

GTEC

Global Test of English Communication



英語4技能検定 GTEC 心理測定学的側面・CEFRとの関係

込山智之* 野澤雄樹**

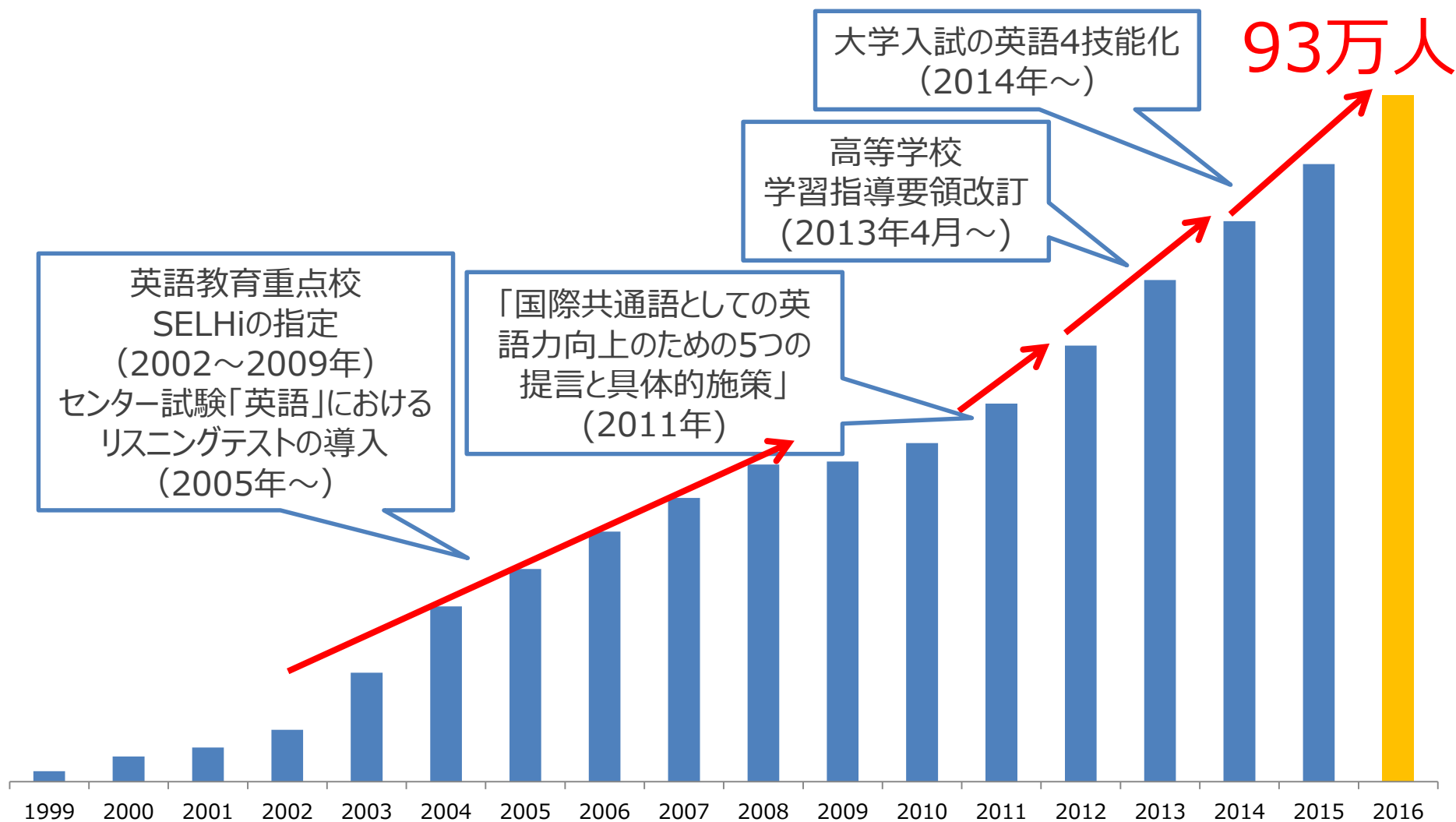
一般財団法人進学基準研究機構 (CEES) *

ベネッセ教育総合研究所**

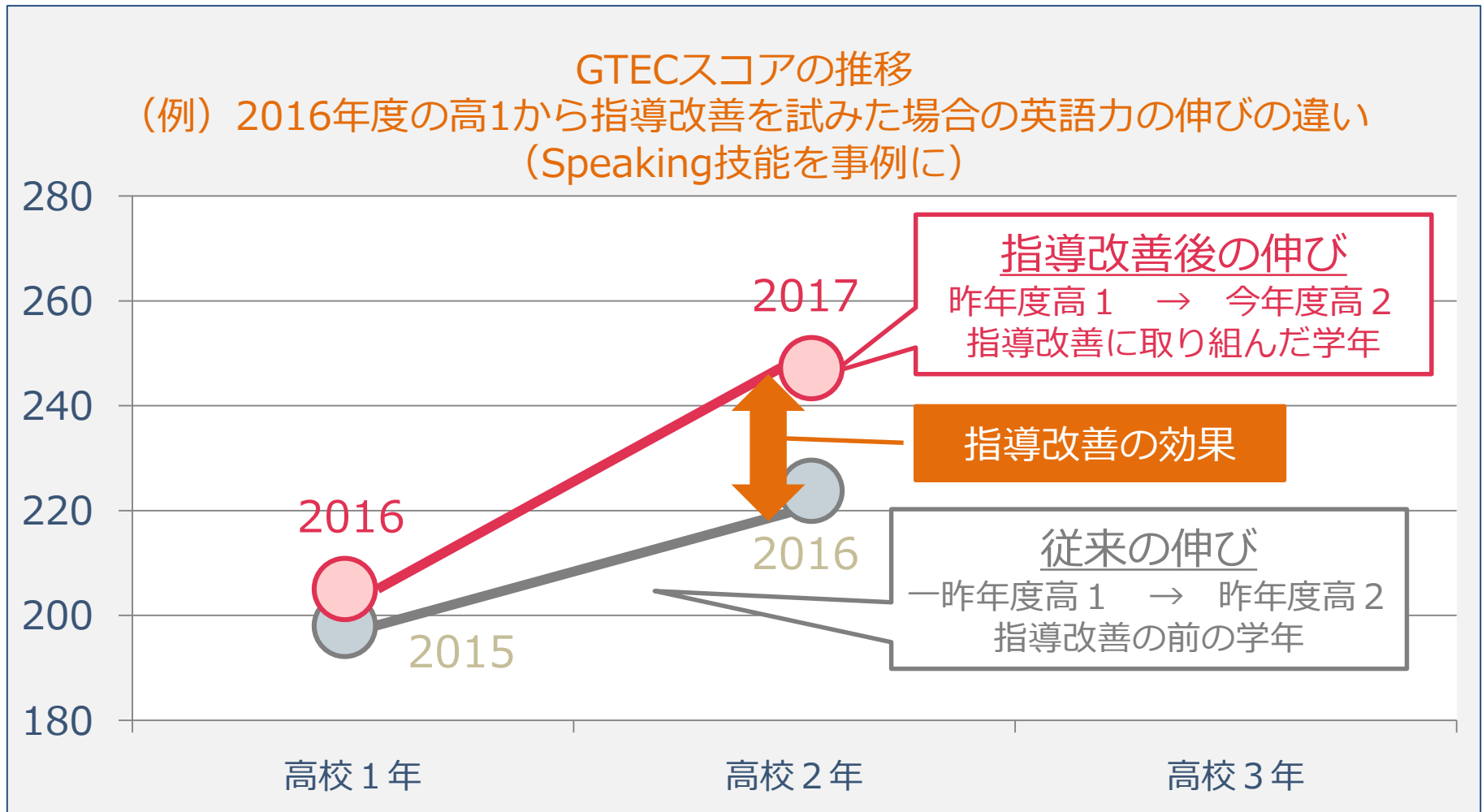
2017年8月20日

日本テスト学会第15回大会 公開シンポジウム1

コミュニカティブな指導の広がりに伴い
中高生において最も受検者の多いスコア型英語検定に



同一生徒の英語力推移を経年比較することが可能。
英語力の伸びからどの技能の指導をどのようにするか検討のベースとなる。



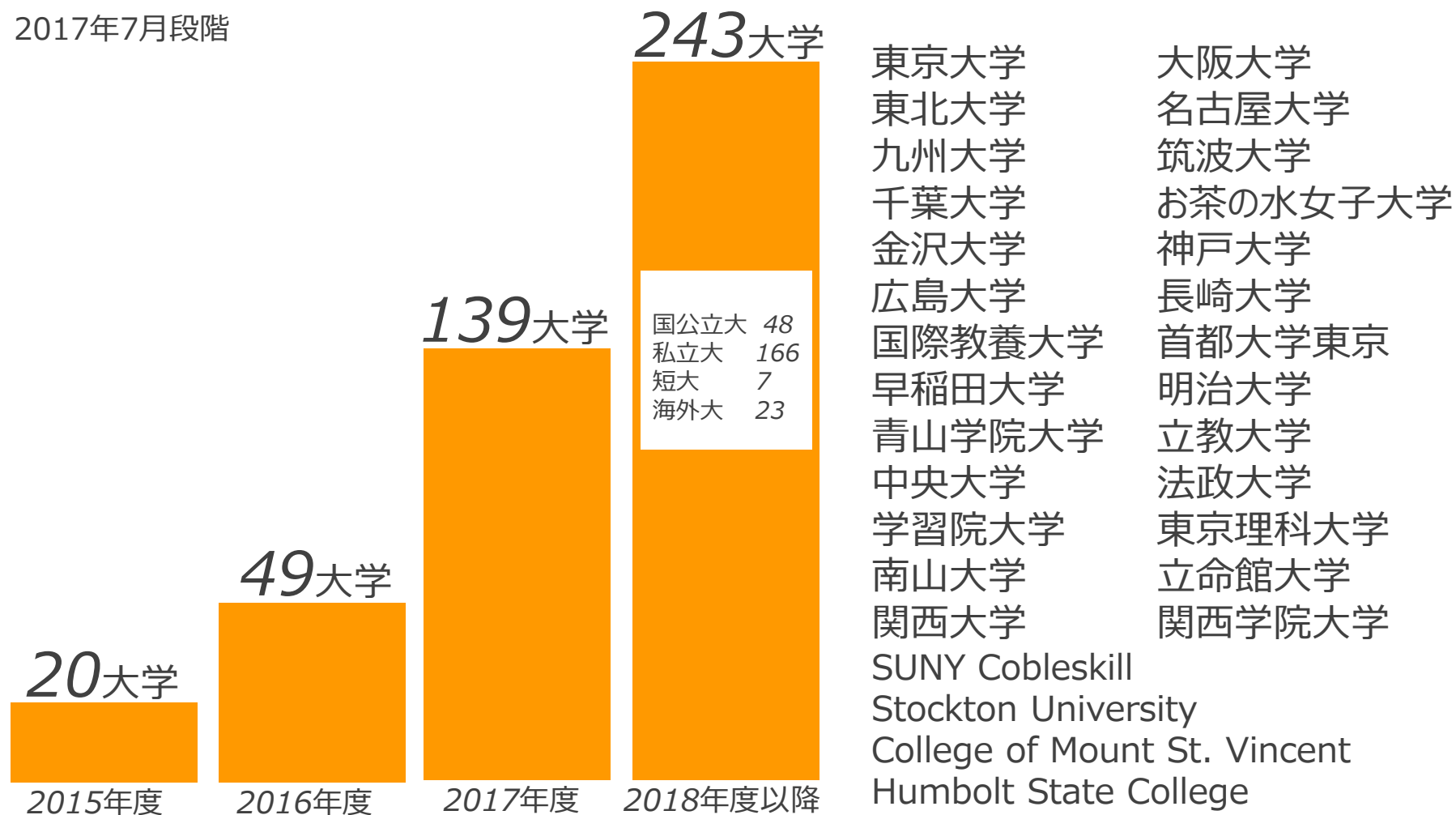
福井県のほぼ全ての中学3年生が、2016年10月14日にGTECを受験。
スピーキングテストも1台/1人のタブレットで同日・同時間で実施。



福井市立安居中学校での実施の様子

高等学校での外部英語検定試験の活用の広がりに伴い
大学入試での外部英活用も大きく広がっている

2017年7月段階



CEFRの各技能6レベルと言語テストの各技能スコアの対応づけのことを示します。Council of Europeが公表しているマニュアルにその手法が詳細に掲載されており、GTECは下記の内容を踏まえてCEFRレベルの閾値スコアを判定しています。

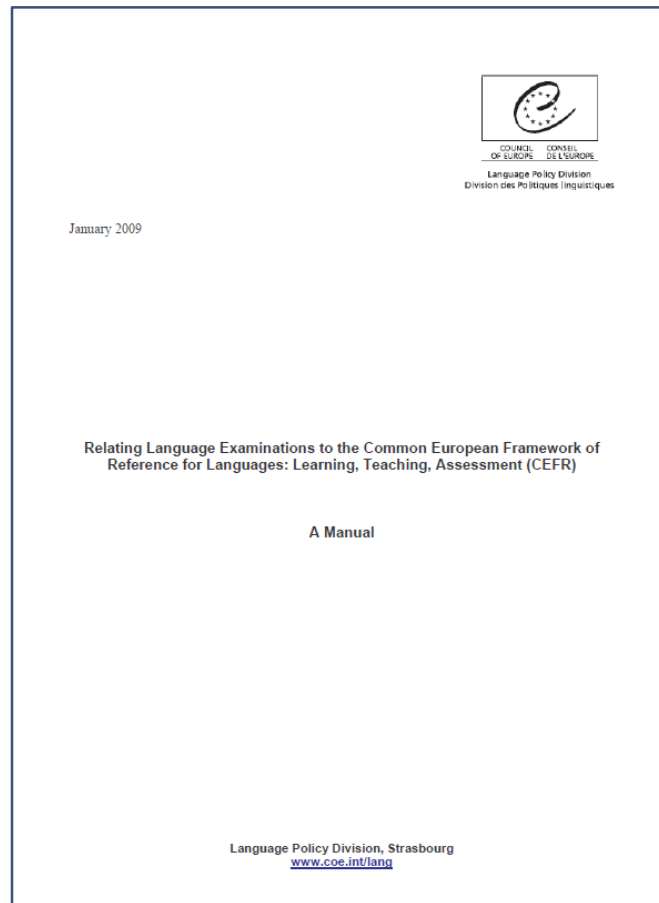
* マニュアルで紹介されている手法例

Table 6.1: Overview of the methods discussed

| Method | Section | Class |
|---|---------|-------|
| Tucker-Angoff | 6.3. | T-C |
| The Yes-No Method | 6.4.1. | T-C |
| The Extended Tucker-Angoff Method | 6.4.2. | T-C |
| The Contrasting Groups Method | 6.5.1. | E-C |
| The Borderline Group Method | 6.5.2. | E-C |
| The Body of Work Method | 6.6. | E-C |
| The Item-descriptor Matching Method | 6.7.1. | T-C |
| The Basket Method | 6.7.2. | T-C |
| The Bookmark Method | 6.8. | IRT |
| A Cito Variation on the Bookmark Method | 6.9. | IRT |

* Council of Europe Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment(CEFR) より引用

GTECでは、受容技能のリーディングとリスニングでは、Bookmark Method (Council of Europe 2009)を用いている。一方、発表技能のライティングとスピーキングではContrasting-Group Method (Council of Europe 2009)を用い、技能ごとの関連付けを行っている。



CEFRとの対応づけ ~総合判定の仕組み~

スタンダードセッティングにより、各技能ごとのCEFR閾値が算出されます。
 GTECではスコア型の検定として各技能の閾値スコアを加算し、
 英語4技能総合の閾値を示しています。

| CEFR Level | GTEC | | | | |
|------------|---------|-----------|---------|----------|-------|
| | Reading | Listening | Writing | Speaking | Total |
| C1 | | | | | |
| B2 | 280 | 290 | 300 | 320 | 1190 |
| B1 | 220 | 220 | 240 | 280 | 960 |
| A2 | 150 | 160 | 190 | 190 | 690 |
| A1 | ~149 | ~159 | ~189 | ~189 | ~689 |

STEP 1

各技能ごとのCEFR閾値を決定

STEP 2

各技能で算出されたものを加算

Ex) 総合B2=280+290+300+320=1190

<受検者ごとの判定例>

| パターン | R | L | W | S | 総合 |
|-------|-----|-----|-----|-----|------|
| 受験者 A | 240 | 250 | 220 | 190 | 900 |
| | B1 | B1 | A2 | A2 | A2 |
| 受験者 B | 300 | 320 | 300 | 300 | 1220 |
| | B2 | B2 | B2 | B1 | B2 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |

使用データ

- ◆GTEC、GTEC CBTともに、分析当時の最新回の本番データを使用。
- GTEC 2016年6月検定の問題セット
- GTEC CBT 2016年7月検定の問題セット

参画メンバー

- ◆パネルはCEFRおよび英語の言語教育、教育測定に精通した研究者6名であった。この6名は全員、2015年度のCEFR関連付け調査にも参画し、CEFRに精通したメンバーであった。
- ◆上記に加えて、GTECの作問・制作に関わっているメンバーや第二言語習得の専門家が参画し、問題形式、採点基準等の質問に答えたり、機器操作、資料の準備や議事録とりなどを行った。

スタンダードセッティングの手法の代表的なものとして下記があげられる。
大きく、①テスト項目中心、②被験者中心、③IRT
これらをベースに分類するものの3つに分けられる。(Council of Europe, 2009)

Table 6.1: Overview of the methods discussed

| Method | Section | Class | |
|---|---------|-------|------------|
| Tucker-Angoff | 6.3. | T-C | = テスト項目ベース |
| The Yes-No Method | 6.4.1. | T-C | |
| The Extended Tucker-Angoff Method | 6.4.2. | T-C | |
| The Contrasting Groups Method | 6.5.1. | E-C | = 被験者中心 |
| The Borderline Group Method | 6.5.2. | E-C | |
| The Body of Work Method | 6.6. | E-C | |
| The Item-descriptor Matching Method | 6.7.1. | T-C | = IRT中心 |
| The Basket Method | 6.7.2. | T-C | |
| The Bookmark Method | 6.8. | IRT | |
| A Cito Variation on the Bookmark Method | 6.9. | IRT | |

Council of Europe. (2009). A Manual: Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment, P.61

Listening Reading ~**Bookmark Method**~

GTECはIRTでスコアを産出しており、テスト項目ごとに困難度の数値が貼りついているため、それを客観的根拠として用い、それに加えてパネルの知見を加えた分析を入れることで、より適正に閾値設定ができると判断。

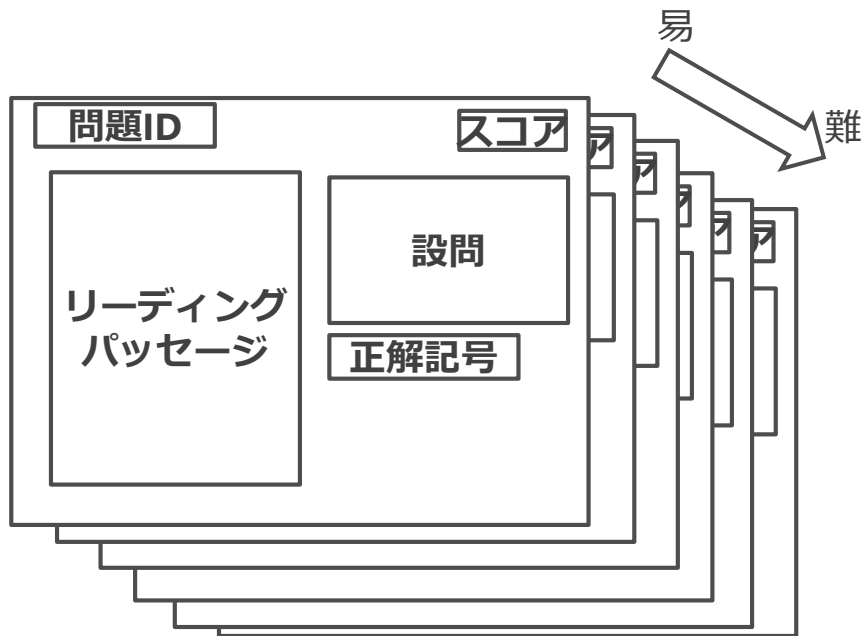
Speaking Writing ~**modified Contrasting-Group Method**~

上記と同じくGTECがIRTを用いてスコア産出している特徴を活かした手法であることに加えて、実際の受検者の産出データを詳細に分析することにより、より現実に応じた閾値設定ができると判断。

Listening Reading ~Bookmark Method~

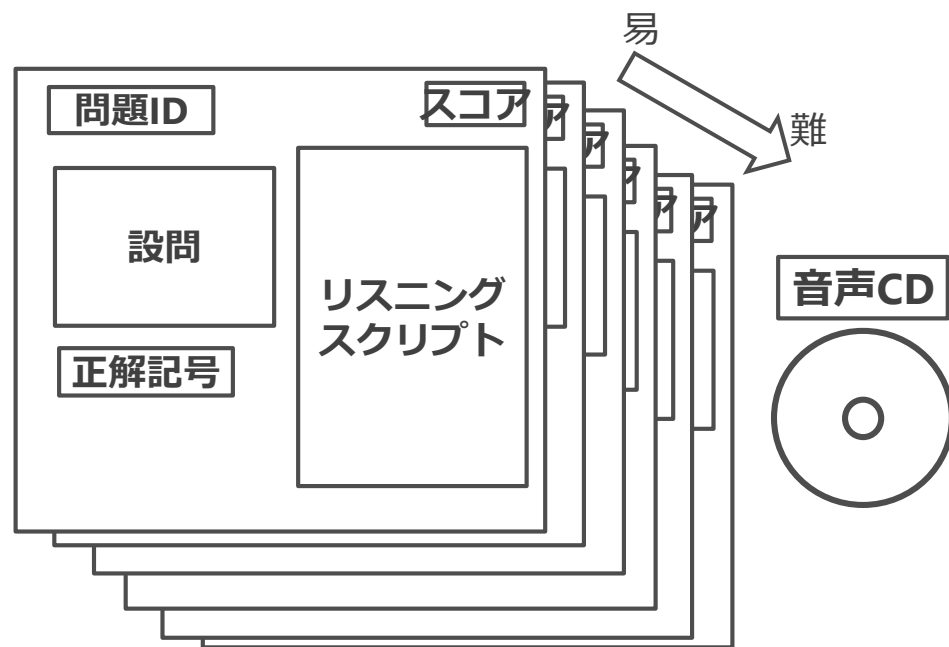
各問題を1ページあたり1問ずつ配置し、IRTの項目パラメータ情報より困難度順に易→難へと並べた小冊子 (OID: Ordered Item Booklet)を作成し、CEFRの各レベルのBorderlineに該当するページに付箋を貼っていきましました。

ReadingのBookletイメージ



※英文素材、スコア、正解記号を記載。

ListeningのBookletイメージ

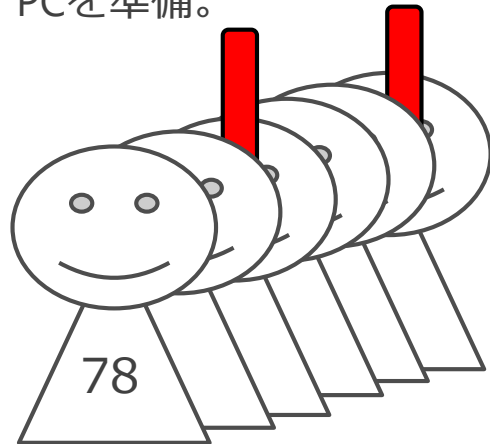


※実際の音声を聞くための音声CDも準備。

Speaking Writing ~*modified* Contrasting-Group Method~

Speaking

受検者の解答を書き起こし、問題IDごと、テスト項目ごとに解答を並べた資料を作成。また、必要に応じて実際の解答音声を聞けるように、PCを準備。



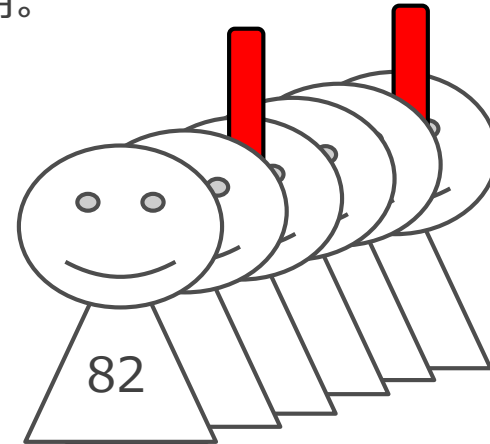
GTEC
Global Test of English Communication

テスト 問題
&
ループリック



Writing

GTEC CBTについては、試験エンジンから抽出した受検者の解答を、GTECについては、マークシートのスキャンしたものを資料として使用。



GTEC
Global Test of English Communication

テスト 問題
&
ループリック

GTEC
Global Test of English Communication

解答用紙
(GTEC)

スタンダードセッティングの流れ（各技能共通）

①

【全体会】

CEFR各レベルのディスクリプタを見ながら、
Borderline Personの英語力のイメージを議論して一致させた。

②

【分科会】

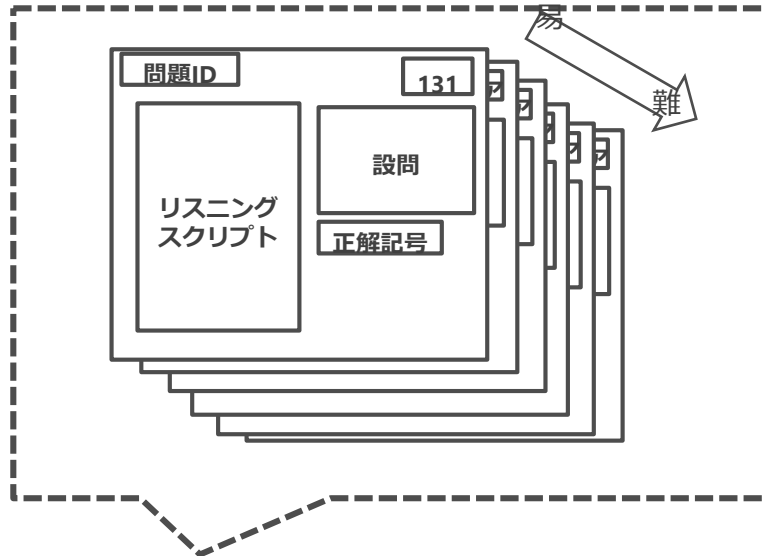
2名ずつ3グループの分科会に分かれ、
資料を見ながら分析を行い、仮閾値を設定。

③

【全体会】

再度全体会に集合して3グループ合同での協議。

▲ グループAの結果
 ▲ グループBの結果
 ▲ グループCの結果



易

難



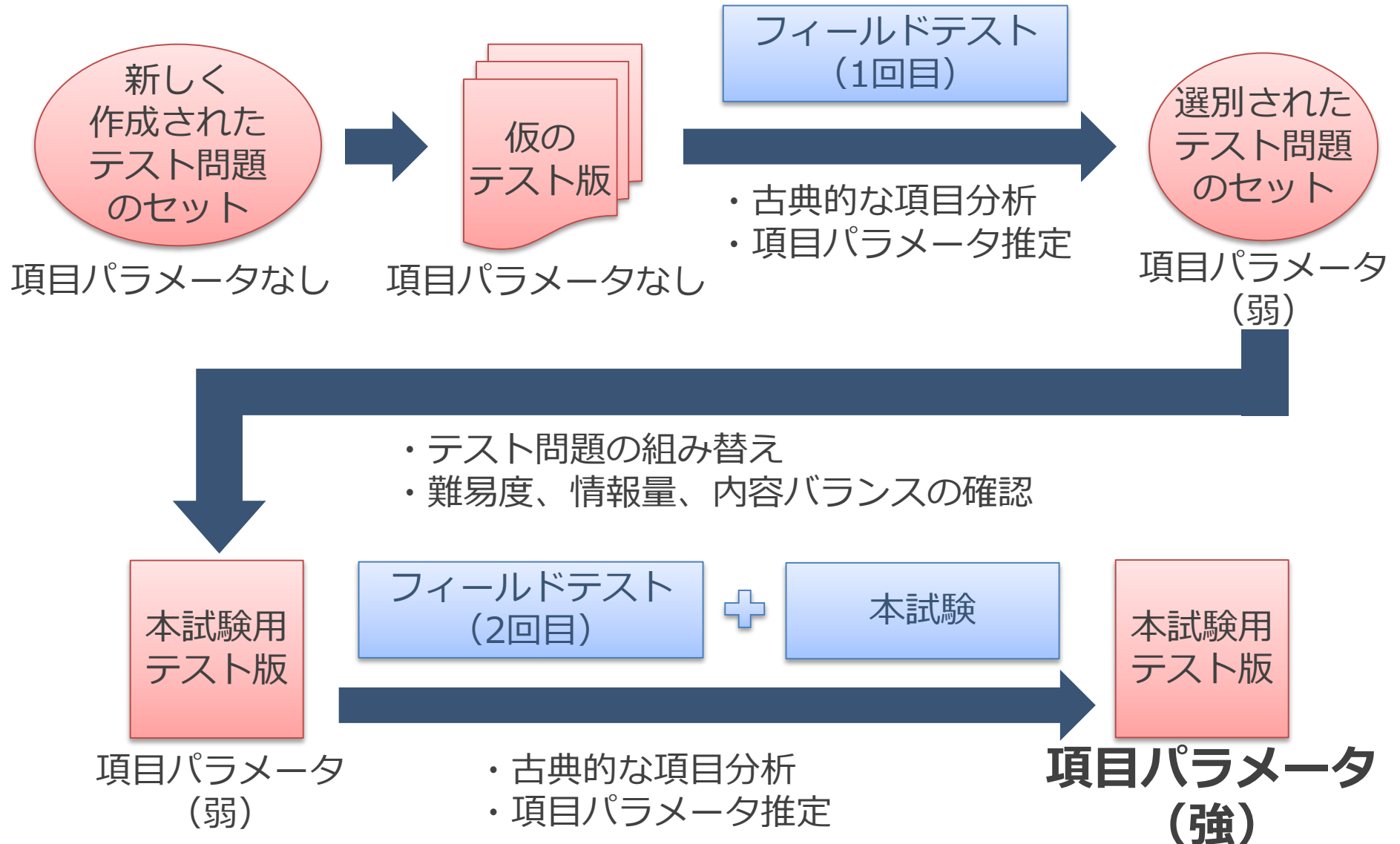
▲ ▲ ▲
 A1/A2の閾値

▲ ▲ ▲
 A2/B1の閾値

▲
▲
▲
 B1/B2の閾値

▲ ▲ ▲
 B2/C1の閾値

項目パラメータ推定の流れ



データ収集デザイン

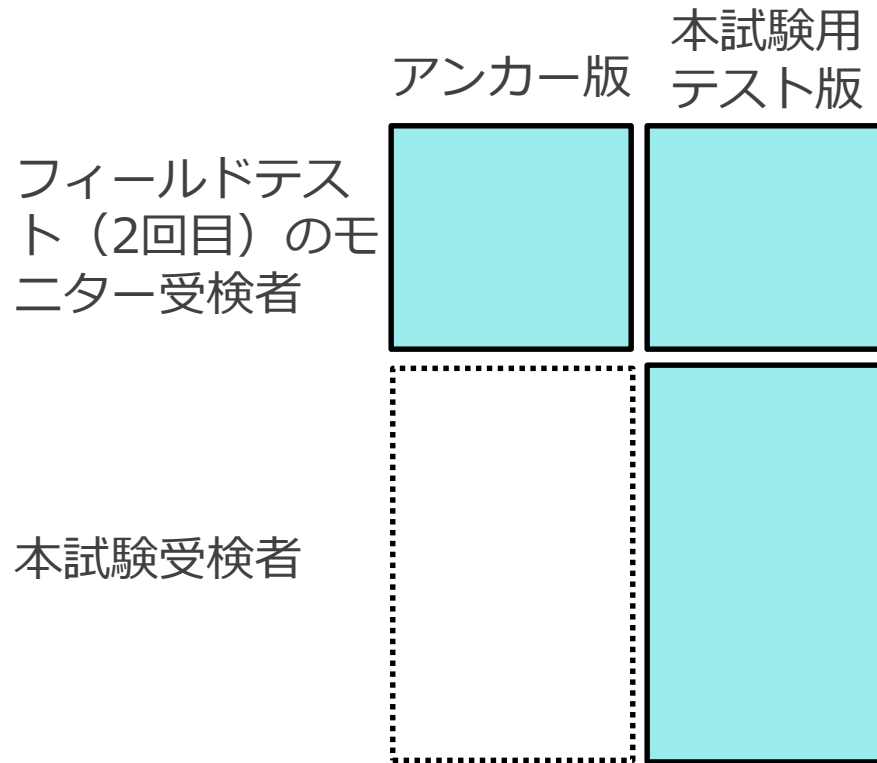
- フィールドテストでは、大学生を中心とするモニター受検者にアンカー版と新版の両方を受検してもらおう **共通受検者デザイン**が基本のデータ収集デザインとなる。

アンカー版・・・項目パラメータが確定しているテスト版

新版・・・項目パラメータを推定したいテスト版

- アンカー版を項目プールとの共通項目とみなせば、**共通項目不
等価群デザイン**と考えることもできる。
- フィールドテスト（2回目）のデータに本試験データを適宜組み合わせ、最終的な項目パラメータを推定している。
 - 本試験データを含めなくても項目パラメータは推定できるが、含めたほうがより正確な推定値になると思われる。

最終パラメータの推定



- 個別推定 + 尺度変換が基本だが、状況に応じて固定パラメータ推定も使用する。デザインの的にマルチグループで推定するほうが適切に思えるが、結論を出すためにはもっと研究が必要。

GTECにおけるスコアリングの流れ

- リスニングとリーディングでは各項目に対する正誤反応、スピーキングとライティングでは各採点項目に対する評定値に基づいて能力値を推定している。
 - 推定された能力値は、各技能内において、受検したテスト問題に関わらず**比較可能**（等化済み）である。
- 推定された能力値は、技能ごとに定める変換式によって**尺度得点**に変換される。各技能の尺度得点を足し合わせたものが**総合得点**となる。
- 尺度得点および総合得点は、スタンダード・セッティングで設定されたCEFR閾値に基づき、技能ごとおよび全体のCEFRレベルに変換される。

発表は以上です。
ご清聴いただきありがとうございました。

連絡先：

込山 智之（こみやま ともゆき）
komiyama@cees.or.jp

野澤 雄樹（のざわ ゆうき）
yuki.nozawa@mail.benesse.co.jp