benchmarking FA scores of LOCSE and establish anchor FA ratings without any dependency problems, which will be used as a basis for precisely assessing longitudinal dependent speech (based on Wright, 2003). This study was conducted in two stages. First, we selected 12 learners and their one-time data (i.e., Times 1, 4, or 8); three raters then decided on benchmarking scores and their rationales after evaluating 12 learners' monologues independently and discussing the divergent ratings. Second, we randomly selected one-time data from the remaining 110 learners (i.e., Times 1 to 8); we requested additional raters to undergo rater training and then evaluate the assigned speeches. We analyzed ratings using many-facet Rasch measurement (i.e., students x tasks x raters x criteria) to understand the fine-grained features of raters, the FA scale, and other aspects (Linacre, 2024).

Preliminary results of the three raters in the first stage suggest that their rater severity was similar, rater consistency was high, and interrater agreement was acceptable within the model expectation, except for Coherence and cohesion (when analyzed separately). Rater bias was detected in Task requirements and Coherence and cohesion, indicating issues in reliably measuring these constructs. We report the results of the expanded data and discuss the practical implications for assessing FA and rater training.

Keywords

Telephone Standard Speaking Test, rater reliability, functional adequacy, many-facet Rasch measurement

**Developing and evaluating a group discussion test:
Focusing on language functions, conversation style, and test-taker perception**

**Yujia Zhou (Tokyo University of Foreign Studies)
Masashi Negishi (Tokyo University of Foreign Studies)
Asako Yoshitomi (Tokyo University of Foreign Studies)**

This study investigated the development of a group discussion test (GDT) to explore the generalizability of scores from a computer-based speaking test to interactive tasks. The computer test is the British Council Tokyo University of Foreign Studies-Speaking Test for Japanese universities (BCT-S), a localization of the Aptis General Speaking component to the Japanese context. The BCT-S has been administered at Individual Achievement Test First Round to applicants to all three Schools of Tokyo University of Foreign Studies (TUFS) since February 2022.

Although different types of validity evidence have been collected regarding the use of the BCT-S (e.g., Zhou et al., 2020, 2021, 2022), the extrapolation inference (Chapelle, 2020) of BCT-S scores has not been explored. Group discussion tasks have been reported to be relevant to the Target Language Use domain for the BCT-S (Zhou et al., 2023). It is therefore critical to understand the extent to which BCT-S performance provides information about test-takers' performance in group discussions. To this end, a GDT was developed, and this study aimed to explore its effectiveness in assessing interactive competence by looking at language functions, conversational style, and test-taker perception.

A total of 94 first-year TUFS students took the GDT in groups of three or four. During the test, students expressed their opinions on two familiar topics in two tasks: a convergent task, where they needed to reach agreement, and a free discussion. After the test, they completed a questionnaire about their perceptions of various aspects of the test. Students' oral responses were transcribed and coded for language functions using Observation Checklist (O'Sullivan et al., 2002). Conversational styles were measured for goal orientation, interactional contingency, and quantitative dominance (Nakatsuhara, 2013); interactional patterns were categorized based on Galaczi (2008).

The results show that the GDT elicited a range of functions, but there were variations in the proportion of the different types of the functions between the two tasks, which echoes the literature (Nakatsuhara, 2013; Van Moere, 2007) and implies the necessity to include both tasks. Variations were also observed in conversation styles among individuals and in interactional patterns among different groups. In general, the students had favorable attitudes toward the GDT. While two-thirds felt nervous, the majority (around 80%) actively participated in discussions and considered the test a fair one with sufficient opportunities to demonstrate their speaking skills. The implications of these results regarding the use of the GDT and future research directions are also discussed.

Keywords

speaking test, BCT-S, group discussion, test perception, language function

CEFR-JSL（日本手話）の開発試案—「手話言語コミュニケーション能力」を中心に—

中野聡子(群馬大学)

浅田裕子(昭和女子大学)

藤田元(上智大学)

日本手話 (Japanese Sign Language: JSL) は、手話言語の1つであり、日本語とは異なる言語体系を有している。聴覚特別支援学校の教師や手話通訳者などの専門職には熟達度の高い日本手話スキルが求められるが、第2言語としての日本手話習得・通訳養成の指導が成功しているとは言い難い。この背景として、日本手話のカリキュラム、教材、試験の多くが手話言語学・手話教育学の基盤をもたない団体や民間で開発されており、学術的包括性、明確性、一貫性を兼ね備えたスタンダードが確立されていないことが考えられる。そこで、著者らは第2言語としての日本手話教育の質の向上を目的として、CEFR に準拠した到達指標である CEFR-JSL の開発に取り組んでいる。

CEFR-CV (2020)では、手話言語にも適用できるように能力記述文の修正や新たなスケールの加筆が行われた。しかし、手話言語の特性が十分に考慮されているとは言い難い。その1つが、【言語コミュニケーション能力】を音声言語・手話言語共通としつつも、【手話能力】を分離していることである。著者らはこの分離が CEFR-CV の手話教育導入への1つの障壁であると考えている。

本発表は、CEFR-JSL において【手話言語コミュニケーション能力】をどのように扱うべきかを提示するものである。具体的には、上記の問題意識をもとに、【手話言語コミュニケーション能力】の中に【手話能力】を統合した。そのうえで下位能力として、「一般的な使用可能言語の範囲」「音韻の使いこなし」「使用語彙の範囲と使いこなし」「文法的正確性」の4つからなる【言語能力】、「社会言語的な適切さと文化的レパートリー」からなる【社会言語的能力】、「テキストの構造 (受容・産出)」「空間設定と視点 (受容・産出)」「言語意識と解釈 (受容)」「存在と効果 (産出)」「発言権の取得・維持 (産出)」「叙述の正確さ (産出)」「手話処理の速度 (受容)」「手話の流暢さ (産出)」の8つからなる【運用能力】を設定した。

Keywords

日本手話, CEFR, 言語コミュニケーション能力, 手話能力

Classroom practice, assessment implementation, and Artificial Intelligence

Myles GROGAN (大阪大谷大学人間社会学部)

Some time has passed since AI became popular in 2022, and many videos, courses, and books are becoming accessible. Although academic literature is being produced, a gap remains in identifying the skills and competencies that a classroom teacher may need for both teaching as assessment. This practice report presents some interim considerations from a short investigation into the potential use of AI in a single university setting. Building on Taylor's (2013) idea of "multiple literacies" for assessment, it explores some practical AI skills a classroom teacher can use to prepare educational materials for teaching and assessment processes. The presenter also considers needs at a broader institutional level. Following a brief overview how AI literature is taking shape in the field of language assessment and education, the presentation opens with some simple use cases for classroom assessment. Prompts will be shared for participants to try these processes. The first of use case demonstrates how short items, such as multiple-choice items, may avoid some of the challenges novice users face with AI. Careful use of prompts allows not only rapid production of items, but also easily produces quizzes and tests that can be uploaded to LMS such as Blackboard or Moodle. Increased online practice of such items allows teachers more opportunity to analyse and find patterns in student practice with classical test analysis techniques. Reverse engineering processes and using prompts as test specification or blueprints also allow classroom teachers greater quality control. The second use case looks at creating custom GPTs for formative assessment in spoken work using a "Shadowing Assistant." This introduces how AI can be used to help provide feedback directly to students in productive practice, creating greater meta-awareness in language learning. This process highlights some of the challenges in creating custom support for speaking and, by extension, writing. It includes narrowing the specifications of an AI to provide consistent support, getting students signed up for accounts, and the differing digital literacy of students. This report shows both merits and practical challenges of using AI for classroom teaching and assessment. After demonstrating the beginnings of practical skills for AI in classroom assessment, the presentation closes by further considering practical and ethical barriers that AI may bring with it. Issues such as teacher training time, use of data, and institutional implementation, as well as the "multiple literacies" of all involved, will be briefly addressed before questions are invited.

Keywords

Artificial Intelligence, material creation, formative assessment

Workshop Information (ワークショップ情報)

題目：

機械学習による言語パフォーマンスの評価

講師 小林 雄一郎 (日本大学)

司会 平井 明代 (筑波大学)

日時：2024年10月6日(日) 9時00分～12時00分(休憩含む)

場所：2号館4階2-406教室

参加費：無料(非会員は1,000円)

定員：なし

申込期限：9月29日(日)

参加条件：機械学習については詳しく説明しますが、基礎的な統計学の知識(相関係数、回帰分析など)を持っていることが望ましいです。**(Rを用いたデモ)**を行う予定ですが、Rに関する知識は必須ではありません

目的：

1. 機械学習による言語パフォーマンスの評価について理解する
2. 決定木分析やランダムフォレストを用いた言語データの分析方法を学ぶ
3. 統計的なパフォーマンス評価を行う際の注意点を知る

手順：

1. 言語パフォーマンスの自動評価
 1. ライティング・スピーキングの自動採点(理論的枠組み)
 2. 学習者コーパス研究(データ構築、特徴量の選定)
 3. 自然言語処理(言語情報の定量化)
 4. 機械学習(自動採点システムの実装)
2. 言語パフォーマンス評価の実例
 1. 機械学習の基礎(線形判別分析、決定木分析、アンサンブル学習など)
 2. ランダムフォレスト分類を用いた実例
 3. ランダムフォレスト回帰を用いた実例
 4. Rによる機械学習のデモ
 5. 言語パフォーマンス評価に機械学習を用いる際の注意点
3. Q&A(諸般の事情により、上記の流れが一部変更になることがあります)

申込方法：

1. 下記URLまたはQRコードより申し込みフォームにアクセスし、必要な情報をご登録ください。
URL: <https://forms.gle/vNLSSjitjHC5twEu5>



2. 上記の申込方法で不具合がある場合はメールでの受け付けも致します。下記の情報を久保佑輔（福岡大学）[kuboy\(at\)fukuoka-u.ac.jp](mailto:kuboy@fukuoka-u.ac.jp) まで e-mail でご連絡ください。

- (1) 氏名・所属・メールアドレス（全員）
- (2) ワークショップの参加理由（全員）
- (3) 講師へのご質問（希望者のみ）

Workshop Information

Evaluating Language Performance with Machine Learning (Conducted in Japanese)

Lecturer: Yuichiro KOBAYASHI (Nihon University)

Chair: Akiyo HIRAI (University of Tsukuba)

Date: October 6, 2024, 9:00–12:00 (with intermissions)

Venue: Building #2, 4th floor, Room 2-406

Fee: Free (1,000 yen for non-members)

Maximum number of participants: No limit

Application deadline: September 29th

Prerequisite: Participants are expected to be familiar with basic statistical knowledge such as correlation coefficients and regression analysis. The workshop will include a detailed explanation of machine learning with demonstrations using R. However, basic data handling in R is not required.

Objectives

1. Understand the evaluation of language performance using machine learning
2. Learn methods for analyzing language data using decision tree analysis and random forests
3. Understand important points of statistical language performance evaluations

Procedure

1. Automatic evaluation of language performance
 1. Automatic scoring of writing and speaking (theoretical framework)
 2. Learner corpus research (data construction, feature selection)
 3. Natural language processing (quantification of linguistic information)
 4. Machine learning (implementation of automatic scoring systems)
2. Examples of language performance evaluation
 1. Basics of machine learning (e.g., linear discriminant analysis, decision tree analysis, ensemble learning)
 2. Analysis examples with random forest classification
 3. Analysis examples with random forest regression
 4. Demonstration of machine learning using R
 5. Key considerations in applying machine learning to language performance evaluation
3. Questions and answer session

How to register

1. To register for the workshop, please go to the following website with the URL or QR code below and fill out the necessary information.

URL: <https://forms.gle/vNLSSjitjHC5twEu5>



2. If there is a problem with online registration, applicants are also accepted by email to Yusuke KUBO (Fukuoka University) at [kuboy\(at\)fukuoka-u.ac.jp](mailto:kuboy@fukuoka-u.ac.jp)

Let us know the following information when you register for the workshop.

- (1) Your name, affiliation, and email address
- (2) Reason(s) for participating in this workshop
- (3) Questions for the instructor, if you have any (Optional)

AI 搭載の英文ライティング自動フィードバックツール

Write & Improve

無料 有料 生徒 教師



Write & Improve は、オンラインでライティング添削をする役立つツールです。ライティングスキル向上のためのヒントとして語彙選択、文法、スペルミスを含むフィードバックを瞬時に提供します。書き直して上達した様子もグラフで確認することができます。スマホ、タブレット、パソコンでご利用いただけます。無料版と有料版がございます。



Write & Improve

無料 サイト内課金あり

使い方は簡単! 今すぐアクセス  動画で確認 

1. トピックを選ぶ
2. 英文を書く
3. 瞬時に添削
4. 結果は CEFR で表示
5. 好きなだけ書き直してレベルアップ!

CEFR 表示

正解を教えるのではなく
気づきを促すフィードバック

有料 **生徒** **教師**

+Class View

2,800 円 (本体+税)
※請求書支払い可能

有料版では、教師は生徒のパフォーマンス（英作文）の進捗状況を確認できます。Write & Improve +Test Zone B2 Firstと IELTS タスク（採点料込）が 12 カ月間ご利用いただけます。

2,800 円 (本体+税) ※請求書での支払いも可能。

教師向け **今すぐアクセス**  **チュートリアル動画**  **無料トライアル** 

英語 4 技能テスト E-Vision モニター受験募集

ELPA から 2025 年発売予定の「英語 4 技能テスト E-Vision」では、「スピーキング、ライティング、リスニング、リーディング」の 4 技能の力をウェブサイト上のテストでコンピュータ、タブレット、スマートフォンにより測ることができ、「主体的・能動的な学習」で必要となる、総合的な英語力がどの程度身についているかを把握することができます。特に大学入学前の英語のクラス分けテストでご利用いただく際にも会場での一斉受験の必要がなく、自宅など、コンディションの整った環境で受験していただけます。また入学前でもまだ学生の手元にコンピュータが用意できていないという場合にも手持ちのタブレット、スマートフォンで受験可能です。

「英語 4 技能テスト E-Vision」のモニターテストを実施していただける学校を募集します。

■英語 4 技能テスト E-Vision 問題概要■

◇仕様：コンピュータ、タブレット、スマートフォンを使用したウェブサイト上の個別受験型テスト	<内容>
◇出題レベル：高校英語総合レベル	1：リスニング（多肢選択）
◇試験時間：リスニング 20 問 25 分	2：リーディング（多肢選択）
リーディング 20 問 25 分	3：ライティングテスト（文再現・整序／自由産出問題出題可能性あり）
ライティング 20 問 10 分	4：スピーキングテスト（文再現・整序／自由産出問題出題可能性あり）
スピーキング 20 問 10 分	
合計 約 70 分（※練習問題、自由産出問題除く）	

■実施概要■

・受験料：無料
・時期：2024 年 12 月まで（※ご相談ください。）
・対象：高等学校／高等専門学校／短期大学・大学 等
・人数：40 名以上
・テスト終了後、引き続きアンケートに答えていただきます。所要時間は約 10 分です。
・受験者名簿を事前にデータでいただきます。
・採点データは、モニターテスト実施後 2 週間程度でお届けします。なお、詳細なスコアレポート等は提供いたしません。

※個人情報、テストの解答およびアンケートのデータに関しては、調査・研究及びデータ分析の目的以外には一切使用いたしません。

NPO と共同でテスト研究・開発・分析しませんか？

●理念を共にする研究者（大学院生含む）募集中●

英語運用能力評価協会（ELPA）は、2003 年 4 月 15 日に東京都知事の認証を得て設立された特定非営利活動法人（NPO）です。

言語能力における測定・評価の研究者、英語教育の研究者、教育現場の先生方および出版社、教育関連企業などが参加した協同作業により、第三者機関として英語教育における公正なテストと評価を提供し、日本の英語教育の現場に対して新しい指導法や学習方法の提言を行ってまいります。

ELPA のテストの共同研究・開発・分析に携わってくださる研究者を随時募集しています。お気軽にご連絡ください。

ELPA は新しい評価システムと テスト教材の研究・開発を通して 日本の言語教育をサポート & デザインしていきます

特定非営利活動法人 英語運用能力評価協会（ELPA）

〒162-0806 東京都新宿区榎町 39-3-501 Tel.03-3528-9891/Fax.03-3528-9892

e-mail: elpa@english-assessment.org

URL: <https://english-assessment.org/>

ELPA Association for
English Language
Proficiency Assessment

 OXFORD
PLACEMENT TEST

Fast, accurate placement testing for
upper secondary and adult learners

高校生・大学生に最適なプレースメントテスト
正確な結果が瞬時に分かります

Test your students' English level quickly and easily
with the *Oxford Placement Test*, which has helped
place millions of students in the right class.

- 完全オンラインのため、学習者は家からでも教室でも
テストを受けることができます。
- すぐに結果が出る自動採点方式によりテスト当日
からクラス分けと指導が可能です。
- 正確な結果が得られるコンピューター適応型テクノロ
ジー。
- 学習管理システムにより、テスト実施から成績管理ま
でがスムーズに。
- 正解率（120点満点）とテストにかかった時間に基づき
CEFR（Pre-A1～C2）レベルを判定

Learn more:

[www.oup.com/elt/
oxford-placement-test](http://www.oup.com/elt/oxford-placement-test)





Versant English Speaking and Listening Test

VERSANT[™]
by Pearson

Versant Speaking Test のアップグレード版 Versant English Speaking and Listening Test が新登場！

AI を活用した採点の精度向上、より優れたテスト体験に加え、新たに Listening と Manner of Speech のスコアを提供します。

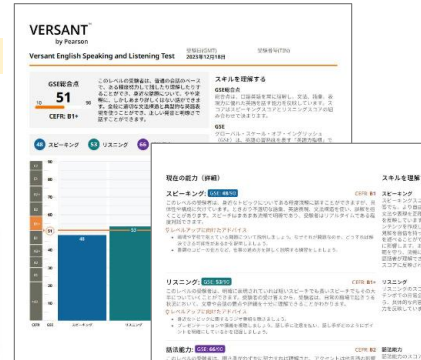
主な特徴

新たな設問タイプが追加(※受験時間は従来のテストと変わりません)
 ・「Passage Comprehension」「Conversations」(正確に話の内容を聞き取る能力を測定)
 ・「Open Questions」(自分の言葉や考えを用いて話す能力を測定) ※採点対象

新たな採点項目が追加
 ・Listening
 ・Manner of Speech(話法能力)

発音のバリエーションが追加
 ・出題音声の英語アクセントのバリエーションが増加

Global Scale of English(GSE)で採点
 ・採点スコアは、Global Scale of English で表示



リモートモニタリング機能(追加オプション)

受験者の音声と顔を認証する「なりすまし防止」機能で、テスト管理者の負担を軽減！
 遠隔で受験状況を確認することができます。受験者の映像と音声にアクセスができるので、受験者へのアドバイスにも役立ちます。

スマホアプリ・ウェブベース対応

Courseware

おすすめタイトル それぞれ独自の特徴を備えた一般英語コースのテキストをご紹介します。



Speak out 3rd Edition

Antonia Clare, JJ Wilson, Frances Eales, Steve Oakes, Lindsay Warwick, Sheila Digney, Lynda Edwards

BBCと共同開発した、ピアソンのベストセラー『Speak out』の最新版。4技能を総合的にカバーしながら、特に学習者のスピーキング能力を伸ばしたいと考える先生方に最適なコースです。人気テレビ番組、街頭インタビューなど、世界中で話されている本物の英語に触れられる BBC 提供のビデオ素材を使用しています。

8 Level British English Print and digital GSE: 22-90 CEFR: A1-C2



Connectivity

Joan Saslow, Allen Ascher

全くの初心者から上級者まで幅広くカバーする、社会的言語能力に焦点を当てた新しいコミュニケーションコース。綿密なスピーキング教育を特徴としており、4技能に加え、雇用に必要なソフトスキルの習得を統合。異文化に対応する能力の育成も含んだ独自のシラバスは、学習者が実社会ですぐに応用できるスキルの向上をサポートします。

6 Level American English Print and digital GSE: 16-84 CEFR: Pre A1-C1



StartUp

Series Consultant: Ken Beatty

革新的な総合英語コースである『StartUp』は、フレキシブルに学習できるよう、実用性が高くメディアリッチなコンテンツを通じて、学習者のモチベーションを高めます。学習者が社会生活、学業、仕事でのコミュニケーションに必要な言語を身につけると同時に、協調性や批判的思考力を養うことができるようデザインされています。

8 Level American English Print and digital GSE: 22-84 CEFR: A1-C1



お問い合わせ ピアソン・ジャパン株式会社 Email: elt.jp@pearson.com

採点時間は約 1 秒！ 人の手を借りず即時採点、即時添削
英語エッセイライティング自動採点サービス



桐原 AI エッセイライティング

powered by PLAYSQUARE

エッセイ自動採点システム〔AEES : Automated Essay Evaluation System〕搭載

学習に喜びと感動を

桐原書店



お問い合わせ・詳細はこちら ▶▶

olbd@kirihara.co.jp



公式 TOEFL® 英単語

ETS / ETS Japan / 笹山尚子



ついに誕生！
10月5日
発売開始！

A5判 560ページ 赤シート付き
ISBN 978-4-7890-1895-1
定価3300円(税込)
発行: the japan times 出版

- ★ ETS が選んだ TOEFL® 受験に本当に必要な見出し語約 2,000 + 派生語約 1,800 を頻出順に完全収録
- ★ 本番テストに基づいた例文を使用
- ★ 英語語義付き。本当の意味 + ニュアンスがわかる
- ★ 音声は本番テストのナレーターによる収録
- ★ TOEFL iBT®, TOEFL ITP® のどちらにも対応

ETS Japan <https://www.officialtestprep.jp/>
株式会社ジャパンタイムズ出版 <https://publishing.co.jp/>

【公式教材ショップ】
で販売中！



著者より



早稲田大学 准教授
笹山尚子先生

本書のおすすめポイント

「公式 TOEFL® 英単語」は、テスト開発元の ETS による世界初の公式英単語集です。ETS が持つ TOEFL® 語彙コーパス(これまでに TOEFL iBT® および関連教材に登場した英単語のリスト)を分析し、出題頻度と重要度が特に高い英単語約 2000 語を収録しています。例文は、過去に出題された問題に基づいて作成されており、音声も TOEFL® のナレーターを模した声優によって録音されているのが特長です。

本書の使い方

TOEFL® のスコアアップには、英単語を文脈の中で学習するのが効果的です。本書に収録された例文を参考に、単語の意味のニュアンスやコロケーション(どんな単語と一緒に使われているか)、発音に注目して、学習を進めましょう。また、英単語を実際のコミュニケーションの中で使えるようになるには、アウトプットが欠かせません。日頃から新しく学んだ英単語を使って、話したり書いたりする機会を探してみてください。本書が皆さんの夢の実現に少しでも役立てることを願っています。

Core (最重要単語) と Expanded (重要単語) の 2 部構成

[CORE]



[EXPANDED]

ETS, the ETS logo, TOEFL®, TOEFL iBT®, TOEFL ITP®, are registered trademarks of Educational Testing Service, Princeton, New Jersey, U.S.A., and used in Japan under license by ETS Japan. Portions of the contents are copyrighted by ETS and used with permission.

英語コミュニケーション能力判定テスト

CASEC
Computerized Assessment System for English Communication

詳しい内容は、
WEBサイトをご覧ください！

CASEC 検索

累計約320万人が受験！
大学・企業に選ばれるCASEC

CASEC(キャセック)は、株式会社 教育測定研究所が開発・運営しているインターネット上で受験できる英語コミュニケーション能力判定テストです。「顔認証機能」(*)による本人確認にも対応しています。

※ 顔認証機能は、オプションサービスとなります。

【主な特徴】

- 短い試験時間 (平均 40 分～ 50 分)
- 試験終了後すぐスコアを表示
- TOEIC®スコア目安・英検®級目安もすぐ分かる
- インターネットでいつでもどこでも受験できる
- 最新のテスト理論に基づいた高い判定精度

JIEM
株式会社教育測定研究所

〒108-0075 東京都港区港南1-8-15 Wビル13階 CASEC学校営業

TEL 03-6897-3752 (平日 9:30～17:30 土・日・祝日・年末年始を除く)

<協賛企業・法人・団体の一覧（50音順）>

ETS Japan 合同会社 *

<https://www.etsjapan.jp>

特定非営利活動法人 英語運用能力評価協会 *

<https://english-assessment.org>

株式会社エル・インターフェース

<https://www.l-interface.co.jp>

オックスフォード大学出版局株式会社 *

<https://www.oupjapan.co.jp/ja>

株式会社 教育測定研究所 *

<https://www.jiem.co.jp/>

株式会社桐原書店 *

<https://www.kirihara.co.jp>

株式会社グローバルキャスト G-TELP 日本事務局

<https://g-telp.jp>

Cambridge University Press & Assessment *

<https://www.cambridge.org>

国際交流基金・日本語試験センター

<https://www.jpf.go.jp> (国際交流基金トップページ)

一般財団法人国際ビジネスコミュニケーション協会

<https://www.iibc-global.org/index.html>

(財) 日本英語検定協会

<https://www.eiken.or.jp>

British Council

<https://englishonline.britishcouncil.org/jp/>

ピアソン・ジャパン株式会社 *

<http://www.pearson.co.jp/>

VELC 研究会（事務局 株式会社金星堂）

<http://www.velctest.org/index.html>

We are deeply grateful to the following companies, corporations, and organizations for their support of our association.

Four companies, corporations, and organizations have Commercial exhibits at the JLTA 27th Annual Conference. We would like to express our sincere gratitude for their contribution to the conference.

当学会は以上の企業・法人・団体様にご支援を頂いております。
第27回全国研究大会においては、7の企業・法人・団体様に展示協賛を
いただいております。本大会へのご貢献に厚く御礼申し上げます。

Commercial Exhibits (展示協賛企業) *

Exhibits are located in the 2-406 and 2-407 on the 4th floor in the Building 2.

展示は、2号館4階2-406, 2-407教室で実施いたしております。

日本語テスト学会 (JLTA) (2024年度) 全国研究大会発表要綱 (簡易版)

Handbook of the 27th Conference of the Japan Language Testing Association

発行日：2024年9月22日

発行：日本語テスト学会 (JLTA) 会長：渡部良典 (上智大学)

事務局：〒960-8516 福島県福島市栄町10-6

久保田恵佑研究室

TEL: 024-581-5533 (直通)

編集：日本語テスト学会 全国研究大会実行委員会
