

# トライアル結果

## リングスキルのテストスコアの正確性と信頼性は？

トライアルから分かったこと：リングスキルのテストスコアは信頼性が高く正確である



**信頼性：**リスニング、リーディング、全体的なテストの信頼性係数はそれぞれ0.92、0.94、0.96。0.90を超える信頼性係数は良好とみなされる。

**正確性：**正確性（精度）の目標レベルはトライアルテストの約90%で達成されました（リスニングテストで91%、リーディングテストでは88%）。目標とする精度に達しなかった試験の大部分は、CEFRの極端なレベルであった：レベルA1以下、C1以上。

### テストスコアの信頼性

リングスキルはコンピュータ適応型テストです。テスト問題は、受検者が前の問題にどれだけうまく答えているかに応じて選択されます（テストは、受検者のレベルに応じたものが出題される）。

クロンバックの $\alpha$ 係数などの信頼性を計算する典型的な方法は、受検者が異なる項目セット（問題）を解くため、適応型テストには使用できません。同様の尺度であるラッシュの信頼性が代わりに使用されます。リスニング、リーディング、および248人のトライアル参加者に基づく全体的なテストのラッシュ信頼性推定値は、すべて0.9を超えています。これは、試験の前に実行されたシミュレーションだけでなく、適応型テストの経験と一致しています。

Table 1. リスニング、リーディング、全体評価のラッシュ信頼性推定値

リスニング信頼性	リーディング信頼性	テスト全体信頼性
0.92	0.94	0.96

### テスト成績の正確性（精度）

テストスコアは、受検者の語学能力の「真の得点」の完全な推定値ではありません。測定の標準誤差（SEM）として知られる、ある程度の統計的誤差が常に存在します。受検者のテストスコアは、その時間の68%の真のスコアが1 SEM以内で、95%の真のスコアが2 SEM以内であることが期待されます。

リングスキルは、固定長テストではなく、固定精度テストです。質問数は受検者ごとに異なりますが、精度（誤差）は多かれ少なかれ一定でなければなりません。すべての得点はほぼ同じSEMでなければなりません。

ターゲットSEMは.44 ロジット（リスニングとリーディングの両方）に設定されています。ロジットは、受検者の能力を推定するために使用される統計単位です。結果を受検者に報告するCambridge English スケールと同じものではありません。

目標とするSEMは、リスニングテストの91%、リーディングテストの88%に到達しました。

目標のSEMに到達しなかった試験は、2つのカテゴリに分けられました：

**1 極度能力：**該当する受検者は、すべて（またはほぼすべて）のアイテムを正しく、または間違っ取得し、その結果、能力の推定値が極端に高くなったり極端に低くなったりします。こうしたケースは予想されており、特に問題にはなりません。

- **リスニング:** 3% が極度能力に該当 (8名の参加者: 極端に高い)
- **リーディング:** 4% が極度能力に該当 (3名の参加者が極端に高く、6名の参加者が極端に低い)

**2. 最長さ (Maximum Length)：**目標SEMが達成されるまでに投与された項目の最大数。

- **リスニング:** 目標SEMが達成されるまでにテストの6%が最長さ (Maximum Length)に達した
- **リーディング:** 目標SEMが達成されるまでにテストの8%が最長さ (Maximum Length)に達した

当該ケースをより詳細に見ると、大半の人は、目標とするSEMの精度（「.44」ロジットではなく「.5」未満）の精度を持ち、CEFRの最上位と最下位レベルであった（レベルC1以上、またはA1以下）

Table 2. 目標SEMに達するテストの割合

目標 SEMに達したか?	リスニング	リーディング
Yes	91%	88%
No – 極度能力	3%	4%
No – 最長さ	6%	8%

つまり、このトライアルテストでは、極限レベル（CEFRレベルC2およびA1）でのテストを改善するための更なる工夫は可能ではあるが、テストが正確性（精度）の目標レベルを確実に達成できるということが明らかになった。