

# Quality perspectives of the TOEIC® Program

# Preface

## はじめに

English-language competency for the workplace and communication in everyday life is central to the global competitiveness of organizations, and to the life prospects of individuals around the world. High quality, accurate, reliable, fair, and valid assessments of English proficiency are key to evaluating and improving such competencies.

Since 1979, the TOEIC Program tests have been trusted by businesses, governments, and educational institutions globally to evaluate real-life workplace and daily English-language communication skills. Millions of TOEIC tests have been administered annually by over 14,000 organizations across more than 160 countries.

As the creator of the TOEIC program, ETS has led the measurement field for more than 75 years in developing and administering high-quality assessments that are built on a foundation of research and evidence.

With the rapid expansion and integration of diverse technologies, including artificial intelligence-based methodologies, into assessment development and scoring, considerations regarding the quality, fairness, and impact of assessments for diverse test-taker populations are more critical now than ever. Optimizing the use of technology while simultaneously advancing these core assessment tenets for individuals from different cultural and language backgrounds has been a sustained focus of research in support of our TOEIC program assessments.

職場や日常生活のコミュニケーションにおける英語能力は、組織の国際競争力や世界中の人々の将来の展望において中核を成すものです。こうした英語能力を評価し、向上させるうえで要となるのが、信頼性・公平性・妥当性のある高品質かつ正確な英語能力テストです。

1979年以降、TOEIC Programは現実世界の職場や日常における英語コミュニケーション能力を評価するテストとして、世界中の企業、政府、教育機関からの信頼を得てきました。TOEIC Programのテストは、世界160カ国余りの14,000以上の企業・団体に活用されており、年間受験者数は何百万人にも及びます。

TOEIC Programの開発団体であるETSは、リサーチとエビデンスを土台として構築された高品質なテストの開発・実施を通じて、測定分野を75年余りにわたって率いてきました。

テスト開発や採点において、AIに基づく手法を含むさまざまな技術の拡張や統合が急速に進む今日では、多様な受験者母集団を念頭に、テストの品質・公平性・影響に配慮することの重要性がかつてないほどに高まっています。最適な形で技術を導入しながら、同時に、文化・言語的背景の異なる個人に配慮したテストの中核的な原則を発展させていくこと——これがTOEIC Programを支える研究の焦点であり続けてきました。

The *Quality Perspectives of the TOEIC Program* frames the foundations and dimensions of quality that are central to the appropriate use of these assessments and the interpretation of scores. It provides score users with essential explanations of the design and development of the TOEIC program as well as key updates on validity, fairness, reliability, washback, and other fundamental quality attributes of these assessments.

Finally, we would like to express our sincere gratitude to the Institute for International Business Communication (IIBC)—a long-term partner with ETS since the development stage of the TOEIC Program, and who administers the TOEIC Program tests and provides related services and research in Japan—for their great support of this booklet project.

本冊子 *Quality Perspectives of the TOEIC Program* は、テストの適切な使用とスコア解釈の中核となる「品質」について、その基盤や諸相を解説するものです。スコア利用者の皆様には、本冊子を通じて、TOEIC Program の設計や開発に関する最も重要な要素のほか、テストの妥当性、公平性、信頼性、波及効果などを含めた基礎的な品質特性に関する主要な最新情報について、ご理解を深めていただければ幸いです。

最後に、本冊子の企画制作にあたり、TOEIC Program 開発時からの長期にわたる ETS のパートナーであり、日本における TOEIC Program の実施運営、関連サービスや調査の提供を担う一般財団法人国際ビジネスコミュニケーション協会 (IIBC) の皆様から多大なるサポートを賜りました。この場を借りて、心より感謝の意をお伝えいたします。

Dr. Kadriye Ercikan  
Vice President  
Research and Measurement Sciences  
ETS

# Contents

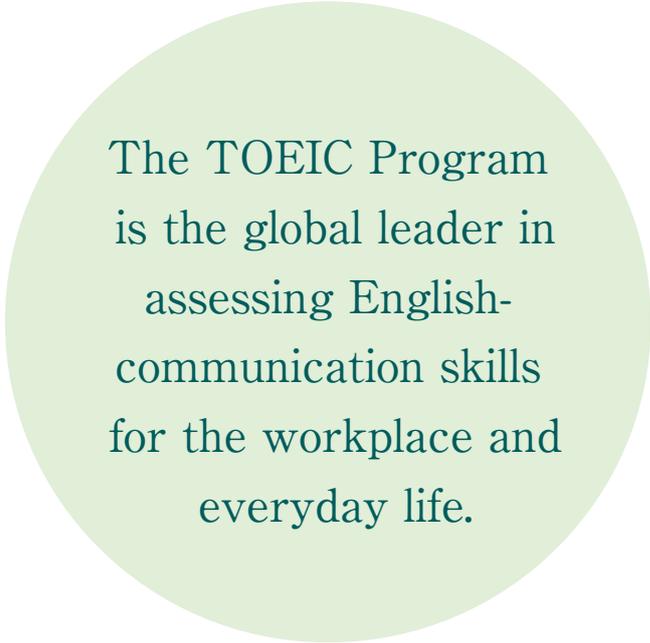
目次

Preface

Contents

About the Contributors: 編集主幹、著者、レビュアーについて

Section 1	TOEIC Programの開発ストーリー	07
Section 2	ETS——世界最大の非営利テスト開発機関	15
Section 3	ETSの品質への取り組み	23
Section 4	TOEIC Bridge TestsとTOEIC Tests	33
Section 5	TOEIC Programを構成する グローバルなコミュニケーションのための英語	45
Section 6	TOEIC Programによるポジティブな波及効果	53
Section 7	「4技能」を測るTOEIC Program	67



The TOEIC Program  
is the global leader in  
assessing English-  
communication skills  
for the workplace and  
everyday life.

# About the Contributors:

The editor-in-chief, authors, and reviewers

編集主幹、著者、レビュアーについて

編集主幹  
Editor-in-chief



Dr. John M. Norris  
主席研究員 (ETS Japan)

東アジア地域を中心に、言語の教授・学習・テストに関する研究およびアウトリーチ活動に従事。タスクに基づく言語教授、テストの妥当性検証、プログラム評価、リサーチシンセシス、メタアナリシスを専門とする。

著者  
Author



Dr. Jonathan Schmidgall  
上級研究員 (ETS)

言語テストの研究者として、TOEIC Programに関する研究の実施・コーディネートに携わる。研究成果を通じて、テストの設計、開発、スコアのマッピング、妥当性検証に寄与している。

著者  
Author



Dr. Saerhim Oh  
研究員 (ETS)

ETSにおいて、言語テスト研究および TOEIC Programに関する研究に従事。研究テーマは第二言語のライティングテスト、テクノロジーを活用した言語テスト、職場における言語能力のテストなど。

著者  
Author



Dr. Jaime Cid  
心理測定・データ分析ディレクター (ETS)

ETSにおいて、TOEIC Programの心理測定品質の監視に従事。データ分析・心理測定業務全般の総監督およびリーダーを務め、TOEIC Programにおける心理測定研究の発展に寄与している。

著者・編集補助  
Author and assistant editor



Reiko Komatsu  
調査研究員、  
アウトリーチスペシャリスト (IIBC)

言語教育・テスト分野の調査研究の分析のほか、調査報告レポートやアウトリーチ記事の執筆に携わる。英語学習者向けの教材開発にも従事。

レビュアー  
Reviewer



Dr. Richard J. Tannenbaum  
テスト開発担当アソシエイト・  
ヴァイス・プレジデント (ETS)

ETSにおいて TOEIC Program 等のテスト設計・開発を統括。妥当性や規準設定に関する幅広いリサーチ経験を有すると共に、スコアマッピングの権威でもある。

レビュアー  
Reviewer



Dr. Feng Yu  
アドバイザー

ETSにおいて19年間にわたり、TOEIC Programのテストおよび関連する教育・学習コースやコンサルティングサービスの企画、設計、開発、実施を率いた。



## Section 1

# TOEIC Programの 開発ストーリー

### 1979年の誕生から「今」、そして未来へ——

グローバル化が加速する1970年代後半、実践的な英語コミュニケーション能力を測るテストの必要性を痛感した3人の日本のビジネスリーダーたち。彼らは世界最大の非営利テスト開発機関であるETSに協力を求めます——これがTOEIC Program誕生への第一歩となりました。ここでは、「世界共通の英語能力テスト」を提供するTOEIC Programの開発ストーリーを紐解いていきます。

---

## TOEIC Program 誕生前夜

---

### 理念を胸に日本から 米国プリンストンへ

「日本のビジネスパーソンに、人間同士の関係を築ける英語コミュニケーション力を身に付けてほしい」——北岡靖男はこの理念を胸に、1977年の風の強い寒い日、米ニュージャージー州プリンストンをひとり訪れていました。世界最大の非営利テスト開発機関であるETSに、後のTOEIC Listening & Reading Test (以下、



北岡靖男  
(1928～1997年)

TOEIC L&R)の開発を依頼するためです。

北岡は1986年に国際ビジネスコミュニケーション協会——現在のIIBCの前身であり、当時は通商産業省(現在の経済産業

省)管轄の公益財団法人——を設立した人物です。北岡は24年間にわたり、世界的に知られる「Time」誌を発行するタイム社に所属し、同社ニューヨーク本社勤務やアジア総支配人を務めるなどの経験を積んでいました。

訪問を受けたETSは「北岡とは誰なのか?」という疑問を抱き、身元照会先のタイム社に問い合わせました。すると同社のプレジデントから「ああ、Kit(北岡の愛称)! ええ、もちろん知っていますよ。彼は元気ですか?」との答えが返ってきたことで、面会の実現に至りました。

北岡から依頼を受けたETSでは、第一世代のTOEFL® Testsをすでに開発済みでした。そのため、彼らは当初、「英語を母語としない人たちを対象とする

英語能力テストとしては、すでにTOEFL Testsがあるのだから、新たにテストを作る必要はない」という考えを持っていました。

ですが、北岡の理念が揺らぐことはありませんでした。そこには日本人として外資系企業に長年勤めながら、グローバル化の波にもまれる日本のビジネスパーソンたちを見つめてきた、彼ならではの思いや体験があったのです。

### 即戦力としての国際要員の 育成が急務に

当時の日本は高度経済成長期を経て、すでにGNPにおいて米国に次ぐ世界第2位にまで上り詰めた経済大国となっていました。一方で、米国を初めとする先進諸国との間に貿易摩擦が生じ、その回避策として現地生産・販売を軸にした海外直接投資型へのシフトが加速していました。こうした海外活動の構造変化によって、現地に派遣する社員の増員が不可避となった企業では、即戦力としての国際要員の育成が急務の課題となっていました。実際に、日本人と現地のビジネスパーソンの間で、言語や習慣の違いを原因とするコミュニケーション上のあつれきが生じる場面も、少なくありませんでした。

北岡はそうした状況をつぶさに見ながら、強い思いを抱いていました。

「海外の人たちとコミュニケーションを図ることなく、単にモノを売ったりお金を投資したりするだけでは、いずれ

日本企業は立ち行かなくなる。心の通ったコミュニケーションを通じて、お互いに信頼し思いやる人間関係を築いていかなければならない。

そのためには、知識としての英語ではなく、実際にコミュニケーションを図るためのスキルとしての英語能力を身に付ける必要がある。それも、海外部門など特定の人だけでなく、多くの日本人が英語コミュニケーションのスキルを磨かなければならない時代がやってくる」

そして、ETSとの交渉の場でこう訴えました。

「異なる言語や文化、伝統の中で生きてきた世界中の人たちの間でビジネスを成り立たせるためには、世界共通の言語として英語を身に付けなければならない。

そのために、ビジネスの現場に必要な英語能力を測る、世界共通のテストが求められている。TOEFL Testsで培ってきたノウハウを、ビジネス

パーソンを対象にしたテスト開発に生かせれば、ETSは世界に対して多大な貢献を果たすことができる」

単にモノを売ったりお金を投資したりするだけでは、  
いずれ日本企業は立ち行かなくなる。  
心の通ったコミュニケーションを通じて、お互いに信頼し  
思いやる人間関係を築いていかなければならない。

この北岡の情熱と揺るぎない理念に、ETSのスタッフたちも深く心を動かされました。ですが、非営利団体であるETSはこれまで政府、教育機関や各種財団からの依頼のみを受け付けており、北岡のような個人からの依頼に対応することが困難な状況でした。

### 事態を切り開いた 日本のキーパーソン2人

この事態を切り開いた重要人物たちがいました。そのひとりとして中核的な役割を果たしたのが、旧通産省出身の渡辺弥栄司です。渡辺は官房長や通商局長などを歴任後、実業家に転身し、日中国交正常化に注力。1972年に日中経済協会を設立し、理事に就任しました。かねて北岡と親交のあった渡辺もまた「これからの日本人は、相手が話し

#### 日米貿易摩擦の主な経緯

1950年代 } 1970年代	貿易摩擦が生じた 輸出製品 ● 繊維 ● 鉄鋼	1972年 …………… 沖縄返還・日米繊維協定 → 日本、繊維製品の対米輸出を自主規制
1980年代 } 1990年代	● カラーテレビ ● 自動車 ● 半導体 ● コンピュータ、etc.	1981年 …………… 日本、自動車の対米輸出の自主規制を開始 1985年 …………… プラザ合意(円高ドル安へ) → 日本からの輸出が減少 1989～1990年 …… 日米構造協議 1993～1994年 …… 日米包括経済協議 1993年 …………… 日本車の米国現地生産台数が対米輸出台数を上回る

ていることを理解したうえで、自分の意見を明確に伝えるといった、生きた英語力を身に着ける必要がある」という考えを持ち、北岡と理念を共にしていました。



渡辺弥栄司  
(1917～2011年)

北岡から、ETSに開発依頼をするためにはテストを実施・運営する組織が必要との相談を受けた渡辺は、このテストの開発は「貿易促進に必要な人材育成」に寄与するものであるため、通商産業省が扱うべきと考え、同省に対し、「一分一秒を争う重要な案件であるから、すみやかに認可してほしい」と働き掛けました。同省も渡辺のミッションに賛同しました。この新しいテストに関わる事業は、日本人や日本企業の国際化に大きく貢献するものであることを認めたのです。

ここに至るまでには、もうひとりのキーパーソンがいました。

た。三枝幸夫です。三枝は北岡のタイム社時代の同僚でした。北岡の構想を聞いた三枝は、「日本の英語教育を変えるなら、ビジネスパーソンの英語コミュニケーション能力を

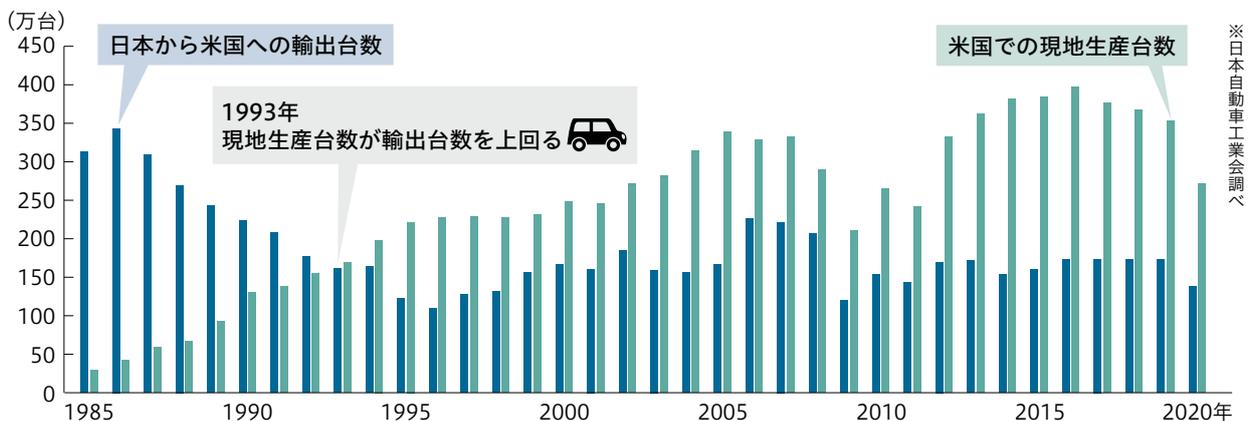


三枝幸夫  
(1931～2005年)

測るモノサシを開発する必要があり、それにはTOEFL Testsを開発しているETSに依頼したらよいのではないかと助言したのです。この助言がなければ、北岡と渡辺がETSと出会うことはなかったかもしれません。後に三枝は早稲田大学教授として言語学の指導に当たりながら、TOEIC Programを基準とした英語能力に関するさまざまな調査・研究に尽力しました。

こうして北岡、渡辺、三枝という3人のキーパーソンの理念と情熱がETSを動かし、1977年、TOEIC L&Rの開発が正式にスタートしました。

### 日本の自動車メーカーによる対米輸出と米国現地生産の推移



※日本自動車工業会調べ

## TOEIC Programの誕生、そして未来へ

### テスト設計の基本コンセプト

ETSに開発を依頼したTOEIC L&Rの基本コンセプトは「英語によるコミュニケーション能力を正確に測定し、目的や目標に照らしてどのレベルに位置するか、それぞれに基準を与えることのできるテストシステム」というものでした（後に発信力を直接測るTOEIC Speaking & Writing Testsも開発されました）。

そして、次の点も重要な要素とされました。

- 初級から上級までの幅広い能力範囲を測定できる
- 特定の国や文化の英語使用に偏らず、国際共通語としての英語能力を測定する
- 出題内容はビジネスシーンを取り上げていても、必要とされる英語力があれば誰でも理解できる言葉で構成する

これらは開発当初から現在まで、TOEIC Program全体の設計思想の中核を成しており、このテストプログラムが「世界共通の英語能力テスト」と言われる所以でもあります。

こうした設計思想を具現化する第一歩として、ETSがテスト開発の着手時に取り組んだのが、フィールドスタディでした。TOEIC L&Rは、国際的な職場や日常生活で実際に必要とされる「国際共通語としての英語を

使ったコミュニケーション能力」の主要な要素を測るテストです。では、「国際共通語としての英語」とは実際にどのようなものなのか？ まずは、その定義を行うためのエビデンスを集める必要があったのです。

ETSは調査団を編成し、さまざまなグローバル企業で英語を使ってビジネスを行う人々を対象に、リスニングとリーディングのタスクや、それらに求められる英語知識に焦点を当てながら、ヒアリングや談話パターンの分析を実施しました。明らかになったのは、主に次のようなことでした。

- 英語圏以外の人たちが参加するミーティングでは、米国人同士の会話で話されるようなイディオムや口語表現は使用されない
- ビジネスシーンで用いられる文章は、学术论文のようなボリュームのあるものでなく、コンパクトにまとまっており、表現も簡潔で明瞭なものである
- プレゼンテーションの場面においては、図表が用いられることが多い

ETSはこれらを含むフィールドスタディから得た知見に基づき、現実のグローバル企業での英語使用の実態から抽出された特徴を組み込んだ言語モデルを「国際共通語としての英語」と定義づけ、テスト問題形式や内容の開発に反映していきました。そのため、テストの開

現実のグローバル企業での英語使用の実態から抽出された特徴を組み込んだ言語モデルを「国際共通語としての英語」と定義づけ、テスト問題形式や内容の開発に反映

発メンバーも、世界各国で英語を教えた経験のある専門家たちが選ばれました。

さらに、初級から上級までの幅広い能力範囲を高い精度で測定するためには、どうしても問題数が多くなり、テスト時間も長くなりがちでしたが、ETSの開発者たちは、テストニングに関する高度な知見、高い専門性と技術を最大限に活かし、テストの信頼性を担保しながら、現実的に運用可能なテストの仕様を設計していきました。また、彼らは幾度もパイロットテストを繰り返し、テスト問題や設計の品質を測り、テストに基づく解釈や意思決定における極めて高い妥当性を担保していきました。

## 「世界共通の英語能力テスト」として進化を続ける TOEIC Program

こうした努力の甲斐があり、1979年12月、TOEIC L&Rは第1回公開テストを迎えます。この第1回は日本全国5都市での実施（受験者数は2,773名）でしたが、1980年には6都市、1990年には20都市、2000年には57都市と、数十年の間に急速な増加の一途をたどり、2022年時点では約80都市で実施されています。1981年には、一般の個人受験者向けの公開テストに加えて、企業や大学などが団体単位で実施する、TOEIC L&R IPテストもスタートしました。

2001年からは、初・中級英語学習者のニーズに対応した TOEIC Bridge Listening & Reading 公開テスト／IPテストを開始。2006年12月からは、発信力の重要性を踏まえ、これまでの TOEIC L&Rに加えて、TOEIC Speaking & Writing Testsの提供が始まり、TOEIC Testsは英語4技能全てを測定できる、包括的なテストプログラムへと進化を遂げました。さらには、TOEIC Bridge Testsも2019年に4技能の測定が可

能なテストにリデザインされています。

そして2023年現在、TOEIC Programの各テスト—— TOEIC Testsおよび TOEIC Bridge Tests—— はアジア、北米、アフリカ、南米、ヨーロッパを含む世界160余りの国・地域において実施されており、利用団体・企業数は14,000を超えます。



TOEIC Programの誕生を報じる記事（朝日新聞、1979年10月4日）  
朝日新聞社に無断で転載することを禁じる 承諾番号：23-3006



第1回 TOEIC L&R公開テストのポスター

こうしたテストの発展に反映されるように、ETSでは、TOEIC Programの各テストが国際的な日常生活や職場における英語コミュニケーションのニーズに即した測定を確実に提供できるよう、常に TOEIC Programのモニタリングと評価を行っています。将来的に必要なに応じて各テストの基礎設計や内容のアップデートを実施する

際には、こうした品質を保証するための取り組みを通じて得られる知見を活用していきます。

当時ETSの交渉窓口を務め、TOEIC Programの初代担当ディレクターとしてTOEIC L&Rの開発を指揮したプロタース・ウッドフォード (Protase Woodford) は、後にこう振り返っています。



プロタース・ウッドフォード  
(Protase Woodford)

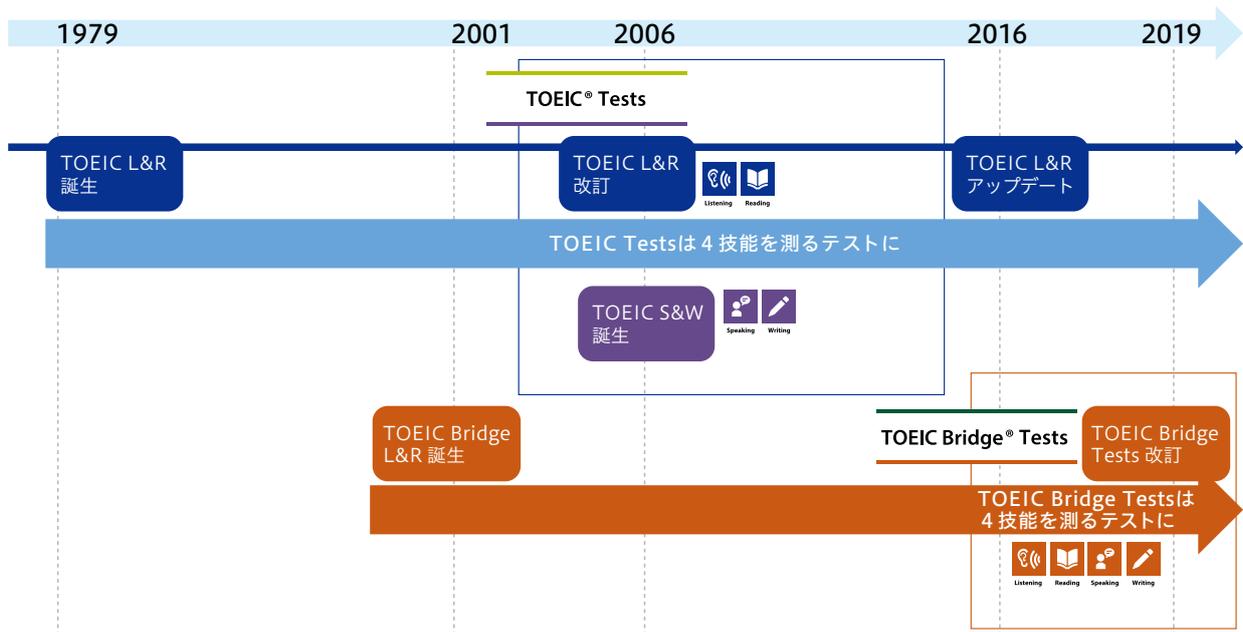
「北岡氏はエネルギーなビジネスパーソンであり、かつ忍耐力のある献身的な理想家でした。北岡氏は国際舞台においてビジネスを展開すること

は、異なる文化と伝統を結ぶ架け橋として役立つと考えていたようです。また、人々をより結びつけるために言葉が演じる中心的役割を十分理解していました。

現在、世界中の数百万という人々に利用されるまでになったTOEIC Programは、彼の夢と理想を実現したものです」

TOEIC Programは、北岡、渡辺、三枝という3人のヴィジナリーたちの理念が、ETSという世界最大のテスト機関と出合ったことで産声を上げました。そして、約40年以上を経た今日もなお、その理想を追求し続けています。

## TOEIC Programの歩み



※個別テスト名などに関しては、現在の表記を使用。IPテストは団体特別受験制度 (IP: Institutional Program) の略称。



## Section 2

# ETS——世界最大の 非営利テスト開発機関

Believing in the life-changing power  
of learning for over 70 years

TOEIC Programの開発を手掛けるETSは、公平で平等な教育テストを求める人々の声に応える形で誕生し、世界最大の非営利のテスト機関として、世界中の教育・学習をサポートしています。ここでは、ETSの創立ストーリーと共に、教育測定分野に大きな影響を与え続ける研究開発活動についてご紹介します。

---

## 1947年の創立以来、テスト分野の先駆けとして

---

### 世界の教育測定・テスト分野を牽引するETS

TOEIC Programを開発したETSは、米ニュージャージー州プリンストンに拠点を置く、世界最大の非営利テスト開発機関です。

創立は1947年。「教育の品質と平等の向上」をミッションに掲げ、世界中に数千名の専門家やスペシャリストを擁し、世界中の学習者や教育者、教育機関、企業、政府機関に調査研究や学習に関するソリューションを提供しています。

ETSが開発するテストはTOEIC ProgramやTOEFL Testsのほか、大学院や専門学校の入学共通試験のGRE® Test、教員資格試験のPraxis® Testsなど、その分野や目的は多岐にわたります。

また、ETSでは経験を積んだ教育者たちと密接に連携しながらテスト問題の執筆やレビューに取り組んでおり、それが適切かつ信用のおけるテストの提供につながっています。

現在、ETSが制作・実施・採点するテスト数は年間数千万本を数え、世界200ヵ国余りの9000を超える地域で受験されています。

このように、世界の教育測定・テスト分野の中心的な役割を担うETSですが、そのコミットメント——最先端の測定理論と実践に基づき高品質なテストを提供する——の根源となったものは、20世紀

初頭に浮上した教育の平等、公平性、機会に関する課題の数々でした。

### 教育の不平等と変革への第一歩

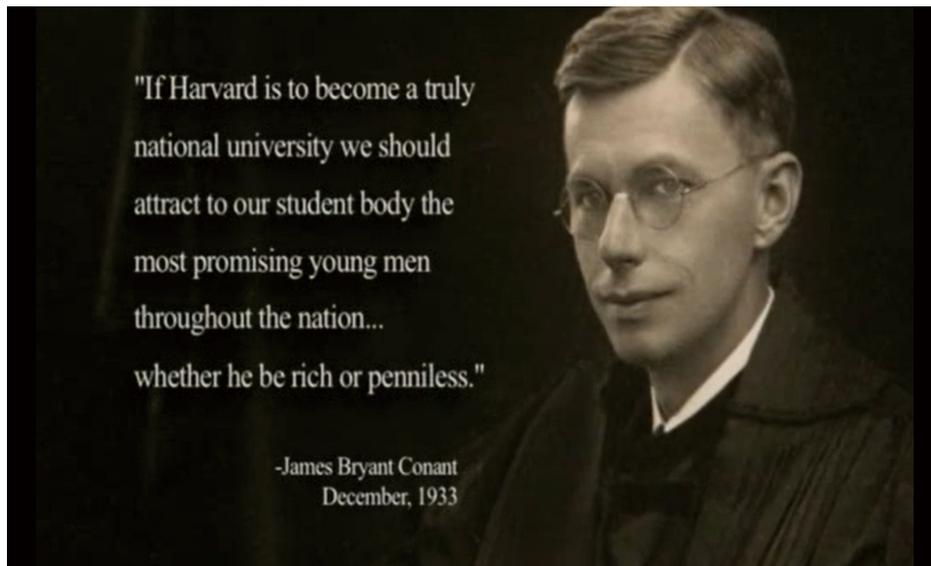
遡ること約100年前の1933年。アメリカは大恐慌の最中にありました。失業率は史上最高値を記録し、見通せない将来への不安が国全体を覆う中、東海岸の上流階級の暮らしぶりは以前と変わらず豊かで、最エリート層の子息の多くはハーバード大学へと進学していきました。

当時、ハーバード大学に門戸が開かれていたのは、最高位の私立学校に通った学生に限られており、そのほとんどが代々富を築いてきた家系の出身でした。

こうした国家の経済危機を目の当たりにし、変革の必要性を胸に立ち上がったのが、若くしてハーバード大学総長に就任したジェイムズ・ブライアント・コナント (James Bryant Conant) 教授でした。

彼は能力評価を主軸にした入学制度により、米国全土を代表する最も優秀な学生をハーバード大学に入学させることで、教育の不平等を是正したいと考えました (それでも当時は変革の対象に女性は含まれていませんでした)。

このコナントの構想は、上流階級からは必ずしも歓迎されたわけではありませんでした。



ジェイムズ・ブライアント・コナント(1893~1978年)

ですが、コナントはハーバード大学の入学制度を抜本的に改革するプロジェクトを立ち上げ、その遂行を託せる人物を探しました。そこで白羽の矢が立てられたのが、ハーバード大学を卒業し、同大学の進学準備校で教えていた青年でした。

コナントはその青年に、ハーバードに戻ってプロジェクトを推進するよう依頼したのです。この青年の名はヘンリー・チョンシー (Henry Chauncey)、後にETS初代プレジデントに就任する人物です。

「ハーバードが真に国を代表する大学となるには、  
米国全土から最も有望な若者を集めなければならぬ...  
富のあるなしにかかわらず」

ジェイムズ・ブライアント・コナント(1933年12月)



ヘンリー・チョンシー(1905~2002年)

## 高等教育への門戸が 米国全土に開放

コナントの命を受けたチョンシーは、黎明期にあった知能テスト分野で革新的な研究を進めていたカール・ブリガム (Carl Brigham) に面会します。

ブリガムはthe College Entrance Examination Board (大学入試委員会; 通称カレッジ・ボード) に所属していました。カレッジ・ボードは1900年に12の大学から成るコンソーシアムを主体として結成され、それまで各大学が個別に開発・実施していた大学入試を共通化することを目指していました。

当時ブリガムはプリンストン大学からの出向という形で、現在のSAT(全米大学入学共通試験)の原型の一部となるthe Scholastic Aptitude Testを開発していました。

1934年、チョンシーはカレッジ・ボードの協力を得て、ブリガムが開発したthe Scholastic Aptitude Testをハーバード大学の奨学金入学試験に導入します。そして、それからわずか数年後の1937年、14の大学が同テストを導入。米国全土の多様な学生に対し、その門戸を開きました。

その結果、これまで富と権力に独占されていた高等教育が広く開放されていき、大学教育を受けた人々が急速に増加し、とりわけ米国の科学研究者のコミュニティは急拡大を遂げます。

第二次世界大戦の終結までには、米国の科学研究者コミュニティは国家の軍事・経済の安全保障において極めて重要な役割を担うようになり、それが高等教育への比類ない規模の投資へとつながっていききました。

## 統一的なテスト機関として ETSが誕生

1946年、入試テスト開発への機運のさらなる高まりを背景に、コナントは統一的なテスト機関の設立を構想し、チョンシーにその実現化を依頼しました。

起業家として成功していたチョンシーは、米国の主要な3つのテスト機関の代表に声を掛けます。1つは、先述のカレッジ・ボードです。2つ目は、the American Council on Education (全米教育評議会) です。同評議会は教員資格試験のthe National Teacher Examinationを開発していました。そして3つ目はthe Carnegie Foundation for the Advancement of Teaching (カーネギー教育振興財団) です。同財団はGRE® Testの開発を手掛けていました。

これら3つのテスト機関はチョンシーの多大な尽力を背景に、長時間に及ぶ交渉を重ね、それぞれが保有するテストプログラムや、資産の一部を統合する方法について合意に至りました。

そして1947年12月19日、ついに3機関を統合した新たな非営利テスト機関として、ETSが誕生したのです。

初代プレジデントにはヘンリー・チョンシーが、そして会長にはジェイムズ・コナントが就任しました。

以来、ETSは今日に至るまで、教育の品質と平等の向上に尽力する科学者や専門家が世界中から集まる場で在り続け、米国のみならず世界中のテスト・教育分野において革新的なソリューションを提供しています。

心理測定・教育テスト分野の発展と向上を牽引する主要な先駆者として歩み続けたETSの70年余りにわたる軌跡は、同分野の黎明期からの歩みと重なります。

次項では、ETSの研究方針や歴代の研究者たちの主要な研究実績にスポットを当てます。



1947年12月にニューヨーク州大学評議会から授与された認可書。



ETSは創立当初、ニュージャージー州プリンストンのナッソー通りの一角で運営していた。

ETS 創立までの歩み		関連する世界の出来事
1934年	<ul style="list-style-type: none"> <li>●ハーバード大学がthe Scholastic Aptitude Test (現SATの原型の一部)を導入</li> </ul>	1929～1930年代 世界大恐慌
1937年	<ul style="list-style-type: none"> <li>●the Scholarship Test (一部にthe Scholastic Aptitude Test を含むテスト式)を14大学が導入</li> <li>●GRE® Test をハーバード大学、イエール大学、プリンストン大学、コロンビア大学が試験的に導入</li> </ul>	1939～1945年 第二次世界大戦
1946年	<ul style="list-style-type: none"> <li>●3つのテスト開発機関「カレッジ・ボード」「全米教育評議会」「カーネギー教育振興財団」が統合に向けて協議を開始</li> </ul>	1944年 米国で軍人社会復帰法が成立 →復員兵への学費援助制度(通称GIビル)により、大学入学者が急増
1947年	<ul style="list-style-type: none"> <li>●ETS 創立</li> </ul>	

Copyright © 2008 ETS. www.ets.org. Items from ETS 60<sup>th</sup> Anniversary video are reprinted by permission of ETS, the copyright owner. All other information contained within this publication is provided by The Institute for International Business Communication (IIBC) and no endorsement of any kind by ETS should be inferred.

---

## 「教育の品質と平等の向上」を追求し続ける

---

### 最も優秀な科学者を 惹きつける研究環境を実現

ETSの初代プレジデントを1970年まで務めたチョンシーは、その就任期間中に、研究活動 (conduct of research) の重要な様式を確立しました。

学術的な自由を重視した彼は、研究者それぞれが独立した考えを持ち、ETSが開発したテストに対して建設的な批判を行うよう奨励しました。そうした風土こそが、ETSによって開発されるテストの品質向上を促すと考えていたのです。

これに加え、研究結果については、企業秘密や機密情報を含まない限りは全て公開するという方針を打ち出します。たとえETSのテストにとって好ましくない結果であっても、それらを隠蔽すれば、ETSが自らの目的のために供する「プロパガンダ」の生産組織になり下がってしまうとの理由からです。また、この公開方針の背景には、教育および測定分野の向上をETSの組織としての義務と捉えたチョンシーの思想もありました。

チョンシーはさらに、テストプログラムの開発資金とは別に、短期的なビジネスニーズに左右されない、長期的な未来を見据えた研究プロジェクトのための資金源を確保することに尽力しました。これにより、テストプログラムの新規開発と、広範なテーマにわたる革新的な調査研究を同時に推進していったのです。

こうした方針に基づく研究活動によって、ETSは最も優秀な科学者たちを惹きつけ続ける研究環

境を築き上げることに成功します。その結果、数々の革新的な研究成果が生まれ、ETSは教育測定という科学分野を牽引する世界的リーダーとしての地位を確立していきました。

### 教育測定分野の基礎に 多大な影響を与えたETS研究員たち

ETSを舞台に活躍した数百名の科学者、心理測定学者、テスト開発者がもたらした目覚ましい研究成果や革新は、教育測定分野の基礎に大きな影響を及ぼすと共に、後に続くETSの研究者や、同分野に携わる世界中の人々にインスピレーションを与えてきました。

主要な功績者の一部を挙げるだけでも、フレデリック・ロード、サミュエル・メシク、マイケル・ケイン、レッドヤード・タッカー、ハロルド・ガリクセン、ノーマン・フレデリクセン、ロバート・ミスレヴィといった、心理測定、統計学、テスト設計および妥当性研究に著しい影響を与え、同分野の歴史に名を刻んだ人物たちが並びます。潜在特性理論、古典的テスト理論、項目応答理論 (IRT)、アイテム分析、妥当性理論と構成概念妥当性、自由解答テスト、エビデンス・センタード・デザイン——これらはETSの功績の中でも、彼らの手によって開発された、あるいは前進がもたらされた主要な理論・実践として挙げられるものです。

その他多くのETS研究員の功績に関する詳細に

については、書籍 *Advancing human assessment: The methodological, psychological and policy contributions of ETS* (Bennett et al., 2017) をご参照ください。

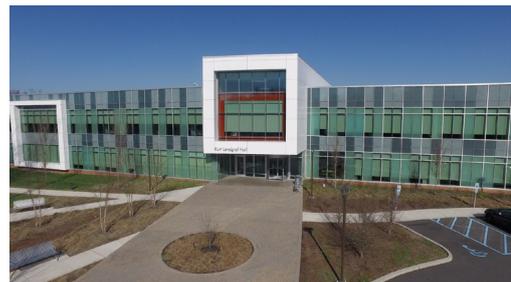
### 広範な分野の専門家集団が 共通のミッションを追求

ETSではこうした重要かつ革新的な功績を称する目的で、リサーチ・チェア制度を長年にわたり設けてきました。リサーチ・チェアには各分野をリードする科学者が一定の任期で就任し、ETSの研究開発活動に対する助言や科学的革新への協力を行います。

2020年時点でのETSリサーチ・チェアは以下の通りです——ロバート・ミスレヴィイ (測定・統計学の Frederic M. Lord チェアに就任)、マイケル・ネトルズ (政策評価・リサーチの Edmund W. Gordon チェアに就任)、マイケル・ケイン (テストの妥当性の Samuel J. Messick チェアに就任)、ランディ・ベネット (テスト

における革新の Norman O. Frederiksen チェアに就任)、アーウィン・カーシュ (大規模テストの Ralph W. Tyler チェアに就任)。

こうした高名なリサーチ・チェアのほかに、ETSでは数百名ものリサーチャー、心理測定学者、テスト開発者、そしてその他多くの広範な分野——応用言語学、教育、心理学、統計学、心理測定学、コンピューターサイエンス、社会学、人文科学——における専門スタッフが活躍しており、そのひとり一人全員が「教育の品質と平等の向上」という共通のミッションを追求し、革新的な調査研究や開発に励んでいます。



スタッフが勤務する現在の ETS のビルのひとつ。



現在の ETS の敷地。大学を思わせる様子から「ETS キャンパス」と呼ばれている。

## Reference

Bennett, R. E., & von Davier, M. (2017). *Advancing human assessment: The methodological, psychological and policy contributions of ETS* (p. 711). Springer Nature.



## Section 3

# ETSの品質への取り組み

妥当性・公平性・信頼性ある  
テスト、リサーチ、関連サービスを届けるために

ETSはミッションドリブンな組織として、商品とサービスにおいて最高水準の品質を実現することにコミットしています。ETSのリサーチャー、心理測定学者、テスト開発者は、テストの妥当性・公平性・信頼性のスタンダードを確立するための取り組みにおいて、先導的な役割を果たしてきました。ETSスタンダードに対するこうした倫理的なコミットメントによって、ETSのテストングおよびリサーチ活動の品質が担保されています。

---

# 『品質および公平性に関するETSスタンダード』

## —— エクセレンスの指標として

---

### 団体方針として スタンダードを設定

ETSは非営利団体として、以下のミッションとゴールの達成に向けて活動しています。

ETSのミッションは、公平性・妥当性あるテスト、リサーチ、関連サービスの提供を通じて、教育の品質と平等の向上に貢献することです。私たちは商品およびサービスを通じて、世界中の全ての人々の知識・技能を測り、学習・パフォーマンスを促進し、教育や職業訓練をサポートします。

この「教育の品質と平等の向上」というミッションを達成するために、1981年、ETS評議会は団体方針として、『品質および公平性に関するETSスタンダード』（*The ETS Standards for Quality and Fairness*）を初採択しました。以来、『ETSスタンダード』は定期的に見直され、教育測定の発展に足並みを揃える形でアップデートを重ねています。

『ETSスタンダード』は品質のベンチマークとして中核的な役割を担っており、テスト、リサーチ、関連商品およびサービスの設計・開発・提供に携わるETSの全てのスタッフによって指針として用いられています。

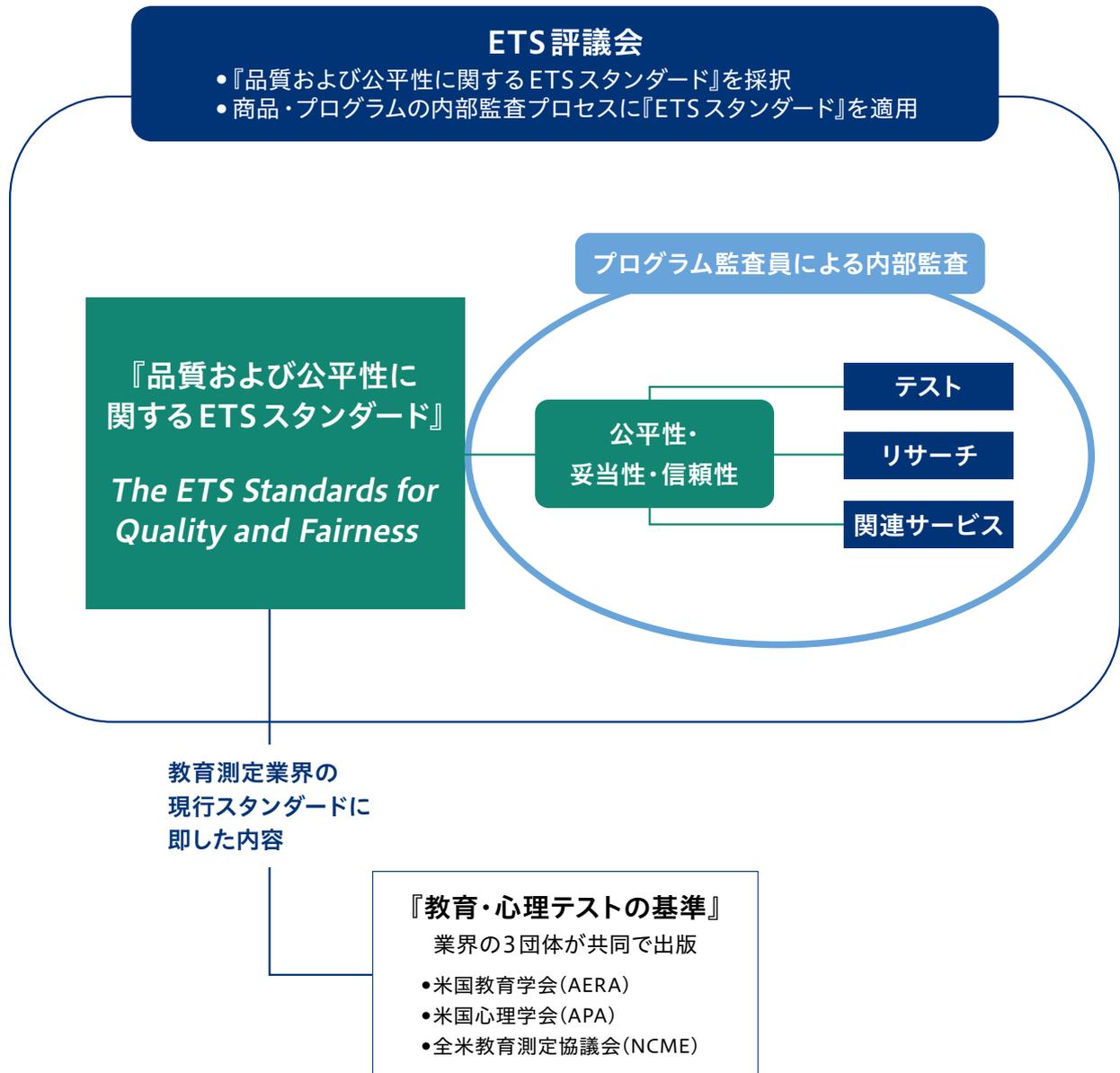
さらに、『ETSスタンダード』はETSのプログラム監査員が商品・サービスを対象とした綿密な内部監査を実施する際の指針としても用いられます。監査のプロセスと結果はETS評議会によって監視され、監査が『ETSスタンダード』に効果的かつ完全に準拠した形で実施されるよう担保されています。

### 業界の現行スタンダードに 即した内容

『ETSスタンダード』は業界の3団体——米国教育学会（AERA）・米国心理学会（APA）・全米教育測定協議会（NCME）——が共同で出版する『教育・心理テストの基準』（2014年版）に明記される、教育測定の現行のベストプラクティスに即しています。

### 世界中のテスト機関の モデルとして

このように、非営利団体であるETSは『ETSスタンダード』を中心に据えた継続的な努力を通じて、測定業界の現行のスタンダードに達する、もしくは上回る水準のテストや商品を提供すると共に、一般社会への説明責任を果たしてきました。その結果、ETSの組織を挙げた『ETSスタンダード』に対する取り組みは、現在では測定のプラクティスや関連スタンダードの改善に尽力する世界中のテスト機関のモデルとなっています。



次項では、テスト品質の必須要素——妥当性・公平性・信頼性——に焦点を当てながら、『ETSスタンダード』に設定された高品質のベンチマークを追求するETSの取り組みの一部をご紹介します。

---

## テスト品質の必須要素—— 妥当性・公平性・信頼性

---

### 妥当性

妥当性はテスト品質の根幹を成すものです。妥当性とは、ETSで活躍したサミュエル・メシクやマイケル・ケインを始めとする科学者たちが長年にわたり構築し洗練させてきた通り、「テストスコアに基づく推論やアクションが、意図された使用に対してどのくらい適切であり、理論・論理・経験上のエビデンスによってどのくらい正当化されるか」を指します。そのため、妥当性検証のプロセスとは、スコアの意図された使用を実証するエビデンスの収集を意味します。

### 妥当性に対する責任

重要な点として、ETSとテスト利用者（顧客、受験者、その他のステイクホルダー）は妥当性を担保する役割を共に担っているということが挙げられます。

テスト開発・提供者であるETSは、意図された使用を満たすテストを設計し、開発し、実施することに対して責任を持ちます。具体的には、ETSでは広範にわたるリサーチを実施し、テストスコアが受験者に対する正確な解釈へとつながるよう担保しています。これにより、テスト利用者が確信を持ってテストを選択し、使用できることを目指しています。

同時に、テスト利用者には、その対象者と目的に合ったテストを選択し、実際の測定内容——妥当性リサーチのエビデンスが示す内容——に基づいてスコアを解釈し、適切な意思決定やアクションのためにスコアを使用していただくことが推奨されます。究極的には、ETSとテスト利用者は、テストの妥当性を裏付けるエビデンスを共に生み出し、テストの利用が、受験者やその他の人々に有益な結果（スコアに基づくより良い意思決定、教育・学習に

対するポジティブな波及効果など）をもたらすことを共に目指していると言えるでしょう。

### テストの目的と対象者について記述する

テストの妥当性担保に関するETSスタンダードの中でも序盤の重要なスタンダードでは、テストに関する以下の要素を明記し、一般公開するよう規定しています。

- 測定対象とする構成概念（知識・技能・その他の特性）
- 各テストの目的
- 受験者に関してテストが主張すること
- スコアやその他テスト結果についての意図された解釈
- 意図された受験者母集団

ETSでは、上記の情報はテスト利用者に対し、ニーズに合ったテストを選択し、スコアを正確に解釈し、適切な意思決定を行うよう促すうえで、極めて重要であると考えています。

### 妥当性を再評価する

また、テスト自体やスコアの妥当性を維持し続けるためには、テストが現実世界の変化に適応していくことが不可欠です。そのため、ETSスタンダードでは、テストに関連する要素（使用する技術、意図された目的、テスト結果の意図された解釈、テスト内容、テストの母集団など）に重要な変更が生じた際には、テストが意図された目的を満たすことを裏付けるエビデンスや、意図された母集団に対するテスト結果の意図された解釈を裏付けるエビデンスを再評価するよう規定しています。また、そうした変更を裏付ける新しいエビデンスを収集・評価するためのリサーチも必要に応じて実施されます。

妥当性は、ETSのテスト品質の重要な基盤を成すものです。ETSはテストの妥当性の実現に尽力しており、広範なリサーチを実施することで、各テストがどのように設計され、スコアが受験者の能力に関する意味のある正確な解釈をどのように裏付けているのかを実証しています。同時に、ETSはテスト利用者に寄り添い、ニーズを理解したうえでスコアを適切に利用できるよう支援しています。

ETSでは、妥当性への取り組みを通じて、受験者、テスト利用者、そして社会に有益な結果をもたらすと共に、意図しないネガティブな結果を最小化することを目指しています。

### 公平性

公平性は、テスト品質におけるもうひとつの重要な要素であり、ETSスタンダードの一部を成しています。公平性は多面的な概念であり、世界の多様化がますます進む現代においては特にその傾向が顕著ですが、測定という文脈で最もわかりやすく定義するならば、「スコアに基づく推論が、さまざま異なる受験者集団にとって、どの程度妥当か」と表現することができるとでしょう。テストが、重要な点（年齢、ジェンダー、地域、母語、障がいの状態など）において異なる受験者によって利用されることを意図している場合、公平性はテスト品質において非常に重要な要素となります。

### 構成概念とは関連性のないスコアの分散を最小化する

テスト開発者は、全ての受験者に公平性あるテストを提供するために、受験者母集団における潜在的な集団に対して、構成概念とは関連性のないスコアの分散による影響を最小化する必要があります。「構成概念とは関連性のないスコアの分散」(construct-irrelevant score variance) という専門用語には難解な響きがありますが、これは単に構成概念——テストが測定する能力——とは無関係に受験者のパフォーマ

ンスに影響を与える要素が存在する、ということの意味しています。

「構成概念とは関連性のない要素」は、特定の人口統計学的背景を持つ受験者（年齢、ジェンダー、地域、民族、人種、母語、障がいの状態においてさまざま異なる受験者）のパフォーマンスに不利な影響を与える傾向があります。ETSで働くテスト開発者、心理測定学者、統計学者、リサーチサイエンティストなどの専門スタッフは、テストプロセスの各段階——設計、開発、実施、採点、事前および事後調査、リサーチ——において指針を示す

ETSスタンダードを遵守することを通じて、公平性を達成すべく厳正に取り組んでいます。

## 公平性をレビューし評価する

具体的な施策として、ETSでは以下の公平性の基本原則に基づき、公平性レビュー、統計分析、リサーチを実施しています。

### <公平性の基本原則>

- 意図された構成概念の重要な要素を測定する
- 受験者の成功に対する構成概念とは関連性のない障壁を回避する
- 多様な受験者集団についての妥当な推論を裏付けるスコアを提供する
- 全ての受験者に敬意を払い、偏りなく対応する

ETSではテストの公平性を達成するために、以下2つのアプローチを含むさまざまな手順を踏んでいます。

(1) テスト開発者は、テスト問題の執筆や内容の精査に当たり、徹底した公平性レビューを複数の段階において実施し、テスト問題や内容に受験者のパフォーマンス発揮を阻む障壁として働く、構成概念とは関連性のない要素（詳細は付録をご参照ください）が含まれていないかを調べ、バイアスや意図しない要素の発生を未然に防いでいます。例えば、英語能力テストの問題が、数学・歴史・科学などの専門的知識を求めるものであった場合には、構成概念とは関連性のない要素を含む問題と見なされ、レビューの段階で除外されます。また、別の例としては、あるトピックについて測定する場合に、複雑なグラフや図の使用の有無にかかわらず同等に測定できるのであれ

ば、最良の測定方法を決定する際には、複雑な視覚資料が特定の障がいのある方のアクセシビリティを妨げる可能性がある事実が考慮されます<sup>\*</sup>。

※上記のようなケースのほかにも、ETSでは受験者の成功に対する「構成概念とは関連性のない身体的障壁」(construct-irrelevant physical barriers)——妥当性のために重要もしくは必須ではないテスト要素が、受験者の知識や「できること」を発揮する能力に干渉する際に生じる障壁——による影響を最小化するために、さまざまな措置を取ることで、身体に障がいのある方に対して公平な受験体験を提供しています。こうした措置には、試験時間の延長、拡大版用紙や点字による受験といった、ETSが規定する特別対応などが含まれます。

(2) テスト実施後には、心理測定学者や統計学者が受験者のデータを用いて、テスト問題やテストが統計的観点から見て公平な振る舞いをしたかどうかを分析しています。問題のひとつ一つが意図した通りに振る舞った状態であることを担保するために、意図通りに機能していない状態——特異項目機能(DIF: differential item functioning)<sup>\*</sup>と言われる現象——が検知されたあらゆる問題は、採点前に除外されます(TOEIC Programに関する公平性レビューについてはSection 5「TOEIC Programを構成するグローバルなコミュニケーションのための英語」もご参照ください)。

※特異項目機能(DIF)  
DIFとは、特定の集団のメンバーにとって、あるテスト問題の難易度が、同等の能力を持つ別集団のメンバーにとってよりも高く・低くなる傾向全般を指します。一般にDIF測定においては、異なる集団のメンバーであっても、受け取った総合スコアが同じであれば、テストが測るものにおいては同等の能力を有すると見なします。ほとんどのDIF分析では、受験者集団は性別、人種、民族性に基づいて定義されます。

ETSでは、公平性に対して非常に真摯に取り組んでいます。テスト受験を希望する多様な集団のあらゆる人々に平等なアクセスを担保するための最先端のプラクティスやスタンダードに従っており、テスト利用者の皆様に信頼していただけるスコアを提供しています。

## 信頼性

信頼性もまたテスト品質に欠かせない要素です。信頼性とは、同一テストの異なるフォーム間、異なるテスト実施回数、異なる採点者間において、スコアが一貫している程度を指します。スコアには意図された解釈を裏付けるだけの十分な信頼性が必要です。ETSでは各テストが求める信頼性の水準を、スコアの意図された使用と、不正確な意思決定がもたらす影響と結果に照らして決定しています。例えば、採用、選抜、入試などのハイスティクスな意思決定に使われるテストであれば、そのテストに求められる信頼性の水準は高くなります。

## 適切な手法を採用

ETSでは、各テストの信頼性の水準を推算する際には、テストの設計・意図された使用・採点方法に応じて、適切な手法を選択し用いています。

スピーキングやライティングのテストで出題されるような自由解答式のタスクについては、口頭・筆記の解答を採点ブルックに基づき採点する人間の採点者、あるいは(TOEFL iBT® Testsなどの一部のテストで採用されている)自動採点システムが示す信頼性が、重要な検討対象となります。ETSでは高度に訓練された人間の採点者や最先端の自動採点技術を採用し、細心の注意を払った採点手順を踏んでいます。採点結果として出されたスコアは、別の採点者によるスコアと持続的・定期的と比較され(e.g., 人間の採点者間での比較や、人間の採点と機械採点との比較)、採点者間の信頼性や一致度が立証されます。

自由解答式タスクのスコアだけでなく、選択解答問題(多肢選択問題など)で構成されるテストセクションのスコアについても、セクションスコアと総合スコアそれぞれにおいて、異なるフォーム間、異なるテスト実施回数で統計上の比較が行われ、信頼性が判定されます。ETSでは、古典的テスト理論や項目応答理論(IRT: item response theory)といった統計的理論を含めた、

さまざまな心理測定技術を駆使し、各テスト問題、各セクション、各フォーム、そして総合スコアの信頼性を確保しています。

## 信頼性の立証と解釈

ETSでは全てのテストについて、ひとつ以上の信頼性係数を算出し、テストの信頼性の数値的エビデンスとして示しています。基本的な信頼性係数は.00から1.00の値を取る絶対値です。1.00という値は完全な一貫性——テストが毎回全く同じように測定している状態——を示します。.00という値は一貫性の完全な欠如——テストの測定が毎回異なるため信頼すべきでない状態——を示します。信頼性係数の値が1.00に近いほど信頼性が高く、テストがあらゆる実施においても受験者の能力を伝えている、ということに対してより大きな信頼を置くことができます。

ETSの統計学者はカットスコアに基づく意思決定の一貫性や正確性を可能な限り担保するために、スコア・スケール上の重要な地点のスコアの信頼性に対して細心の注意を払っています。さらに、ETSの心理測定学者は全てのテスト実施を常時モニタリングし、今後のテスト問題やテストフォームの開発に向けたフィードバックを提供しています。

## 情報提供と分析内容の記録

ETSはテスト利用者に対し、テストスコアの信頼性について適切なエビデンスを提供するだけでなく、信頼性の分析に関する詳細な情報を定期的に記録し公開しています。

こうした情報には、採用した統計や統計手法、信頼性統計の推算対象とした母集団、自由解答式テストで用いられた採点手順や採点者間の一致度といった内容が含まれます。

ETSではこうした情報公開の取り組みを通じて、専門家による分析結果の評価や——さらに重要な点として——分析の再現を促し、それにより測定分野の科学的

発展に寄与していきたいと考えています。

ここでは割愛したテスト品質の要素はほかにも多くありますが、テスト品質に欠かせない要素——妥当性・公平性・信頼性——に対する上述のETSの取り組みは、「教育の品質と平等の向上」というミッションを追求するETSの組織を挙げた努力を示すものです。

ETSはそのスタンダードと共に、過去70年がまさにそうであったように、これからも進化し続けていきます。妥当性・公平性・信頼性ある商品とサービスを常に提供するため、そして教育測定分野を発展させるために、現在進行中のコミットメントを今後も続けていきます。

## References

Educational Testing Service. (2014). *The ETS Standards for Quality and Fairness*. Author.

Educational Testing Service. (2022). *ETS Guidelines for Developing Fair Tests and Communications*. Author.

Livingston, S. A., Carlson, J., & Bridgeman, B. (2018). Test reliability-basic concepts. (*Research Memorandum No. RM-18-01*). Educational Testing Service.

## 付録

2022年版の『公平なテスト開発とコミュニケーションのためのETSガイドライン』（*ETS Guidelines for Developing Fair Tests and Communications*）では、受験者の成功に対する「構成概念とは関連性のない障壁」をカテゴリ別にリストアップしています。その一部である「構成概念とは関連性のない感情的障壁」（construct-irrelevant emotional barriers）として、以下を挙げています。

### 受験者の成功に対する「構成概念とは関連性のない感情的障壁」

#### ■避けるべきトピック：

非常に物議をかもし・屈辱的・扇動的・動揺を招くようなトピックについては、妥当性において重要でない場合には、回避することが最も望ましい。

#### ■注意を要するトピック：

弁論

アバター（ジェンダー、人種、民族の多様性を表現する、もしくはそうした特性のないものを使用）

伝記

ブランド名

対立

暗号化された素材

障がい

進化（人類の進化や、人類と霊長類の類似性）

集団間の差異

ユーモア、皮肉、風刺

贅沢品

地図（紛争地域の地図）

集団に対する誤った扱い

私的な質問

宗教

ロールプレイ（感情的なストレスを与えるロールプレイなど）

性的な行動

奴隷

ステレオタイプ

暗黙の想定

暴力や苦しみ

視覚素材（明確な目的なく用いる場合）



## Section 4

# TOEIC Bridge Tests と TOEIC Tests

初級から上級の英語学習者を  
対象とする“テストファミリー”

TOEIC Program は、日常生活や職場における英語コミュニケーション能力を包括的に評価します。初・中級者を対象とする TOEIC Bridge Tests、中・上級者により適した TOEIC Tests の提供を通じて、英語を学び始めたばかりの方から、上級レベルの英語力の習得を目指す方まで、幅広い学習者のラーニングジャーニーをサポートしています。

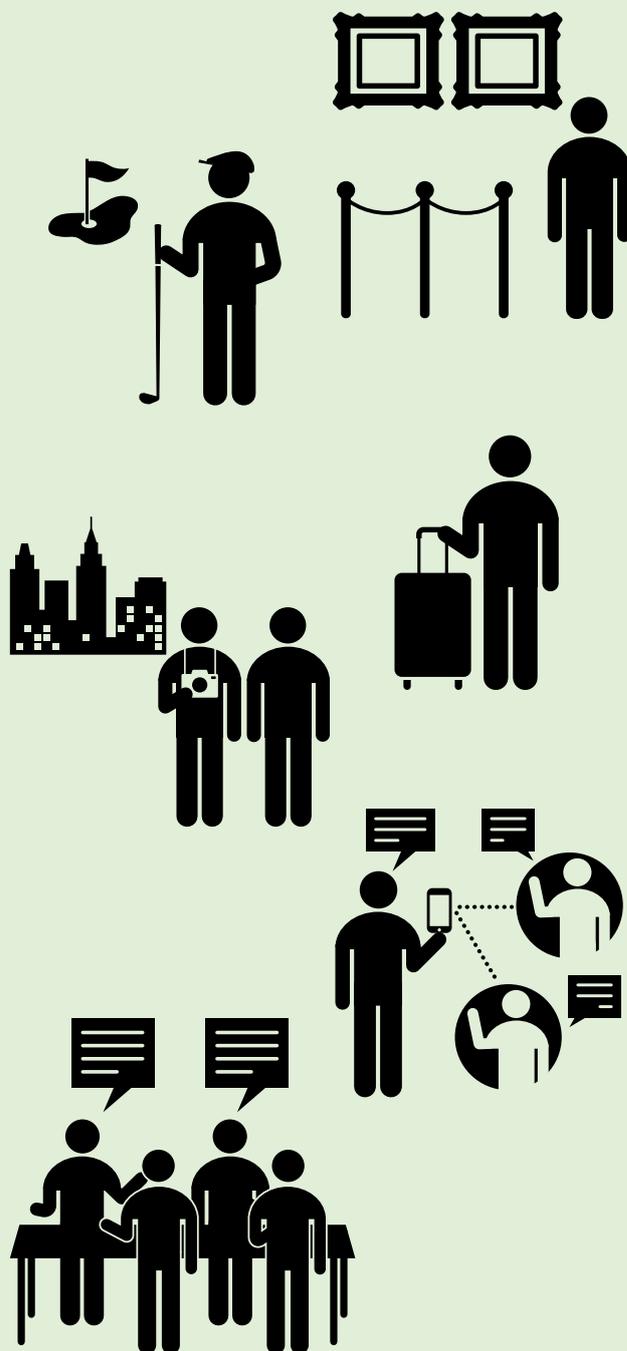
# TOEIC Program——日常生活や職場における 英語コミュニケーション能力を測る“family of assessments”

## 国際社会の 共通言語としての英語

英語は、世界中のビジネスや国際的なコミュニケーションにおいて、最もよく使われる言語となりました。世界で最も学習されている外国語、インターネットと情報化の時代における主要言語、母語の異なる人々がコミュニケーションを図るための「国際社会の共通言語」(global lingua franca)——それが英語という言語です。

英語を学び始めたばかりの方は、英語力を身に着けることで、得られる情報の範囲や、海外旅行といった活動の幅を広げていくことができます。たとえ英語能力レベルがさほど高くなくとも、日常生活のさまざまな場面——趣味やスポーツなどの余暇、移動や旅行、映画や音楽や博物館などの娯楽、英語を母語としない人同士での交流——において、英語力は大いに役立ちます。

そのため、多くの初・中級者は、上級レベルの英語能力の習得を長期的な目標に据えているかもしれませんが、初・中級レベルの英語力をすぐに実践的に活用できる場面は、社交会話、趣味、移動や旅行、娯楽など、日常生活の中にたくさんあります。もちろん、職場においても、オフィスでの打ち合わせや、同僚とのテキストメッセージのやり取りなど、初・中級者が力を発揮できる場面が見つかることでしょ。



中・上級レベルの英語能力は、日常や職場における幅広い場面で活用することができます。職場での活用機会としては、業務に関連する情報を読み、聞いて理解する場面のほか、国をまたいで、同僚と共に働く、顧客とコミュニケーションを図る、ビジネス相手と交渉するといった場面が挙げられます。

英語学習者の中には、中級レベルの英語能力を習得できれば日々の業務に必要な基本情報にはアクセスできるという方もいれば、さらに上級のレベルを最終目標とし、将来的にグローバルビジネスで活躍するキャリアを展望している方もいます。企業で働く人々や起業家が英語を学習するのは、この時代において、コミュニケーションを円滑に図ることがいかに重要であるかを実感しているからではないでしょうか。

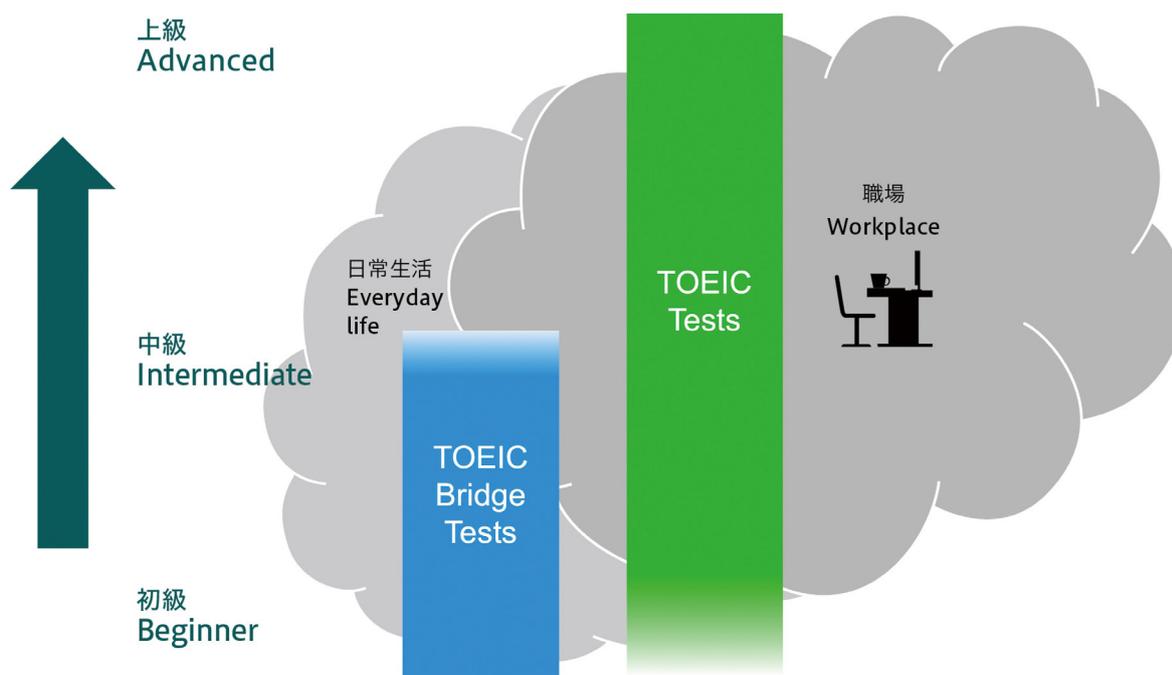
国際的な職場で特に英語力が求められる分野としては、企画、財務、製造、人事、各種技術などが挙げられます。



## “ファミリー”としての TOEIC Bridge Tests と TOEIC Tests

下図のように、TOEIC Programは現実世界の日常生活と職場におけるコミュニケーションの場面で必要となる、初級から上級レベルの英語力を包括的に評価しています。TOEIC Programは初・中級レベルの英語能力を測定するTOEIC Bridge Testsと、中・上級レベルにより重点を置いたTOEIC Testsによって構成されています。

TOEIC Bridge TestsとTOEIC Testsはいずれも、英語の4つの技能(聞く・読む・話す・書く)を直接評価するよう設計されており、これら4技能の測定を通じて包括的なアプローチを提供し、受験者個々人の英語力の全体像をより完全な形で描き出します。また、TOEIC Programは、英語能力レベルがさまざまな受験者に関わりのある現実世界の場面やタスクに重点を置くことで、彼らが実際の生活で「遂行できると思われること」を反映した評価を提供するよう設計されています。



**TOEIC Bridge Tests : 初級から中級レベルを測定**

TOEIC Bridge Listening & Reading Tests  
TOEIC Bridge Speaking & Writing Tests

**TOEIC Tests : 中・上級レベルに重点を置いて測定**

TOEIC Listening & Reading Test  
TOEIC Speaking & Writing Tests

## TOEIC Bridge Tests

### ——日常生活における英語力に重点

TOEIC Bridge Testsは初級から中級レベルの基礎的な英語力（聞く・読む・話す・書く）を測定し、日常生活における英語コミュニケーション能力について、信頼性・妥当性・公平性ある評価を提供しています（Schmidgall et al., 2019; Schmidgall et al., 2021）。リスニングのタスクでは、会話やトークなどの短い話し言葉を理解する能力などを測り、リーディングのタスクでは、短い書き言葉を理解する能力などを測ります。スピーキングのタスクでは、簡単な用件を伝える（e.g., 単純な質問をする、依頼をする）能力、基本的な情報を与える能力、個人的な関心事について、明瞭に話す相手とやり取りする能力などを測ります。ライティングのタスクでは、簡単な用件についてメモを取ったり、メッセージを書いたりする能力を測ります。TOEIC Bridge Testsでは、これらの能力を、初・中級の英語学習者に関わりのある適切なタスクによって測定しています。

TOEIC Bridge Testsのタスクは、中等教育の学生や大人にとって身近なコミュニケーションの場面やトピックに設定されており、よく見られる個人的・公的な場面のほか、一般的で身近な職場の場面も出題されます。個人的な場面としては、例えば家族の集まりや、個人的な趣味や興味に関連するものが挙げられます。公的な場面としては、移動や旅行、娯楽イベント、ショッピングなどが挙げられます。職場の場面は、例えばオフィスでの打ち合わせ、雑談、テキストチャットといった、専門知識がいっさい求められない最も一般的なものに限られます。TOEIC Bridge Testsでは、こうした日常生活において比較的よく出合う場面やトピックを用いることにより、初・中級レベルの受験者が「聞く・読む・話す・書く」の4技能を発揮しやすいリアルな設定を提供しています。

## TOEIC Tests

### ——職場における英語力に重点

TOEIC Testsは職場や日常生活における英語能力を測ります。これらの場面で必須とされる「聞く・読む・話す・書く」の技能に焦点を当て、受験者が英語能力をさまざまな形で発揮する機会を提供しています（ETS, 2022a; ETS, 2022b）。

TOEIC Listening & Reading Test（以下、TOEIC L&R）のリスニングセクションには、陳述、質問、会話、トークを多様な英語（e.g., 米国、英国、カナダ、豪州およびニュージーランドのアクセントを採用）で録音した音声から作成された問題が出題され、リーディングセクションには、さまざまな種類の文書の理解を求める問題が出題されます。

TOEIC L&Rのリスニングセクションでは、「短い会話や長めの会話において、要点を推測できる、詳細が理解できる」といった能力や、「フレーズや文から、話し手の目的や暗示されている意味が理解できる」といった能力を測定します。リーディングセクションでは、「具体的な情報を見つけて理解できる」、「情報を関連付けることができる」、「文書中の情報を基に推測できる」、「文脈を踏まえて語彙・文法を理解できる」といった能力を測定しています。

TOEIC Speaking Testのタスクは、「英語に堪能な話者に対して、理解しやすい言葉で話すことができる」、「社交や仕事上の定型的なやり取りができる」、「一般的な職場にふさわしい筋道の通った発話を続けられる」といった能力を示す機会を提供します。TOEIC Writing Testのタスクは、「文法的に正しい文を作成できる」、「情報を伝えるために複数の文で構成される文章を作成できる」、「複雑な考えを表すために複数の段落から構成される文章を作成できる」

といった能力を示す機会を提供します。

TOEIC Testsが測定するこうした能力はいずれも、現実の世界において——とりわけ職場において——効果的にコミュニケーションを図るうえで欠かせないものです。

### 「社会人」「学生」間で 有利・不利の生じないテスト

TOEIC Testsのタスクは、日常生活や職場のコミュニケーション場面に設定されていますが、特定の職種や職場に特化した知識やスキルを測るのではなく、一般的な能力を測ることに焦点を当てています。TOEIC L&Rには、日常の活動（旅行、娯楽、健康、外食など）に関連した場面や状況のほか、一般的な職場の場面（企画、財務、製造、オフィス、人事、購入など）が出題されます。TOEIC Speaking & Writing Testsには、多様な仕事の場面で求められるタスクや、あらゆる文化や文脈に共通する身近な日常のタスクが出題されます。TOEIC Testsでは、特定の業界に固有のビジネス知識を評価あるいは求めることはなく、日常的な仕事で用いられる範囲を超えたビジネスや技術分野の専門的な語彙の知識を求めることもありません。

重要な点は、全てのテスト問題が、内容と公平性の観点から複数のステップにわたるレビューを受けているということです。この「内容レビュー」および「公平性レビュー」はテスト問題の中に、異なる受験者集団の間——「社会人・学生」、「年長者・若年者」、「女性・男性」など——において不公平な形で有利・不利に働く内容が含まれる可能性を最小化するために行われます。

最近、ETSの研究者や心理測定学者たちは、TOEIC L&Rが学生の受験者に比べて社会人に対

して不公平に有利に働くテストであるかどうかの調査を実施しました (Schmidgall et al, in press)。調査では、あらゆるテスト問題から「特異項目機能」(DIF: differential item functioning)を検出するよう設計された統計手法が用いられました。このDIF分析のプロセスによって、テスト問題やスコアにおける潜在的なバイアスが検出されます。DIFとの関連が示された場合、そのテスト問題には、2つの受験者集団間 (e.g., 「社会人・学生」)において、いずれか一方に対して不公平な形で有利に働く可能性があることが示唆されます。そのため、当該テスト問題は、コンテンツの専門家パネルによって精査され、一方の受験者集団にとって不公平に有利に働く内容が含まれているかどうかについて、判定が下されます (公平性およびDIF分析についてはSection 3「ETSの品質への取り組み」もご参照ください)。

先の調査では、TOEIC L&Rの9本のテストフォーム (合計1,800問)をDIF分析し、学生と社会人 (いずれもフルタイム)のパフォーマンスを比較しました。最初の段階でDIFの可能性が示された問題は、1,800問中14問 (0.8%)のみでした。つまり、一方が他方に対して不公平な形で有利に働いた可能性のある問題は非常に少なかったのです。結果を受けて、専門家パネルが該当する14問のレビューを行ったところ、最終的に14問全てにおいて、学生よりも社会人に有利に働く (あるいは社会人よりも学生に有利に働く)、明確なバイアスは検知できないという結論に至りました。

大量のサンプル数に対して、統計分析によりDIFの可能性が示されたテスト問題の割合が極めて低かった (1%未満)こと、また、当該テスト問題において社会人および学生の受験者に作用するバイアスが明確に示されなかったとする専門家パネルの結論——これらの調査

結果は、TOEIC L&Rの内容が、学生・社会人のいずれか一方に対して、不公平な形で有利に働くようには制作されていないという点について、受験者、教員、スコア利用者に改めて保証するものと言えます。

TOEIC Program が提供する TOEIC Bridge TestsとTOEIC Testsは、学生・社会人を問わず、英語レベルの異なる多様な英語学習者のニーズに対応しています。受験者それぞれの目標や目的——目指したい英語能力のレベルや、英語能力の測定対象として重点を置きたい場面——に合わせて、ご活用いただければ幸いです。

---

## TOEIC Program と歩む Learning Journey

---

### 英語学習の旅を照らす指標として

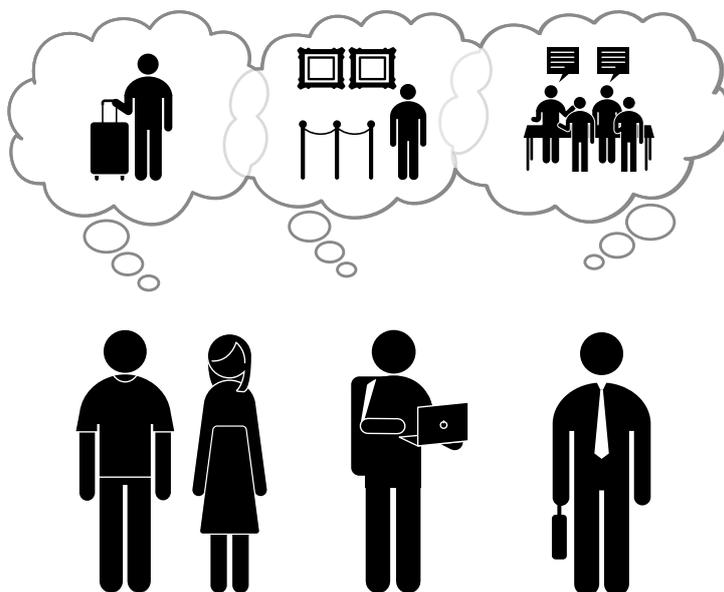
TOEIC Bridge Tests および TOEIC Tests は、日常生活や職場における初級から上級レベルの英語能力を測る、信頼性・妥当性・公平性あるテストを求める利用者——学習者、教員、組織——に有効にご活用いただけるよう設計されています。各テストは、学習者が英語の4技能を横断的に伸ばしていく道を進む中で、有効な指標としての役目を果たし、学習者が learning journey をどれだけ歩むことができたのか、その軌跡を照らすベンチマークとなります。

### 初・中級者が「できること」を発揮しやすい TOEIC Bridge Tests

TOEIC Bridge Tests は、初・中級レベルの中等教育の学生や大人の学習者が英語の知識やスキルを習

得していく過程において、包括的かつ有効な評価指標となります。

TOEIC Bridge Tests のタスクは、TOEIC Tests のタスクよりも難易度がいくらか易しく設定されています。これは TOEIC Bridge Tests のタスクが初・中級者に適したものとなるよう、専門家たちが細心の注意を払って設計しているためです。初・中級者にとって、相対的に難易度が高い TOEIC Tests の受験はチャレンジングな体験となるかもしれません。対して、TOEIC Bridge Tests のタスクは、初・中級者が現実世界で遂行できると想定されるコミュニケーションタスクを反映しています。そのため、初・中級の受験者は本来の持っている知識やスキルをより発揮しやすく、英語で「実際にできること」を示すことができます。

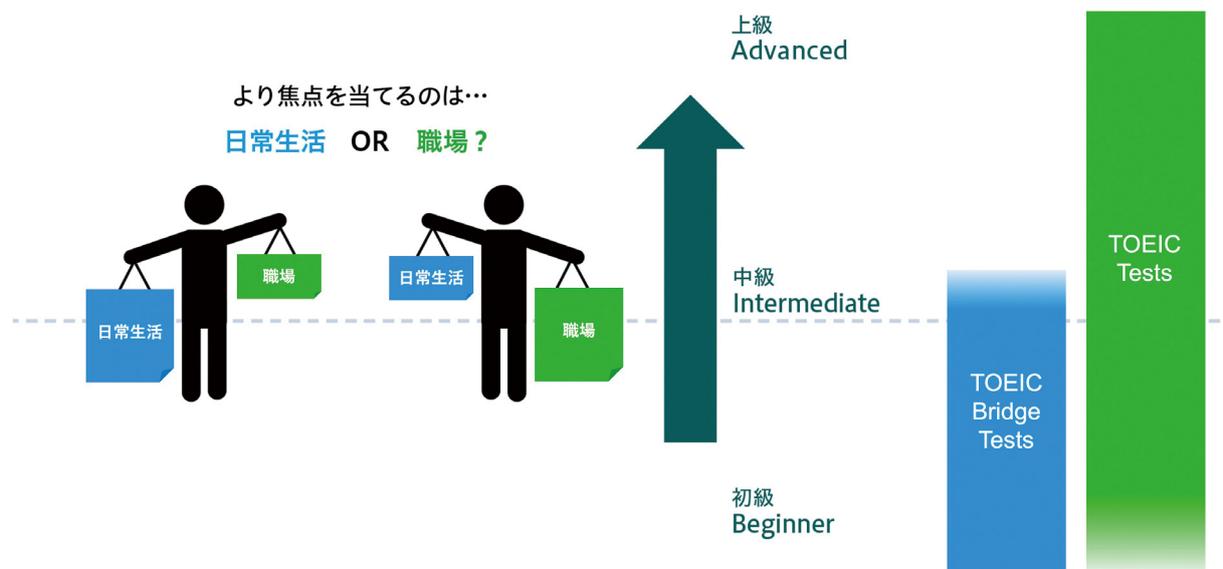


## 中・上級者は焦点やニーズに応じた使い分けを

中級者にとっては、TOEIC Bridge Tests、TOEIC Testsのいずれもが受験対象となるでしょう。適したテストを見極めるために重要となるのが、そのテストが「どのような場面の英語力を測定対象としているのか」という点です。

中級者が日常的な使用場面（個人的・公的な場面な

ど）により焦点を当てて測定するならば、TOEIC Bridge Testsの方がより適切な選択となり得るでしょう。日常生活の場面には、上級レベルの言語知識やスキルが必要となる複雑なコミュニケーションタスクは概して登場しないためです。一方で、中級者が一般的な職場の場面により焦点を当てて測定する場合には、TOEIC Testsの方がより適しているでしょう。



また、国際的な職場での活躍を目指す中・上級者は概して、英語学習の計画やコミュニケーション上のニーズに対して、TOEIC Testsが非常に適していると感じることでしょう。ここで改めてお伝えしたいのが、TOEIC Testsは日常や職場における英語能力に関する情報を提供するよう設計されていますが、先述の通り、いかなる専門的なビジネス知識も求めず、また、ビジネス経験があ

る方にとって有利に働くこともない、という点です。全般として、TOEIC TestsのタスクはTOEIC Bridge Testsのタスクよりも難しく複雑ですが、これは出題されるコミュニケーションタスク自体の性質に起因しており、職場に関連した難しいピックや場面が設定されているわけではありません。

TOEIC Testsのスコア(幅広い能力範囲を対象とする各4技能のスコア)は学生・社会人を問わず、学習者が多様な一般的コミュニケーションの目的のために「いかに効果的に英語を使用できているか」「学習目標をどこまで達成できているか」を正確に示す指標として、ご利用いただけます。

ここまで、英語学習の道のりをサポートする適切なテストを求める学習者の視点からTOEIC Programのご紹介をしてきましたが、ここで改めて、TOEIC Testsが初級者も含めたあらゆる英語レベルの学習者に有用な情報を提供するという点についても、お伝えしておきたいと思います。

TOEIC Testsは、例えば、英語学習を始めたばかりの学習者が、「自分の現在地と最終目標である上級レベルとの“距離”を知りたい」というケースや、初・中級者が「現実世界の職場におけるコミュニケーションタスクを

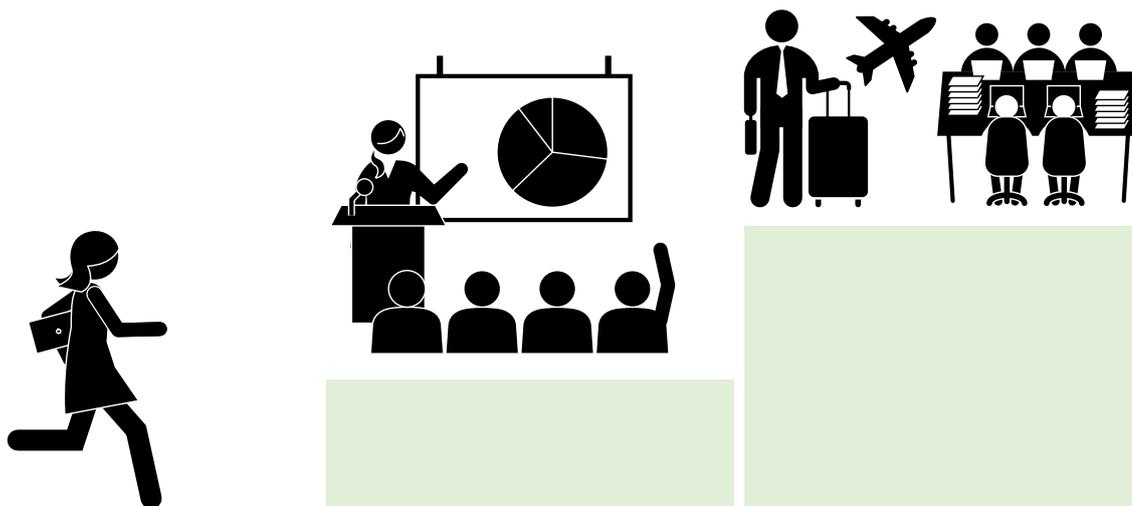
体験して、自分の改善点を見つけたい」というケースにもお役に立ていただけます。さらに、団体利用者(言語学習プログラムなど)が、全レベル(初級から上級まで)の学生を適切なクラスやコースにプレイングする際や、企業の利用者が、英語レベルがさまざまに異なる全従業員(初級者も含む)のひとり一人に適切な配置を行う際にも、効果的かつ効率的なソリューションとなります。

TOEIC Bridge TestsおよびTOEIC Testsは丹念に設計された、信頼に値する適切なテストです。英語学習者やその指導者におかれましては、TOEIC Programのスコアを学習進捗の記録や学習のフィードバックにご利用いただけることでしょう。また、雇用者や団体利用者におかれましても、TOEIC Programのスコアを特定のプログラムや役職に求められる英語力を有する人材の選抜や把握にお役に立ていただければ幸いです。



## References

- Educational Testing Service. (2019a). *TOEIC Bridge® Speaking and Writing Tests: Examinee handbook*. Author.
- Educational Testing Service. (2019b). *TOEIC Bridge® Listening and Reading Test: Examinee handbook*. Author.
- Educational Testing Service. (2022a). *TOEIC® Speaking and Writing Tests: Examinee handbook*. Author.
- Educational Testing Service. (2022b). *TOEIC® Listening and Reading Test: Examinee handbook*. Author.
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge® Tests (*Research Report No. RR-21-20*). Educational Testing Service. <https://doi.org/10.1002/ets2.12335>
- Schmidgall, J., Huo, Y., Cid Carreno, J., & Wei, Y. (in press). Do full-time employees have an unfair advantage over full-time students? Investigating fairness claims for a general-purpose assessment of English proficiency for the international workplace (*Research Report*). Educational Testing Service.
- Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests—Framework paper (*Research Report No. RR-19-30*). Educational Testing Service. <https://doi.org/10.1002/ets2.12267>



## 付録

### ■ TOEIC TestsとTOEIC Bridge Testsに出題されるテーマ(場面や状況)の例

テーマ	TOEIC Tests	TOEIC Bridge Tests
企画	調査、製品開発	—
外食	ビジネスランチ、インフォーマルなランチ、宴会、歓迎会、レストラン予約	ランチ、ディナー、レストラン、予約、ピクニック
娯楽	映画、劇、音楽、美術、展示会、美術館、マスメディア	映画、劇、音楽、美術、美術館
財務と予算	銀行業務、投資、税金、経理、請求書	—
一般業務	契約、交渉、合併、マーケティング、販売、保証書、事業計画、カンファレンス、労務	銀行、図書館、郵便局、予約、広告
健康	医療保険、通院、歯科医、診療所、病院	健康全般と健康管理
住居／法人不動産	(住居／法人不動産) 工事、仕様、購入と賃貸、電気・ガスサービス	(住居) アパート、住居、購入と賃貸、修理
製造	組立ライン、工場管理、品質管理	—
オフィス	役員会議、委員会、レター、メモ、電話、ファックスとEメールメッセージ、オフィス機器と家具、事務手続き	レター、電話、Eメールとテキストメッセージ、オフィス機器と家具、オフィスの雑談
人事	採用活動、雇用、退職、給料、昇進、求人申込、求人広告、年金、賞与	—
購入／買い物	(購入) 買い物、備品の注文、出荷、請求書	(買い物) 食料雑貨類、衣服、ネットショッピング
技術	電子機器、技術、コンピュータ、実験室と関連器具、技術仕様	コンピュータと技術
旅行・その他	電車、航空機、タクシー、バス、船、フェリー、切符、予定、駅や空港のアナウンス、レンタカー、ホテル、予約、遅延、キャンセル	道案内、電車、航空機、タクシー、バス、船、フェリー、切符、予定、駅や空港のアナウンス、レンタカー、ホテル、予約 (その他) 趣味、スポーツ、日常の活動、余暇の活動、天気予報、報道番組、新聞、etc.

## Section 5

# TOEIC Programを構成する グローバルなコミュニケーション のための英語

## 英語能力テストのグローバルスタンダードを確立

TOEIC Program は英語が学ばれ、使われている世界中の日常生活や職場における英語コミュニケーション能力を測定しています。各テストでは、ネイティブスピーカーのような英語能力や専門的なビジネス英語は求められません。測定対象とするのは、ノンネイティブスピーカーが習得すべきと期待される、「国際的な場面で効果的にコミュニケーションを図る」ための英語力です。

---

## ノンネイティブスピーカーの英語コミュニケーション能力を測定するグローバルなテスト

---

### 進化し続ける

#### 英語コミュニケーションに対応

効果的な英語コミュニケーション能力は、グローバル社会における「真の必需品」です。とりわけ今日のように高度につながるインターネット時代においては、世界中のビジネスや教育その他の場面で、英語はコミュニケーション言語——「国際社会の共通言語」(global lingua franca)——として、さまざまな目的、多様な文脈で用いられています。英語の使用目的や話者も多様化しており、英語コミュニケーションの在り方も徐々に進化しつつあります。

TOEIC Programでは、進化を続ける国際的な英語コミュニケーションの性質を認識し、国際的な職場や日常生活における英語使用の実態をより精度高く把握するためのリサーチを、過去数十年にわたって推進してきました。

これまでに実施したリサーチの一部として、以下が挙げられます——「ビジネス英語」の使用 (Dudley-Evans & St. John, 1996)；国際的な職場における英語スピーキング能力 (Schmidgall & Powers, 2021)；国際的な職場における英語ライティング能力 (Schmidgall & Powers, 2020; Lee & Schmidgall, 2020)；国際的な職場における英語リスニングおよびリーディング能力 (Schmidgall et al., 2021b)；国際的な英語コミュニケーションに関するより広範な理論に関連付けられる TOEIC Programにおける英語能力の定義 (Schmidgall, in press)——最後に挙げたリサーチでは、国際的な場面で効果的にコミュニ

ケーションを図るうえで求められる英語能力の実像を捉えるべく、職場で働く人々を対象としたニーズ分析サーベイや、リサーチ文献レビューを含めたさまざまな手法が活用されています。

### ネイティブスピーカーのような英語や専門的な英語を評価対象としないテスト

TOEIC Programは、こうしたリサーチから得られたエビデンスに基づく理解の下、ノンネイティブスピーカーが英語を使う場面で求められる英語コミュニケーション能力を測定する、信頼の置けるテストを提供しています。ネイティブスピーカーの英語力にどれだけ近づけたかという点は評価の対象にしておらず、また、「アメリカ英語」や「イギリス英語」といった特定の種類の英語のみに重点を置くこともありません。さらに、特定の職種やビジネス場面で出合うような、専門的あるいは例外的な英語を測定の対象とすることもありません。

TOEIC Programは、一般的な英語コミュニケーションタスクの遂行能力を証明する必要がある、世界中のあらゆる学習者の期待に応える英語能力テストを提供します。

### テストタスクは世界中で見られる典型的な状況に設定

TOEIC TestsやTOEIC Bridge Testsのタスクは、世界中のさまざまな場面で共通して見られる英語使用を反映するよう設計され、最も一般的な日常生活

や職場における英語使用の実態を精緻に分析した結果に基づき、厳選されています (e.g., Hines, 2010; Lee & Schmidgall, 2020)。

例えば、TOEIC Bridge Listening & Reading Testsのリーディングテストには、「日本の国際線空港で、会議の手配をするために同僚間で交わされるテキストメッセージを読む」といったタスクや、「オーストラリアの図書館の公共サービス通知に目を通す」といったタスクが出題されるかもしれません。また、TOEIC Speaking Testには、「カナダの市場調査会社によるテレビ視聴習慣に関するインタビューに答える」といったタスクが出題されるかもしれません。

こうしたリーディングやスピーキングの例に限らず、リスニングやライティングも含めて、TOEIC Programが主眼とするのは、英語力が求められる一般的な状況や活動において効果的にコミュニケーションを図る能力を測定することです。それが世界のどこで起きていることなのか、ということとは関係ありません。TOEIC Programは真にグローバルな英語テストを提供します。

### 受験や解釈の地域を問わず 比較可能なスコアを提供

グローバルなテストであるTOEIC Programは、同じテストであればフォームが異なっても、いどこで受験しても比較可能なスコアを提供します。メキシコで受験した方と、フランスで受験した方がいた場合、両受験者のスコアが同じならば英語能力レベルも同じであると解釈することができます。

実際にETSの心理測定学者はさまざまな統計技術を駆使して、受験者のパフォーマンスがTOEIC Programのスコア・スケール上に確実に正確に換

算されるよう尽力しており、「等化」と呼ばれる統計処理を行うことで、異なるテストフォーム間やテスト実施回数におけるスコアの比較可能性を担保しています。

さらに、さまざまに異なる受験者のテスト体験が非常に等しいものであり、スコアが実際の英語コミュニケーション能力を公平かつ正確に示していることについても、信頼してご利用いただけます。

一例として、TOEIC Speaking & Writing Tests (以下、TOEIC S&W)のスコアは、世界中の国際的な職場で働くプロフェッショナルが受験者の英語コミュニケーション力を評価した場合の結果を予測できることが、ETSの研究によって示されています。つまり、TOEIC S&Wのスコアが高い受験者ほど、国際的な職場環境でより効果的なコミュニケーターとして認識されることが分かったのです (Schmidgall & Powers, 2020, 2021)。

また、TOEIC Programのスコア・スケールは、英語学習者の能力について、最大に有意義かつ解釈可能な情報を提供するように設計されています。これは英語力の尺度として、世界中のスコア利用者に共通の理解をもって利用していただくためです。TOEIC Programは初級から上級まで、多くのレベルを識別するスコアを提供しており、スコアの意味についても、公式認定証や補足資料を通じて、詳細にお伝えしています。

■ TOEIC Programのスコアに関する資料



[https://iibc.me/  
qual\\_persp\\_tp\\_01](https://iibc.me/qual_persp_tp_01)

---

## 世界の英語コミュニケーションの「今」を反映する 高品質かつ公平なテストを提供

---

### 定期的なテスト設計のアップデートで 英語コミュニケーションの変化に対応

TOEIC Programはグローバルな英語能力テストとして、受験者のコミュニケーションのニーズや今日の一般的な日常生活や職場での英語使用を反映したタスクや内容を出題しています。

TOEIC Programでは、世界中のスコア利用者に対し、最も正確かつ有意義な測定結果を確実に提供し続けるために、各テストの基礎設計を定期的にレビューしており、必要に応じてアップデートを実施しています (e.g., Ashmore et al., 2018; Park & Bredlau, 2018)。

例えば、2019年にはTOEIC Bridge Testsがリデザインされました。リデザインを担当したETSの研究者、テスト開発者、心理測定学者から成るチームは設計プロセスの第一歩として、測定対象となる英語能力の特徴——初・中級者が日常生活で求められる英語力——の定義に取り組みました。

まず研究者たちが、日常のコミュニケーションで求められる言語知識や技能の特徴や、初・中級者が遂行すべきと期待されるコミュニケーションタスクの特徴について調査および定義し (Schmidgall et al., 2019)、それを受けて、テスト開発者たちが新たなテストタスク——現実世界において重要なコミュニケーションタスクに該当し、かつ、日常のコミュニケーションで求められる言語知識と技能を測定するタスク——を開発していきました (Everson et al., 2019)。

### テスト内容やタスクは 英語使用に関する最新調査を基に作成

さらに、TOEIC Programでは、一般的な日常生活や職場での英語の使われ方に関する最新状況の把握に基づき、常に新しいコンテンツを制作しています (e.g., Yoon et al., 2017)。

ETSのテスト開発者たちは、TOEIC Programが測る基本的な英語能力を反映する典型的なコミュニケーションタスクを調査し、今日の世界各地での英語の使われ方を公平に代表できるよう力を尽くしています。

例えば、出題されるタスクには、Eメール、スマートフォンのテキストメッセージ、インターネット上での言語使用など、多様なコミュニケーション媒体を反映したものが含まれます。

右記のTOEIC Bridge Writing Testのサンプル問題では、コンピュータやスマートフォンを介したテキストチャット環境におけるライティングのタスクが出題されています。

また、テストの内容は身近なトピック (旅行、娯楽、健康、外食、ニュースなど) や、一般的なビジネスや職場でのコミュニケーション上の要求 (国際的なビジネスや職場でよく見られるタスク) に関する最新情報を反映しています。

このように、TOEIC Programのテスト内容やタスクは、現実世界の活動やコミュニケーション方法の変化に常に対応しています。

**Directions:** Respond clearly and fully to the message from your friend Chris. You have 8 minutes to prepare and write.

- Suggest one dish you like, and
- Briefly explain to Chris how to make it

**Chris**

**Chris:**  
Hope you're doing well! My sister's visiting me tomorrow and I'm thinking of cooking something nice for her (don't know what!). I wonder if you have any ideas...

■公式サイトから  
サンプル問題を  
ご覧いただけます。



[https://iibc.me/  
qual\\_persp\\_tp\\_02](https://iibc.me/qual_persp_tp_02)

### 世界中のあらゆる受験者に 高品質かつ公平なテストを提供

TOEIC Programは世界中の受験者に、高品質かつ公平なテストを提供しています。テストの品質と公平性を担保するため、テスト問題は実際のテストで使用される前段階において、品質および公平性に焦点を当てた厳格なレビューのプロセスを通過しなければなりません。

まず、各問題は複数回にわたる内容レビューを受けます。このプロセスでは、テスト開発者が独立した立場から、各問題が確実に適切かつ仕様に沿った

ものとなっているかを精査します。

内容レビューを通過した問題は、公平性レビューとセンシティビティ・レビューを受けます。ここでは、各問題の内容に、あらゆる属性(e.g., ジェンダー、言語、文化)の受験者集団に対しても、有利・不利に働くバイアスが生じていないことを確認します。

ETSのテスト開発者たちは、ETSの公平性に対するコミットメントの一環として、こうした包括的な取り組みを行っています。ETSではスコアの公平性について、以下のように主張を表明しています(Schmidgall et al., 2021b, p. 12)。

- テストタスクは、一部の受験者に対して不適切に有利・不利に働く可能性のある解答形式や設問を用いないものとする
- テストタスクは、受験者が不快に感じる可能性のある内容を含まないものとする

テスト実施後の採点プロセスでは、問題（およびテスト全体）が適切に機能していることを確認するための統計分析が行われます（詳しくはSection 4「TOEIC Bridge TestsとTOEIC Tests」をご参照ください）。

ETSの心理測定学者は測定を専門とする科学者として、テストタスクやテスト内容の品質を絶えずレビューしており、世界中で英語能力の証明を必要としている多様な多くの英語学習者集団にとって、適切なテストを確実に届けられるよう取り組んでいます（e.g., Yoo et al., 2019）。なお、ETSが設定する品質

および公平性に関するスタンダードについてより詳しくは、Section 3「ETSの品質への取り組み」をご参照ください。

ETSの研究者、テスト開発者、心理測定学者たちによるこうした連携を通じて、TOEIC Programは今日における最も一般的な英語使用——学習者が習得すべきと想定される英語——を公平かつ理解しやすい形で示しています。また、トピックについても、あらゆる学習者にとって十分に身近なものとなっています。

TOEIC Programは、世界中の日常生活や職場において有用な英語の「学び」を測り、学習者や教員の焦点がそうした学びに向かう一助となるテストを提供していきます。

## References

- Ashmore, E., Duke, T., & Sakano, J. (2018). Background and goals of the TOEIC® Listening and Reading update project. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 3.1-3.8). Educational Testing Service.
- Dudley-Evans, T., & St. John, M. J. (1996). Report on business English: A review of research and published teaching materials (*TOEIC Research Report No. 2*). Educational Testing Service.
- Everson, P., Duke, T., Garcia Gomez, P., Carter Grissom, E., Park, E., & Schmidgall, J. (2019). Development of the redesigned TOEIC Bridge® tests (*Research Memorandum No. RM-19-10*). Educational Testing Service.
- Everson, P., & Hines, S. (2010). How ETS scores the TOEIC® Speaking and Writing test responses. In D. Powers (Ed.), *TOEIC® Compendium* (1st ed., pp. 8.1-8.9). Educational Testing Service.
- Hines, S. (2010). Evidence-centered design: The TOEIC® Speaking and Writing tests. In D. Powers (Ed.), *TOEIC® Compendium* (1st ed., pp. 7.1-7.31). Educational Testing Service.
- Lee, S., & Schmidgall, J. (2020). The Importance of English Writing Skills in the International Workplace (*Research Memorandum No. RM-20-07*). Educational Testing Service.
- Park, E., & Bredlau, E. (2018). Expanding the question formats of the TOEIC® Speaking test. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 2.1-2.8). Educational Testing Service.
- Schmidgall, J. (in press). Validity considerations for testing speaking in the workplace: TOEIC. In L. Davis (Ed.), *Challenges and innovations in speaking assessment*. Routledge.
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021b). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge® Tests (*Research Report No. RR-21-20*). Educational Testing Service. <https://doi.org/10.1002/ets2.12335>
- Schmidgall, J., Lee, S., & Kim, M. (2021a, November). *Developing specific-purposes workplace listening assessments for commerce and science students in Japan: Design considerations and validity evidence*. Paper presented at the 7th annual international conference of the Asian Association for Language Assessment (AALA), Seoul, South Korea.
- Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests—Framework paper (*Research Report No. RR-19-30*). Educational Testing Service. <https://doi.org/10.1002/ets2.12267>
- Schmidgall, J., & Powers, D. E. (2020). TOEIC® Writing test scores as indicators of the functional adequacy of writing in the international workplace: Evaluation by linguistic laypersons. *Assessing Writing*, 46, 1-13. <https://doi.org/10.1016/j.asw.2020.100492>
- Schmidgall, J., & Powers, D. E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC® Speaking test scores from linguistic laypersons. *Language Testing*, 38(2), 302-325. <https://doi.org/10.1177/0265532220941803>
- Yoo, H., Manna, V. F., Monfils, L. F., & Oh, H.-J. (2019). Measuring English language proficiency across subgroups: Using score equity assessment to evaluate test fairness. *Language Testing*, 36(2), 289-309.
- Yoon, S.-Y., Lee, C.M., Houghton, P., Lopez, M., Sakano, J., Loukina, A., ... Madnani, N. (2017). Analyzing item generation with natural language processing tools for the TOEIC® Listening test (*Research Report No. RR-17-52*). Educational Testing Service. <https://doi.org/10.1002/ets2.12183>



## Section 6

# TOEIC Programによる ポジティブな波及効果

オーセンティックなタスクを通じて、  
実践的な英語コミュニケーション能力の習得を促進

TOEIC Program では、受験者の実際の英語コミュニケーション能力について意味ある解釈が可能なスコアを提供するために、オーセンティックな英語コミュニケーションに重点を置いています。テストの内容、タスク、採点の基準は現実世界における英語の使用に即しており、世界中の英語教育および学習にポジティブな波及効果をもたらすよう設計されています。

# TOEIC Programのオーセンティックなタスク ——現実世界のコミュニケーション場面を反映

## 英語コミュニケーション能力とは 英語を“使って”物事を遂行する能力

TOEIC Programは、受験者が国際的な日常生活や職場の場面で効果的にコミュニケーションを図るために、自らの英語の知識や技能をいかに上手く活用できているかを直接測定しています。

コミュニケーション能力(communicative competence)——使用する能力(ability for use)とも呼ばれます——を評価するテストであるTOEIC Programが測定対象とするもの、それは単に英文法や語彙の知識や、英単語の聞き取り、英語音素の発音、英文作成といった言

語を理解し産出するための基礎的なスキルに留まりません。

確かに、これらを含めた言語知識や基礎スキルはコミュニケーション能力の重要な構成要素です。ですが、それら自体が実際のコミュニケーション能力を代表するわけではありません。

「効果的なコミュニケーション」には、言語を使って物事を遂行する行為が伴います。英語学習者は、英語で理解したり意味を伝えたりするために、読む・聞く・話す・書く際に自らの持てる言語知識や基礎スキル

を駆使します。その場面は多岐にわた  
り、文章や聞き手、目的も  
さまざまです。



※1 「ディスコース」とは、一貫性ある話し言葉・書き言葉の文章を形成する単語や文の集合を指します。

※2 「プラグマティクス」とは、特定の状況や文脈、受け手に対応する言語の使われ方を指します。

## コミュニケーションタスクを遂行する中で 言語はオーセンティックに使用される

英語を使う能力について考えるために、コミュニケーションを伴う「タスク」を想像してみましょう。コミュニケーションタスクを遂行するためには、言語能力を用いる必要があります。タスクの例として、以下が挙げられます。

- 重要な情報を知って対応を決められるよう、アナウンスを聞く
- いつ・どこで物事が行われるのかを理解するために、スケジュールを読む
- 食べたいものを注文するために、飲食店で発話する
- 相手に考えやニーズや目的を伝えるために、Eメールを書く

これらのタスクの例からは、コミュニケーションを図るために言語がオーセンティック (authentic: 実際の) な形で——つまり、目的、状況、文脈や、時には特定の相手に応じた形で——使われていることが分かります。TOEIC Tests および TOEIC Bridge Tests では、英語学習者が現実世界でこうしたタスクに取り組む際に必要となる、オーセンティックな形でコミュニケーションを図る能力を測定しています。

### 現実世界の日常生活や職場の場面から 抽出されたタスクを通じて測定

学習者の英語コミュニケーション能力をオーセンティックに測定するために、TOEIC Tests と TOEIC Bridge Tests には、現実世界の日常生活や職場の場

面から抽出されたタスクが用いられています (e.g., Ashmore et al., 2018; Park & Bredlau, 2018)。こうしたタスクは実際のコミュニケーションでの英語の使われ方を反映するために、受験者に対して、文脈や目的に応じて英語を解釈する (読む・聞く)、あるいは発信する (話す・書く) ことを求めます。

例えば、TOEIC Speaking Test のタスクでは、下図のような、オーセンティックな情報が提示される場合があります。図には、新規ビジネスの開始に関する一連のセミナーについての情報が掲載されています。

受験者のタスクは、まずは図に示された情報を読んで理解し、セミナー参加に興味を持って電話をかけてきた話者からの一連の質問に答えることです。

*(Narrator): Hello, I'm calling about a conference on May 27 I saw advertised in the newspaper. It's about starting your own business. I was hoping you could give me some information.*



#### **STARTING AND MANAGING YOUR OWN BUSINESS**

**Date:** May 27

**Location:** Bristol Office Building

**Seminars:** 9:00 A.M. "Financing Your Business," Room 210 — *Martha Ross, Certified Public Accountant*  
11:00 A.M. "How to Promote Your Own Business," Room 312—*Howard Brown, Brown Publishers*

**OR**

11:00 A.M. Planning for Profit, Room 318 — *John Phillips, Phillips Associates*  
1:00 P.M. Lunch\*

2:00 P.M. Sales Techniques Workshop, Room 246 — *Helen King, West Side Consultants*  
4:00 P.M. General Discussion

**Registration Fee:** Individuals, \$95.00  
Members of the Business Information Center, \$75.00

\*Not included in registration fee

質問の例としては以下のようなものがあるでしょう。

Could you tell me what time the conference starts and how long it will last?

セミナーは何時に始まって、何時に終わりますか？

タスクに取り組む受験者は、実際の職場の場面を体験している状態となります。情報を処理し、潜在的な顧客からの質問に正確に答えて、効果的にやり取りをしなければなりません。このタスクでは、語彙や文法の知識のほかに、「読む・聞く」能力、そして最後には「話す」能力が求められます。これはまさに、現実世界において、さまざまな言語知識と技能を駆使して遂行するタスクに取り組んでいる状態と言えます。

■公式サイトから  
サンプル問題を  
ご覧いただけます。



[https://iibc.me/qual\\_persp\\_tp\\_03](https://iibc.me/qual_persp_tp_03)

## 採点基準も 現実世界に即して設定

受験者のパフォーマンスについても、現実世界における期待、あるいは基準に応じて採点されます。例えば、上記のタスクに対するパフォーマンスの場合、採点の対象となるのは、いかに上手に適切な語彙や文法を活用できたか、いかにわかりやすく明瞭に話すことができたか、だけではありません。質問に対してどれだけ正確に答えられたか、返答がどのくらい完成されていたか、どのくらい社会的に適切であったか、という点も採点対象となります。

また、受験者がネイティブスピーカーのように言語

を用いているかという点は採点の対象にはされていません。英語に堪能な話者（ネイティブもしくはノンネイティブ）にとって、返答が理解しやすいか、コミュニケーションが効果的かどうか、という点に基づき採点されます。

## オーセンティシティを重視する理由 (1) ——コミュニケーション能力を 直接測定するテストに“不可欠”なため

ではなぜ、TOEIC Programはオーセンティックな英語コミュニケーションをこれほど重視するのでしょうか？ そこにはいくつかの重要な理由があります。

第一に、テスト利用者（企業・団体・教育関係者など）が得たい情報は、受験者が英語を使って「できること」と、それを「いかに適切にできるのか」であり、受験者が英語の言語システムや構造に関してどのような知識を持っているかという点だけではない、ということです。

テスト利用者は英語能力テストのスコアを、受験者のコミュニケーション能力——英語を「使って」コミュニケーションを図る能力——の指標として捉えています。各4技能においてオーセンティックなコミュニケーションタスクを出題しない英語テストは、最良の場合でも、英語を使ってタスクを達成する能力を、間接的かつ不完全な形で測定するに留まります。

倫理と責任を果たす言語テストとして、受験者のコミュニケーション能力に対する意味ある解釈を裏付けるためには、現実世界での言語使用を反映したオーセンティックなコミュニケーションタスクを出題しなければならぬのです (Norris & East, 2021)。

## オーセンティシティを重視する理由 (2)

### ——教育・学習にポジティブな 影響を与えるため

TOEIC Programがオーセンティックな英語コミュニケーションを重視する第二の重要な理由は、英語学習者や教員に有意義な学習目標を提供し、信頼の置けるテストによって学習進捗を測定していただくためです。

TOEIC Programで好成績を収めるために言語知識と技能を伸ばすことは、同時に英語コミュニケーション能力の習得という有意義な目標に取り組むことでもあります。結果として、学習者はオーセンティックなコミュニケーションタスクを達成できるようになるでしょう——これこそが、言語教育における望ましい成果ではないでしょうか。

このTOEIC Programと、英語教育・学習の関係については、次項でさらに詳しく見ていきます。

## 適切なテストの選択がポジティブな波及効果につながる

### より適切で高品質なテストの選択がより良い成果につながる

TOEIC Programの主要な目的のひとつが、質の高い意思決定の裏付けとなる言語テストの提供です。下図のように、言語テストは言語技能に関する情報を提供しています。こうした情報は個人に関する重要な意思決定に用いられることがあります。例えば組織が候補者の「採用・昇進」の判断材料としたり、研修機関が生徒の「プレースメント」の決定に用いたりするケースが挙げられます。

テストスコアに基づく意思決定はポジティブあるいはネガティブな結果（成果）をもたらします。例えば、組織が言語テストのスコアに基づいて、言語

技能が求められるポジションに社員の採用・昇進を行い、その社員が新たに就いたポジションで言語技能を発揮できた場合には、その意思決定はポジティブな（もしくは有益な）結果をもたらしたと言えるでしょう。その社員が必要な言語技能を持たず、それが仕事の遂行能力に影響を及ぼす場合には、その意思決定はネガティブな成果を招いたと言えます。

また、例えば研修機関が言語テストのスコアに基づいて生徒（受験者）を正確に分類し、適切なコースに振り分けることができた場合には、その意思決定はポジティブな成果をもたらしたと言えるでしょう。



プレースメントにおける正確な意思決定は、効果的な教育・学習に寄与する可能性があります。また、当然ながら、適切なコースに生徒を振り分けることができなかつた場合には、教育・学習の効果が下がってしまう可能性があります。こうした影響はテストが教育・学習にもたらす「波及効果」(washback)と呼ばれています。この「波及効果」については後ほど詳しく見ていきます。

言語テストの選択は意思決定がもたらす成果に影響を及ぼし得るため、テスト利用者はいくつかの重要な点に留意する必要があります。まず何よりも、言語テストが提供する言語技能に関する情報が、意思決定の材料となる言語技能・知識・能力に関連した情報である必要があります。多くの場合、テスト利用者は英語コミュニケーション能力に関心がありますが、こうした能力は(限定的・間接的なタスクによって英語コミュニケーション能力を予測するテストと比べて)、「聞く・読む・話す・書く」技能を直接測定するテストによって最も良く証明されます。

次に、スコアの品質が意思決定の重要性に対して十分に高い必要があります。2つの異なるテストが同じ技能を測定すると主張している場合——例えば、2つのテストがいずれも「英語スピーキング能力」を測定する場合など——であっても、対象技能の定義や測定方法はそれぞれのテストによって必然的に異なります。また、測定品質(信頼性・妥当性・公平性)の水準も異なるでしょう(測定品質について詳しくはSection 3「ETSの品質への取り組み」をご参照ください)。究極的には、質の高い意思決定につながるべ

ストケースシナリオとは、より直接的なアプローチで最も関連性のある技能を高品質に測定するテストを識別することなのです。

### テストによる教育・学習への「波及効果」

教育的観点から、言語テストの選択が言語の教育・学習に影響を及ぼす可能性が指摘されています(e.g., Hsieh, 2017)。とりわけ、採用や人材選抜などの極めて重要な意思決定を伴う場合、テストは教育・学習に大きな影響をもたらし得ることが、リサーチにより分かっています。

教員や学習者は、重要なテストの準備のために多くの時間を費やす場合がありますが、その影響として、彼らはそのテストが評価対象としている言語技能の向上に注力することになります。上記でもすでに触れていますが、重要な意思決定にテストを用いることによるこうした影響も、テストが教育・学習に与える「波及効果」(washback)と呼ばれます。

### 波及効果は「ポジティブ」にも「ネガティブ」にも作用する

ここで重要なのが、波及効果は「ポジティブ」にも「ネガティブ」にも作用するという点です。言語テストがオーセンティックで適切なタスク——現実世界においてオーセンティックにコミュニケーションを図る能力を測定するよう設計された現実的なタスク——を出題し、適切な言語技能に焦点を当てている場合は、教育・学習にポジティブな影響をもたらす可能性があります。一方で、言語テストがオーセンティックでない不適切なタスクを出題している場合には、現実世界での英語の使われ方を反映しない

タスクに注力するよう教員や学習者を促すことになり、教育・学習にネガティブな影響をもたらす可能性があります (Norris, 2018の議論も参照)。

また、言語テストが言語知識や技能の非常に限られた側面に焦点を当てている場合にも、やはり教育・学習にネガティブな影響をもたらす可能性があります。教員や学習者が幅広いコミュニケーション能力に目を向ける代わりに、言語能力の限られた側面に注力するように促されてしまうためです。

では、ここで波及効果について、2つの視点——(a)タスクのオーセンティシティ；(b)測定対象とする技能——から見ていきましょう。

## 波及効果に影響を与える要素：

### (a) タスクのオーセンティシティ

まず、なぜタスクのオーセンティシティが重要なのでしょうか？ それは、言語テストがオーセンティックなコミュニケーションタスクをどの程度含むかによって、波及効果はポジティブなものにも、ネガティブなものにもなり得るためです。

例えば、リスニングとリーディングのテスト問題のほとんどが文法や語彙の知識を測るものだった場合、学習者は文法や語彙の暗記に注力した方が良いと考えるかもしれません。文法や語彙の知識は言語能力の重要な基礎を成すものですが、現実世界でリスニング力やリーディング力を用いて言語を理解するためには、より幅広い知識や技能が求められます。

オーセンティックなテストタスク——現実世界のコミュニケーション場面を効果的にシミュレートし

たタスク——は教員や学習者を、現実世界のコミュニケーションで求められる技能に注力するよう促します。TOEIC TestsおよびTOEIC Bridge Testsはこの根本方針を念頭に、ポジティブな波及効果を促進するよう設計されています (Schmidgall et al., 2021)。

例えば、右記のTOEIC Listening & Reading Testのリーディングセクションのサンプル問題をご覧ください。

このタスクは職場の場面に設定されており、受験者は限られた時間を使って複数の文書を参照し、設問に解答します。3つの文書に目を通しながら必要な情報を集めていけば、解答にたどり着ける問題ですが、こうした情報処理——限られた時間内でさまざまな資料から目的の情報を見つける作業——はまさに現実生活の場面、とりわけ職場でよく見られるタスクです。タスクに設定されたコミュニケーションのゴールを達成するには、必要な情報を捉えて一定時間内に処理する能力が不可欠となりますが、このようにオーセンティックなタスクをテストに出題することにより、ポジティブな波及効果を期待することができます。テストの準備のための学習において、こうしたタスクを練習することで、実社会で役立つスキル——英文を効率的に読んで情報を迅速に処理する方法——を学ぶことができるためです。

■本サンプル問題の全ての設問は、以下サイトでご確認いただけます。



[https://iibc.me/qual\\_persp\\_tp\\_04](https://iibc.me/qual_persp_tp_04)

**Sparky Paints, Inc.** 

Sparky Paints, Inc., makes it easy to select the right colors for your home. Browse through hundreds of colors on our Web site, [www.sparkypaints.com](http://www.sparkypaints.com). Select your top colors, and we'll send free samples right to your door. Our color samples are three times larger than typical samples found in home-improvement stores and come with self-adhesive backing, allowing you to adhere them to your walls so you can easily see how colors will coordinate in your home. When you're ready to begin painting, simply select your chosen colors online, and we'll ship the paint of your choice to arrive at your home within 3-5 business days, or within 2 business days for an additional expedited shipping fee.

\*Actual colors may differ slightly from what appears on your monitor. For this reason, we recommend ordering several samples in similar shades.

<http://www.sparkypaints.com/shoppingcart>

**Sparky Paints, Inc.**  

Order Summary #3397		Customer: Arun Phan	
Item	Size	Quantity	Price
Caspian Blue SP 237	n/a	1	\$0.00
Deep Sea Blue SP 298	n/a	1	\$0.00
Stormy Blue SP 722	n/a	1	\$0.00
Misty Gray SP 944	Gallon	2	\$50.00
Tax (8 percent)			\$4.00
Expedited shipping			\$18.99
<b>Total</b>			<b>\$72.99</b>

[Proceed to Checkout](#)

<b>From:</b>	Arun Phan <arun.phan@tnet.com>
<b>To:</b>	Customer Support <support@sparkypaints.com>
<b>Date:</b>	March 12
<b>Subject:</b>	Order #3397

Hello,

Thanks for sending my order #3397—it arrived this morning. Unfortunately, the paint was not the one I had asked for. I had selected color SP 944 but received SP 945 (Ocean Waves). They appear right next to each other on your Web site, so the two may have been confused at your end. Could you send me the correct paint, along with additional samples that are close in color to SP 722? That sample worked well in my house; the others looked too green on my walls.

Thank you,  
Arun Phan

199. Which color does Mr. Phan indicate that he likes?
- (A) Caspian Blue
  - (B) Deep Sea Blue
  - (C) Stormy Blue
  - (D) Misty Gray

解答: (C) Stormy Blue

## 波及効果に影響を与える要素：

### (b) 測定対象とする技能

次に、どういった技能に焦点を当て、測定の対象とするかを検討することが、なぜ重要なのかについても見ていきましょう。

多くの場合、テスト利用者はコミュニケーション能力に関心を持っていますが、ここで波及効果の前提である「教育・学習の目標に対して適切なテストを使用する」という点を踏まえて考えてみましょう。

例えば、学習者が総合的な英語コミュニケーション能力の習得を目指しているとします(右記のCASE 1-1をご参照ください)。この場合、英語の4技能(読む・聞く・話す・書く)の全てを測定することが必要となりますが、この「4技能受験」が受信・発信の能力をバランスよく教育・学習することにつながります。つまり、このケースでは、テストによる教育・学習への波及効果がポジティブに作用している、と言えるでしょう。

また、テストの選択は、教員や学習者に対して、どのようなコミュニケーション能力のモデルに価値を置くのかを伝えるメッセージでもあります。CASE 1-1に示した通り、「4技能モデル」は、総合的な英語コミュニケーション能力(4技能)の習得を目指す教育・学習に対して、よりポジティブな波及効果を与える可能性があります。テスト受験に向けて準備をする教員や学習者が、言語能力のあらゆる側面に目を向けるように促されるためです。長期的に見れば、教員や学習者がコミュニケーションを効果的に図るために必要な幅広い能力に注力するようになり、英語能力の習得という点で、より高い成果を生み出すことにつながるでしょう。

さらに、右記のCASE 1-2のように、リーディングとリスニングの技能を最優先で習得したいケースでは、リーディング力とリスニング力のみを(オーセンティックなタスクを用いたテストによって)測定した場合においても、これら2技能の教育・学習につながるため、ポジティブな波及効果が期待できます。

一方で、4技能は互いに関連し合っているため、究極的には4技能受験はリスニングやリーディングの技能の向上にも効果的です。この点についてはSection 7「[4技能]を測る TOEIC Program」をご参照ください。

では、p. 64の図解CASE 2-1のように、総合的なコミュニケーション能力(4技能)の習得に関心がありながらも、リーディングとリスニングの技能の測定結果のみを用いた場合に起こり得ることを考えてみましょう。

リーディングとリスニングの技能に関する情報のみを用いると、総合的な言語能力に関する情報が限られ、コミュニケーション能力の全体像が見えにくくなります。その結果、言語評価の質が下がること、意思決定の誤りが生じやすくなり、ポジティブな成果を得る機会が減ってしまう可能性があります。さらに、波及効果として、教員や学習者に対して、コミュニケーション能力の「4技能モデル」に価値を置いていないというメッセージを送ることにもなってしまいます。

また、p. 64の図解CASE 2-2のように、リーディングとリスニングの技能を習得したいと考えながらも、語彙や文法のみを測定した場合においても、波及効果はネガティブに作用する可能性があるでしょう。

## CASE 1

教育・学習の目標に  
対して適切なテスト  
を使用

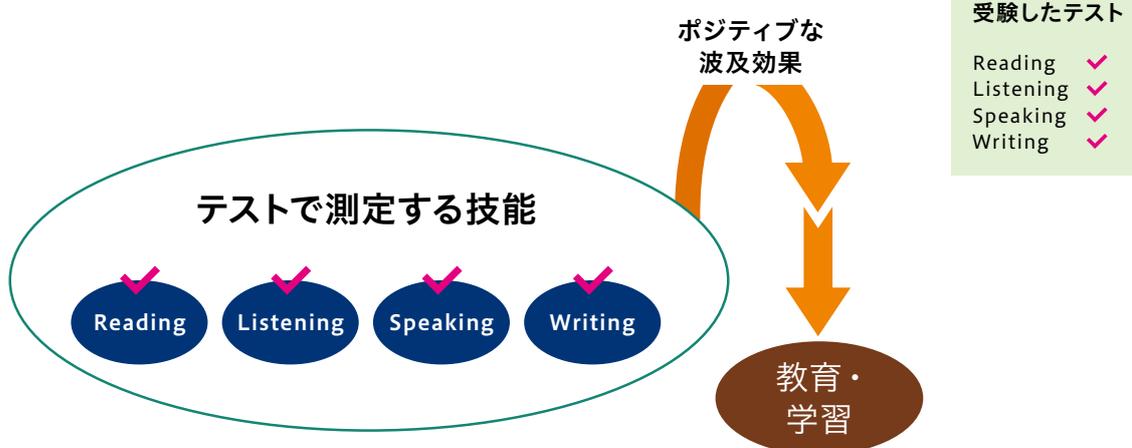
ポジティブな  
ウォッシュバック効果

適切な教育・学習が  
促される

## CASE 1-1



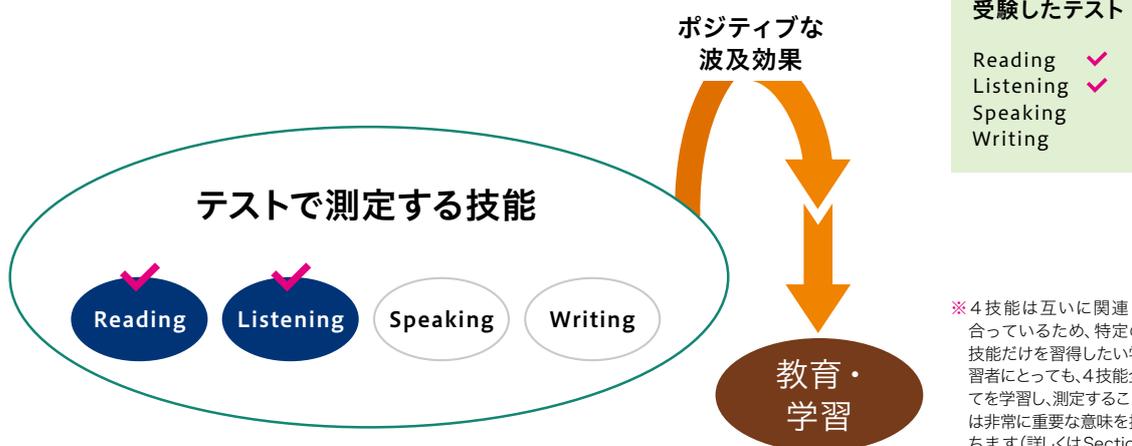
学習開始時から、4技能の習得を目指す受験者  
→4技能全てのテストを受験



## CASE 1-2



限られた期間で「読む・聞く」技能を優先したい受験者※  
→2技能「読む・聞く」のテストを受験



※ 4技能は互いに関連し合っているため、特定の技能だけを習得したい学習者にとっても、4技能全てを学習し、測定することは非常に重要な意味を持ちます(詳しくはSection 7「4技能」を測るTOEIC Program)をご参照ください。

## CASE 2

教育・学習の目標に対して適切でないテストを使用

ネガティブな  
ウォッシュバック効果

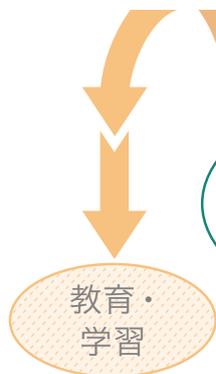
適切な教育・学習に  
つながらない可能性

### CASE 2-1



学習開始時から、**4技能の習得を目指す**受験者  
→一部の技能(例:「読む・聞く」のみ)のテストを受験

ネガティブな  
波及効果



テストで測定する技能



受験したテスト

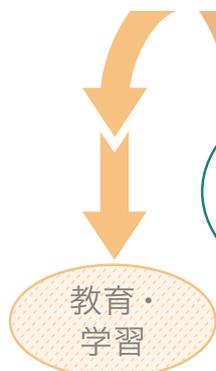
Reading ✓  
Listening ✓  
Speaking  
Writing

### CASE 2-2



限られた期間で「読む・聞く」技能を優先したい受験者  
→語彙と文法の知識のみを測るテストを受験

ネガティブな  
波及効果



テストで測定する技能



受験したテスト

Reading  
Listening  
Speaking  
Writing

〈言語知識と基礎スキル〉

語彙 ✓  
文法 ✓  
発音  
ディスコース  
プラグマティクス、etc.

※1 「ディスコース」とは、一貫性ある話し言葉・書き言葉の文章を形成する単語や文の集合を指します。  
※2 「プラグマティクス」とは、特定の状況や文脈、受け手に対応する言語の使われ方を指します。

ここまでお伝えした通り、(a) タスクがいかにオーセンティックであるか；(b) どういった技能に焦点を当て、測定の対象とするか、という両視点から、適切なテストを使用することが、教育・学習へのポジティブな波及効果につながります。TOEIC Tests および TOEIC Bridge Tests は、4 技能をオーセン

ティックなタスクを通じて測定することで、教員や学習者が現実世界のコミュニケーションにおいて求められる技能に注力するよう促しています。そして、これこそが TOEIC Tests および TOEIC Bridge Tests の設計の根本方針でもあります。

## References

- Ashmore, E., Duke, T., & Sakano, J. (2018). Background and goals of the TOEIC® Listening and Reading update project. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 3.1-3.8). Educational Testing Service.
- Hsieh, C.-N. (2017). The case of Taiwan: Perceptions of college students about the use of the TOEIC® tests as a condition of graduation (*Research Report No. RR-17-45*). Educational Testing Service. <https://doi.org/10.1002/ets2.12179>
- Norris, J. M. (2018). Task-based language assessment: Aligning designs with intended uses and consequences. *JLTA Journal*, 21, 3-20.
- Norris, J. M., & East, M. (2021). Task-based language assessment. In M. Ahmadian & M. Long (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 507-528). Cambridge University Press.
- Park, E., & Bredlau, E. (2018). Expanding the question formats of the TOEIC® Speaking test. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 2.1-2.8). Educational Testing Service.
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge® Tests (*Research Report No. RR-21-20*). Educational Testing Service. <https://doi.org/10.1002/ets2.12335>



## Section 7

# 「4技能」を測る TOEIC Program

## 総合的な英語コミュニケーション能力を 高精度に評価

TOEIC Programは英語コミュニケーションの4技能「聞く・読む・話す・書く」を評価し、各技能のスコアを通じて、受験者の能力を解釈するための有効な情報を提供しています。本セクションでは、「総合的な英語コミュニケーション能力」——英語のさまざまな側面における全般的な能力——の習得という観点から、TOEIC Programが4技能の測定を推奨する理由や背景について解説します。

## TOEIC Programと4技能

——「読む・聞く・話す・書く」はつながっている

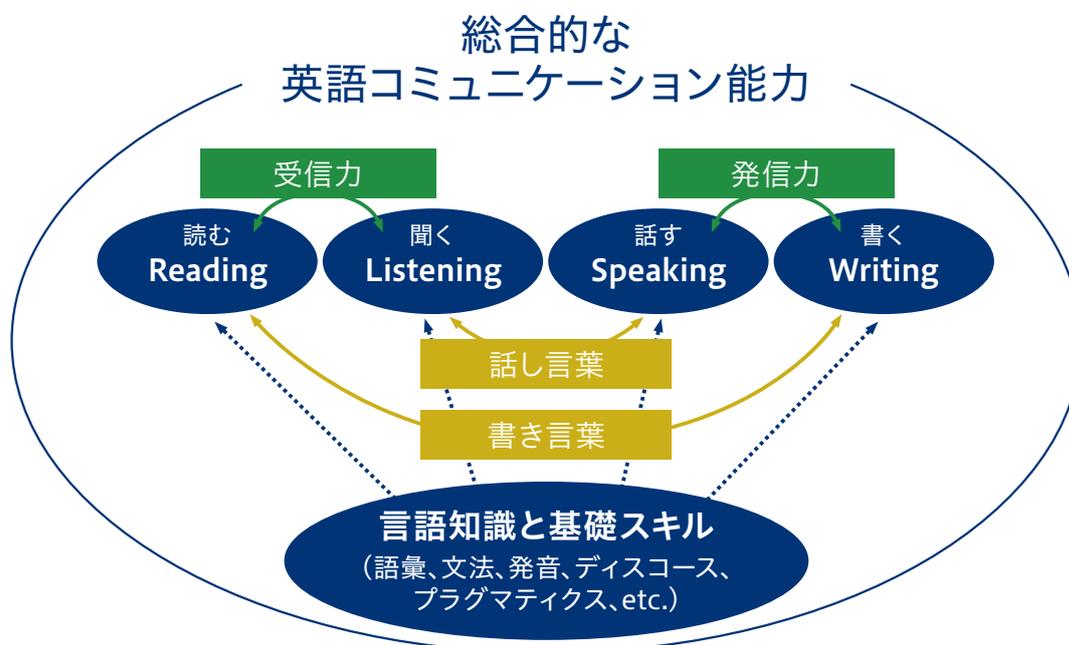
### 「4技能」アプローチとは？

第二言語能力は多数の要素で構成される——これは第二言語習得と言語テストに携わる研究者たちの大方の共通理解となっています(Bachman & Palmer, 2010; Canale & Swain, 1980)。つまり、私たちが第二言語能力の総合力として捉えているものは、実際には、相互に関連する知識や技能が集まったものなのです。

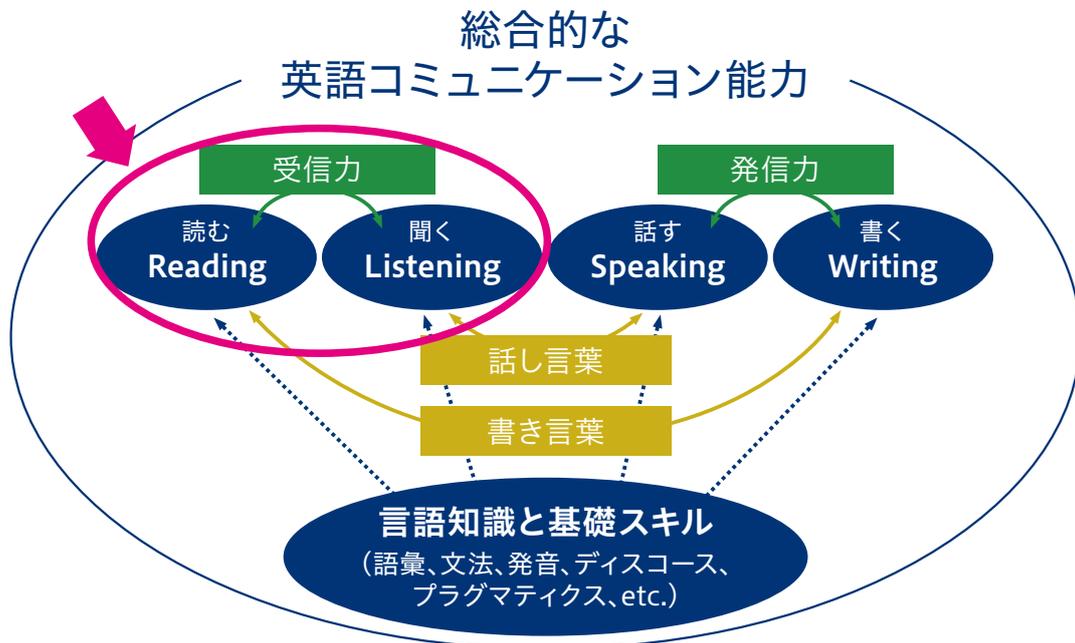
コミュニケーション能力を構成する多様な言語使用を

簡潔に整理するために、研究者たちが見出した最も有意義な手法のひとつが「4技能」アプローチです。4技能アプローチでは、「読む・聞く・話す・書く」という4技能を、多様な言語使用のモダリティ(コミュニケーションのモード)においてコミュニケーションを図る能力を構成する必須の言語技能としています。

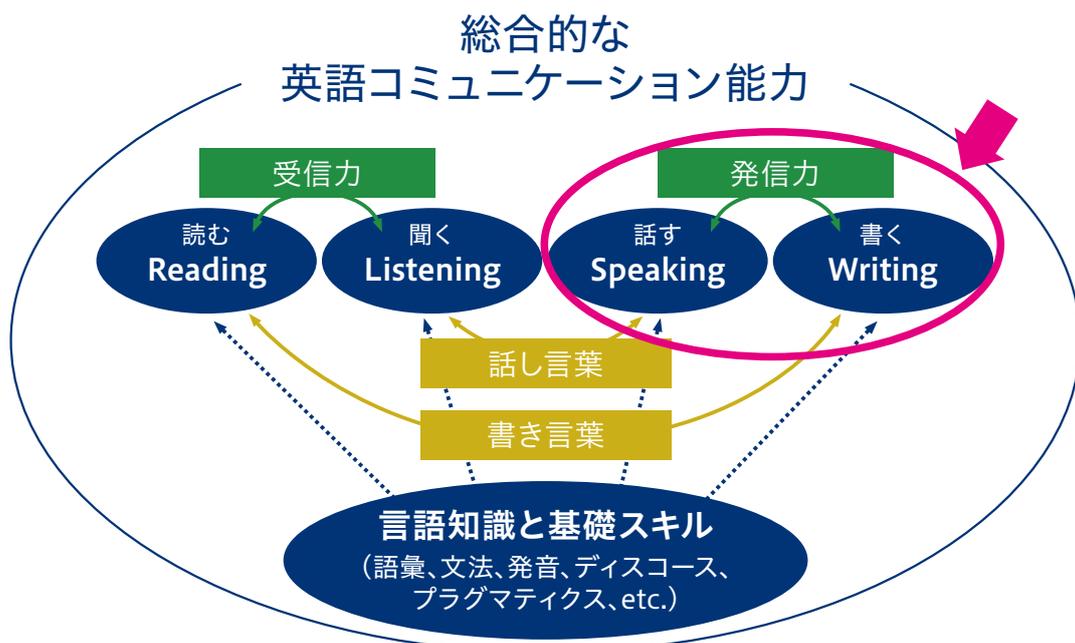
下図のように、4技能はつながっているが、それぞれが別々の技能でもあります。



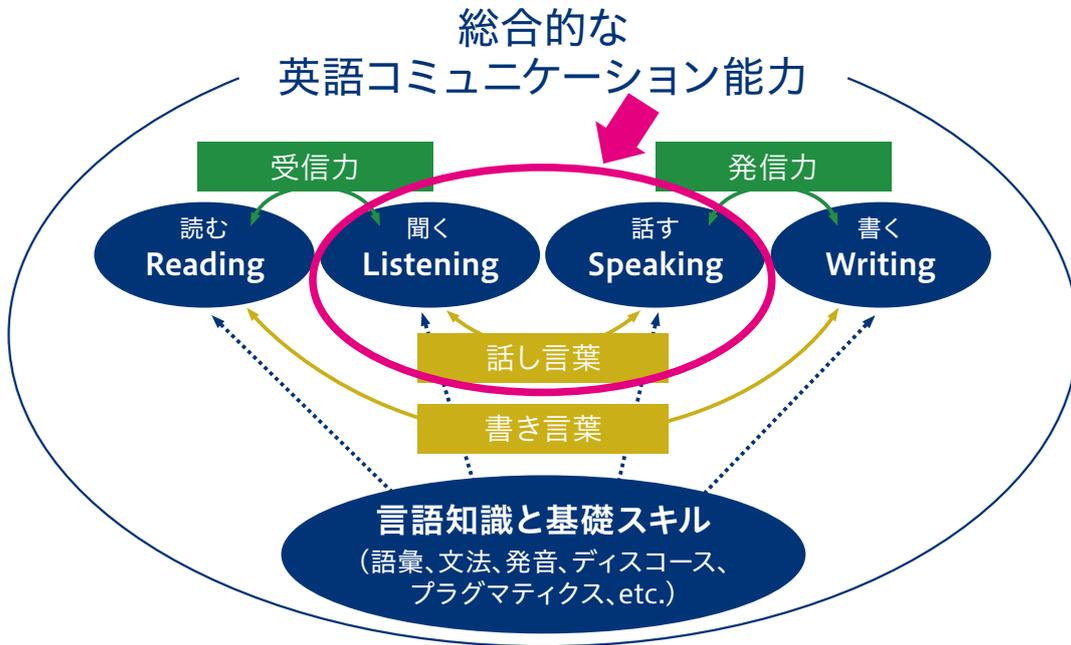
「読む・聞く」は受信力、つまり目や耳を通じてインプットした言語から意味を組み立てる能力です。



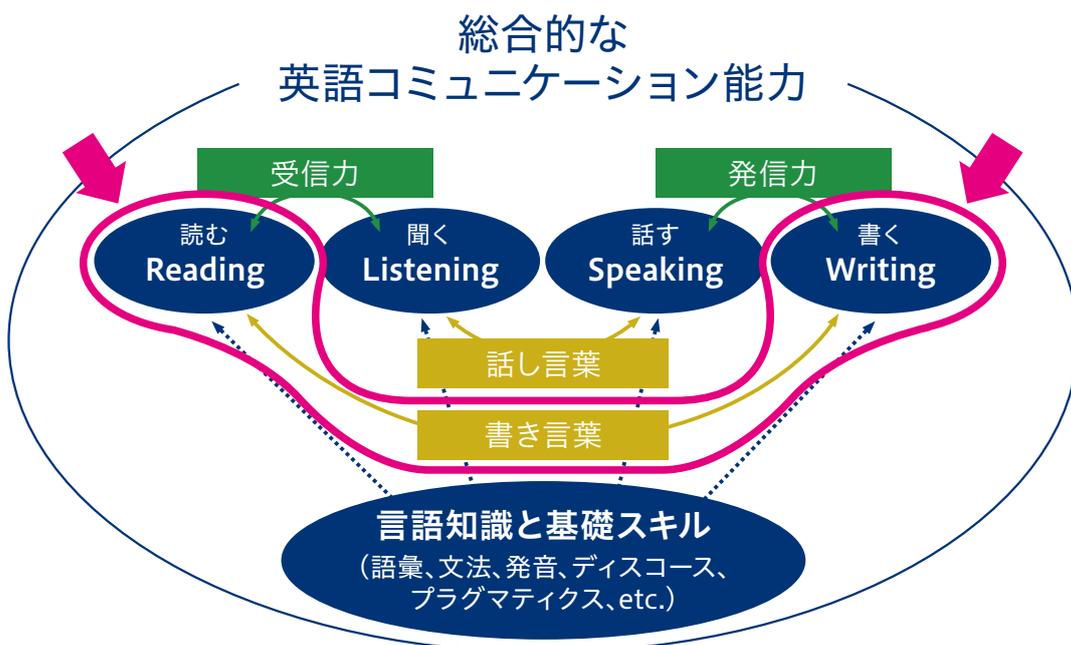
「話す・書く」は発信力、つまり受け手に向けてメッセージを作り出す能力です。



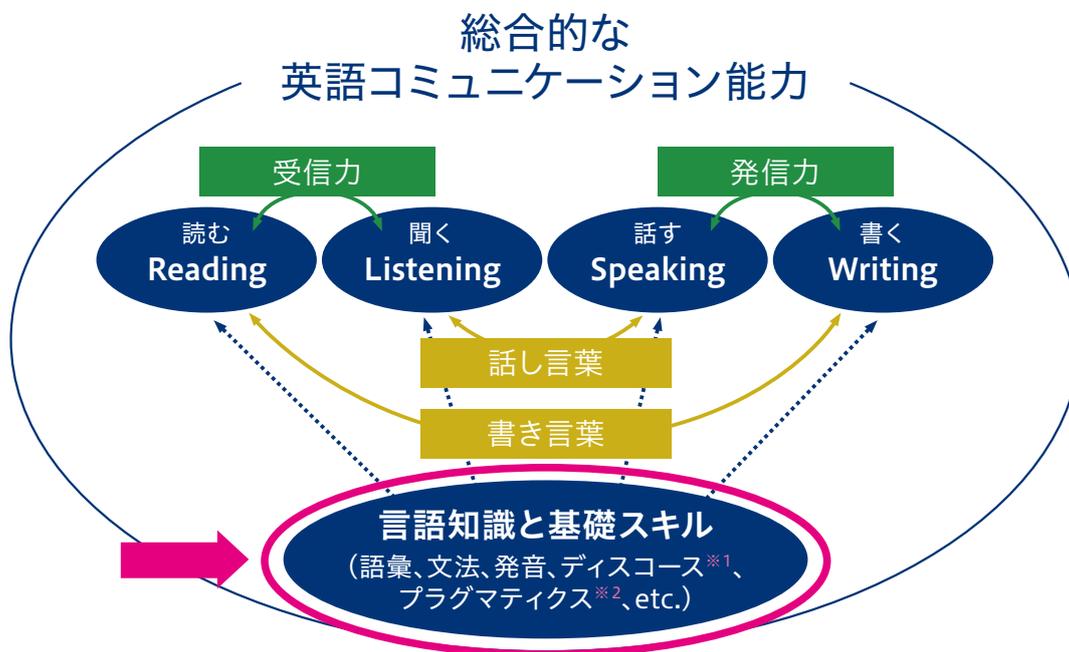
また、「聞く・話す」はいずれも「話し言葉」という点で、つながっています。



「読む・書く」はいずれも「書き言葉」という点で、つながっています。



さらに、コミュニケーションの4技能（読む・聞く・話す・書く）は、土台となる言語知識と基礎スキル（語彙、文法、発音、ディスコース※1、プラグマティクス※2など）を共有しています。



※1 「ディスコース」とは、一貫性ある話し言葉・書き言葉の文章を形成する単語や文の集合を指します。  
 ※2 「プラグマティクス」とは、特定の状況や文脈、受け手に対応する言語の使われ方を指します。

### コミュニケーションタスクの遂行には 言語知識と基礎スキルに加えて コミュニケーションの技能が必要

この土台となる「言語知識と基礎スキル」は、個々の「コミュニケーション技能」（読む・聞く・話す・書く）と共に活用されます。つまり、「言語知識と基礎スキル」と「コミュニケーション技能」の組み合わせが、コミュニケーションタスクのパフォーマンスを左右します。現実世界でもほとんどの場合において、コミュニケーションタスクを遂行するには、言語知識、基礎スキル、コミュニケーション技能を組み合わせる必要があります。

テストによって特定のコミュニケーション技能を評価する際、その測定対象となるのは、「言語知識と基礎スキ

ル」と「コミュニケーション技能」を組み合わせ、コミュニケーションタスクを遂行・達成する能力です。例えば、TOEIC Speaking Testによって学習者を評価する際、テスト利用者が知りたいのは、その受験者がスピーキングタスクを遂行する中で、(a) 言語知識と基礎スキル (e.g. 語彙や文法の知識、わかりやすく発音する能力、プラグマティクスやディスコースの観点から適切な文や文章を組み立てる能力) を活用しながら、(b) 言語コミュニケーション技能 (e.g. 描写する能力、質問に応える能力、意見を述べる能力) を発揮できているか、という点なのです。

先の図から推察される通り、TOEIC Programの各テストは、受験者の「言語知識と基礎スキル」と「測定対象の技能を用いて効果的にコミュニケーションを図る能力」の両者を捉えています。

## TOEIC Programは 4技能を個別に評価

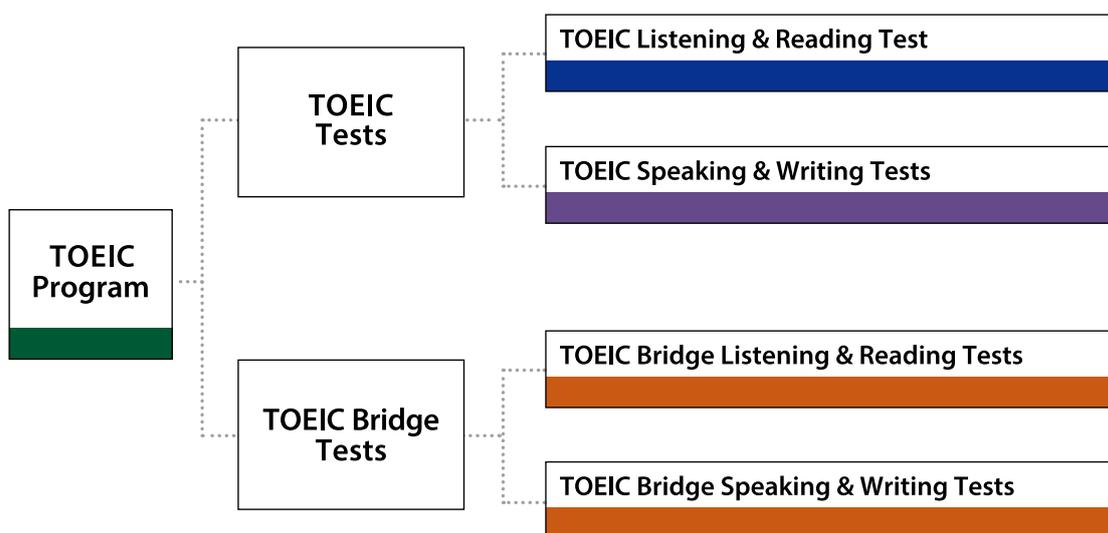
TOEIC Programは4技能全てを個別に測定し、技能ごとにスコアを提供しています。スコアのレポートも技能別に提供しており、特定の技能に絞って測定したいという利用者のニーズに応えています。そうしたニーズの例としては、雇用者が、4技能全てではなくその一部を求められる仕事や職種において人材を選抜する際に、対象とする仕事に必要な技能のスコアを主に利用するといったケースが挙げられます。また、学習者や教員は、技能別のテストとスコアによって、学習者の英語コミュニケーション能力の強みや伸ばすべき点を把握することができます。

下図の通り、TOEIC Tests、TOEIC Bridge Testsのいずれをご利用いただいても、4技能を個別に測るこ

とができます。

職場や英語学習の状況によっては、一部の技能を個別に測定することが重要な意味を持つこともあります。一方で、多くの場合、言語テストの利用者が第一に求めているのは、総合的なコミュニケーション能力を評価することではないでしょうか。

4技能受験は総合的なコミュニケーション能力を最適な形で評価し、その評価結果の有効性は一部の技能のみを測定する場合に比べてより高いものとなります。英語の4技能はつながっていながら、それぞれが別々の技能でもあります。そのため、全4技能を個別に測定することにより、英語コミュニケーション能力をより詳細に把握しながら総合的に評価し、土台となる言語知識と基礎スキルをより高い精度で評価することが可能となります。



## TOEIC Programと総合的な英語コミュニケーション能力の構築

### 4技能に偏りのない学習者は より高いパフォーマンスを発揮する傾向に

目的やニーズによっては、4技能の一部を評価することが現実的なケースもあるでしょう。一方で、学習者や教員にとっては、総合的な英語コミュニケーション能力の構築に注力することも重要です。

概して、4技能をバランスよく習得した学習者は、英語力に偏りのある学習者（例えば、「受信力」の方が「発信力」よりも高いなど）に比べて、英語を介したコミュニケーション環境で高いパフォーマンスを発揮しています。そのため、例えば学習者の総合的な英語コミュニケーション能力を把握するためにTOEIC Listening & Reading（以下、TOEIC L&R）のみを用いる場合には、4技能受験の結果を用いるほどの有効性は期待できない可能性があります。TOEIC L&Rで高スコアを取得している受験者であっても、発信力はそれほど高くないというケースがあり得るためです。そのため、学習者の総合的な英語コミュニケーション能力の把握を目的とする場合には、TOEIC Programの各テストによって提供されるような「全4技能の包括的な評価」を用いることを強く推奨します。

TOEIC Programが4技能の測定による総合的な英語コミュニケーション能力の評価を推進する背景には、いくつかの重要な理由があります。

### 現実世界のコミュニケーションタスクでは 同時に複数の技能が求められる

ほとんどの職場環境において不可欠とされ、多くのTOEIC Programの利用者が最も重視しているもの、それは総合的な英語コミュニケーション能力であり、個別の技能自体ではありません。

コミュニケーション能力とは基本的に、モダリティ横断的に言語を統合して用いることであり、現実世界のコミュニケーションタスクでは概して複数の技能を同時に活用します。例えば、テスト利用者が、ビジネス会議で円滑に英語プレゼンテーションを遂行できる人材を選抜する場合、候補者のスピーキング力を測定することは必要かつ適切でしょう。ですが、こうしたスピーキングタスクはたいていの場合、十分なスピーキング力のみで円滑に対応し切れるものではありません。関連する背景情報を読んで理解できるリーディング力や、聴衆からの質問を理解できるリスニング力も求められます。

そのため、4技能のうちのひとつだけを測定、あるいはいくつかの技能だけを測定しては、受験者のコミュニケーション能力の全体像を捉えきれない可能性があるのです。

このことは、初級者も含めた全レベルの受験者に当てはまります。例えば、TOEIC Bridge Testsに出題されるコミュニケーションタスクの多くは、初級者のパフォーマンスを効果的に引き出し、彼らが持てる能力を存分に示すことができるよう設計されています。

初級者が「書ける・話せること」は限られているかもしれませんが、テスト利用者が受験者の総合的な英語コミュニケーション能力に基づいて意思決定を行う際には、受信力（聞く・読む）だけでなく、発信力も含めて測定することが重要です。

TOEIC Programの各テストは個々の技能を詳細に測定し、受験者の能力に関する有用な情報を提供しています。同時に、4技能の測定を通じ、英語を用いて効果的にコミュニケーションを図る能力をより完全な形で評価するという、多くのTOEIC Program利用者の目的に応えています。

### 関連技能の測定結果を活用することで より精度の高い評価が可能に

さらには、測定対象の技能だけでなく、その関連分野の技能についても測定することで、受験者の技能をより正確に評価できることが、ETSの研究によって分かっています。

例えば、ある候補者のスピーキング力を評価するというケースでも、その関連技能（e.g., リスニング、リーディング、ライティング）に関する情報も組み合わせることで、スピーキング力をより精度高く評価することができます。4技能は互いに絡み合い、土台となる言語知識や基礎スキルは4技能横断的に用いられるため、関連技能の測定結果を共に活用することが、対象とする技能をより詳しく、微細かつ正確に評価することにつながります。

例えば、ある受験者が4技能全てを測定した結果、リーディングとライティングのスコアが比較的高い一方で、リスニングとスピーキングのスコアが低めだった場合、重点的に学習する必要がある分野は「話し言葉」に関するスキルであると解釈することができます。

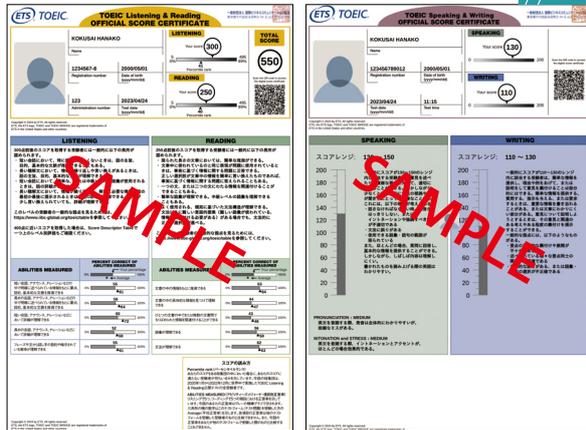
また、リスニングとリーディングで高スコアを取得しながらも、スピーキングとライティングのスコアが低めだった場合には、「発信力」を伸ばす練習を増やす必要があると解釈することができます。

さらには、4技能は互いに絡み合い、土台となる言語知識と基礎スキルを共有しているため、4技能別のスコアレンジ別評価を解釈する際に、ある特定の知識（例えば「語彙」など）に焦点を当ててみることで、その知識の全体像や学習の改善点を把握することができます。

例えば、右記のCASEのように、ある受験者がTOEIC Testsのスコアレンジ別評価をチェックして、リーディングの該当箇所では「中級レベルの語彙が理解できる」とあり、スピーキングでは「使用できる語彙・語句の範囲が限られている」と記述されていたとします。この場合、その受験者は「なるほど、自分には中級レベルの語彙知識があるけれど、多様な語彙を使って話す練習が必要なのだな」と考えるかもしれません。つまり、複数の技能を測定することで、日常タスクの遂行可能性や重点的に伸ばすべき点を予測するうえで、相互補完の効果が期待できるのです。

# CASE

## TOEIC Testsの公式認定証



TOEIC Testsで4技能を測定した結果が出そろった。こうして4技能を一覧してみると、自分のコミュニケーション能力の全体像がよく分かるし、改善ポイントも細かくチェックできる。

学習計画を立てるためには、スコアをもっと詳細に理解したい。

よし、自分のスコアレンジに該当する評価一覧も参照・比較してみよう。

まずはリーディングの語彙について見てみよう。長所は「中級レベルの語彙が理解できる。文脈中の難しい語彙、よく使用される単語の例外的な意味、慣用的な使い方が理解できることもある」とある。弱点には改善ポイントが示されているな。

## TOEIC Testsのスコアレンジ別評価一覧(一部抜粋)

### リスニング

スコア	長所	弱点
370~275	<ul style="list-style-type: none"> <li>一般的に以下の長所が認められます。</li> <li>短い会話において、特に語彙が難しくないときは、話の主旨、目的、基本的な文脈が推測できることもある。</li> <li>長い聴解文において、情報の繰り返しや言い換えがあるときは、話の主旨、目的、基本的な文脈が理解できる。</li> <li>短い会話において、簡単な、または中級レベルの語彙が使用されるときは、話の詳細が理解できる。</li> <li>長い聴解文において、情報が繰り返され、解答に必要な情報が話の最初か最後に提示されるときは、話の詳細が理解できる。情報が少し言い換えられていても、詳細が理解できる。</li> </ul>	<ul style="list-style-type: none"> <li>一般的に以下の弱点が認められます。</li> <li>短い会話において、応答が間接的だったり、簡単に予測できないとき、もしくは語彙が難しく、話の主旨、目的、基本的な文脈の理解が困難である。</li> <li>長い聴解文において、広い範囲にわたって情報を関連付ける必要があるときは、もしくは難しい語彙が使用されるときは、話の主旨、目的、基本的な文脈が理解できない。</li> <li>短い会話において、構文が複雑なときや、難しい語彙が使われている場合は、話の詳細が理解できない。否定構文が使用されるときは、詳細が理解できないことが多い。</li> <li>長い聴解文において、広い範囲にわたって情報を関連付ける必要があるとき、もしくは情報が繰り返されないときは、話の詳細が理解できない。言い換えられた情報、または難しい文法的な構造はほとんど理解できない。</li> </ul>

### スピーキング

スコア	長所	弱点
110~120	<ul style="list-style-type: none"> <li>一般的にスコアが110-120のレンジ内に該当する受験者は、ある程度、意見を述べる、または複雑な要求に応えることができる。ただし、応答には以下のような問題がみられる。</li> <li>言葉が不正確、あいまい、または同じ言葉を繰り返している</li> <li>聞き手の立場や状況をほとんど、またはまったく意識していない</li> <li>間が長く、躊躇することが多い</li> <li>考えを表現すること、またいくつかの考えを関連づけて表現することに限界がある</li> <li>使用できる語彙・語句の範囲が限られている</li> </ul> <p>また、ほとんどの場合、質問に回答し、基本的な情報を提供することができる。しかしながら、しばしば内容は理解しにくい。</p> <p>書かれたものを読み上げる際の英語は概ねわかりやすいが、自らが考えて話をするときは、発音、イントネーション、強調すべき部分に時々誤りがある。</p>	

### リーディング

スコア	長所	弱点
420~325	<ul style="list-style-type: none"> <li>一般的に以下の長所が認められます。</li> <li>文章の主旨や目的が推測できる。詳細が推測できる。</li> <li>意味を読み取ることができる。言い換えがあっても、事実に基づく情報が理解できる。</li> <li>文章に使用されている語彙や文法が難しいときでも、文章の限られた範囲内では情報を関連付けることができる。</li> <li>中級レベルの語彙が理解できる。文脈中の難しい語彙、よく使用される単語の例外的な意味、慣用的な使い方が理解できることもある。</li> <li>規則に基づいた文法構造が理解できる。また、難しく、複雑で、あまり使用されない文法的な構造が理解できる。</li> </ul>	<ul style="list-style-type: none"> <li>一般的に以下の弱点が認められます。</li> <li>文章内の広い範囲にわたる情報を関連付けることができない。</li> <li>難しい語彙、よく使用される単語の例外的な意味、または慣用的な使い方が理解できないこともある。似たような意味で使われる複数の単語は、区別できないことが多い。</li> </ul>

### ライティング

スコア	長所	弱点
140~160	<ul style="list-style-type: none"> <li>一般的にスコアが140-160のレンジ内に該当する受験者は、簡単な情報を提供する、質問をする、指示を与える、または要求することが的確にできるが、理由や例をあげて、または説明をして、意見を裏付けることは部分的にしかできない。</li> <li>簡単な情報を提供する、質問する、指示を与える、または要求するときは、明確で、一貫性のある、的確な文章を書くことができる。</li> <li>意見について説明しようとするときは、その意見と関連のある考えやある程度の裏付けを提示することができる。</li> </ul> <p>一般的な弱点には、以下のようなものがある。</p> <ul style="list-style-type: none"> <li>要点の具体的な裏付けや展開が不十分である</li> <li>述べられている様々な要点同士の関連が不明確である</li> <li>文法的な誤りがある、または語彙・語句の選択が不正確である</li> </ul>	

おお、スピーキングでは「使用できる語彙・語句の範囲が限られている」とある。でも、大丈夫、中級レベルの語彙知識があることはリーディングの結果で分かっているから、自分に必要なのは、多様な語彙を使って話す練習をすることだ！

### 参考資料：スコアレンジ別評価一覧

<p>TOEIC L&amp;R</p> <p><a href="https://iibc.me/qual_persp_tp_05">https://iibc.me/qual_persp_tp_05</a></p>	<p>TOEIC S&amp;W</p> <p><a href="https://iibc.me/qual_persp_tp_06">https://iibc.me/qual_persp_tp_06</a></p>	<p>TOEIC Bridge L&amp;R</p> <p><a href="https://iibc.me/qual_persp_tp_07">https://iibc.me/qual_persp_tp_07</a></p>	<p>TOEIC Bridge S&amp;W</p> <p><a href="https://iibc.me/qual_persp_tp_08">https://iibc.me/qual_persp_tp_08</a></p>
---	---	--	--

## 4 技能の相関性は“完全”ではない

### ——技能別の測定が必要

4技能の間には関係性があり、ひとつの技能の測定結果がそのほかの技能の少なくとも間接的なエビデンスとなる可能性があります。それでもやはり、4技能は個別の技能であり、それぞれを別々に測定する必要があります。本セクションで前述した通り、言語能力は多数の要素で構成されており、そのうちの特定の能力は4技能それぞれでのみ用いられるというのが、研究者たちの大方の共通理解となっています。

また、ETSの研究者たちが、TOEIC Testsを受験した12,000人余り（うち約7,500人が全4技能を受験）のデータを分析した結果、4技能のテスト結果（リスニング、リーディング、スピーキング、ライティングの各スコア）の間には、中程度の相関はあるものの完全な相関はない、ということが分かっています（Liao et al., 2010）。

つまり、4技能の各テストはそれぞれが英語コミュニケーション能力について重要な情報を付加することが示唆されたわけです。

### TOEIC Testsスコア（リスニング、リーディング、スピーキング、ライティング）間の相関性

	Listening Score	Reading Score	Speaking Score	Writing Score
Listening Score	1.00			
Reading Score	.76	1.00		
Speaking Score	.66	.57	1.00	
Writing Score	.59	.61	.62	1.00

### テストは「何をどのように教え、学ぶのか」を伝えるメッセージ

また、テストは学生や現在・未来の働き手に届ける、英語学習に関するメッセージでもあります。そのため、それが社会・教育にもたらし得る影響も考慮する必要があります。

TOEIC Programのように、ハイステイクなテスト（e.g., 採用、証明、入学、卒業などの目的で利用されるテスト）が「何を測るのか」は、「何をどのように教え、学ぶのか」といった側面にも影響を与える可能性があります。というのも、ハイステイクなテストの受験に臨む教員や学習者はテストで測る技能に、より大きな力を注ぐためです。

この現象は、テストがもたらす教育・学習への「波及効果」（washback）と呼ばれます。波及効果は、テストが適切な教育・学習モデル（例えば、4技能の習得など）を後押しし、テストが提供する情報の適切な使用法を教員が心得ている場合には、ポジティブに作用することがあります。

一方で、テストの影響によって、教員や学習者の注力する方向性が、総合的な英語コミュニケーション能力の構築につながらない言語知識や技能へと傾いた場合には、ネガティブに作用することがあります。特定の技能のテストのみを選択して受験することにより、より力を注ぐ技能とそうでない技能とが生まれ、4技能のバランスが偏ってしまうかもしれません。例えば、リスニングとリーディ

ングのみを受験対象としていると、スピーキングやライティングに注意があまり向かなくなる恐れがあり、これらの技能のレベルがほかの技能よりも低くなっていた、という結果を招く可能性があります（詳しくはSection 6「TOEIC Programによるポジティブな波及効果」をご参照ください）。

#### 4技能の相互補完性を重視し 効果的な教育・学習を

さらに、長期的な視点で見た場合、特定の技能のみを受験する学習者は、全4技能を受験した場合に得られていた可能性のある利点——言語学習速度の加速——を享受できない場合があります。4技能は相互補完的な関係にあり、また、先述の通り、土台となる言語知識や基礎スキルを共有しています。そのため、4技能受験により、言語学習の速度が加速する可能性があるのです。例えば、スピーキングのテスト受験のために発音の練習をした受験者は、それがリスニング力の向上につながっていることに気が付くかもしれません。また、4技能受験を視野に入れた学習者は、例えば語彙を学習する際、意味の理解だけでなく、コミュニケーションで実際に使うために幅広い側面（品詞、音声、発音、綴り、プラグラティクスやディスコースに関連する特徴）を学ぶ必要があることに気が付くでしょう。このように、4技能受験のための

学習をする中で、受験者はひとつの技能の向上がほかの技能の向上に寄与すると共に、それらの技能の土台となる言語知識や基礎スキルの向上にもつながっていくことを実感するでしょう。

以上から、受験者にはTOEIC Programによる4技能受験を視野に入れた学習を推奨します。特定の技能のみを受験するために学習する場合よりも、ポジティブな波及効果が期待できるでしょう。全4技能を測定することにより、教員や学習者はTOEIC Bridge TestsやTOEIC Testsが提供する学習者に関する情報——総合的な英語コミュニケーション能力および各技能の習得状況——を存分に活用しながら、人材選抜のための意思決定、学習進捗の把握、フィードバックや学習指導の提供、プログラム改善に当たることができます。

英語コミュニケーション能力——関係し補完し合う個々の言語技能の集まり——がTOEIC Programの利用者の関心事であること、また、テストが教育・学習に波及効果をもたらすことを踏まえ、TOEIC Programでは全4技能の測定を通じて、受験者の総合的な英語コミュニケーション能力を評価すると共に、その構築を促していくことを強く推奨します。

## References

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
- Liao, C.-W., Qu, Y., & Morgan, R. (2010). The Relationships of test scores measured by the *TOEIC®* Listening and Reading Test and *TOEIC®* Speaking and Writing Tests. *TOEIC® Compendium* (1st ed., pp. 13.1-13.15). Educational Testing Service.

# IIBC

あなたが世界をつなぐ  
あなたと世界をつなぐ

一般財団法人 国際ビジネスコミュニケーション協会  
The Institute for International Business Communication

IIBC公式サイト <https://www.iibc-global.org>

本資料の無断転載・複製を禁ず

本冊子は公式サイトでもご覧いただけます。

[https://www.iibc-global.org/toeic/corpo/  
official\\_doc.html](https://www.iibc-global.org/toeic/corpo/official_doc.html)



[https://iibc.me/  
qual\\_persp\\_tp\\_h4](https://iibc.me/qual_persp_tp_h4)