

FAQ

——よくあるご質問と回答集——

Quality perspectives of
the TOEIC® Program

【本冊子について】

本冊子では、TOEIC Program のご利用者様から寄せられたご質問を取り上げ、TOEIC Program を開発する ETS の研究員による回答を掲載しています。本冊子が皆様のより良いテスト選択、英語教育・学習の一助となりましたら幸いです。

編集主幹・著者・レビュアー

編集主幹

Dr. John M. Norris

主席研究員 (ETS Japan)

著者

Dr. Jaime Cid

心理測定・データ分析ディレクター (ETS)

Dr. Jonathan Schmidgall

上級研究員 (ETS)

Dr. Saerhim Oh

研究員 (ETS)

著者・編集補助

Reiko Komatsu

調査研究員、アウトリーチスペシャリスト (IIBC)

レビュアー

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)

Dr. Feng Yu

アドバイザー

Frequently Asked Questions

- TOEIC Listening & Reading Test (以下、TOEIC L&R) や TOEIC Bridge® Listening & Reading Tests (以下、TOEIC Bridge L&R) で最高点を取るには、全問正解する必要があるのですか? …… p. 04
- TOEIC Programのスコアと「偏差値」はどう違うのですか? …… p. 05
- TOEIC Programのスコアの有効期限は、なぜ2年とされているのですか? …… p. 08
- なぜ解答用紙に性別や年齢などを記入する必要があるのですか? …… p. 10
- TOEIC Speaking & Writing Tests (以下、TOEIC S&W) と TOEIC Bridge Speaking & Writing Tests (以下、TOEIC Bridge S&W) では、人工知能 (AI) による自動採点を採り入れていますか? また、人による採点と、AIによる採点の違いを教えてください。 …… p. 12
- TOEIC Program のリスニングテストに4種類のアクセントが使われている理由を教えてください。 …… p. 14
- アカデミック英語を学んできた学習者でも、TOEIC Program の各テストで実力を発揮できるでしょうか? …… p. 16
- TOEIC Program のリスニングテストでは、なぜ問題用紙に書き込んだり、メモを取ったりしてはいけないのですか? …… p. 18
- なぜ TOEIC Program のテストでは問題用紙の持ち帰りができないのですか? …… p. 19
- TOEIC Program のテストの難易度は上がりましたか? (難化しているように感じます) …… p. 20
- TOEIC Program のテストの制作と採点には、項目応答理論 (IRT) が用いられているのですか? …… p. 23

Question

TOEIC Listening & Reading Test (以下、TOEIC L&R)やTOEIC Bridge Listening & Reading Tests (以下、TOEIC Bridge L&R)で最高点を取るには、全問正解する必要があるのですか？

Answer

どちらのテストにおいても、最高点、つまり最大スコア (TOEIC L&Rの各セクションで495点、合計で990点; TOEIC Bridge L&Rの各テストで50点、合計で100点) を取るのに、必ずしも全問を正解する必要がない場合があります。

TOEICのリスニングとリーディングの換算スコアはテストが実施されるたびに、ETSのテスト開発および統計の専門家によって算出されます。同じ種類のテストであっても、テストフォームが違えば出題される問題も異なるため、フォーム間で全体的な難易度にわずかな差が生じる可能性があります。ETSではTOEIC L&RとTOEIC Bridge L&Rが実施されるたびに、異なるフォームから換算したスコアが全ての受験者にとって比較可能かつ公平なものになるよう、統計学の専門家が統計的な処理を行い、一定の換算スコアに達するために必要な正答数をテストフォームの難易度に基づき調整しています。そのため、例えばあなたが今月受験したTOEIC L&Rのリスニングセクションの難易度がいつもよりもわずかに高かった場合、全問を正解しなくとも、495点 (TOEIC L&Rの各セクションの最大スコア) を取得する可能性がある、ということになります。

Question

TOEIC Programのスコアと「偏差値」はどう違うのですか？

Answer

TOEIC Programのスコアは「偏差値」とは明らかに異なります。TOEIC Programのスコアは英語能力レベルを示すスコア・スケール——「モノサシ」——に基づきレポートされ、それ自体が英語能力の指標として有意義に解釈が可能です。一方で、「偏差値」は平均点を50ポイント、その平均からの1標準偏差を±10ポイントに設定し、あるヒト・モノの立ち位置を相対的に示すものです。そのため、平均値からの距離との関連性においてのみ解釈が可能です。

初めに、TOEIC Programのスコアの特徴について簡単にご説明いたします。TOEIC Programのスコアは、英語能力レベルを示すスコア・スケールに基づきレポートされます。スコアは「素点」を基に算出されます。「素点」とは、リスニングやリーディングでは正答した問題の数、スピーキングやライティングではタスクごとに設定された採点スケール上での評価点を指します。TOEIC Programではこれらの「素点」を統計等の手法を用いて、初級から上級まで全レベルの英語能力を示すよう設計されたスコア・スケール上のスコアに換算しています。こうした換算プロセスを含め、TOEIC Programではテストの設計・開発・実施・採点の全工程を通じて、スコア基準の一貫性を常に担保しています。

TOEIC Programのスコアは、英語学習者の日常生活や職場でのコミュニケーション能力の程度を表すよう意図されており、スコア・ディスクリプターには全てのスコアレンジについて、スコアが示す意味や英語学習者の「強み・弱み」の理解に役立つ詳細な解説が掲載されています。

■スコアレンジ別評価一覧

TOEIC L&R



https://iibc.me/qual_persp_tp_faqs_01

TOEIC S&W



https://iibc.me/qual_persp_tp_faqs_02

TOEIC Bridge L&R



https://iibc.me/qual_persp_tp_faqs_03

TOEIC Bridge S&W



https://iibc.me/qual_persp_tp_faqs_04

また、TOEICスコアは、受験者のスコアに基づき遂行可能と予想されるコミュニケーションタスクの特徴という観点からご理解いただくことができます。ETSの研究者たちはさまざまなCan-do調査を行い、TOEIC Programの受験者に対して、日常生活や職場のコミュニケーションタスクをどの程度遂行できているかについて質問し、その結果をまとめています。TOEICスコアの意味をより良く理解するための資料としてご参照ください (e.g., Powers, et al., 2009; Powers, et al., 2010; Schmidgall, 2020)。

さらに、TOEIC S&Wのスコアは、職場における「話し手・書き手」としての受験者のコミュニケーション能力の指針として解釈することができます。ETSの研究者たちは、TOEIC Speaking TestとTOEIC Writing Testを対象に調査を行いました。日本を含む世界中の職場で働く人々に依頼して、TOEIC Speaking Test (あるいはTOEIC Writing Test) の受験者のコミュニケーション能力を評価してもらったところ、TOEIC Speaking TestとTOEIC Writing Testの各スコアは、国際的な職場で働く人々を相手にコミュニケーションを効果的に図る能力を予測すると解釈できる結果が得られました (e.g., Schmidgall & Powers, 2020; Schmidgall & Powers, 2021)。

以上から、TOEIC Programのスコアは独自で具体的、かつ意味のある解釈が可能な受験者の英語能力の指標とすることができます。TOEIC Programのスコアはそれ自体が英語能力について実質的で信頼価値のある情報を提供しています。そのため、個々人のスコアを他者との関係性において解釈したり、ほかのテストのスコアや指標と比較したりせずとも、受験者の英語能力を理解することが可能です。

次に、TOEIC Programのスコアと「偏差値」の違いについてですが、両者は明らかに非常に異なります。「偏差値」は平均点を50ポイントとして設定し、平均からの1標準偏差を±10ポイントに設定しています。そのため、「偏差値」は実際にはスコアではなく、相対値もしくは比較値を表す手段であり、あるヒト・モノの立ち位置を、ほかのヒト・モノとの比較において示すものです。ここで重要な点は、「偏差値」はいかなる能力分野 (英語やその他) においても個人の実際の能力や知識や習熟度について、具体的に解釈可能な意味を持たないということです。「偏差値」は常に相対的なものであり、平均値からの距離との関連性においてのみ解釈が可能です。

さらに、TOEIC Programのスコア・スケールは、意味のある規準集団に基づき設定された不変の尺度であり、実施回ごとに異なる受験者集団の能力分布の影響を受けることはありません。そのため、受験者のスコアの経時的な推移を分析することで、長期にわたってパフォーマンスを評価することが可能です (例えば、TOEIC L&Rのトータルスコアが2年間で500点から700点に上昇した場合、TOEIC L&Rが測定対象とする英語能力が2年間で相当程度上昇したと解釈できます)。一方で、「偏差値」は実施回ごとに異なる

る受験者集団の能力分布の影響をその都度受けるため、「偏差値」に基づいて能力の推移を直接的に解釈することはできません(例えば同じテストであっても、実施回ごとに受験者集団が異なるため、ある回の「偏差値60」と別の回の「偏差値60」が必ずしも同じ能力を示すとは限りません)。

TOEIC Programのスコアは、一貫性あるスコア基準に基づく、具体的かつ非常に有意義な受験者個々人の英語能力の指標としてご活用いただけます。「偏差値」は、実施回ごとに異なる受験者集団における受験者個々人の相対的な位置を示す指標です。そのため、日本では、高校や大学などへの入学を希望する学生の相対的な能力レベルを推定する際などに用いられています。

References

- Powers, D. E., Kim, H.- J., & Weng, V. Z. (2010). The redesigned TOEIC® Listening and Reading Test: Relations to test taker perceptions of proficiency in English. In D. Powers (Ed.), *TOEIC® Compendium* (1st ed., pp. 6.2-6.13). Educational Testing Service. <https://www.ets.org/Media/Research/pdf/TC-10-06.pdf>
- Powers, D. E., Kim, H.- J., Yu, F., Weng, V. Z., & VanWinkle, W. (2009). The TOEIC® Speaking and Writing Tests: Relations to test-taker perceptions of proficiency in English. In D. Powers (Ed.), *TOEIC® Compendium* (1st ed., pp. 11.2-11.40). Educational Testing Service. <https://www.ets.org/Media/Research/pdf/TC-10-11.pdf>
- Schmidgall, J. (2020). The Redesign TOEIC Bridge® Tests: Relations to test-taker perceptions of proficiency in English. (*Research Report No. RR-20-07*). Educational Testing Service. <https://onlinelibrary.wiley.com/doi/10.1002/ets2.12288>
- 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>

Question

TOEIC Programのスコアの有効期限は、なぜ2年とされているのですか？

Answer

外国語の能力が数年あるいはより短期間のうちに、大きく向上・低下する可能性があることなどを踏まえ、スコアに基づく意思決定の公平性を重視するETSの方針により、2年という有効期限が設定されています。

ETSでは、スコアの取得時から意思決定に利用するまでの有効期限を2年とすることを推奨しています。この方針は、高品質で正確な意思決定を促進するETSの倫理的な取り組みの一環であると同時に、業界水準に即しており、その他の主要な英語能力テスト提供者やスコア利用者が掲げるスコアの有効性に関する方針にも沿っています。また、外国語能力の向上・低下の経過を調査したりサーチにも基づいています。

TOEIC Programのスコアは、テスト受験時点での言語能力を示すものです。いかなる英語能力テストであっても、そのスコアに基づく意思決定は、それを下す時期がスコアの取得時点から近いほどより公平で信頼に足るものとなります。

実際に、外国語の能力は数年、あるいはより短期間のうちに大きく変化する可能性があります。例えば、短期間であっても外国で集中して学習した経験によって、急速に言語能力が向上する可能性があり、とりわけリスニングとスピーキングの即応力に顕著な伸びが見られることがあります。また、言語能力は低下する可能性もあります。この現象は「言語の喪失」(language attrition)と呼ばれています。

過去数十年の調査からは、言語能力の要となる要素(単語を思い出す能力や流暢に話す能力など)は、長期間にわたってその言語が学ばれない、あるいは活発に使われない状況が続くと、低下しやすいことが示唆されています。実際に、2年を超えた時点で認識できるレベルの「外国語の喪失」が開始した事

例がたびたび観察されています。

2年という期限は以上の実情——言語能力が大きく向上・低下する可能性；テスト実施から一定期間を経た英語能力テストのスコアの有用性を主張すべきでないとするETSの方針——に鑑みて設定されています。

TOEIC Programでは、スコアの有効期間を2年間に限定することで、意思決定者が学習者の英語能力を過剰評価することを防ぐと共に、有意義な学習期間を経て言語能力が向上した受験者の能力の正確な評価を促しています。

Question

なぜ解答用紙に性別や年齢などを記入する必要があるのですか？

Answer

あらゆる受験者集団に対してスコアの公平性を担保し、特定の受験者に対するバイアスを防止するためには、さまざまな受験者集団のパフォーマンスを評価し、モニタリングする必要があります。そのため、TOEIC Programでは受験者の性別や年齢、その他背景情報(学歴や雇用状況など)を収集しています。

ETSではTOEIC Programの内容とスコアに、受験者の特徴による有利・不利をもたらすバイアスが決して生じないように尽力しています。この方針は『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*)※に明記されており、同文書にはTOEIC Programのテスト内容の作成やテスト実施前レビューの方法に関する情報が掲載されています(『品質および公平性に関するETSスタンダード』についてより詳しくは*Quality perspectives of the TOEIC Program* 本編のSection 3「ETSの品質への取り組み」をご覧ください)。

※『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*)



https://iibc.me/qual_persp_tp_faqs_05

テストの作成過程においては、公平性を担保するためのあらゆる予防策が講じられますが、ETSが求める公平性の水準を確実に満たすために、ETSの心理測定学者がさらに追加調査を実施しています。あらゆる受験者集団に対してスコアの公平性を担保し、特定の受験者に対するバイアスを防止するためには、さまざまな受験者集団のパフォーマンスを評価し、モニタリングする必要があります。そのため、TOEIC Programでは受験者の性別や年齢、その他背景情報(学歴や雇用状況など)を収集しています。背景情報——個人単位ではなく集団単位でのみ使用されます——はスコアの妥当性に関する重要なフィードバックをもたらすため、背景の異なる全ての受験者に対してスコアの公平性を担保するために役立てられています。例えば、性別情報は、採点対象の問題に性別上のバイアスが生じておらず、意図する内容と異なる側面を測定していないことを担保するために欠かせない統計分析に用いられま

す。意図した通りに機能していないことが明らかになったいかなる問題も、ETSによる公平性担保の方針に従い、採点対象から除外されます。

さらに、受験者が回答した性別や年齢、その他背景情報はReport on Test Takers Worldwideにも活用されています。本レポートはTOEIC Programの受験者やスコア利用者向けに発行されており、受験者の背景情報のほか、TOEICスコアや英語能力の向上に影響を与える要素などについて、より詳しく知ることができます。

2022 Report on Test Takers Worldwide:
TOEIC® Listening & Reading Test

(日本語版)



https://iibc.me/qual_persp_tp_faq_06



https://iibc.me/qual_persp_tp_faq_07

2022 Report on Test Takers Worldwide:
TOEIC® Speaking & Writing Tests



https://iibc.me/qual_persp_tp_faq_08

以上から、ETSではテストスコアの公平性およびTOEIC Programの品質を担保するために必須のプロセスとして、性別や年齢などの背景情報を収集しています。なお、TOEICスコアの品質を検証するための調査研究において受験者の背景情報が実際にどのように活用されているのかについてご興味がある方は、以下の参考文献をご参照ください。

References

- Wei, Y., & Low, A. (2017). Monitoring score change patterns to support TOEIC® Listening and Reading test quality. (*Research Report No. RR-17-54*). Educational Testing Service. <https://doi.org/10.1002/ets2.12186>
- 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. <https://doi.org/10.1177/0265532218776040>
- Wei, Y. (2013). Monitoring TOEIC® Listening and Reading test performance across administrations using examinees' background information. In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies: Volume II* (pp. 11.1-11.28). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/chapter/2013/jrou.html
- Qu, Y., Huo, Y., Chan, E., & Shotts, M. (2017). Evaluating the stability of test score means for the TOEIC® Speaking and Writing tests. (*Research Report No. RR-17-50*). Educational Testing Service. <https://doi.org/10.1002/ets2.12180>

Question

TOEIC Speaking & Writing Tests (以下、TOEIC S&W) と TOEIC Bridge Speaking & Writing Tests (以下、TOEIC Bridge S&W) では、人工知能 (AI) による自動採点を採り入れていますか？ また、人による採点と、AIによる採点の違いを教えてください。

Answer

TOEIC S&W と TOEIC Bridge S&W の解答は、高度に訓練され認定を受けた複数の人間の採点者によって採点されています。人間の採点者は、これらのテストに出題されるコミュニケーションタスクを達成するための言語使用の多様な側面を非常に効果的に評価できることが、調査により明らかになっています。ETSではAIによる採点技術の探究と向上に継続的に取り組んでおり、AI採点の正確性および信用性が十分な水準に達した段階でさまざまなテストに搭載し使用しています。

現在、TOEIC S&W と TOEIC Bridge S&W の解答は、評価の信頼性を最大化するために、高度に訓練を受けた複数の人間の採点者によって採点されています。TOEIC S&W と TOEIC Bridge S&W にはスコアの信頼性に加え、タスクが求める情報を受験者がどのくらい効果的に提供できたかを評価する特徴があり、それが世界中の多くの組織から支持される理由のひとつとなっています。人間の採点者を用いることで、タスクを通じて測定するスピーキング能力とライティング能力の全ての構成概念を、確実に採点プロセスに含めることができます (Hines, 2010)。

各タスクは包括的なルブリック——受験者のパフォーマンスの多様な側面をスケール上の能力レベル別に示したもの——に基づき採点されます。多様なタスクやルブリックは、受験者の現実世界における実践的なコミュニケーション能力を捉えることを目的としており、受験者がタスクを遂行する際に発揮したパフォーマンスも現実世界での成功や達成を示す基準に従って採点されます。例えば、一部のスピーキングのタスクでは、ある状況や受け手に対してプラグマティクスの観点から見て適切な（つまり、丁寧さや改まり度の観点から適切な言語を使用して）応答をすることが求められます。一部のライティングのタスクでは、考えを論理的に構成することや、文や段落間のつながりに配慮することが求められます。また、スピーキングとライティングのいずれにおいても、与えられた情報に触れながら意見や考えを伝えることで「達成した」と見なされるタスクが出題されます。

受験者のパフォーマンスにおけるこうした特性の全てが高度な解釈を必要とします——タスクの目的、話し手・書き手の意図、受け手の存在を踏まえながら、複雑な要素や微妙な差異を捉える必要があるのです。自動採点によってこうした特性を評価すること——自動採点の測定能力によって人間の採点を正しく予測すること——は困難であることが明らかになっています。ですが、こういった特性こそが、現実世界でのスピーキングやライティングのタスク遂行能力を示すうえで、非常に重要なのです。TOEIC Programではこうしたタスクを重視し、多様な問題を出題しているため、人間の採点者によって受験者のパフォーマンスを採点することがより望ましいと考えています。また、これまでの経験や調査結果から、人間の採点者は訓練を受けることで、こうしたタスクを成功させるパフォーマンスにつながる、言語使用の多くの多様な側面を捉えられるようになることが示されています (Lin, et al., 2019; Qu, et al., 2013; Qu, et al., 2018; Schmidgall, 2018)。

ETSでは長年にわたり AI採点の性能の開発に取り組んでおり (e.g., Zechner & Evanini, 2019)、すでに一部の言語能力テストにAIを搭載し、多くのケースで人間による採点と組み合わせて使用しています。AI採点の性能が引き続き向上し、TOEIC S&WやTOEIC Bridge S&Wに出題されるタスクの採点において正確性が認められると判断された際には、これらのテストにも採点方法の一部として導入される可能性があります。

References

- 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. <https://www.ets.org/Media/Research/pdf/TC-10-07.pdf>
- Lin, P., Cid, J., & Zhang, J. (2019). Field study statistical analysis for the redesigned TOEIC Bridge® tests (*Research Memorandum No. RM-19-09*). Educational Testing Service.
- Qu, Y., Huo, Y., & Chan, E. (2018). Evaluating the stability of test score means for the TOEIC Speaking and Writing tests. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 8.1-8.17). Educational Testing Service.
- Qu, Y., & Ricker-Pedley, K. L. (2013). Monitoring individual rater performance for the TOEIC Speaking and Writing tests. In D. Powers (Ed.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume II* (pp. 9.1-9.9). Educational Testing Service.
- Schmidgall, J. (2018). The consistency of TOEIC Speaking scores across ratings and tasks. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies: Volume III* (pp. 7.1-7.12). Educational Testing Service.
- Zechner, K., & Evanini, K. (Eds.). (2019). *Automated speaking assessment: Using language technologies to score spontaneous speech*. Routledge.

Question

TOEIC Program のリスニングテストに 4 種類のアクセントが使われている理由を教えてください。

Answer

TOEIC Program のリスニングテストでは、受験者の英語学習背景によって有利・不利が生じない状態を保ちながら、テストを現実世界の状況により近い内容にするために、一般的かつ非常に理解しやすく、職場や日常生活で触れる可能性が概して最も高いと想定される、4 地域のアクセントを採用しています。

2006年に TOEIC L&R がリデザインされた際、リスニングセクションで用いられるアクセントの地域は米国・英国・カナダ・豪州およびニュージーランドに拡大しました。これらの地域のアクセントが選ばれた理由は、世界中の英語学習コミュニティで最も一般的に使用され、教えられていたためです。

2019年に TOEIC Bridge Tests がリデザインされた際にも、TOEIC L&R と同じ理由から、リスニングテストで用いられるアクセントの種類は上記の 4 地域に増えました。

また、リスニングテストで用いられるアクセントの重要な特性として挙げられるのが、その強度設定です。アクセントには穏やかものから強いものまでありますが、リスニングテストには穏やかなアクセントが採用されています。強いアクセントには特有の性質があり、その影響でほとんどの聞き手にとって——英語のネイティブスピーカーにとってさえも——理解が難しくなってしまう可能性があります。

つまり、4 地域の穏やかなアクセントを採用した論拠としては、受験者が国際的な職場や日常生活の場面で触れる可能性のある、一般的かつ非常に理解しやすいアクセントを用いるため、ということになります。

TOEIC のリスニングテストには、話者の話を理解するために、特定の種類の英語を習得していることが

必要となる問題は出題されません。アクセントの種類を増やした目的は、受験者が異なるアクセントを聞き分けられるかを判定するためではありません。英語学習背景によって有利になる受験者がいない状態を保ちながら、テストを現実世界の状況により近い内容にするためです。

TOEIC L&Rのリデザインに伴い4地域のアクセントが採用されて以降も、英語は世界中のますます多くのさまざまな場面で用いられています。多くの種類の英語（インド英語、ナイジェリア英語、シンガポール英語など）がより使われるようになり、英語のアクセントも著しく多様化しています。こうしたおびただしい種類の多様なアクセントをテストに導入することは現実的ではありませんが、テスト内容が現実世界での言語使用の重要な側面をどの程度反映しているかを定期的に見直すことは重要です。

TOEIC Programについても、将来的なリデザインや改訂の際には、受験者の英語学習背景による有利・不利が発生しないよう、品質と公平性の水準を維持しながら、国際的なコミュニケーションにおける英語の使われ方の変化に対応することになるでしょう。

Question

アカデミック英語を学んできた学習者でも、TOEIC Programの各テストで実力を発揮できるでしょうか？

Answer

アカデミック英語には、日常生活や職場で求められる一般的な英語コミュニケーションの技能に共通する部分が多くあるため、アカデミック英語能力と実践的な英語コミュニケーション能力を併せて習得している学習者は、TOEIC Programのテストでも十分に力を発揮できるでしょう。一方で、アカデミック英語を特徴づける特定の知識や技能を測定する場合には、アカデミック英語に特化して設計されたテストが推奨されます。

TOEIC Programは、初級から上級の英語能力レベルの学習者を対象に、日常生活と職場の場面における一般的な英語コミュニケーション能力を測定するよう設計されています。

そのため、テストのトピックやタスクは多様なビジネス・職場や、誰もが行う日常の活動において見られる、実際のコミュニケーション上の要求に関連したものとなっています。

日常生活や職場のコミュニケーション能力の習得を言語教育の目標に設定している大学や高校の英語教育プログラムにおいて、TOEIC Programは学生の英語能力の伸長をプログラムから独立した形で測定する有効な手段となります。また、英語環境の日常生活や職場でコミュニケーションを図ろうとする学生のレディネスを把握する際にも、有効にご利用いただけます。

アカデミック英語 (English for academic purposes) には、日常生活や職場での一般的なコミュニケーションの技能——現実世界における多様なタスクの遂行能力も含めて——に共通する部分が多くあります。アカデミック英語の能力を習得してきた学習者は、実践的なリーディング、リスニング、ライティング、スピーキングの技能も併せて習得していれば、TOEIC Programのテストでも十分に力を発揮できるはずです。実際、TOEIC Programのテストに出題される類の日常生活や職場のタスクを遂行す

る能力は、英語環境の大学での生活や勉強を検討している方を含め、さまざまな英語学習者にとって重要なものです。

もちろん、大学の授業や学術的な場面でのコミュニケーションで必要となる、アカデミック英語を特徴づける特定の知識や技能を測定する場合には、TOEFL® iBT Testのようにアカデミック英語に特化して設計されたテストが推奨されます。

テスト利用者は、どの英語能力テストが最も自らのニーズに即しているのかを慎重に検討し、状況に応じて選択するとよいでしょう。

Question

TOEIC Programのリスニングテストでは、なぜ問題用紙に書き込んだり、メモを取ったりしてはいけないのですか？

Answer

TOEIC L&RのリスニングセクションやTOEIC Bridge Listening Testは多肢選択問題によるテストのため、さまざまな手法でカンニングを試みようと考える受験者が存在する可能性があります。そのため、ETSではこうしたカンニング行為を防止し、全ての受験者にとって公平なテストを提供するために、問題用紙への書き込みを禁止することをグローバルポリシーとしています。

同様の理由から、問題用紙を使ってメモを取ることも許可されていません*。

また、TOEIC Programのリスニングテストで受験者が聞く描写文、会話、説明文は比較的短く、多くの情報を記憶保持することは求められません。現実世界に照らしても、こうした情報を聞いてメモを取らない方が自然かつオーセンティックと言えます。例えば、相手から「会議室はどこですか？」と聞かれて、その質問をメモに取ることはありません。

以上の理由から、問題用紙への書き込みやメモを取る行為は禁止されています。

※ETSでは、TOEIC L&Rのリーディングセクション、TOEIC Bridge Reading Testにおいても、カンニング行為の防止の観点から、問題用紙への書き込み・メモ取りの禁止をグローバルポリシーとしています。

Question

なぜTOEIC Programのテストでは問題用紙の持ち帰りができないのですか？

Answer

TOEIC L&RとTOEIC Bridge L&Rのテストフォームは複数のテスト実施にわたって使用されるため、常にセキュアな環境で管理されなければなりません。問題や解答が一般に公開されてしまえば、TOEIC Programの完全性(integrity)が損なわれてしまうでしょう。

ETSでは、スコア基準の一貫性を確実に保つために、「等化」(equating)という手法を導入しています。等化とは、簡単に言うと、TOEIC Programのリスニングとリーディングのテストのスケールスコアを異なるテストフォーム間で比較可能な状態にするための統計的なプロセスです。等化のプロセスでは、ETSの心理測定学者が、既存のテストフォームでの受験者のパフォーマンスの分析結果と、新たに作成したテストフォームでの受験者のパフォーマンスの分析結果とを比較し、フォーム間の難易度の差異を調整しています。このプロセスを実施するために、テストフォームを新たに作成する際には、過去の実施回に出題された「アンカー問題」(anchor item)と呼ばれる問題を新規フォームに組み入れています。新規フォームに組み込まれたアンカー問題に対する受験者のパフォーマンスを分析することで、既存フォームと新規フォームの難易度を比較することができるため、新規フォームの難易度レベルが適切であるかどうかを検証することができます。テスト問題と解答が一般公開されてしまうようなことがあれば、こうした等化に使用されるアンカー問題に対する受験者のパフォーマンスに、受験者の真の能力が反映されなくなり、等化プロセスは機能不全に陥ってしまいます。

また、TOEIC L&RとTOEIC Bridge L&Rのテスト問題と解答は、TOEIC Programの著作権保持者であるETSの知的財産として、著作権法により保護されています。

こうした理由から、ETSが保有するTOEIC L&RとTOEIC Bridge L&Rのテスト問題と解答は、ETSによる書面での許諾なしには、いかなる目的においても公開・使用することはできません。

Question

TOEIC Programのテストの難易度は上がりましたか？（難化しているように感じます）

Answer

TOEIC Programの各テストの全体的な難易度が時を経て変化する、ということはありません。TOEIC Programでは、同じテストについては、全てのフォームが同一のコンテンツ設計と統計仕様に沿って制作されています。また、テストの改訂やアップデート時には、ETSのテスト開発や統計学の専門家によって、新バージョンのテストが旧バージョンのテストと同じ難易度・品質を確実に保つように設計・開発され、妥当性検証のための調査も事前・事後だけでなく、常時実施されています。

TOEIC Programでは、同じテスト（設計とスコア・スケールが同一のテスト）については、全てのフォームが同一のコンテンツ設計と統計仕様に沿って制作され、組み立てられ、実施されています。

TOEIC Programでは、テストのアップデートや改訂を行う際には、パイロット調査やフィールド調査を実施して、新バージョンのテストが以前と同じ構成概念や技能を測定していること、そして難易度が同じであることを担保しています。新バージョンのテストは、ETSのテスト開発や統計学の専門家によって、旧バージョンのテストと同じ難易度・品質を確実に保つように設計・開発されます。そのため、TOEIC Programでは、設計とスコア・スケールが同一のテストにおいて、全体的な難易度が時と共に変化する、ということはありません。

例えば、TOEIC L&Rは2006年にリデザインされ、2016年にアップデートされましたが、その裏付けとなる調査レポートがそれぞれ公開されており、ウェブサイト「TOEIC Research」(<https://www.ets.org/toEIC/research.html>) から簡単に入手することができます。レポートにはテスト内容の比較、問題に関する統計情報、旧バージョン（リデザイン前もしくはアップデート前）のテストと新バージョンのテストにおける受験者のパフォーマンスに関する情報などが掲載されています。

これらの調査によって示されたエビデンスは、構成概念と心理測定の観点から、リデザインやアップデートをまたいでスコアを連続的に解釈することが可能であるとする主張の論拠となっており、リデザインおよびアップデート後も TOEIC Program のスコア・スケールが同じ能力レベルを反映していることを担保しています (Liao, et al., 2010; Cid, et al., 2017)。

さらに、テストのリデザインやアップデートの際には、受験者やテスト利用者が新旧バージョンのテストに対して「どう感じているか」という観点からの検証も行われます。具体的には、TOEIC Program のテストのリデザインやアップデートを実施する際、ETS のテスト開発の専門家は受験者やスコア利用者などのステークホルダーの「テストに対する知覚」(perceptions of testing) を、テスト設計や検証のプロセスに要素として組み入れています。受験者の知覚に関する情報は、テスト開発過程で実施するパイロットテストやフィールドテストの段階で、認知インタビューやサーベイを通じて収集され、テストのユーザビリティ(指示の明確さ、準備・解答時間の的確さなど)、テストに対するさまざまな認識(オーセンシティ、難易度など)、テスト問題ごとの検討課題(視覚素材の有用性など)についての把握に役立てられます (Everson, et al., 2021; Schmidgall, 2022)。

もう一つ例を挙げます。TOEIC L&R の採点プロセスは、心理測定学者によって常にモニタリングされており、異なるテストフォーム間やテスト実施回間でのスコアの一貫性が保たれ、意味のある公平な言語能力の解釈が可能となるよう担保されています。モニタリング事例として以下に2件を挙げます。1件目は異なるテスト実施回間におけるスコアの安定性の精査(Wei, 2013)、2件目はリピーター受験者の経年的なスコア変化パターンの調査を実施したものです(Wei & Low, 2017)。

ETS ではこうした調査を通じて、TOEIC Program の品質管理へのコミットメントへの理解を促すと共に、スコアの信頼性・妥当性を裏付ける経験的証拠を提供しています。

References

- Cid, J., Wei, Y., Kim, S., & Hauck, C. (2017). Statistical analyses for the updated TOEIC® Listening and Reading test (*Research Memorandum No. RM-17-05*). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/report/2017/jyfa.html
- Everson, P., Duke, T., Garcia Gomez, P., Carter Grissom, E., Park, E., & Schmidgall, J. (2021). Development of the Redesigned TOEIC Bridge® Tests. In J. Schmidgall (Ed.), *The research foundation for the Redesigned TOEIC Bridge® Tests: A compendium of studies: Volume IV* (pp. 2.1-2.53). Educational Testing Service.
- Liao, C.-W., Hatrak, N., & Yu, F. (2010). Comparison of content, item statistics, and test-taker performance for the redesigned and classic TOEIC Listening and Reading tests. In D. Powers (Ed.), *TOEIC® Compendium* (1st ed., pp. 4.2-4.9). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/report/2010/itjs.html
- Schmidgall, J. (2022). The Role of Test Taker Perceptions in Test Development and Validation. Paper presented at the 8th annual international conference of the Asian Association for Language Assessment (AALA), Chennai, India.
- Wei, Y. (2013). Monitoring TOEIC® Listening and Reading test performance across administrations using examinees' background information. In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies: Volume II* (pp. 11.1-11.28). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/chapter/2013/jrou.html
- Wei, Y., & Low, A. (2017). Monitoring score change patterns to support TOEIC® Listening and Reading test quality. (*Research Report No. RR-17-54*). Educational Testing Service. <https://doi.org/10.1002/ets2.12186>

Question

TOEIC Programのテストの制作と採点には、項目応答理論 (IRT) が用いられているのですか？

Answer

TOEIC Programでは手法のひとつとして、適切なデータが使用可能な場合に項目応答理論 (IRT: item response theory) をテストの制作、テストフォームの構築、採点に用いています。

IRTに基づく分析は個々の問題の属性——難易度や、受験者間の能力差異をどの程度有効に弁別するか——を評価する際に特に有効な役割を果たします。また、IRTを用いることにより、各能力範囲におけるスコア精度を推定することができます。この推定情報は、信頼性あるテストフォームを構築し、受験者の能力を精密に推定するスコアを生み出すために利用されます。これらのIRTの特性は、異なるテストフォーム間やテスト実施回間においてスコアの一貫性を確実に担保するために欠かせないものとなっています。

一方で、古典的テスト理論に基づく手法も、各テスト問題やテストフォームのパフォーマンスを評価する際に用いられています。このように、TOEIC Programでは高品質な測定を確実に提供するために、テストフォームおよびテスト実施ごとに包括的な分析を行っています。

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FAQ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://iibc.me/qual_persp_tp_faq_h4