



# GTEC スコア統合 調査報告

2018年2月

## 1. 調査結果サマリ

今回の調査結果を踏まえ、最終決定した GTEC の同一尺度における各問題タイプの上限值および CEFR 対応表は以下となる。

### <GTEC 各タイプの上限值>

問題タイプ	リーディング	リスニング	ライティング	スピーキング	トータル
CBT	350	350	350	350	1400
Advanced	320	320	320	320	1280
Basic	270	270	270	270	1080
Core	210	210	210	210	840

### <GTEC の CEFR 対応表>

CEFR	リーディング	リスニング	ライティング	スピーキング	トータル
C1	330	340	340	340	1350
B2	280	290	300	320	1190
B1	220	220	240	280	960
A2	150	160	190	190	690
A1	60	70	60	80	270

## 2. 調査研究プロセス

GTEC および GTEC CBT は項目反応理論に基づいて運用されている英語 4 技能検定である。どちらの検定も、テスト問題（項目）に対する反応パターンから能力値（ $\theta$  値）を推定し、その値をスコアに変換している。GTEC および GTEC CBT の各項目に対するパラメータが同一の能力尺度上に推定されているならば、2 つの検定から得られる能力値は比較可能となる。また、推定された能力値を同一の変換式でスコアに変換するならば、2 つの検定から得られるスコアは比較可能となる。

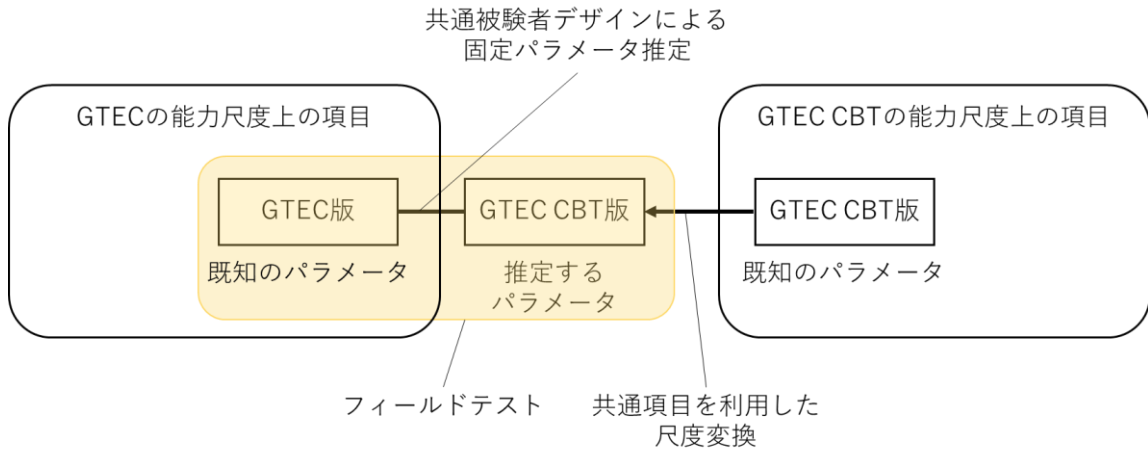
リーディングおよびリスニングにおいては、GTEC および GTEC CBT の各項目のパラメータは同一の能力尺度上に推定されているため、両検定から得られるスコアは比較可能である。一方、スピーキングおよびライティングにおいては、異なる能力尺度上に項目パラメータが推定されているため、スコアは比較可能でない。この問題を解決するために、GTEC CBT の各項目のパラメータを GTEC の能力尺度上に変換するための調査研究（フィールドテスト）を行った。

フィールドテストでは、大学生および社会人等のモニターを 1,100 人程度募集し、GTEC の項目で構成されるテスト版（GTEC 版）と、GTEC CBT の項目で構成されるテスト版（GTEC CBT 版）の両方を受検してもらう共通被験者デザインを用いる。

GTEC 版の各項目は、GTEC の能力尺度上に推定されたパラメータを持っている。フィールドテストデータに基づいて項目パラメータを推定する際に、GTEC 版の各項目のパラメータをこれらの値に固定することで、GTEC CBT 版の各項目のパラメータを GTEC の能力尺度上に推定することができる。

GTEC CBT 版の各項目は、GTEC CBT の能力尺度上に推定されたパラメータを持っている。そのため、フィールドテストデータに基づいて項目パラメータ推定を行った後は、GTEC CBT の能力尺度上に推定された項目パラメータと、GTEC の能力尺度上に推定された項目パラメータの 2 セットを持つことになる。このとき、前者の項目パラメータが後者の項目パラメータに近くなるような線形変換を求めることで、GTEC CBT の能力尺度上の項目パラメータを GTEC の能力尺度上に移すための変換式を得ることができる。

## ■統合デザインの概要図



### 3. 調査研究結果の詳細

#### (1) 調査に使用したデータ

##### ①テスト版

■GTEC：2017年12月本番試験で出題したテスト版

■GTEC CBT：2017年7月本番試験で出題したテスト版

##### ②データ収集デザインと被験者

日本国内に居住する大学生および社会人を中心にモニター受検者を募集し、GTEC版とGTEC CBT版の両方を受検してもらう共通被験者デザインを用いた。不適切な受検者（片方のテスト版しか受検しなかった，など）を除外した結果，最終的に分析に含めた受検者はライティングが1122人，スピーキングが1104人だった。

#### (2) 分析担当者

野澤雄樹

アイオワ大学大学院教育学研究科博士課程修了 Ph.D. [教育測定学]

元 ACT, Inc. 研究員

#### (3) 結果の詳細

項目反応理論を用いるための事前準備として，ライティングおよびスピーキングの各技能について，GTECの項目とGTEC CBTの項目が同じ能力を測定していることの検証を行った。R言語のpsychパッケージ（Revelle, 2017）を使用して平行分析および因子分析を行ったところ，どちらの技能においても，完全な一因子構造とは言えないものの，GTECの項目とGTEC CBTの項目は概ね近い能力を測定していることが確認された。

このことから、一次元の項目反応モデルを適用して同一の能力尺度上に表現することに大きな支障はないと判断した。

項目パラメータの推定は、専用のソフトウェアである IRTPRO (Scientific Software International, 2016) を用いて行った。技能ごとに、GTEC 版の項目パラメータを固定し、受検者群の能力分布の平均と標準偏差を自由に推定することで、GTEC CBT 版の項目パラメータを GTEC の能力尺度上に推定した。

GTEC の能力尺度上に新しく推定された項目パラメータを旧版、GTEC CBT 版の項目プールにおける既知パラメータ (GTEC CBT の能力尺度上) を新版として、後者を前者に移動させるための尺度変換を技能ごとに求めた。具体的な計算は STUIRT (Kim, & Kolen, 2004) を使用して行った。STUIRT によって求められた 4 つの方法 (Mean/Mean 法, Mean/Sigma 法, Haebara 法, Stocking-Lord 法) に基づく尺度変換に大きな違いはなかったため、過去の研究結果との整合性を考慮して、最も適切と思われる方法 (どちらの技能も Mean/Sigma 法) を採用した。

GTEC CBT スコアを GTEC CBT の能力値に変換した後、上記の手続きで得られた尺度変換を使用して GTEC の能力値に変換し、さらに GTEC スコアに変換することで、任意の GTEC CBT スコアを GTEC スコアに対応付けることができる。GTEC CBT スコア尺度上に定義された C1 の閾値 (350 点) は GTEC スコアに変換したところ、ライティングとスピーキングのどちらにおいても、340 点前後になることがわかった。そのため、GTEC スコア尺度上における C1 の閾値は、ライティング・スピーキングとも 340 点となった。

同様に、ライティング・スピーキングのどちらの技能においても、GTEC CBT のスコア上限 (350 点) は GTEC スコア尺度における 340 点に対応することになる。しかし、リーディングおよびリスニングと上限値が異なることによる混乱を避けるため、ライティングとスピーキングの測定範囲を引き上げ、GTEC スコア尺度における 350 点が上限になるように設定した。測定範囲の引き上げが可能であることは、過去の受検者の実データから確認している。

#### 4. 引用文献

Scientific Software International. (2016). *IRTPRO 4.0 for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International.

Kim, S., & Kolen, M. J. (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models* [Computer software]. Iowa

City, IA: Authors. Available from <https://education.uiowa.edu/centers/center-advanced-studies-measurement-and-assessment/computer-programs#irt>

Revelle, W. (2017). *psych: Procedures for psychological, psychometric, and personality research* [Computer software]. <https://CRAN.R-project.org/package=psych> (R package version 1.7.8)