

時事英語力の測定を目指したオンラインテスト 「CNN GLENTS」の開発と妥当性検証の試み

笹尾 洋介（京都大学）

小川 洋一郎（朝日出版社）

@The JACET 60th Commemorative International Convention (Online)

(2021.08.29)

発表概要

CNN GLENTSとは？

CNN GLENTSの開発

CNN GLENTSの妥当性の検証

発表概要

CNN GLENTSとは？

CNN GLENTSの開発

CNN GLENTSの妥当性の検証

CNN GLENTSとは

CNN Global **EN**glish **T**esting **S**ystem

CNN素材を用いたオンラインテスト

CNN GLENTSが
測定したい能力

(できる限り)
実践的な
英語運用能力

CNN GLENTSの特徴①

— Point 01

作られた英語ではなく、 生の英語ニュースが素材

リスニング問題、リーディング問題、いずれも世界最大のニュース専門放送局CNNの英語ニュースから出題。政治、経済、科学、教育、文化など幅広いジャンルから、世界の「いま」を理解するために必要な英語力を測定します。実際のニュース映像を使った「動画視聴問題」も導入しています。

<https://www.asahipress.com/special/glents/> より

CNN GLENTSの特徴②

— Point 02

場所を選ばず受験できる オンライン方式

コンピューターやスマートフォン、タブレットなどの端末とインターネット接続環境があれば、どこでも好きな場所で受けられるオンラインテストです。

※ スマートフォン、タブレットで受験される場合はアプリのダウンロードが必要です。

※ iOS版は今後対応予定です。

CNN GLENTSの特徴③

— Point 03

自動採点で結果をすぐに表示、 国際指標CEFRにも対応

テスト終了後、自動採点ですぐに結果がわかります。国際的な評価基準であるCEFRとの対照レベルやTOEIC[®] Listening and Reading Testの予測スコアも表示され、現在の英語習熟度を知ることができます。受験可能期間終了後、英文トランスクリプトなどが閲覧でき、復習すれば確実に英語力アップにつながります。

※ CEFRとは、Common European Framework of Reference for Languages(外国語の学習・教授・評価のためのヨーロッパ言語共通参照枠)を指し、外国語運用能力を示す国際指標の1つです。

<https://www.asahipress.com/special/glents/> より

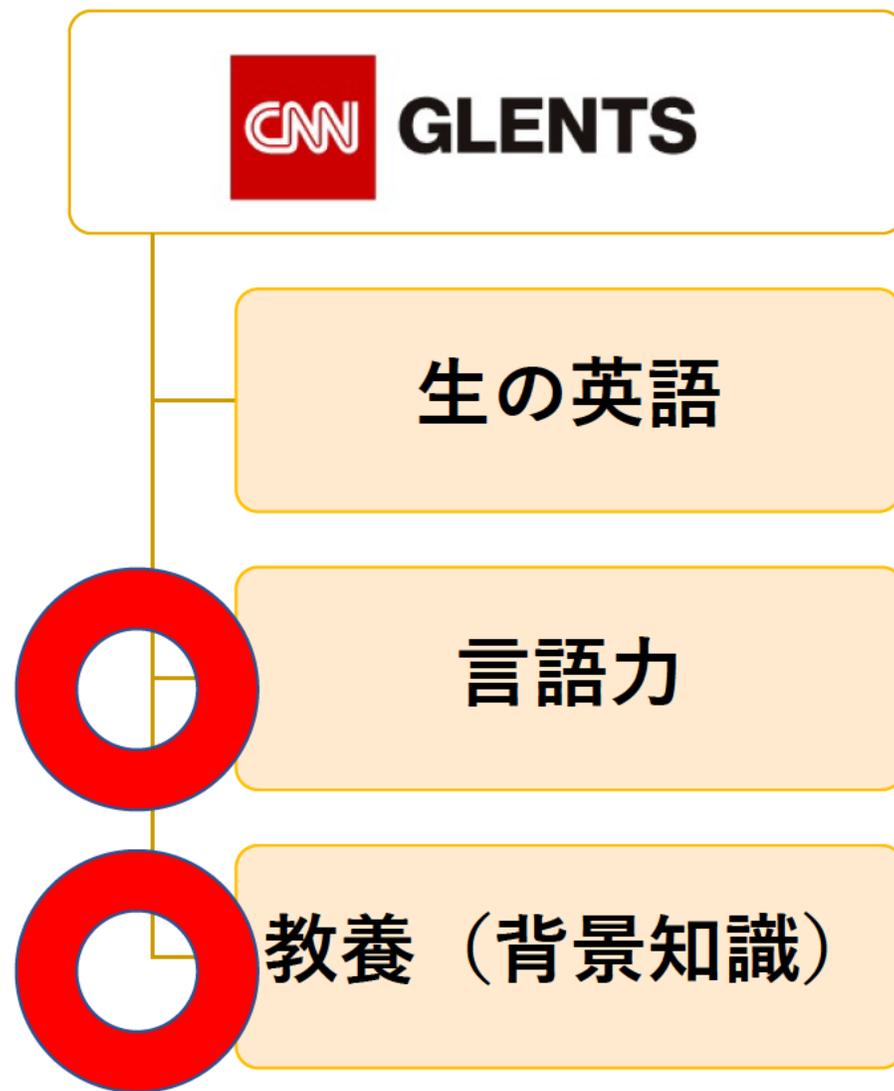
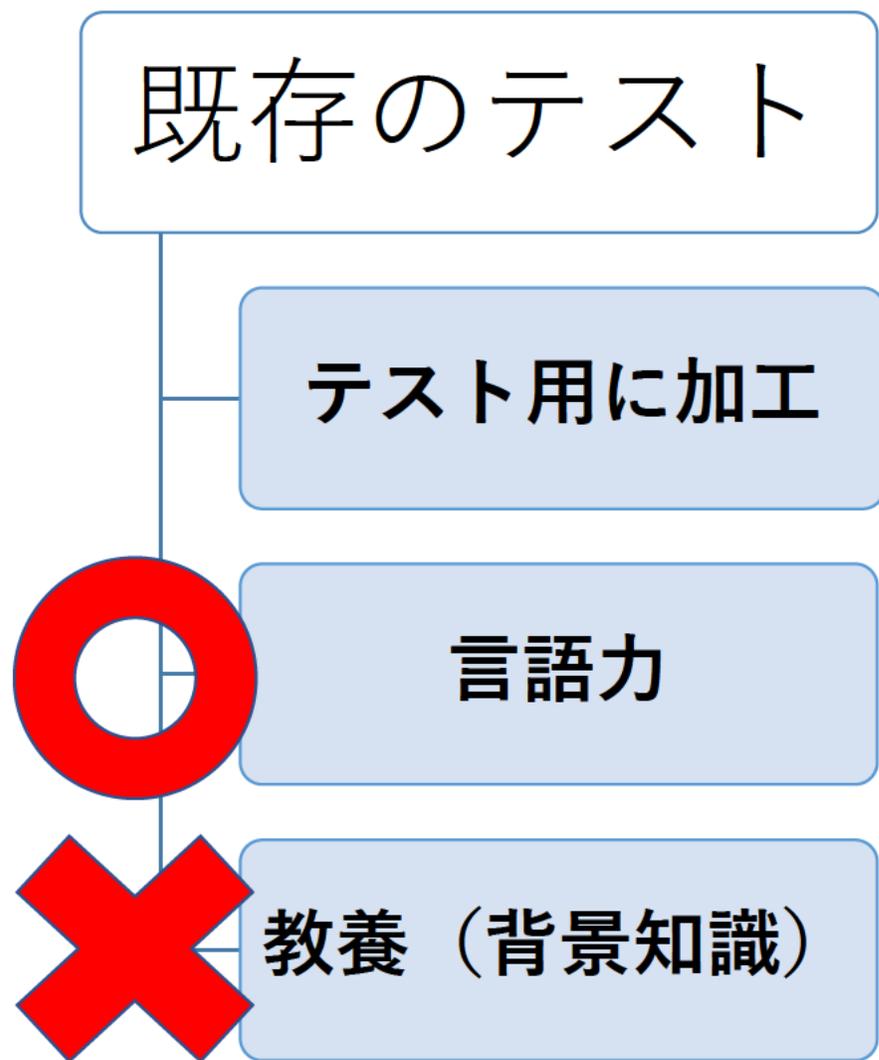
CNN GLENTSの特徴④

— Point 04

コミュニケーションに 必要な社会・文化知識にも配慮

世界の人々と円滑にコミュニケーションを図るには、一般的な英語力に加えて、近年話題になっている政治・経済・科学などの問題に関する背景知識が必須です。「国際教養セクション」では、知っておくべき世の中の情勢や今話題になっている事柄を題材にした様々な問題が登場します。

既存の英語テストとの違い（まとめ）



現実のコミュニケーション

背景知識

top down

理解

bottom up

言語知識（語彙・文法など）

発表概要

CNN GLENTSとは？

CNN GLENTSの開発

CNN GLENTSの妥当性の検証

CNN GLENTS の種類



① 「**CNN音声版**」 CNNの生音声

③ 「**CNN音声版**」 CNNの生音声

約140~180 wpm

② 「**ナレーター音声版**」 ナレーターの録音音声

④ 「**ナレーター**」の録音音声

約120~140wpm

(参考)
大学入学共通テスト
約140 wpm



一般公開テスト

問題構成

	PART	測定する能力	設問数
全30問 リスニングセクション CNN音声版(約30分)と ナレーター音声版(約35分) から選択可能	1. 書き取り	書き取り	4
	2. ショートニュース	大意把握、詳細理解	6
	3. ロングニュース	大意把握、詳細理解	8
	4. インタビュー	大意把握、詳細理解	8
	5. ビューイング (動画視聴)	動画理解 (大意把握、詳細理解)	4
全20問 リーディングセクション 25分	1. 語彙	語彙	4
	2. ショートニュース	大意把握、詳細理解	6
	3. ロングニュース	大意把握、詳細理解	10
全10問 国際教養セクション 15分	ニュースの理解に必要な 背景知識を問う問題		10

※ リスニングセクションの所要時間は目安です。試験によって変わります。

	PART	測定する能力	設問数
	1. 書き取り	書き取り	4
30～40秒 × 3	2. ショートニュース	大意把握、詳細理解	6
2分程度 × 2	3. ロングニュース	大意把握、詳細理解	8
3分程度 × 2	4. インタビュー	大意把握、詳細理解	8
30秒程度 × 2	5. ビューイング (動画視聴)	動画理解 (大意把握、詳細理解)	4
全20問	1. 語彙	語彙	4
200語程度 × 3	2. ショートニュース	大意把握、詳細理解	6
500～600語程度 × 2	3. ロングニュース	大意把握、詳細理解	10
全10問 国際教養セクション 15分	ニュースの理解に必要な 背景知識を問う問題		10



一般公開テスト

問題構成

三肢択一

(正しいものを選ぶ)

四肢択一

(誤っているものを選ぶ)

※ リスニングセクションの所要時間は目安です。試験によって変わります。

	PART	測定する能力	設問数
全24問 リスニングセクション CNN音声版と ナレーター音声版から選択可能 約20分	1. 書き取り	書き取り	4
	2. 短文理解	語彙	4
	3. ショートニュース	大意把握、詳細理解	4
	4. ニュース	大意把握、詳細理解	4
	5. インタビュー	大意把握、詳細理解	4
	6. ビューイング (動画視聴)	動画理解 (大意把握、詳細理解)	4
全16問 リーディングセクション 20分	1. 語彙	語彙	4
	2. ショートニュース	大意把握、詳細理解	4
	3. ニュース	大意把握、詳細理解	8



団体受験用テスト

問題構成

	PART	測定する能力	設問数
全24問 30～40秒 × 2 1.5分程度 × 1 1分程度 × 2 30秒程度 × 2	1. 書き取り	書き取り	4
	2. 短文理解	語彙	4
	3. ショートニュース	大意把握、詳細理解	4
	4. ニュース	大意把握、詳細理解	4
	5. インタビュー	大意把握、詳細理解	4
	6. ビューイング (動画視聴)	動画理解 (大意把握、詳細理解)	4
全16問 200語程度 × 2 300～400語程度 × 2	1. 語彙	語彙	4
	2. ショートニュース	大意把握、詳細理解	4
	3. ニュース	大意把握、詳細理解	8



団体受験用テスト

問題構成

三肢択一

(正しいものを選ぶ)

発表概要

CNN GLENTSとは？

CNN GLENTSの開発

CNN GLENTSの妥当性の検証

妥当性とは...

Validity is an **overall evaluative judgment** of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of **interpretations and actions** based on test scores or other modes of assessment.
(Messick, 1989, p.13)

妥当性の検証

1. Content
 - Relevance
 - Representativeness
 - Technical quality
2. Substantive
3. Structural
4. Generalizability
 - Item calibration invariance
 - Person measure invariance
 - Reliability
 - Invariance across administrative contexts
5. External
6. Consequential
7. Responsiveness
8. Interpretability

(American Psychological Association et al., 1999; Bachman, 1990; Chapelle, 1999; Kane, 2011; Messick, 1989, 1996; Smith Jr., 2004)

妥当性の検証

1. **Content**
 - Relevance
 - Representativeness
 - Technical quality
2. Substantive
3. Structural
4. **Generalizability**
 - Item calibration invariance
 - Person measure invariance
 - **Reliability**
 - Invariance across administrative contexts
5. **External**
6. Consequential
7. Responsiveness
8. **Interpretability**

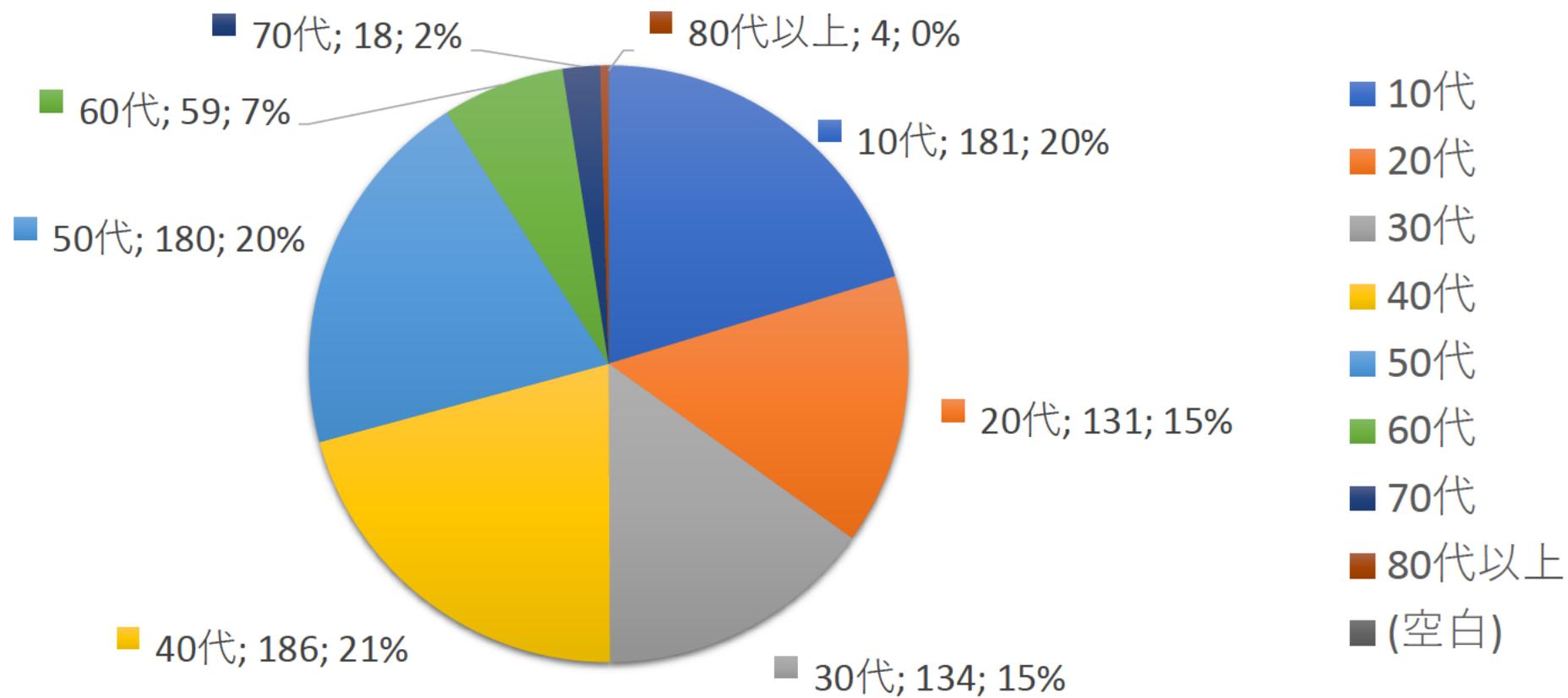
(American Psychological Association et al., 1999; Bachman, 1990; Chapelle, 1999; Kane, 2011; Messick, 1989, 1996; Smith Jr., 2004)

Study 1：第1回プレテスト

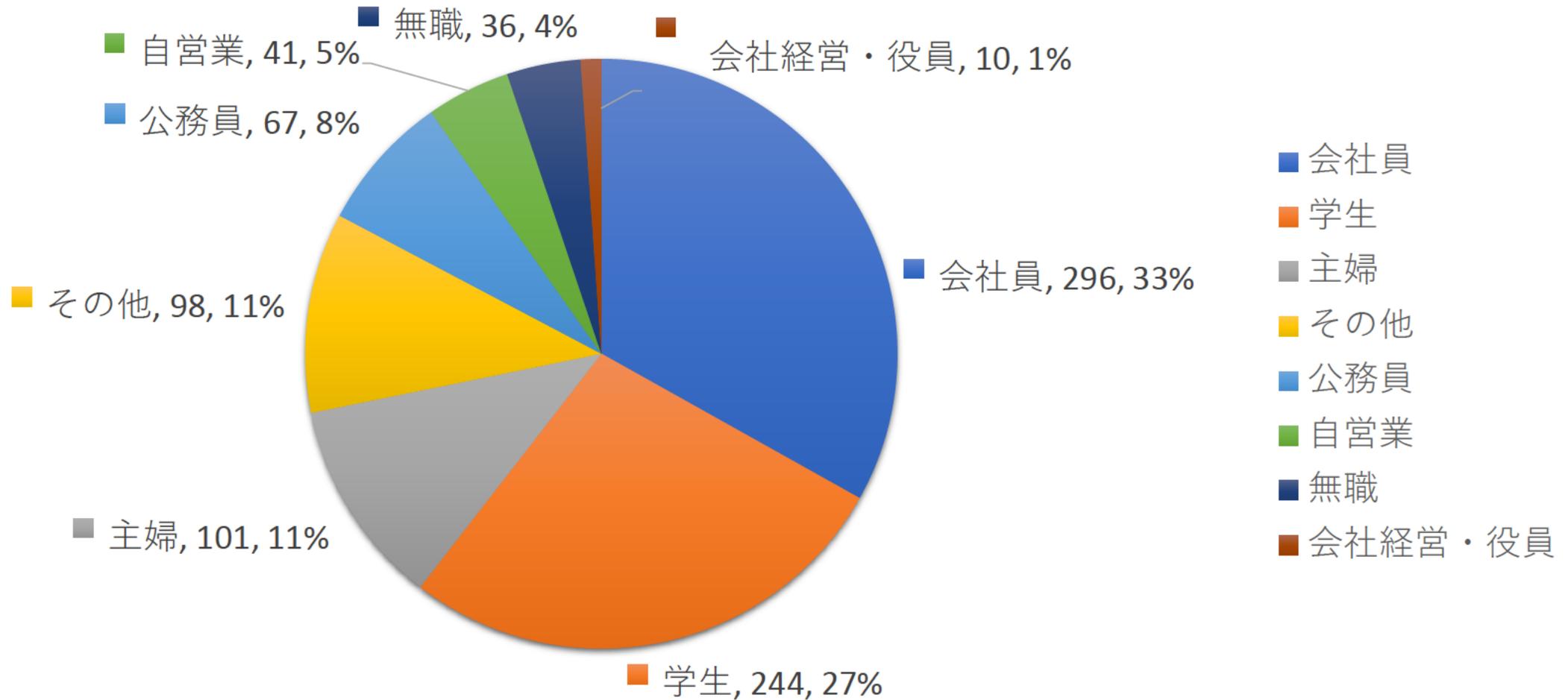
22

- 時期：2020年3月～5月
- 受験者：994名
- 素材：一般テスト（CNN音声版）
※動画問題なし
- 方法：オンライン

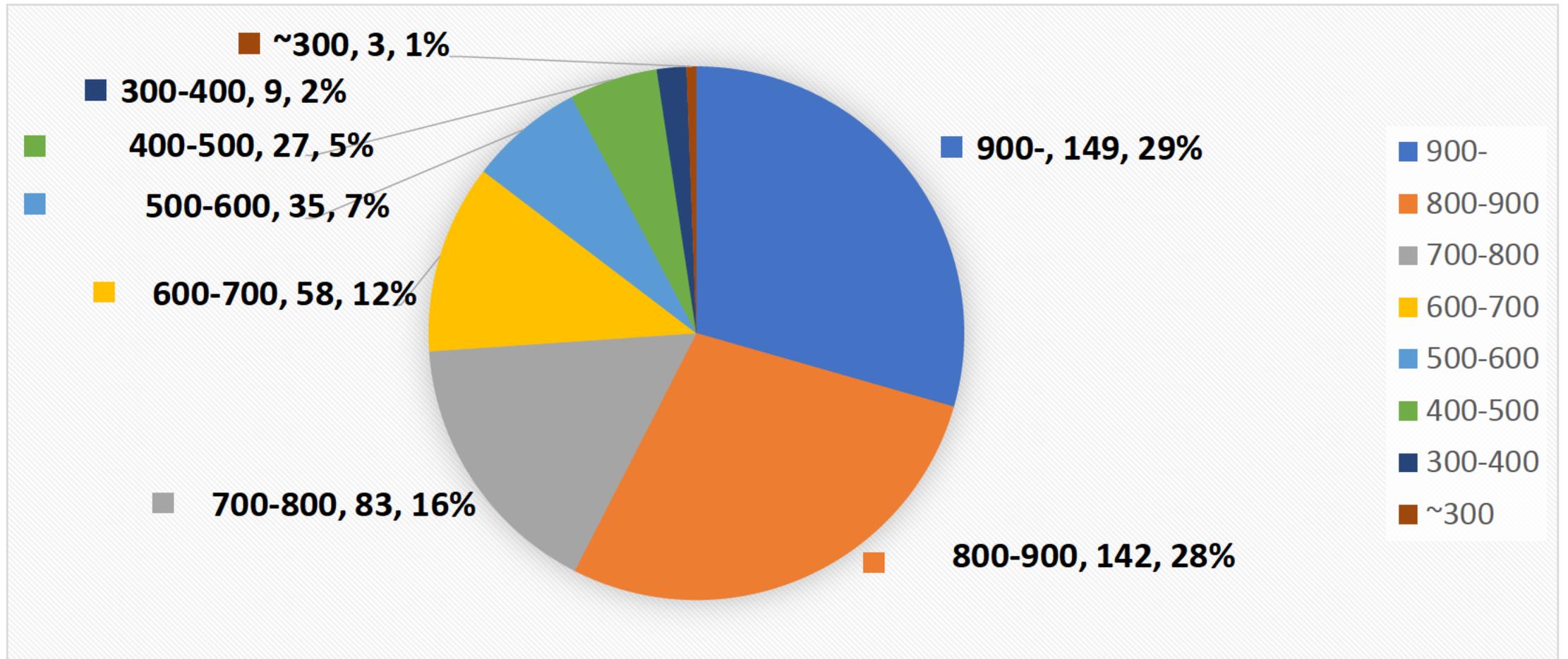
年齡



職業



TOEIC®L&R スコア分布



記述統計

全体 (N=994)

	平均	得点率	標準偏差	最大	最小
全体	59.2	59.2%	16.6	100	8
L	35.0	58.3%	10.2	60	2
R	24.2	60.6%	8.0	40	2

学生 (N=244)

	平均	得点率	標準偏差	最大	最小
全体	54.4	54.4%	15.8	92	8
L	32.0	53.3%	10.0	58	6
R	22.2	55.5%	7.8	40	2

1. Content aspect (内容的側面)

< Relevance (関連性) >

実践的英語運用能力

- ニュース素材の理解度
- 修正を施さない

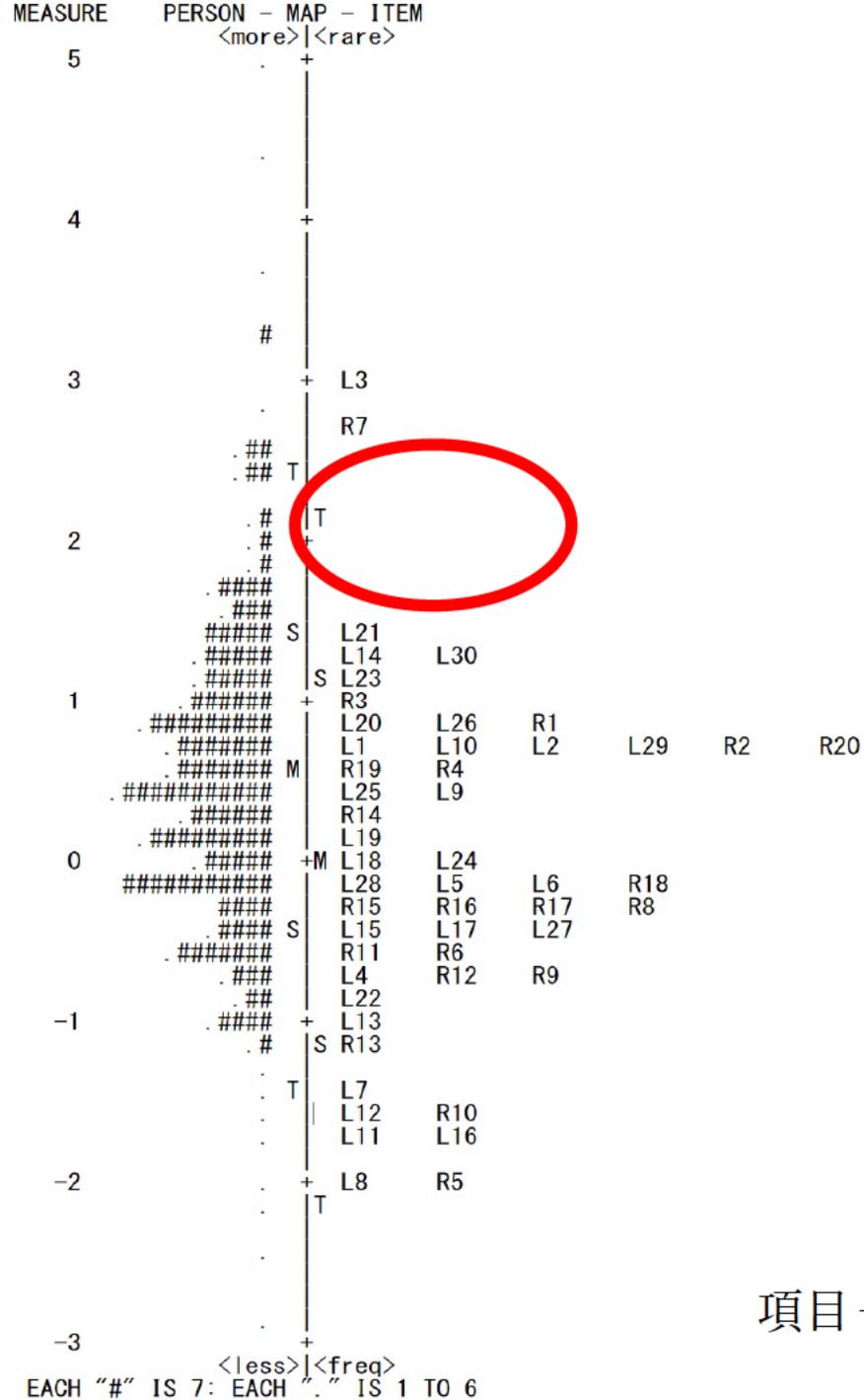
分析的測定

- 語彙
- 要旨理解
- 詳細理解



1. Content aspect (内容的側面)

< Representativeness
(代表性) >



項目－受験者マップ

1. Content aspect (内容的側面)

< Technical quality >

- Outfit
 - outlier-sensitive fit statistic.
 - sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them (and vice-versa).
- Infit
 - inlier-pattern-sensitive fit statistic.
 - sensitive to unexpected patterns of observations by persons on items that are roughly targeted on them (and vice-versa).
- Outfit MnSq > 1.5 : 1項目 (R7)
- Infit MnSq > 1.5 : なし

(Linacre, 2021)

項目 (R7)

	Data count	Data %	Ability Mean	Ability SD	Infit MNSQ	Outfit MNSQ
Correct	138	14	0.93	1.36	1.4	1.8
Incorrect 1	168	17	-0.04	0.73	0.6	0.6
Incorrect 2	681	69	0.59	0.87	1.1	1.2

4. Generalizability aspect (一般化可能性) Reliability

Cronbach alpha	Rasch person reliability	Rasch item reliability
.87	.87	.99

5. External aspect (外的側面)

自己申告のTOEIC®L&Rスコアとの相関 (Pearson r)

	L	R	合計
TOEIC	.652**	.549**	.671**

$N=506$

CNN検定結果に基づくTOEIC予測スコア (回帰式)

$$\text{予測TOEICスコア} = 6.5 \times (\text{CNNスコア}) + 360$$

6. Interpretability aspect (解釈可能性)

得点の解釈ができることが重要

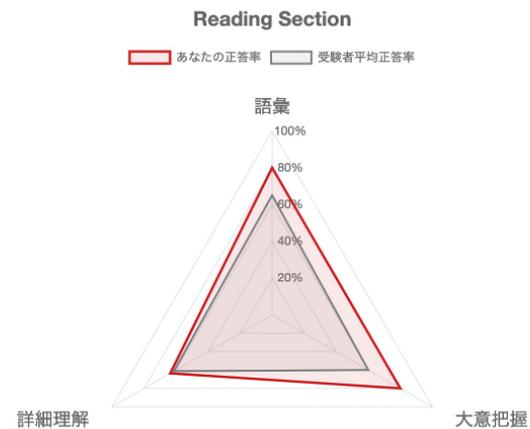
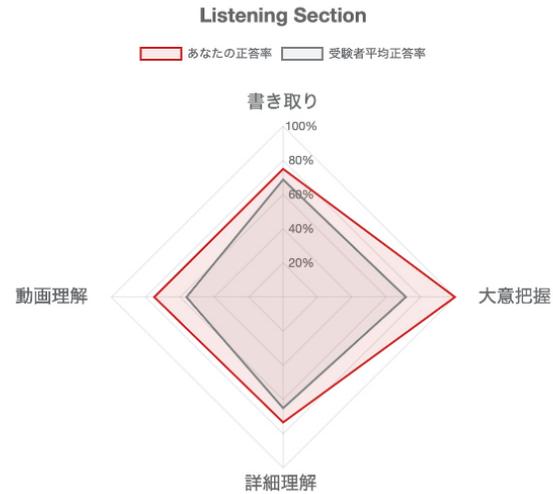
フィードバック例①

測定された能力のレーダーチャート

リスニングセクションとリーディングセクションは以下の能力を測定する設問で構成しています。

- 書き取り（リスニングセクション）／語彙（リーディングセクション）
- 大意を把握する能力
- 内容を詳細に理解する能力
- 動画を視聴して理解する能力

それぞれの能力について、あなたの成績を表すレーダーチャートは以下の通りです。



<https://www.asahipress.com/special/glents/score/> より

1. レベル診断

GLENTS スコア	66 ~ 89点
スコアのレベル	★★★☆☆ (3)
相当するCEFRのレベル	B2

社会生活で求められるコミュニケーションを行うことができる

- ・ 様々な文章を読む、または聞いて、ある程度正確に理解することができる。
- ・ 時事英語で使用される語彙・表現に関する豊富な知識を持っている。
- ・ 様々なジャンルの時事情報に精通している。
- ・ 様々な方言で話される英語をかなりの程度理解できる。

学習のヒント

語彙・表現に関する知識を広げること、様々なジャンルの素材に触れて時事知識を深めることをお勧めします。また、基礎的な英語力は十分ですので、「生の英語」に触れる機会をできるだけ多く持つことを心掛けてください。

フィードバック
例②

<https://www.asahipress.com/special/glents/score/> より

フィードバック 例③

2. CNN GLENTS (L & R Total Score) と予測されるTOEIC® Listening and Reading Testスコアとの関係

CNN GLENTS (L & R) スコア	TOEIC® L & R Test予測スコア※1
90～100点	945～990
80～89点	880～940
70～79点	815～875
60～69点	750～810
50～59点	685～745
40～49点	620～680
30～39点 ※2	350～615
0～29点 ※2	345以下

TOEIC® Listening and Reading Test 予測スコア

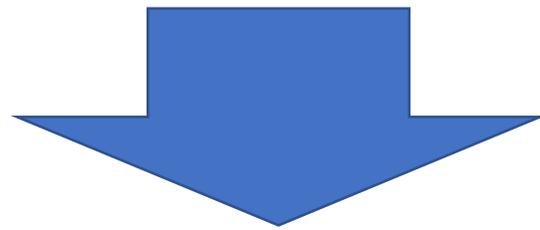
815～875

<https://www.asahipress.com/special/glents/score/> より

- **目的：NT版とNR版による能力推定値の違い（得点調整）を明らかにする**
 - ※NT版（ナチュラル音声），NR版（ナレーター音声）
- 時期：2020年7～10月
- 受験者：合計3,345名（NT版1,377名，NR版1,968名）
- 素材：一般テスト
 - ※動画問題なし
- 方法：オンライン

実験デザイン

	L_NT版 (30問)	L_NR版 (30問)	R (20問)	合計
NT版	30問	(欠損値)	20問 (共通項目)	50問
NR版	(欠損値)	30問	20問 (共通項目)	50問



項目難度と受験者能力 (推定値) を比較可能

NT版とNR版の比較

項目難度

	度数	平均値	標準偏差	最大値	最小値
NT版	30	0.149	1.09	2.86	-2.08
NR版	30	-0.047	0.86	1.92	-1.61

相関係数 (Pearson $r = .922$)

対応のあるt検定

$t(29)=2.423, p=.022$

NR版の補正

リスニング項目（各30問）だけを再分析し、NT版とNR版それぞれの受験者のRasch能力値を推定

NR版受験者の能力値に対して、一律0.19633を減じる（項目難度の平均値の差）

「① NT版受験者の能力値」と「② NR版受験者の補正済み能力値」を比較し、②が①を最小限上回る得点を求める

スコア補正

奇数：上下の能力値の平均

リスニング点	①NT版	②NR版補正
60	5.23	4.79
59	4.59	4.17
58	3.95	3.16
57	3.56	2.77
56	3.16	2.54
55	2.91	2.30
54	2.66	2.14
53	2.47	1.98
52	2.28	1.81
51	2.13	1.64
50	1.97	1.2
49	1.84	1.0
48	1.72	1.2
47	1.59	1.15
46	1.46	

1点減

3点減

NR版のスコア - 3 =
NT版換算スコア

今後の方針

1. 4つの版の換算式の作成



- ① 「**CNN音声版**」 CNNの生音声
- ② 「**ナレーター音声版**」 ナレーターの録音音声
- ③ 「**CNN音声版**」 CNNの生音声
- ④ 「**ナレーター音声版**」 ナレーターの録音音声

2. 動画セクション, 国際教養セクションの項目の分析

3. TOEIC®スコアとの関係, CEFRとの関係の精査

引用文献

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). Standards for Educational and Psychological Testing. APA.
- Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford University Press.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Kane, M. (2011). Validating score interpretations and uses. *Language Testing*, 29(1), 3-17.
<https://doi.org/10.1177/0265532211417210>
- Linacre, J. M. (2021). Winsteps® Rasch measurement computer program User's Guide. Version 5.1.1. Winsteps.com.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Smith Jr., E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch Measurement: Theory, Models and Applications* (pp. 93-122). JAM Press.



(c) 2021 Cable News Network. A WarnerMedia Company. All Rights Reserved.