

## CBT 方式によるスピーキングテストの現状

篠原亜紀・夷石寿賀子・石田華奈子・李文鑫

### 1. 背景

近年、外国語教育において、人と人との円滑なコミュニケーションを目標とする学習が重視されるようになり、教育現場では、「聞く」「話す」「読む」「書く」の4技能を育成するための指導や評価が求められている。

日本国内の英語教育においては、平成29～30年に学習指導要領が改訂されたが、たとえば、高等学校学習指導要領の改訂のポイントに、「目的や場面、状況などに応じて外国語でコミュニケーションを図る力を着実に育成」と示されている（文部科学省 2018a）。また、新指導要領では、「話す」を「やり取り」と「発表」の2領域に分け、目標を「聞く」「読む」「書く」「話す（やり取り）」「話す（発表）」の4技能5領域とし、これらを総合的に扱うことを一層重視する科目と、「話すこと」「書くこと」による発信の能力の育成を強化する科目が新設されている（文部科学省 2018b）。大学入試においても、2020年度の「大学入学共通テスト」より、英語科目で従来の2技能に「話す」「書く」を合わせた4技能を測ることになり、スピーキングテストが導入されることが決定した<sup>(1)</sup>。

日本語教育においても、コミュニケーションが重要視されるようになってから久しい。しかしながら、大規模のスピーキングテストは未だ存在しない。国際交流基金は、「日本語能力試験」へのスピーキング科目導入の可否を探るため、1998年から2002年にかけて調査と試行試験を実施した（日本語能力試験企画小委員会口頭能力試験調査部会 2003）が、スピーキング科目の増設には至らなかった。また、2014年には「JFS 準拠ロールプレイテスト」（国際交流基金 2014）が公開されたが、本テストはプレースメントテスト等、教育現場での活用を想定しているものであり、公式に合否判定やレベル認定をするものではない。

スピーキングテストは、他の技能よりも実施にも採点にも時間と手間がかかり（小泉 2015：55）、大規模テストの実施は容易ではない。しかし、CBT（Computer Based Testing）方式を導入することによって、一度に多数の受験者を対象としたスピーキングテストの実施が可能となる。CBT方式のスピーキングテストは、ヘッドセットを装着してコンピューターで受験するもので、受験者は音声や画像によって出題された問題に対し、マイクを通して回答する。回答は録音され、後に採点者に送られる。CBT方式は、TOEICや英検など英語のスピーキングテストですでに導入されており、大学入試においても、京都工芸繊維大学（羽藤・神澤 2015）

や佐賀大学 (早瀬ほか 2018) のように、CBT 方式によるスピーキングテストを独自に開発し、英語科目の試験として実施している例がある。

コミュニケーションを重視した教育が進む中、今後、日本語教育においても大規模スピーキングテストの需要はますます高くなることを見込まれる。そこで、筆者らは、現在行われている CBT 方式によるスピーキングテスト (以下、CBT) がどのようなものか、現状を把握するための調査を行った。本稿では、その調査結果について報告する。

## 2. CBT の現状

調査対象とした CBT は、資料1の通りである<sup>(2)</sup>。対象言語は限定せず、CBT として実施されているものを調査対象とした。スピーキングテストとして独立しているものだけでなく、多技能を測るテストの一部としてスピーキングテストがある場合も対象とした。また、テストがレベル別に分かれている場合には、学習者・受験者が多いと思われる A2~B1相当のレベルを中心に調査した。

各テストについて、公式 Web サイトやテスト問題集などを閲覧し、公開されている情報を洗い出して整理した。また、Web サイト等にサンプルが掲載されている場合は視聴・体験した。調査した CBT の概要を、以下に項目を整理して述べる。

### 2.1 問題数・時間

多くの CBT において、問題数は10問前後であるが、中には3問程度のもの (Avant STAMP、以下、STAMP) から、60問以上になるもの (Versant) まであり、多様である。所要時間は、スピーキングが多技能テストの一部として設置されている場合は10~20分程度が主流である。単独のスピーキングテストの場合は、最大で60分になることもあるが (ONiT)、20~30分程度のものが多い。

### 2.2 問題形式

CBT の問題形式はテストによって多様であるが、英検 CBT/英検 S-CBT (以下、英検) の問題形式は他の CBT でもよく採用されており、CBT の特徴を表しているといえる (表1)。

英検では、ウォーミングアップの後、文章とイラストが表示され、20秒で黙読するよう指示される。準備時間の後、音読をし、その後、音読した文章についての質問に答える。さらに、イラストについての質問に答え、最後は、受験者自身について答える。大問4の2問目は、質問が再生された後、「Yes」と「No」のボタンが表示され、10秒以内にどちらかをクリックすると、次の質問が再生される仕組みである。

表1 英検（3級）のテスト構成

大問	問題数	内容	回答時間	出題例
—	—	Warm-up	10秒	What is your favorite food? What did you eat this morning?
1	1問	音読	45秒	Rugby is a popular team sport around the world. Rugby players have to play well together, so they need to practice very hard. Japan will hold the Rugby World Cup in 2019.
2	1問	音読した内容について質問に答える	30秒	Please look at the passage. Why do Rugby players need to practice very hard?
3	2問	イラストを見て質問に答える	15秒	①Please look at the picture. What is the old man reading? ②Please look at the woman with glasses. What is she going to do?
4	2問	自身について質問に答える	15秒	①What do you like to do on weekends? ②Have you ever been to a foreign country? Yes. → Please tell me more. No. → What country would you like to visit?

CBT の一例として英検を挙げたが、英検の問題形式を含め、CBT でよくみられる問題形式について、以下に整理する。

### 2.2.1 音読・復唱

CBT には、表示された文字を読み上げる音読や、聞こえてきた音声を繰り返す復唱のタスクを取り入れているものが多い（英検、TOEIC Speaking、Versant、HSKK ネット試験等）。音読や復唱の形式は、実生活のコミュニケーション活動とは異なるが、文章を文字または音声で提示するだけで済むため、複雑なシステムは必要なく、CBT において出題しやすい形式だといえる。また、正答が決まっているため、採点もしやすいと考えられる。

### 2.2.2 描写・説明

写真や絵を見て、その内容を描写したり、説明したりするタスクも CBT には多い。英検の例（表1）にも、イラストを見て質問に答えるものがあるが、たとえば、TOEIC Speaking（以下、TOEIC）では、市場やスーパーなどの売り場の写真を見て、様子をできるだけ詳しく説明するというタスクがある。また、連続した4つの絵を見て、そのストーリーを語るもの（SJPT）もある。

図1は、TEAP CBT（以下、TEAP）の例である。絵に描かれていることを描写するという点では上述のテストと同様であるが、TEAP では、ロールプレイのように場面や状況を与え

(大学のキャンパスで鞆をなくしたため学生センターに来た)、相手と会話するような設定にすることで(学生センターのスタッフと話す)、実際のコミュニケーションに近いタスクとなっている。

また、「あなたは今、学校で新しい交換留学生に会いました。その生徒はあなたの友だちについて知りたがっています。あなたの友だちを1人以上説明しなさい。その友だちの名前、年齢、住んでいる場所、そして何をするのが好きか教えましょう」(STAMP)のように、イラストではなく、自分のことを説明するものもある。



図1 TEAP の例

### 2.2.3 意見を述べる

与えられたテーマについて、自分の意見を述べるタスクも多い(SJPT、TOEIC、TOEFL iBT、GTEC CBT、オランダ語国家試験等)。描写・説明のタスク同様、一方向の発話であり、一人で長く話すことも可能なため、CBTに適している問題形式だといえる。以下は、TOEFL iBT(以下、TOEFL)の例である。どちらの意見に賛成か、理由を挙げて意見を述べるのが求められており、高めのレベルの能力を測ることも可能である。

Some people enjoy taking risks and trying new things. Others are not adventurous; they are cautious and prefer to avoid danger. Which behavior do you think is better? Explain why.

### 2.2.4 質問に答える

自分自身のことについて質問されて答えるという問題形式は、多くのテストでレベルを問わず取り入れられている。「暇なとき、何をしますか」(SJPT)、「What did you do last weekend, and where did you do it?」(TOEIC)というように身近な話題について質問される。質問は通常、音声で提示されるが、インタビュー形式の場合、質問者である面接官が画面上に表示されることもある。たとえば、英検や市民統合テストでは、実際の面接官の動画が再生され、面接官から直接質問される。また、OPIc-Jでは、アニメーションで動く面接官のアバター(Mika)が常時画面に表示され(図2)、「Mika」の質問に答える形式でテストが進められる。

受験者が質問に答えるタスクに加え、OPIc-Jには、「私が住んでいるところについて、あなたがすべてを知ることができるように、3つか4つ質問をしてください」といった、受験者から面接官に質問をするタスクもある。アバターを表示させることで、仮想とはいえ、相手がいることを示し、受験者から質問をするという形式がCBTでも可能になっているといえよう。

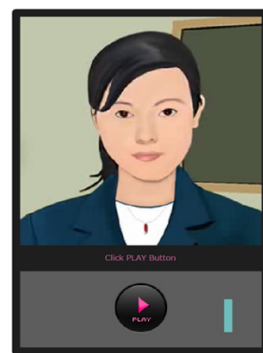


図2 OPIc-Jのアバター

しかし、CBTでは、画面上の「相手」が受験者の質問に答えることはない。実生活のコミュニケーションでは、相手の発話に対して何らかの反応や、さらなる質問を投げかけることによって会話が進行するが、CBTでは、受験者の発話に反応しながらやりとりを続けることは、現状ではシステム上不可能である。しかしながら、2往復以上の会話に近い形を実現したCBTもある（英検、GTEC CBT、AP Japanese等）。英検は、上述のとおり、Yes/Noをボタンで選択することで、次の質問が再生され、2往復のやりとりに近い形を実現している。また、GTEC CBT（以下、GTEC）では、以下のような問いかけが、回答時間を設けた後で続けて再生される。

Today is your first day at a new English school in England. Listen to the teacher,  
and answer his questions.

What's your name?

Where are you from?

What is your favorite city?

Tell me about the city.

Where do people like to travel in your country?

When is your best time to visit in your country?

2つめ以降の質問は、前の質問の回答を踏まえたものではなく、書き起こしてみると詰問のようにも見えるが、トピックは一貫しており、実際に答えてみると、6往復の会話をしているような形になる。

### 2.2.5 多技能統合

画面上の資料を読んでからそれについて述べたり、説明を聞いてから要約をしたりするような、他の技能とスピーキングを合わせたタスクを出題しているテストもある（ONiT、TOEIC、TOEFL、GTEC、TEAP等）。たとえば、ONiTには、メールを読んでから内容を取引先に確認する問題や、取引先から受けた電話の内容を上司に伝言するといった問題がある。これらの

テストは、アカデミック場面や、ビジネス場面での言語運用力を測ることが目的であるため、現実の場面に基づく多技能統合型のタスクが積極的に取り入れられている。

## 2.3 評価・採点

採点方法には、採点者が採点しているもの（以下、人採点）と、音声認識や AI の技術によって自動的に採点されるもの（以下、自動採点）がある。それぞれについて分けて記述する。

### 2.3.1 人採点

現状では、ほとんどの CBT において、人採点が行われている。人採点には、採点者の主観が影響する（日本語テスト学会 2016：207）といった課題も挙げられているが、各テストで、採点者の質を担保するための方策がとられているようである。たとえば、「全回答の20%を第2評価者が評価」（STAMP）、「2名で採点し、採点結果が異なる場合は上位採点者が確定」（GTEC）といった情報が公開されている。また、採点者の資格について、OPIc-J や SJPT は、ACTFL 公認の OPI テスターが採点している。その他のテストでも、「訓練を積んだ評価者」「認定された評価者」といった記述が目立ち、採点者の資格等を公開することで、採点の信頼性を主張している。

評価の観点には、「発音」「流暢さ」「語彙」「文法」といったスピーキングを支えるための言語能力的な項目を含めているのが一般的である。また、課題が達成できたかどうかを観点に含めているテストも多く（AP Japanese、JLCAT、JSST、ONiT、GTEC、市民統合テスト等）、実際のコミュニケーションに近いタスクを出題し、「どの程度できたか」が評価されている。

先述した「JFS 準拠ロールプレイテスト」も、課題が達成できたかどうかを中心に評価している。このテストは、「JF 日本語教育スタンダード」に準拠したものであるが、「JF 日本語教育スタンダード」では、どのような文法を知っているか、単語をいくつ知っているかではなく、「日本語を使って何ができるか」という「課題遂行能力」をレベルの指標にしている（国際交流基金 2017）。コミュニケーションを重視するテストであれば、課題が達成できたかという観点も重要であろう。

### 2.3.2 自動採点

現状では、Versant と TOEFL が自動採点を行っている。TOEFL は人採点と自動採点を併用しているため、完全に自動採点のみで判定をしているのは Versant だけである<sup>(3)</sup>。人採点は、採点にかかる時間やコストが大きいのが、自動採点によってそれを軽減することが可能である。しかし、ほとんどのテストが人採点を行っているという現状から、自動採点を採用することが容易でないことが推測できる。

Versant のテスト構成を見ると、他のテストと比べて明らかに異なり、その問題形式や問題数が自動採点のために設計されていることがわかる（表2）。

表2 Versant のテスト構成

大問	問題数	内容	出題例
1	8問	音読	This weekend is Ann's birthday.
2	16問	復唱	Leave town on the next train.
3	24問	質問	Would you get water from a bottle or a newspaper? (回答例： From a bottle.)
4	10問	文の構築	(音声) was reading, my mother, her favorite magazine (回答例) My mother was reading her favorite magazine.
5	3問	話の要約	3つの短いストーリーが読まれます。ストーリーは一度ずつ読まれます。ピーッという音の後に話の要約を30秒以内に英語で言ってください。状況、登場人物、行動、結末など詳しく語るようにしましょう。
6	2問	自由回答	家庭生活や好みに関する2つの質問が読まれます。質問は2回ずつ読まれます。ピーッという音の後に40秒以内で質問に答えてください。

2.1でも述べたが、Versant の問題数は60問以上あり、調査対象とした CBT の中では最多である。自動採点で信頼し得る判定を出すにはある程度の問題数が必要であることがわかる。

自動採点は、コンピューターが受験者の回答をネイティブスピーカーモデルと比較し、どの程度一致しているかを分析するという仕組みである。大問1～4は正答が決まっており、また、大問5についても、要約に使われる語句や表現がある程度特定されるため、ネイティブスピーカーモデルとの比較が可能になる。大問6は、Versant の中で唯一、受験者が自身のことを自由に回答する問題であるが、採点対象外とのことである<sup>(4)</sup>。つまり、現状では、自動採点は自由回答の形式には対応していない。

Versant の評価の観点は、「話し方（発音、流暢さ）」と「内容（語彙、文章構文）」である。「話し方」については、音声を分解し、答え始めるまでの時間、回答の長さ、ためらい、休止、詰まり、音声のピッチとトーン、音声のアクセントとイントネーションを分析することで測定される。「内容」は、音声を単語、句、文章に変換し、ネイティブスピーカーモデルと比較してどの程度一致しているかを分析することで測定される。これらの観点は、人採点のテストにも含まれていたが、自動採点は、課題が達成できたかどうかを測ることができないという点で人採点と異なる。自身のことについて自由に話すタスクや、実際のコミュニケーションに近いタスクの採点も不可能である。また、日本言語テスト学会（2016：110）は、「Versant は厳密な意味ではスピーキングテストではなく、『聞かれたことを素早く理解し、返答する能力』を測定するテストである」と述べている。自動採点でコミュニケーションに必要な能力の一部を

測ることはできても、コミュニケーションの力を総合的に測ることは現時点では難しい。

### 3. まとめ

今回の調査によって、CBTの現状について以下のようなことがわかった。

まず、問題形式については、音読・復唱、絵や写真を見て描写・説明をするもの、自身の意見を述べるもの、質問に答えるもの等がよく使われている。ビジネスやアカデミック場面での能力を測るテストでは、読んで話す、聞いて話すといった多技能統合型のタスクもよく用いられている。音読や復唱は、発音などの言語能力を測ることに留まるが、それ以外の問題形式では、場面を設定したり、自身のことを話させたりすることで、実際のコミュニケーションに近いタスクを出題することができる。CBTでは、双方向のやりとりをすることが困難であるという問題点もあるが、工夫をすれば、実際のやりとりに近づけることは可能である。

ただし、上述のような実際のコミュニケーションに近いタスクは、人による採点のみ可能である。自動採点のテストは問題形式の制限があり、音読・復唱や、選択肢のある問題等、正答が決まっているような問題形式は可能であるが、実際のコミュニケーションに近いタスクや、自身のことを自由に話すようなタスクの出題はできない。コミュニケーションに必要な言語能力を部分的に測るのであれば、自動採点の方法が可能であるが、コミュニケーションができるかどうか、つまり、課題遂行能力を測るのであれば、現状では人採点の方法をとることになる。人採点には採点者の主観が入るという問題点もあるが、採点者資格や訓練により、採点の信頼性を高めることも可能であろう。

いくつかの課題はあるものの、CBTの可能性は大きい。CBTによって大規模の日本語スピーキングテストが実現することは、日本語教育にとっても大きな発展である。今後、コミュニケーションを的確かつ効率的に測ることのできるCBT方式のスピーキングテストの開発を期待したい。

#### 〔注〕

- <sup>(1)</sup> スピーキングテストは、複数の民間試験を利用した導入が計画されていたが、地域・経済格差や各民間試験の測る目的が異なる等の指摘を受け、2019年11月に導入延期が発表された。
- <sup>(2)</sup> SJ-CATは現在、運用を停止しているが、調査時は運用していたためリストに含めている。JSSTは電話で受験する方式でありCBTとは異なるが、音声による自動出題形式であるため、調査対象とした。表の「CEFR対応」は、CEFRに準拠しているもの、または、CEFRレベルとの対照表を公開しているものを○とした。
- <sup>(3)</sup> TOEFLは公式Webサイトに採点者による採点と自動採点を併用しているとの記載があるが、詳細は不明。SJ-CATは自動採点を行っていたが、現在は運用を停止している。英検は2019年度から導入を予定していたが ([https://www.eiken.or.jp/eiken/info/2018/pdf/20181017\\_pressrelease\\_aisaiten.pdf](https://www.eiken.or.jp/eiken/info/2018/pdf/20181017_pressrelease_aisaiten.pdf))、その後の詳細は不明。
- <sup>(4)</sup> VersantのValidation Reportには、“This task is used to collect longer spontaneous speech samples.



## CBT 方式によるスピーキングテストの現状

Candidates' responses to items in this section are not scored, but are available for review by authorized listeners.” (6-7) と書かれている (<https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/Versant/ValidationSummary/Versant-English-Test-Description-Validation-Report.pdf>)。

### 〔参考文献〕

- 小泉利恵 (2015) 「スピーキングの評価—スピーキングテスト作成・実施を中心に—」 望月昭彦・深澤真・印南洋・小泉利恵 (編著) 『英語4技能評価の理論と実践—CAN-DO・観点別評価から技能統合的活動の評価まで—』、43-57、大修館書店
- 国際交流基金 (2014) 『JF 日本語教育スタンダード準拠 ロールプレイテスト テスター用マニュアル』、国際交流基金
- 国際交流基金 (2017) 『JF 日本語教育スタンダード【新版】利用者のためのガイドブック』、国際交流基金
- 日本語テスト学会 (2016) 『日本語テスト学会誌』 19 (20周年記念特別号)
- 日本語能力試験企画小委員会 口頭能力試験調査部会 (2003) 『口頭能力試験科目の創設に向けて』、国際交流基金
- 羽藤由美・神澤克徳 (2015) 「CBT 英語スピーキングテストの開発と実施—入試への導入にむけた試みの検証—」 『京都工芸繊維大学情報科学センター広報誌』 34、30-48
- 早瀬博範・林裕子・江口誠 (2018) 「四技能を問う英語 CBT 入試開発に向けた取組み」 『LET Kyushu-Okinawa BULLETIN』 18、15-29
- 文部科学省 (2018a) 「高等学校学習指導要領の改訂のポイント」  
<[https://www.mext.go.jp/content/1421692\\_2.pdf](https://www.mext.go.jp/content/1421692_2.pdf)> (2020年8月25日)
- 文部科学省 (2018b) 「【外国語編 英語編】高等学校学習指導要領 (平成30年告示) 解説」  
<[https://www.mext.go.jp/content/1407073\\_09\\_1\\_2.pdf](https://www.mext.go.jp/content/1407073_09_1_2.pdf)> (2020年8月25日)

資料1 調査したスピーキングテストの概要

テストの名称	対象言語	目的	問題数・時間	構成	評価の観点	採点	CEFR 対応
AP Japanese	日本語	・米国における大学入試、単位認定等 ・米国の中高生対象 ・大学レベルの能力を有していることを証明する	5問 10分	1. 会話(4問、回答20秒) 2. プレゼン(1問、回答2分、準備4分)	・Task Completion ・Delivery ・Language Use	・大学教員、APプログラム教師が採点 ・結果は①5段階のAPスコア、②APスコアに基づく5段階の推薦(Extremely well qualified～No recommendation)、③5段階の大学の評価(A～N/A)	
Avant STAMP (4S)	日本語	米国の学校教育における言語能力の評価	3問 20～25分	産出(与えられた状況で説明する)	ACTFLガイドラインに準拠したSTAMPルーブリック ・Text-Type ・Functions/Complexity ・Vocabulary ・Accuracy/Comprehensibility	・認定された評価者がルーブリックで採点 ・全回答の20%を算し評価者が評価し、評価者の正確さと信頼性を確認 ・結果はSTAMP Level1～8(Novice-Low～Advanced-Mid/High)	
JLCAAT (日本語コミュニケーション能力測定試験)	日本語	日本語でのコミュニケーション能力を客観的に測定する	4問 20分	1. 発表(1問、回答2分、準備60秒) 2. 対話(2問、回答2分30秒) 3. 意見(1問、回答2分、準備2分)	問題ごとの観点(評価基準はCEFRに準ずる) 1. 構成、課題、音素、流暢さ、語彙、表現、文法 2. やりとり、音素、流暢さ、語彙、表現、文法 3. 構成、課題、音素、流暢さ、語彙、表現、文法	・熟練した日本語教員が複数で採点 ・結果は①技能別の得点(各100点満点)、②技能別CEFRレベル、③能力別CEFRレベル(表現能力)、④総合CEFRレベル	○
JSST	日本語	日本語の運用能力を測る	10問 14分	1. 質問に答える(7問、回答45秒) 2. 説明する(2問、回答50秒、準備15秒) 3. プレゼンする(1問、回答60秒、準備15秒)	・総合タスク(日本語を使って何ができるか) ・発話の形(使いこなせる日本語「形」) ・正確さ(文法、語彙、発音、流暢さ) ・内容(どのような内容について話せるか) ・待遇表現(敬語が適切に使えるか)	・訓練を積んだ評価者が3名で採点 ・5つの評価基準を総合的評価 ・結果は10段階(Level1～10)	
ONIT (口頭ビジネス日本語試験)	日本語	ビジネス場面に沿った実践的な口頭能力を測定する	12問 75分(注意点説明15分、テスト60分)	1. 状況説明(回答40秒) 2. 指示確認(回答40秒) 3. 伝言(回答60秒) 4. 状況連絡(回答40秒) 5. 確認(回答60秒) 6. 報告・相談(回答60秒) 7. 意見説明(回答60秒) 8. 報告(回答60秒) 9. 敬語運用(回答40秒) 10. 反対意見説明(回答90秒) 11. 解決方法提案(回答120秒) 12. 意見・根拠説明(回答120秒)	・課題達成+6つの観点による分析的評価 ・発音 ・文法 ・語彙 ・流暢さ ・構成 ・聞き手への配慮	・徹底したトレーニングを積んだ評価者が複数で採点 ・結果は①課題達成能力評価(0～300点、0～7レベル、問題ごとにS.A.B.C.Fの5段階)と、②各観点の分析的評価(1～4の4段階、問題ごとにA.B.C.Fの4段階)	
OPIc-J	日本語	日本語コミュニケーション能力を総合的に測定する	12～15問(受験者のレベルによって異なる) 60分(オリエンテーション20分、テスト40分)	＜オリエンテーション＞ 1. 出題のための事前アンケート 2. 難易度決定のための自己評価 ＜テスト＞ 1. レベル、関心分野に基づいた質問に答える 2. ロールプレイ	・コミュニケーション総論能力(一貫して、安定して、継続して、臨機応変に対応できる言語的タスク遂行能力) ・文章構成力(文章の長さ、構成能力) ・状況に応じた表現力(テーマと状況に対する表現能力) ・質問意図の把握能力(質問意図の把握) ・文法、語彙、流暢さ、発音	・ACTFL公認のOPIc RaterがACTFLガイドラインに基づいて測定 ・5つの観点を総合的評価 ・結果はOPIc level1～7 (Novice LOW～Advanced LOW)	○

CBT 方式によるスピーキングテストの現状

SJ-CAT	日本語	日本語の総合的なスピーキング能力を測定する	4問(受験者の回答の正誤によって異なる) 15~30分	1. 文読み上げ(10~15秒) 2. 選択肢読み上げ(10~15秒、準備5秒) 3. 文生成(回答10秒、準備5秒) 4. 自由発話(回答30秒、準備5秒)	問題ごとの観点 1. 発音・イントネーションが自然か 2. 提示された場面の状況を理解して、正しい選択肢を選ぶことができたか 3. 正しい文を発話しているか 4. 流暢さ、正確さ、内容、表現力	・自動採点 ・100点満点	
SJPT	日本語	・日本人との自由な意思疎通能力を測る ・韓国人対象	26問 50分(オリエンテーション20分、テスト30分)	1. 自分について答える(4問、回答10秒、準備3秒) 2. 絵を見て簡単に答える(4問、回答6秒、準備3秒) 3. 日常的な場面で敏速に答える(5問、回答15秒、準備2秒) 4. 身近な質問に答える(5問、回答25秒、準備15秒) 5. 意見を述べる(4問、回答50秒、準備30秒) 6. 与えられた状況で話す(3問、回答40秒、準備30秒)場面設定 7. ストーリーを話す(1問、回答90秒、準備30秒) 8. 最後に一言(自由発話)(30秒)	・文法 ・語彙 ・発音 ・流暢さ	・日本語ネイティブのOPIテスター資格保持者 ・2~3名が採点 ・4つの観点を分析的評価 ・3段階で採点 ①Levelの付与 ②4つの観点を評価 ③Quality Control ・結果は①10段階のレベル(Level1~10)、②観点別の評価	
英検CBT 英検S-CBT (3級)	英語	(使える英語)の力を測る	6問 15分	(Warm-up)簡単な質問に答える 1. 音読(1問) 2. 音読した内容について答える(1問) 3. イラストについての質問に答える(2問) 4. 自分について答える(2問)	・程度(身近な英語を理解し、使用する) ・審査領域(身近なことについてやりとりする) ・音読(個々の単語の発音や意味の区切りなどに注意して読む) ・Q&A(与えられた情報を理解し、適切な表現を使って答える、自分の考えを論理的に述べる) ・アライチョード(積極的にコミュニケーションを図ろうとする態度)	・トレーニングを受けた採点者が採点 ・英検CSEスコアに基づき合格を判定 ・結果は①合格、②英検CSEスコア、③合格ポジションを示す英検バンド、④正答数・得点、⑤学習アドバイス	○
GTEC CBT	英語	「アカデミックな英語力」を測定する	7問(小問含め全12問) 20分	1. 会話応答(6問、2分) 2. 情報伝達・照会(3問、回答60秒、準備3秒) 3. 意見展開(3問) ①回答2分、準備90秒 ②回答3分、準備3分 ③回答1分	問題ごとの評価基準 1. 質問の意図に沿って、適切な応答ができていますか 2. 内容に基づいて、相手に伝わるような描写説明、問いかけができていますか 3. ①自分の意見が述べられているか、②自分の意見をサポートする理由が添えられているか、③流暢さ	・海外在住の英語話者が2名で採点 ・採点者が異なる場合は上位採点者が確定 ・結果は①技能別GTECスコア(0~350点)、②総合スコア(0~1400点)、③CEFRレベル、④技能別のCEFRレベルに基づくコメント、⑤4技能のスコアバランス	○
TOEFL iBT	英語	アカデミックな場面で必要とされる英語運用能力を測定する	4問 17分	1. 独立型問題(1問、回答45秒、準備15秒) 2. 統合型問題(3問、回答60秒、準備20/30秒)	・全体(General Description) ・話し方(Delivery) ・言語使用(Language Use) ・話の展開(Topic Development)	・米国ETSで訓練を受けた複数の採点者と、自動採点システムを併用して採点 ・結果は①総合スコア(0~120)、②4技能各スコア(0~30)、③5段階のレベル(基礎~上級)	○

TOEIC Speaking	英語	英語でコミュニケーションするために必要な能力を直接測る	11問 20分	1. 音読(2問、回答45秒、準備45秒) 2. 写真描写(1問、回答45秒、準備45秒) 3. 応答(3問、回答15秒または30秒) 4. 提示された情報に基づく応答(3問、回答15/30秒、準備45秒) 5. 解決策を提案する問題(1問、回答60秒、準備45秒) 6. 意見を述べる問題(1問、回答60秒、準備30秒)	問題ごとの評価基準 1. 発音、イントネーション、アクセント 2. 発音、イントネーション、アクセント、文法、語彙、一貫性 3~6. 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	・ETSの認定を受けた採点者が採点、採点結果はScoring Leaderとテスト開発者が全て確認 ・各観点を分析の評価 ・問題1~4は4段階(0~3)、問題5~6は6段階(0~5)で採点し、合計をスコアに変換 ・結果は①スコア(0~200)、②スコアに基づく能力レベル(1~8)、③発音、イントネーション、アクセント3段階(HIGH、MEDIUM、LOW)	○
TEAP GBT	英語	・実践的な英語運用力を測定する ・アカデミック	8問 30分	(Warm-up) 1. 質疑応答(3問、回答45秒) 2. 伝言、描写(2問、指示文15秒、準備45秒、回答60秒) 3. 矛盾点指摘(1問、指示文15秒、読解・準備120秒、回答90秒) 4. 要約、要旨(①1問、指示文15秒、読解30秒、聴解後準備30秒、回答60秒、②1問、準備30秒、回答60秒)	・発音 ・文法の範囲と正確さ ・語彙の範囲と正確さ ・流暢さ ・受け答え	・最低2名で採点、結果に大きな差があった場合は上位採点者が再採点 ・各受験者の回答を出題される問題単位で分割し、採点者に割り当て ・結果は①スピーキングスコア(0~200)、②CEFRのレベル、③Estimated TEAP Score、④各評価基準の評定とアドバイス、⑤TEAP GBT CSEスコア	○
Versant	英語	実践的な英語力を測定する	63問 20分	1. 音読(8問) 2. 音唱(16問) 3. 質問(24問) 4. 文の構築(10問) 5. 話の要約(3問) 6. 自由回答(2問)	問題ごとの評価基準 1. 発音、流暢さ 2. 発音、流暢さ、文章構文 3. 語彙 4. 流暢さ、文章構文 5. ~6. 発音、流暢さ、語彙、文章構文	・自動採点 ・内容(語彙、文章構文)と話し方(発音、流暢さ)を評価し、総合スコアを算出 ・サブスコアの比重: 語彙20%、文章構文30%、発音20%、流暢さ30% ・結果は、20~80点のサブスコアと総合点	○
HSKネット試験(初級)	中国語	基本的なコミュニケーションを行うことができる能力を測る	27問 17分	1. 音唱(15問、回答7秒) 2. 質問を聞いて答える(10問、回答10秒) 3. 質問を読んで答える(2問、回答90秒、準備7分)	非公開	・結果はスコア(100点満点、合格80点)	○
市民統合テスト	オランダ語	オランダへの移住者(移民)がオランダの言語と文化を理解しオランダ社会と統合して生活できること	24問 35分	1. 質問に答える(12問) 2. 聞いて選択式から選ぶ(12問)	・受験者の言うことが理解できる ・発話が出された質問に対する適切な反応である ・発話が「A2レベルの言語を使用」の定義を満たしている	・採点者がスピーキング能力評価モデルを用いて採点 ・総合的に評価 ・結果は合否	○
オランダ語国家試験(B1レベル)	オランダ語	日常的な会話ができる最低限レベルのオランダ語運用能力を認定する	16問 30分	1. 与えられた状況で話す(8問、回答20秒) 2. 意見を述べる(8問、回答30秒、準備15秒)	・何が重要であるかを明確に述べる ・論理性 ・単語と文の構造(文法) ・言葉の選択 ・接続語を使っているか ・発音 ・話すペース	・訓練を受けた評価者が2名で採点 ・結果は合否判定 ・技能毎に合否が表示され、4技能全て合格すれば認定される	○