

第49回全国英語教育学会福岡研究大会

自動採点を前提とした学内版英語 スピーキングテストの開発と予備 的試行

望月正道(麗澤大学)・千葉庄寿(麗澤大学)

mmotizu@reitaku-u.ac.jp

2024年8月25日 福岡工業大学

1. はじめに

学習指導要領：目標

5領域で英語を運用できる能力を身に
着けさせる

指導と評価

産出的技能では、実技試験(パフォーマンステスト)が必要

2. 先行研究

小泉(2022): 教育課程の中でのST作成方法について詳述

深澤(2022): 高校英語コミュニケーションIでのST実践例

Koizumi & Hirai (2012)

物語を再話させるSTがどのような言語機能を抽出し、標準化されたSTとの関係を調査

Jamieson, Wang, & Church (2013)

**留学生に対するEAPプログラムのプレース
メントテストとして学内のST(PIE)の作成**

標準STと比較

近藤・石井(2017)

教育現場で頻繁に実施されない理由

- ① 人的・時間的コストが膨大である。
- ② 採点者間信頼性に問題がある。
- ③ 受験者への即時のフィードバックが難しい。

➡ 自動採点で解決

Cox & Davies (2012)

自動採点は、発話された単語をテキストに変換する音声認識プロセスを利用

音声の波形のパターンを解析し、音声モデルと一致させ、単語を特定

問題：自然な発話は認識しづらい

解決策：発話の領域を限定することで認識精度を上げることができる

Cox & Davies (2012)

誘発された口頭応答(Elicited Oral Response) 発話を限定

外国語を聞いて、それを繰り返して言う

受験者は外国語の文を聞き、その意味を理解し、それを繰り返すために外国語を処理をする必要

文や文章の音節数、文法、語彙を高度化させることで、テスト項目の難易度を変化

Cox & Davies (2012)

採点方法:「各文の単語を別々の項目として認識し、正しく発話された単語の数を評価するようにプログラムする」(p.605)

EORを音声認識による自動採点した結果は、限られた目的のためのスピーキング能力を予測することができる

Versant Speaking Test (Pearson, 2008)

自動採点システムのみで採点
電話またはコンピュータで受験
テスト時間は約20分
6つのパート

A「音読」、B「復唱」、C「質問」、D「文の構築」、E「ストーリーリテリング」、F「自由回答」

パートAからEが自動採点

自動採点のアルゴリズムは非公表

TOEIC® Speaking Test(Educational Testing Service, 2006)

自動採点システムと人間による採点を併用

コンピュータで受験。テスト時間は約20分
6つの問題形式

「音読問題」、「写真描写問題」、「応答問題」、「提示された情報に基づく応答問題」、「意見を述べる問題」

採点規準は、発音、イントネーション、アクセント、文法、語彙、一貫性、内容の妥当性、内容の完成度

ひとつの疑問

**Versant形式のSTを
音声認識の単語一致率のみで採点し
たら
受験者のスピーキング能力をどの程度
測定することができるのだろうか？**

研究課題

研究課題1:

Versant形式のSTを音声認識の単語一致率のみで採点した結果は、標準STの結果とどのような関係にあるのか

研究課題2:

Versant形式のSTの設問は、受験者をどのように弁別するのか

研究課題

研究課題2

Versantの「音読」、「復唱」、「質問」、「文の構築」、「ストーリーリテリング」という出題形式が、TOEIC® Speaking Test(Educational Testing Service, 2006、以下TOEIC_ST)の得点とどのような関係にあるのかを模索する

3.1学内STの作成

3.1.1 Versant 問題の分析

Versant の練習問題集(Test Prep Guides, 2022)の10回分のテスト問題をコーパス化し、パートAからEの特徴を分析

分析ツールはNew Word Level Checker (水本、2022)でCEFR-Jを分類語彙表に指定

表1 Versant Practice Examsのパートごとの平均文長(語数)とCEFR分類(%)

	文長	A1	A2	B1	B2	PN	NA
A	10	76.8	12.7	5.2	1.7	2.5	1.1
B	9.8	78.5	11.4	5.2	1.6	1.2	2.2
C	13.2	76.5	9.1	5.6	2.6	3.1	3.1
D	8.8	81.2	11	4.9	1.6	0.6	0.8
E	64.1	81.6	11.2	3.3	0.6	2.4	0.8

***PN = Personal Noun NA = Not Available**

Versant 練習問題集の分析

問題文の平均語数

パートA, B	10語
パートC	13語
パートD	9語
パートE	64語ほど

語彙レベル

CEFRのA1、A2レベルで約90%

B1が5%

B2が2%程度

3.1.2問題項目の作成

Versantの問題例を参考に、各パートの問題をつぎの平均文長になるように作成

パートA(音読): 5問 文長平均10語

パートB(復唱): 5問 文長平均10語

パートC(質問): 7問 文長平均10語

パートD(文の構築): 7問 文長平均9語

パートE(ストーリーリテリング): 1問 文章
長 65語

3.1.3 受験の手順

受験者はCALL教室で、マイク付きのヘッドセットを装着

マイクで録音が可能であることをチェック

ヘッドセットからの指示に従って解答

パートA: 音読

指示文のあとに5つの英文がコンピュータ
画面上に提示

「問題文1を音読してください」という指示
のあと、10秒以内に音読

10秒後、つぎの指示

第5問まで終わると、「これでパートAは終了
です」の指示

10秒後にパートBの指示文

パートB: 復唱問題

読み上げられる英文を10秒以内に復唱

10秒後に、つぎの英文が読み上げられる

第5問まで終わると、「これでパートBは終了です」の指示

10秒後にパートCの指示文

パートC: 質問

読み上げられる英語の質問に10秒以内に解答

10秒後に、つぎの質問が読み上げられる

第7問まで終わると、「これでパートCは終了です」の指示

10秒後にパートDの指示文

パートD: 文の構築

読み上げられる3つの英語のまとまりを10秒以内に正しい語順に並べ替えて構築した英文を言う

10秒後に、つぎの3つの英語のまとまりが読み上げられる。

第5問まで終わると、「これでパートDは終了です」の指示

10秒後にパートEの指示文

パートE:ストーリーリテリング

読み上げられる英語の文章の内容を30秒以内にできるだけ多く再生

30秒後に、「これでパートEは終了です」の指示

10秒後に、すべてのスピーキングテストが終了したことが告げられる。

パートE

指示文と英文はすべて第1著者が読み上げた。

Zoomのレコーディング機能を使い、パワーポイントのスライドに各パートの指示文を示しながら、指示文と英文を読み上げるのを録画した。

3.2データ収集と処理

文科系の大学1年生から4年生25名

TOEIC SWテストを2024年1月に受験
23名は、同日の午前または午後に、
2名は翌日に学内STをCALL教室で受験
学内STの受験時間は約11分

音声ファイルは、Office365のWordの
Transcribe機能を利用してテキスト化

3.3 学内STの採点

単語の一致率をもとに採点

パートA、B、Dは、1文の語数のうち正確に再生できた語数の割合を得点

パートB: 問題文 Can you help me with this math problem?

解答: Can you help me with math program? 得点 0.75

パートDは、文構築の問題だが、位置に関係なく、語が再生されていれば、得点に含めた

パートC: 質問

解答した発話に質問の答えとなるキーワード1語が含まれていれば、1点、含まれていなければ0点

質問: If inspiration strikes you, do you have an exciting idea or feel bored?

予想される解答”I have an exciting idea.”, “An exciting idea”, “Exciting”など

解答に**exciting**が含まれていれば、1点

パートE:ストーリーリテリング

どれだけ正確に聞き取った英文の内容を再生できたかが問題

問題文の**内容語の異なり語数のうち正確に再生できた語数の割合を得点**

問題文は延べ語数79、内容語は29語

例: 解答の発話のうち原文で使われている内容語が7語ならば、

得点: $7 \div 29 = 0.24$

パートE:ストーリーリテリング

パートA～D:10語程度の文が5～7問あり、
5～7点の配点

パートE:79語の文章を再生する課題

パートEは8点分に相当すると考え、上の
式で得られた得点を8倍

学内STの満点は30点

解答の音声データは、WordのTranscribe
機能でテキスト化し、第1著者が、採点

4. 結果

4.1 学内STとTOEIC_STの得点

学内ST

平均 13.0、 最高 22、 最低 6

TOEIC_ST

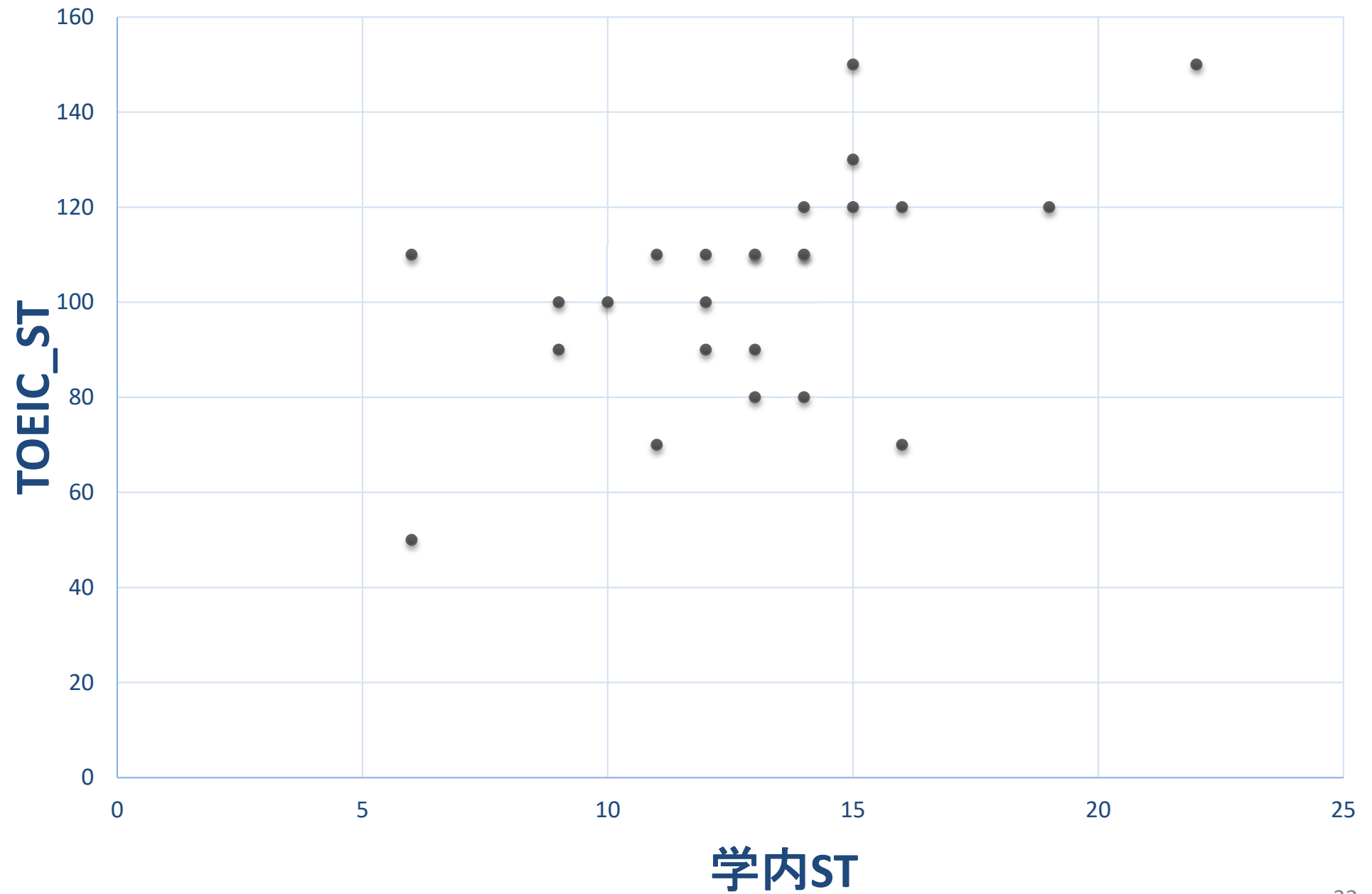
平均 104、 最高 150、 最低 50

スピアマン順位相関係数、0.56

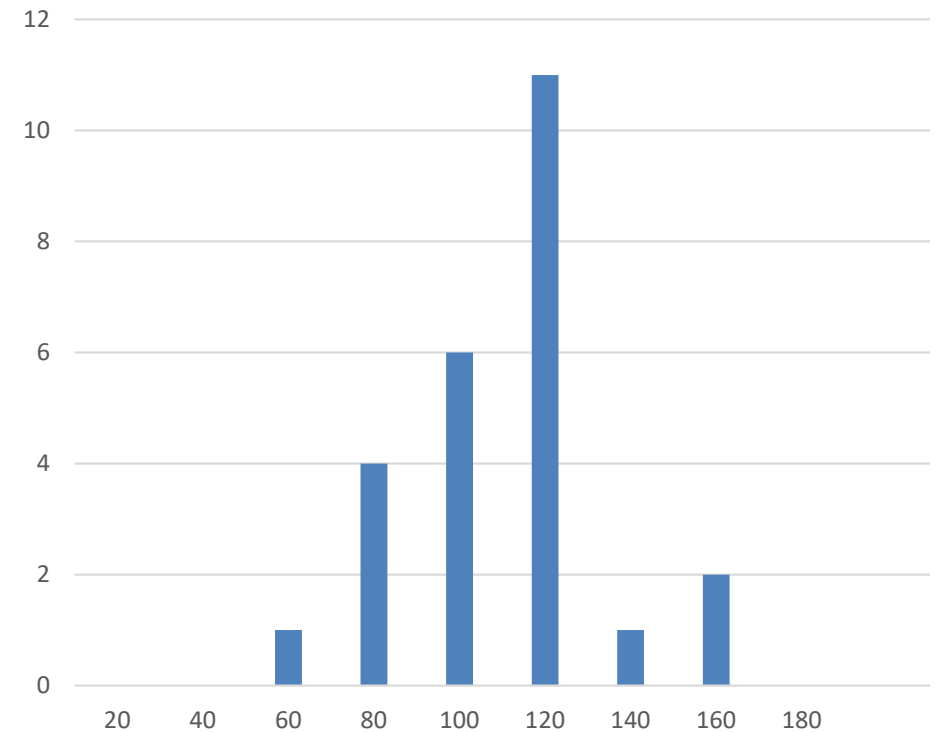
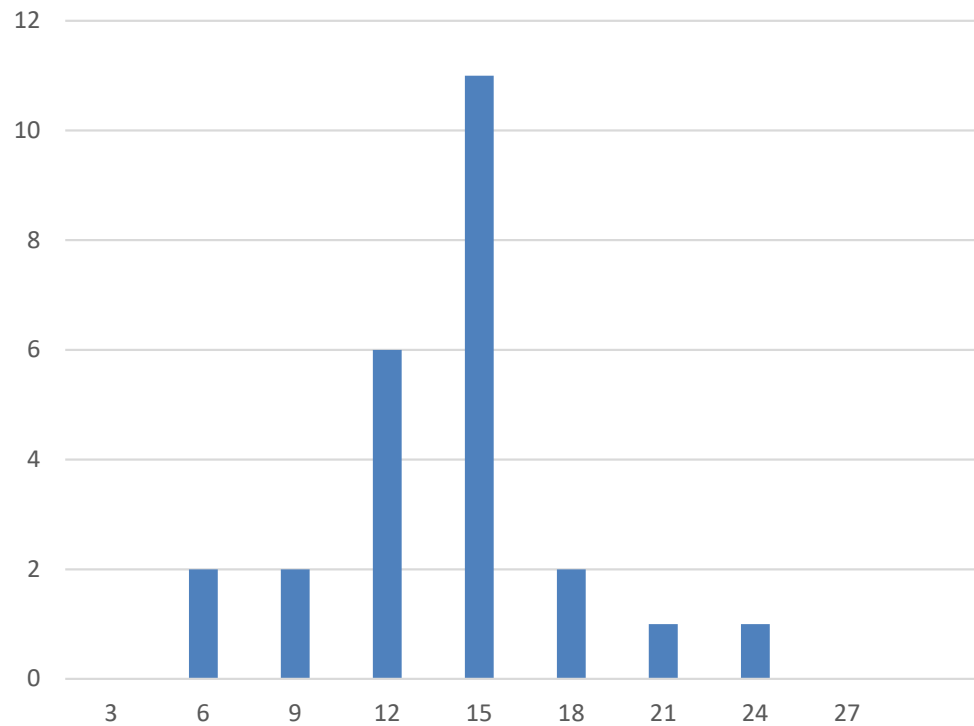
表2 学内STとTOEIC_STの平均・標準偏差

N=25	学内ST	TOEIC_ST
平均	13.0	104
標準偏差	3.5	23
最大	22	150
最小	6	50

学内STとTOEIC_STの得点散布



学内STとTOEIC_ST得点の分布



4.2学内STのパートごとの平均・標準偏差・正答率

N=25	A	B	C	D	E
平均	4.3	3.2	1.5	2.4	1.5
標準偏差	0.8	0.8	1.3	0.8	1.2
正答率	0.86	0.64	0.21	0.48	0.19

学内STパート得点とTOEIC_ST得点の相関係数

N=25	A	B	C	D	E
スピアマン順位相関係数	0.24	0.12	0.4	0.54	0.56

5. 考察:研究課題1

2つのSTの相関係数: 0.56

2つのSTの得点分布は、きわめて似かよっている

Versant形式のSTを音声認識の単語一致率のみで採点した結果は、受験者のスピーキング能力をある程度測定していると考えられる。

これは、Cox & Davies (2012)の主張を裏づけるものといえる。

5. 考察:研究課題1

本研究では、自動音声認識機能によってテキスト化された解答を人手によって採点したが、テキスト化された答案を原文と機械により一致率を算出できるようになれば、まさに” a cost-effective alternative to conducting costly face-to-face speaking proficiency interviews”となりえる可能性があるだろう。

5. 考察:研究課題1

原因1

学内ST はすべての問題形式に音声で解答するわけだが、その発音が正確でないと、自動音声認識システムは正しい単語に変換できない。

学内STの得点は、受験者のスピーキング能力の発音の正確さにもっとも依存している

発音は、TOEIC_STにおいても重要な構成要素として利用されているので、発音の良し悪しの点で、2つのSTは中程度の相関関係を築いていると考えられる。

5. 考察:研究課題1

原因2

パートBの「復唱」、Dの「文の構築」、Eの「ストーリーリテリング」は、聞いた内容を記憶して、再生する必要がある。第二言語の記憶容量は、その熟達度とともに発達していく

解答が第二言語の記憶容量に依存するために、学内STは受験者のスピーキング能力を測定し、TOEIC_ST得点と中程度の相関が生じていると考えることができる。

5. 考察:研究課題2

学内ST正答率 低い

TOEIC_ST得点との相関係数 高い

学内ST正答率 高い

TOEIC_ST得点との相関係数 低い

5. 考察:研究課題2

「ストーリーリテリング」は、79語の英文を聞き、その内容を30秒間で伝える課題である。

TOEIC_STの「応答問題」や「提示された情報に基づく応答問題」に求められる技能と近いものが求められる「提示された情報に基づく応答問題」では、3つめの質問は、テキストの複数の情報を統合して解答しなければならない

「ストーリーリテリング」に求められる能力と近いものが必要とされると考えられる。

「ストーリーリテリング」はTOEIC_ST得点ともっとも高い相関関係にあると推察される

5. 考察:研究課題2

「文の構築」の相関係数 0.54

受験者は、ばらばらに聞こえてくる語句から文全体の意味を把握して、元の文に復元しなければならない

瞬時の理解、語句の記憶、文の構築力が求められる

5. 考察: 研究課題2

「質問」の相関係数 0.40

原因1

この課題は質問文を理解して、その答えを言わなければならないために、語彙力が問われる

If land is barren, does it produce crops a lot or a little?という質問であるが、barrenとcropという単語の意味を知らなければ、正解することはできない。

より多くの単語を知っていることが重要

そのために、パートAやBよりも相関係数が高かったと考えられる。

5. 考察:研究課題2

原因2

パートDやEほど高くなかった理由としては、解答を2つの選択肢から選ぶ形式が多く、2分の1の確率で正解できてしまう

これら2つの原因から、「質問」の相関係数は、0.40と中程度のものだったと推定される。

5. 考察

**Versant形式に基づく学内STは、ある程度
受験者のスピーキング能力を測定できる
手軽なスピーキングテストになりえる可能性
がある**

5. 考察

本研究では、音声データをテキスト化する工程、テキスト化されたデータを採点する工程は人的作業によってなされたが、それらの工程も機械によって処理することができれば、まさにa cost-effective alternative to conducting costly face-to-face speaking proficiency interviews” (Cox & Davies, 2012, p. 614)ということができるだろう。

5. 考察

スピーキングテストとして、スピーキング能力の測定に利用できるだけでなく、採点の自動化と即時のフィードバックシステムが可能になれば、学習者が日々スピーキングの練習をするのに活用できる

6. 結論

限界1

参加者数が25名と少ない

限界2

受験者が聞き、再生が求められる英文が
非英語母語話者によって録音されたもの

6. 結論

今後の研究の方向性

1) 自動採点システムの確立

2) 異なる版を作成し、等化

参考文献

小泉利恵(編著)(2022)『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

国際ビジネスコミュニケーション協会(2024)「TOEIC® Speaking Test」

<https://www.iibcglobal.org/toEIC/test/speaking/about/format.html>

近藤悠介・石井雄隆(2017)「英語学習者の発話自動採点システムの開発と英語教育プログラムへの導入可能性の検討」
Language Education and Technology, 54, 23-40.

深澤真(2022)「ST具体例」小泉利恵(編著)(2022)、83-96.

水本篤 (2022).「New Word Level Checker の概要」『外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会報告論集』12, 1-24. <https://doi.org/10.31219/osf.io/whr9a>

ALC Press (2010)Standard Speaking Test. ALC Press.

参考文献

Cox, T.L., & Davies, R.S (2012) Using automatic speech recognition technology with elicited oral response testing. CALICO Journal, 29(4), 601-618.

Educational Testing Service (2006) The TOEIC® Speaking and Writing Tests. Educational Testing Service.

Jamieson, J., Wang, L., and Church, J. (2013) In-house or commercial speaking tests: Evaluating strengths for EAP placement Journal of English for Academic Purposes, 12, 288-298.

Koizumi, R., and Hirai, A. (2012) Comparing the Story Retelling Speaking Test With Other Speaking Tests. JALT Journal, 34, 35-59.

Pearson (2008) Versant Speaking Test. Pearson.

Test Prep Guides (2022). Versant English Speaking Voice Test Practice Exams with Sample Responses, Free Recordings, and Exam Tips. Exam Study Aids & Media.

資料1:学内スピーキングテスト

指示:パートAは音読の課題です。指示にしたがって、回答してください。では文1を音読してください。(10秒ポーズ)

- (1) Food loss is a big problem worldwide.
- (2) The couple hopes to find a job in the country.
- (3) Starting a new business can be really challenging.
- (4) I need to concentrate more on my job to get promoted.
- (5) In the evenings, he attends Italian classes at a community college.

資料1:学内スピーキングテスト

指示:パートBは復唱の課題です。聞こえてくる英語をそのまま繰り返して言ってください。たとえば、I like playing tennis.と聞こえてきたら、I like playing tennis.と繰り返して言ってください。それでは始めます。

- (1) It is very cold here in winter. (10秒ポーズ)
- (2) Can you help me with this math problem?
- (3) You need to turn right at the second traffic light.
- (4) How many people do you expect to attend the contest?
- (5) I enjoyed window-shopping with no intention of buying anything.

以上でパートBは終了です。

資料1:学内スピーキングテスト

指示:パートCは語彙の課題です。聞こえてくる英語の質問に簡単に答えてください。たとえば、What day comes after Sunday?という質問には、Mondayと答えます。それでは始めます。

(1)What would you use to cut vegetables?

(2) When you sleep, are your eyes open or closed

(3) What do hens lay?

(4)Is the result of something a consequence or a concern?

(5) Which do you use when you sleep, a pillow or a ladder?

(6) If inspiration strikes you, do you have an exciting idea or feel bored?

(7) If land is barren, does it produce crops a lot or a little?

以上でパートCは終了です。

資料1:学内スピーキングテスト

指示:パートDは並び替えの課題です。聞こえてくる英語の語句を並び替えて英文を作ってください。たとえば、1000 meters / swim / I can'tと聞こえたら、I can't swim 1000 meters.と答えます。それでは始めます。

(1) always comes/ he / late to school

(2) to chew / are too hard / these biscuits

(3) until Friday / will be postponed / the meeting

(4) all employees / a special bonus next month / will receive

(5) is the most exciting / that book / I've ever read

以上でパートDは終了です。

資料1:学内スピーキングテスト

指示:パートEは聞いた話を再現する課題です。聞こえてくる英語の話を30秒間でできるだけたくさん再現してください。たとえば、登場人物、場面、状況などを話してください。英語の話が終わったらビープ音がなります。それから30秒間で話してください。それでは始めます。Lisa likes classical music and tries to learn the titles of tunes she listens to. One day she heard beautiful music on her radio. She didn't know the title of the tune, but she was quite sure she had heard it before. She wondered why. Several weeks later, she came across her father's collection of records. She found the title of the tune she had heard on the radio. Her father must have played it when she was little. (30秒ポーズ) 以上でパートEは終了です。