

## 共通テストと民間試験導入

### CEFR、波及効果その他の課題を実証研究の成果をみながら検証する

大学入試のあり方に関する検討会議 (第6回)

2020年(令和2年)4月23日

渡部良典

#### 前提となる教育評価の知識の確認

1. 適正な教育評価には満たすべき条件がある。
  - **信頼性** = 測定の安定性
  - **妥当性** = 測定したい能力を測定しているか・結果から対照能力がわかるか
  - **実用性** = **実行可能性** (feasibility)
  - **公平性** (fairness)
  - **真正性** (authenticity)
  - **教育効果** (educationally beneficial effects) ≡ 波及効果 (**washback effects**)

現実にはこれらすべての条件を満たすことは不可能である。大学入学試験では**信頼性**、**実行可能性**、**公平性**が特に重要である。( \*厳密な意味では妥当性は常に最も重要な課題ではあるが、ここでは素朴な意味で、テストの内容が測定したい対象能力と一致しているかどうかという意味で使用する。)

2. **過去を検証する到達度評価と将来を予測する熟達度評価**
  - カリキュラムの学習内容の習得状況を確認するために使うテストを**到達度テスト (achievement test)** という。センター試験・共通テスト日本の学習指導要領を基盤にして開発・作成された到達度テストである。
  - 将来必要となる課題をこなす知識や力があるかどうかを調べるために使うテストを**熟達度テスト (proficiency test)** という。TOEFL や IELTS は後者である。英語圏の高等教育で必要とされる能力 (ニーズ) を分析して開発・作成された熟達度テストである。
3. **論点は、各種外部試験が共通テストの替わりとなり得るかどうかこの一件だけである。**
  - **4 技能測定のためのテスト ≠ 共通テスト / センター試験 ≠ 民間試験**
  - 各種テストの質を検証するものではない。
  - 各大学で行われている民間テストの使用については扱わない。
  - 国公立大学の2次試験で行われているスピーキング・テストを対象としているのではない。

## 論点 1 : CEFR による各テストの配列換算表はどこまで信頼できるのか

- A1、A2、B1、B2、C1、C2 各レベルは価値観を示しているわけではない。
  - 配列換算表のテストはそれぞれが異なる方法で英語能力の異なる側面を測定対象としている。
  - 各テストの得点は必ずしも can-do で記載されている能力ができることを保証しない。
  - 測定値には必ず誤差が含まれる。各レベルの分割点は不安定である。
1. 「各資格・検定試験と CEFR との対照表」(参考資料 A、3 頁) では到達度テストと熟達度テストの区別がされておらず、各テストの目的が全く区別されずに並べられている。
  2. CEFR の各レベルは価値判断を示すための指標ではない。C2 は C1 よりも優れている、C1 は B2 よりも優れている、…A1 は最低である、ということを示しているわけではない。A1 レベルに記載されている (can-do) ことができることを示しているのであり、たとえば、A1 は A2 より不完全である、劣っているという意味ではない。どのレベルの測定を狙っているかはテストによって異なる。例えば、A1 (「具体的な欲求を満足させるためのよく使われる日常的表現、基本的な言い回しは理解し…」) レベルを TOEFL は測定できないというよりも、測定対象としていないのである。従って、A1～C2 それぞれを測定の対象とするならば、それぞれのレベルの到達度を示す必要がある(渡部改訂試作版、参考資料 B、4 頁)。
  3. 今回の改革の焦点スピーキング・テストについて。スピーキングと言っても理解した内容を口頭で伝える、思いを口頭で伝える、音声言語をつかってやり取りをする、依頼する等々多様な種類がある。各テストは方法が違い、測定対象の能力も違う(参考資料 C、5 頁)。例えばケンブリッジ英検は受験者どうしに対話をさせる。TOEFL iBT は読んで、聞いて理解した内容を PC に録音させる。IELTS は英語の母語話者との面接、TEAP は対面式、英検と G-TEC は音読と内容に関する質疑応答、絵の描写などである。英検は 2 次試験において対面式で実施され、G-TEC は録音式である。かりに A2 レベルと判断されてもその A2 の指し示す話す能力は各テストによって内容が違う。(参考資料 D、6 頁)。
  4. CEFR は言語能力を規定するために、can-do statement、すなわち「～ができる」など言語の機能・運用能力を対象としている。多くの研究者が客観的尺度の作成に取り組んでいるが、採点者の解釈によってレベル設定も異なりうる。レベルを示す得点は極めて不安定である(Green, 2018; Papageorgiou, 2015 他)(参考資料 E、6 頁)。テストの得点は Can-Do で記載されていることができることを必ずしも保証しない。
  5. 1 点刻みの合否判定の替わりとして CEFR を用いた資格試験とすることが改革の主旨の一つとなっているようであるが、CEFR のレベル設定にテストの得点を使う限りは、1 点刻みの得点を使わざるを得ない。

- CEFR は**複言語主義** (plurilingualism) の理念に基づいて、ヨーロッパ諸言語の間に共通の指標を設けるために開発された言語能力の基準である (Council of Europe, 2001) 。
- 欧米の大学では B2 レベルを入学許可基準とすることが多い。

**参考資料 A**

**各資格・検定試験とCEFRとの対照表**

文部科学省 (平成30年3月)

CEFR	ケンブリッジ 英語検定	実用英語技能検定 1級-3級	GTEC Advanced Basic Core CBT	IELTS	TEAP	TEAP CBT	TOEFL IBT	TOEIC L&R/ TOEIC S&W
<b>C2</b>	230   200			9.0   8.5				
<b>C1</b>	199   180	3299   2600	1400   1350	8.0   7.0	400   375	800	120   95	1990   1845
<b>B2</b>	179   160	2599   2300	1349   1190	6.5   5.5	374   309	795   600	94   72	1840   1560
<b>B1</b>	159   140	2299   1950	1189   960	5.0   4.0	308   225	595   420	71   42	1555   1150
<b>A2</b>	139   120	1949   1700	959   690		224   135	415   235		1145   625
<b>A1</b>	119   100	1699   1400	689   270					620   320

○ 表中の数値は各資格・検定試験の定める試験結果のスコアを指す。スコアの記載がない欄は、各資格・検定試験において当該欄に対応する能力を有していると認定できないことを意味する。  
 ※ ケンブリッジ英語検定、実用英語技能検定及びGTECは複数の試験から構成されており、それぞれの試験がCEFRとの対照関係として測定できる能力の範囲が定められている。当該範囲を下回った場合にはCEFRの判定は行われず、当該範囲を上回った場合には当該範囲の上限に位置付けられているCEFRの判定が行われる。  
 ※ TOEIC L&R/ TOEIC S&Wについては、TOEIC S&Wのスコアを2.5倍して合算したスコアで判定する。  
 ※ 障害等のある受検生について、一部技能を免除する場合等があるが、そうした場合のCEFRとの対照関係については、各資格・検定試験実施主体において公表予定。

参考資料 B

	A1	A2	B1	B2	C1	C2	
	基礎段階の言語使用者		自立した言語使用者		熟達した言語使用者		
5. ほぼ完璧な程度まで達成	5	5	5	5	5	5	
4. かなりの程度まで達成	4	4	4	4	4	4	
3. 半分程度の達成度で	3	3	3	3	3	3	
2. ある程度までは達成	2	2	2	2	2	2	
1. 最低限のレベルは満たしている	1	1	1	1	1	1	
言語能力一般	具体的な欲求を満足させるための、よく使われる日常的表現と基本的な言い回しは理解し、用いることもできる。自分や他人を紹介することができ、どこに住んでいるか、誰と知り合いか、持ち物などの個人的情報について、質問をしたり、答えたりできる。もし、相手がゆっくり、はっきりと話して、助け船を出してくれるなら簡単なやり取りをすることができ	ごく基本的な個人的情報や家族情報、買い物、近所、仕事など、直接的関係がある領域に関する、よく使われる文や表現が理解できる。簡単に日常的な範囲なら、身近で日常の事柄についての情報交換に応じることができる。自分の背景や身の回りの状況や、直接的な必要性のある領域の事柄を簡単な言葉で説明できる。	仕事、学校、娯楽で普段出会うような身近な話題について、標準的な話し方であれば主要点を理解できる。その言葉が話されている地域を旅行しているときに起こりそうな、たいていの事態に対処することができる。身近で個人的にも関心のある話題について、単純な方法で結びつけられた、脈絡のあるテキストを作ることができる。経験、出来事、	自分の専門分野の技術的な議論も含めて、抽象的かつ具体的な話題の複雑なテキストの主要な内容を理解できる。お互いに緊張しないで母語話者や取り取りができるくらい流暢かつ自然である。かなり広汎な範囲の話題について、明確で詳細なテキストを作ることができ、さまざまな選択肢について長所や短所を示しながら自己の視点を説明できる。	いろいろな種類の高度な内容のかなり長いテキストを理解することができ、合意を把握できる。言葉を探しているという印象を与えずに、流暢に、また自然に自己表現ができる。社会的、学問的、職業上の目的に応じた、柔軟な、しかも効果的な言葉遣いができる。複雑な話題について明確で、しっかりとした構成の、詳細なテキストを作ることができる。その	聞いたり、読んだりしたほぼ全てのものを容易に理解することができる。いろいろな話し言葉や書き言葉から得た情報をまとめ、根拠も論点も一貫した方法で再構成できる。自然に、流暢かつ正確に自己表現ができ、非常に複雑な状況でも細かい意味の違いを表現できる。	
他に「読む」、「聞く」、「書く」、「口頭でやり取り」、「口頭で表現」の記述がある。ここでは省略							
	A1	A2	B1	B2	C1	C2	
	基礎段階の言語使用者		自立した言語使用者		熟達した言語使用者		
ケンブリッジ英検	ケンブリッジA2 Key						
	ケンブリッジB1 Preliminary				ケンブリッジB2 First		
	ケンブリッジC1 Advanced					C2 Proficiency	
	英検3級						
	英検準2級			準1級			1級
Core							
Basic							
Advanced							
CBT							
IELTS			4.0 - 5.0	5.5 - 6.5	7.0 - 8.0	8.5 - 9.0	
TEAP	135 - 224		225 - 308	309 - 374	375 - 400		
TEAP CBT	235 - 415		420 - 595	600 - 795	800		
TOEFL iBT			42 - 71	72 - 94	95 - 120		

**参考資料 C**

名称	年間受験者数	得点の有効期限	実施頻度・時期	主たる試験会場	受験料	スピーキング・テストの方法と測定対象能力
ケンブリッジ英検	550 万	制限なし	2-3 回	札幌、仙台、東京、静岡、長野、名古屋、大阪、広島、福岡	11,880 円~9,720 円	受験者同士の対話 写真について質疑応答、ディスカッション等
実用英語技能検定	339.4 万	制限無し 留学用は 2 年間	3 回	(周辺エリアを含む) 北海道 16, 東京 23 以上 (「その周辺」)、他 227 カ所以上 (「その周辺」)	5,800 円~4,800 円	録音式：音読、内容理解の質疑応答、絵画描写等
GTEC	126 万	2 年間	2 回から 5 回	4 技能：札幌 2、仙台、船橋、新宿、飯田橋、武蔵小杉、横浜、名古屋栄、東梅田、大阪、岡山、広島、福岡	5,980 円~9,900 円	録音式：音読、内容理解の質疑応答、絵画描写等
IELTS	3.7 万 全世界 290 万	2 年間	40 回	東京テストセンター：札幌、仙台、埼玉、東京、横浜、長野、金沢、静岡 // 大阪テストセンター：名古屋、京都、大阪、神戸、岡山、広島、福岡、熊本	25,380 円	面接官との対話、インタビュー、スピーチ、ディスカッション
TEAP	2.4 万	2 年間	3 回	26 都道府県 (北海道、宮城、秋田、福島、茨城、栃木、群馬、埼玉、千葉、東京、神奈川、新潟、富山、石川、長野、静岡、愛知、京都、大阪、兵庫、広島、香川、福岡、長崎、熊本、沖縄)	6,000 円 (2 技能) ~15,000 円 (4 技能)	面接官との対話、ロールプレイ、スピーチ、等
TEAP CBT	700	2 年間	3 回	全国 13 都道府県 (北海道、宮城、埼玉、千葉、東京、神奈川、長野、愛知、京都、大阪、兵庫、広島、福岡)	15,000 円 (4 技能)	録音式：質疑応答、要約、読解・図表等。
TOEFL iBT	非公表	2 年間	40-45 回	北海道札幌 3, 東北 (秋田 1, 宮城-仙台 2)、東京 21, 中国 (広島 2)、四国 (香川 1、高知 1) 他 44 カ所、	\$ 235	録音式：与えられたテーマについて、読解・聴解についてまとめ等

参考：英語 4 技能資格・検定試験懇談会運営 (n.d.) . 英語 4 技能試験情報サイト [http://4skills.jp/qualification/cambridge\\_English.html](http://4skills.jp/qualification/cambridge_English.html) 202

0 年 2 月 10 日

参考資料 D

異なる能力を測定している可能性が高い

IELTS と TOEFL iBT の相関

Score	相関係数
IELTS Listening & TOEFL Listening	0.63
<b>IELTS Speaking &amp; TOEFL Speaking</b>	<b>0.57</b>
IELTS Reading & TOEFL Reading	0.68
<b>IELTS Writing &amp; TOEFL Writing</b>	<b>0.44</b>
<b>IELTS Total &amp; TOEFL Total</b>	<b>0.73</b>

ETS (2010) [https://www.ets.org/s/toefl/pdf/linking\\_toefl\\_ibt\\_scores\\_to\\_ielts\\_scores.pdf](https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf)

参考資料 E

レベルを示す得点は不安定である

	文部科学省 (2019)	ETS (2015)	ETS (2008)
<b>C2</b>			
<b>C1</b>	120 – 95	120 – 95	120 – 110
<b>B2</b>	94 – 72	94 – 72	109 – 87
<b>B1</b>	71 – 42	71 – 42	86 – 57
<b>A2</b>			
<b>A1</b>			

Pagageorgiou, et al., (2015) <https://www.ets.org/Media/Research/pdf/RM-15-06.pdf>

ETS (2008) [http://language.sakura.ne.jp/icnale/images/about/toefl\\_mapping.pdf](http://language.sakura.ne.jp/icnale/images/about/toefl_mapping.pdf)

各レベルの得点 (cut-off scores) が  
2015 年では 2008 年の公表値よりも低く  
なっている。

## 論点 2 : 民間テスト代行案採用で公平性は保てるのか

- 民間テスト代行案では公平性が確保できない。(参考資料 C)
  - 受験会場が限られていて、地域による格差が大きい。
  - 受験料が異なる。
  - 実施回数が異なる。
  - スピーキング・テストの測定対象能力がことなる。
  - 障がい者への配慮が一律ではない。(特に配慮したスピーキング・テストは実施されているのだろうか。)
- 民間テストでは緊急時に対応が困難である。
- 本来は大学当局が各テストの目的等の詳細を熟知し、適正に使用しなければならぬにもかかわらず、公開されている情報は受験者向けである。

### 論点3：「よい」テストは「好ましい」波及効果をもたらすか？

(小職は、博士論文のテーマとして以来、国内外の研究を広く検証してきました。)

- 実証研究の示すところ大規模テストの波及効果 (washback effects) は極めて限定的である。
  - テストの質と質の高いテスト運用は別次元の課題である。
  - 波及効果には負の効果もありうる。スピーキング・テスト対策として文法事項を教えパターン化した指導が紹介されている例がある。
  - 効果は必ず指導者、学習者、使用者を媒介としてもたらされる。受験指導でも工夫をして「使える英語」を指導している実例は数多くある。
  - 社会的影響 (societal impact) としてすでにできる生徒に利益になりうることもありうる。
1. 改革の要点の一つにテストを変えることによって教育指導学習が向上するという期待がある。波及効果 (Washback effects) は 1990 年代から世界各地で実証研究が行われている。
  2. どの研究もテストを変えることによって得られる効果は極めて限定的であることを示している。
  3. 例示：検証はテストの分析 → 効果の予測 → 事前面談 → 授業観察 → 事後面談 → 授業の記録 (文字起し等) → カテゴリー別に頻度・継続時間・型を比較分析 → 必要があれば統計検定 → 結果の解釈、という手順で進む (Watanabe, 2004)。
  4. 高等学校の受験対策授業・通常の授業、予備校の夏期講習・冬期講習を含めた総計 24 時間の授業観察。観察の前後に授業担当の先生方に面談を行った。授業は、「受験技術への言及」、「文法用語使用頻度」、「語の和訳」、「句・節の和訳」、「文の和訳」、「生徒の発話の訂正」、「英語の音声指導」、「発話の練習」、「英語の使用時間」、「日本語の使用時間」、「授業の雰囲気 (笑いの頻度)」、「講義形式の時間」、「ペア・ワークの時間」、「グループ・ワークの時間」、「生徒の口頭発表」、「個人作業」、これらのカテゴリーについて分析した。

図 1 高等学校の授業—波及効果が存在すると仮定された場合の条件

	教員 A		教員 B
受験を意識した授業	↑	← [同じ] →	↑
	[異なる]		[異なる]
通常授業	↓	← [同じ] →	↓

図 2 予備校における授業—波及効果が存在すると仮定された場合の条件

	教員 A		教員 B
大学 C の入試問題	↑	← [同じ] →	↑
	[異なる]		[異なる]
大学 D の入試問題	↓	← [同じ] →	↓



## 結果

図2 高等学校の授業

	教員 A		教員 B
受験を意識した授業	↑	←[同じ]→[違う]	↑
	[異なる] [似ている]		[異なる] [似ている]
通常授業	↓	←[同じ]→[違う]	↓

図2 予備校における授業

	教員 A		教員 B
大学 C の入試問題	↑	←[同じ]→[違う]	↑
	[異なる]		[異なる]
大学 D の入試問題	↓	←[同じ]→[違う]	↓

5. テストの種類の違いよりも教員の指導方法の違いが大きい。同様の報告が世界各地からなされており、その中には米国の NCLB（おちこぼれゼロ政策）等外国語以外の研究もある（Fullan, 2010）。すなわち、テストの効果は必ずテスト使用者を媒介ともたらされる。
6. テストと指導法の間には直接の関係はない。波及効果はテストの質とは直接関係がない。質の高いテストが好ましい効果をもたらすわけではない。テストの質と質の高いテスト運用は別次元の課題である。
  - 事例 1) すべて英語で出題されているテストのための指導でもすべて英文和訳を行い、詳細な文法説明を行っている指導、2) 英文を日本語で要約させる問題を生徒のレベルにあわせて問題を作り、生徒間でやりとりをさせながら理解を深めさせる指導、3) 和文英訳の問題を使って、生徒にトピックセンテンスを与え、グループで沢山の関係のある文を作らせ、黒板に書かせ、それをつかってまとまりのある英文を書かせる、4) 時間を与え過去問を解かせ答え合わせ、5) 対話文をつかって実際にやりとりの練習をさせる、等々。
7. 各国で行われているどの実証研究も同様の結果を示している。
  - (香港) “... the washback effect of the new examination on classroom teaching was limited despite expectations to the contrary” (Cheng, 2005). 新テストの波及効果は期待に反して極めて限られていた。
  - (ギリシア) “... many other factors beyond the test, the teachers or students ... need to be taken into account when studying the washback effect of a high-stakes exam to explain why washback takes the form it does in a given context” (Tzagari, 2009). 重要なテストの波及効果を考察するためには教員や学習者を含めたテスト以外の様々な要因を考慮しなければならない。
    - ➔ Washback is a highly complex phenomenon ... simply changing the contents or methods of an examination will not necessarily bring about direct and desirable changes in teaching and learning (Tzagari, & Cheng, 2017, p.364) 波及効果は極めて複雑な事象であり、単にテストを変えるだけでは指導や学習に期待した効果は望めない [ことを実証研究の結果は示している。]

8. テストの効果には負の効果と正の効果がある。負の効果をもたらすか正の効果をもたらすかは教員の指導次第である。（例：ネットでつけたあるスピーキング・テスト「対策」サイト：「トピックに一般動詞が入っている場合：トピックの例：The classes h takes.→ **質問文の例**：What classes do you take?**質問文への直し方**：What 名詞 do you 一般動詞?※ここでは質問形式になり、主語が You となるので一般動詞に s を付けないようにしましょう」(<https://www.fourskills.jp/teap-speaking-template>)。<-- これでもスピーキング・テスト養成といえるのか。期待した波及効果といえるのか。
9. ではどうすれば好ましい波及効果が得られるか。現在の波及効果の研究はこの段階にあり、主にイノベーションの普及理論と動機付けの理論の枠組みで解釈が行われている。

④  
**論点4：日本人は英語ができない、とくにスピーキング力が弱い、というのは本当か。**

- TOEFL, IELTS 等の国際比較は各国の教育の質を比較したものではない。
  - テストによって各国のランクは異なる。日本の成績は必ずしも常に低いわけではない。
  - そもそも学校教育が理由で日本人は英語が話せないというエビデンスは存在するのかが疑問である。
1. TOEFL iBT や IELTS の国別比較を示して日本国籍の受験者の成績の低さを指摘されることが多い。さらに、だから日本の公教育における英語教育はなっていないと言われることがある。しかし、どちらのテストも各国で無作為抽出をして比較をしているわけではなく、受験者数も受験者の年齢も国によって大きく違う。得点から各国の高等教育前教育の質を判断することはできない。
  2. あえてスピーキングとライティング（2009年版）を例に国別比較を試みる。（表1参照。）
  3. TOEFL iBT の成績はアジア圏で最下位である。一方、IELTS の成績はそれほど低くはない。

**表1：TOEFL iBTとIELTS（2009年実施）のスピーキング、ライティング  
 アジア諸国の結果比較表**

		TOEFL		IELTS 2009			
		Speaking	Writing	スピーキング		ライティング	
1	パキスタン	23	23	1	マレーシア	6.92	6.17
2	フィリピン	22	23	2	香港	6.64	5.77
3	インド	22	22	3	フィリピン	6.51	6.25
4	スリランカ	21	22	4	インドネシア	6.24	5.5
5	バングラデシュ	21	22	5	ミャンマー	6.04	5.61
6	マレーシア	20	23	6	ベトナム	5.99	5.56
7	香港	20	22	7	韓国	5.89	5.29
8	インドネシア	19	21	8	スリランカ	5.88	5.78
8	ネパール	19	20	9	<b>日本</b>	<b>5.86</b>	<b>5.35</b>
9	ミャンマー	18	19	10	タイ	5.8	5.29
10	韓国	18	20	11	台湾	5.8	5.24
11	タイ	18	19	12	中国	5.7	5.12
12	台湾	18	19	13	インド	5.58	5.51
13	中国	18	20	14	ネパール	5.43	5.42
14	ベトナム	17	20	15	パキスタン	5.41	5.45
15	<b>日本</b>	<b>16</b>	<b>18</b>	13	バングラデシュ	5.25	5.33

## 論点5：話すことは簡単な技能なのか、4技能はそれぞれ独立しているのか？

- 話すことは言語能力のみならず、身体能力、認知能力、情意要因を同時に機能させて可能となる極めて高度な言語運用能力である。
  - ある技能ひとつだけが突出してできない、あるいはできるという国はない。スピーキングも例外ではない。
  - 日本人学習者に欠けているのは、読んで聞いた内容を解釈し、目的に合った内容を発話することだと推測できる。
1. 話すこと（スピーキング）は運動神経と情意要因と認知プロセスを同時並行で行う複雑な作業である。Levelt（1989）のモデルによると、概念形成→概念に合致した意味を持つ単語を探す→合致した意味を表す単語がなければ最も近い意味をもった単語を探す→単語の文法的特性に応じて文構造を作り出す（例えば、他動詞ならば、主語＋動詞＋目的語）→語形変化を整える→音韻の法則に合わせて発話の準備（played は/id/、pressed は/t/など）→運動神経を使って発話、というプロセスをとっている。
  2. TOEFL iBTの結果を考察する。（表2参照。）
    - このデータから、スピーキング「だけ」が出来ない国はなく、日本も例外ではないことが見て取れる。
    - さらに、2020年度実施全国学力・学習状況調査—英語（参考資料F、14頁）の結果を基に考察すると、理解に基づいた意味のある会話力に課題があることが推測される。
    - あえて一般化すれば、高等教育の英語教育が機能していない可能性はある。

表 2 : TOEFL iBT 2017 年 アジア諸国の結果

TOEFL 受験者—アジア圏諸国	読む	聞く	話す	書く	合計
Singapore	24	25	24	25	97
India	23	24	24	24	94
Malaysia	22	23	22	24	91
Pakistan	22	23	24	23	92
Korea, Republic of	22	21	20	21	83
Philippines	21	22	23	23	89
Hong Kong	21	22	22	23	88
Indonesia	21	22	21	22	85
Korea, Democratic People's Republic of	21	21	20	21	83
Taiwan	21	21	20	20	82
China	21	19	19	20	79
Sri Lanka	20	22	22	22	85
Viet Nam	20	20	20	22	82
Myanmar	19	20	20	21	80
Thailand	19	20	19	20	78
<b>Japan</b>	<b>18</b>	<b>18</b>	<b>17</b>	<b>18</b>	<b>71</b>
Cambodia	16	17	19	19	72

参照 [http://www.ets.org/s/toefl/pdf/94227\\_unlweb.pdf](http://www.ets.org/s/toefl/pdf/94227_unlweb.pdf) 一部省略

参考資料 F

2020 年度実施全国学力・学習状況調査—英語

- 「話すこと」の問題の一部とその結果
- 問いが音声とともに表示され、生徒は応答を PC に録音

英語② 即興でやり取りをする

あなたは、ユイコとアラン先生と話しています。まず、ユイコとアラン先生が、2人で話している場面から始まります。そのあと、あなたが尋ねられたら、2人のやり取りの内容を踏まえて、会話が続いていくように英語で応じてください。解答時間は20秒です。それでは始めます。

アラン先生が見せている写真



即興と言うよりも会話を理解して場面にあった会話力が求められている。

A: Look at this picture of my family.  
This is my favorite picture.  
Y: Nice! Who is she?  
A: Oh, she is my mother, Nancy. And he is my brother, Tom. He can cook very well.  
Y: I see. What kind of work does your mother do?  
A: She is a teacher.  
Do you have any other questions about them?

- 10.5%が正答
- 20.7%は無回答

英語③ まとまりのある内容を話す

あなたの学校で、海外のあるテレビ局が「世界の子どもたちの夢」というテーマで番組を収録しています。画面にある話してほしい内容の①、②について、英語で話してください。1分間、内容を考えたあと、30秒で話します。それでは、内容を考えましょう。それでは、30秒で話してください。

話してほしい内容

(①、②のどちらか話してください)

- ①あなたの将来の夢、または、将来やってみたいこと
- ②その実現のために頑張っていること、やるべきこと



自分の思いを述べる能力が問われている。

- 45.8%が正答
- 4.6%は無回答

## 要点のまとめ

1. 「各資格・検定試験と CEFR との対照表」は我が国の受験制度に適用するには不十分である。（\*各大学が自主的に使用するためのガイドラインとしては使用できる可能性がある。しかし、その際にも特に公平性を保つために、受験回数、試験実施会場の場所、受験料、目的、採点のプロセス、信頼性の検証結果、妥当性の検証結果等々詳細な情報を提供し、使用者は各教育機関のアドミッション・ポリシーとの整合性を図ることが必須である。また文部科学省は厳しく検証する必要がある。）
2. **センター試験・共通テスト**は学習指導要領を基盤にして開発作成された**到達度テスト**であり、実施にも細心の注意が払われている。さらに、公平性、信頼性、妥当性をはじめとして大規模テストの満たすべき条件を満たしている。危機管理等においても対処ができる。万が一不足の事態に遭遇した場合でも一斉に対応が可能である。民間試験はセンター試験・共通テストの替わりにはならない。
3. 実証研究の結果の示すところ大規模テストの波及効果は極めて限定的である。
4. 話すことはその他の 3 技能と切り離して独立した能力として扱うことはできない。十全な英語力の姿をゆがめる可能性が高い。特に読解力と聴解力を前提としてこそ豊かな会話力になる。

## 提案

1. スピーキング・テストが必要であると判断された場合、大学入試センターが中心となって開発することが望ましい。録音型はすでに全国学力・学習状況調査—英語で開発実施してる。文部科学省、国立教育政策研究所、研究者との協力のもとで学習指導要領にそって作成開発された。これをモデルとして、たとえば大学入試センターが研究者とチームを組んで高校の教員と協力しながら、開発することは十分に可能である。
2. 実行年度の設定は必要ではあるが、性急に進めるべきではない。試行実験を繰り返し、実行可能性、信頼性、妥当性、公平性等々のテスト実施の基本的な条件が満たされてから実施すべきである。これについても、過去の蓄積、特に大学入試センター試験にリスニング・テスト、全国学力・学習状況調査等の例が参考になる。
3. 海外の事例を収集し成功例、失敗例を検証し、我が国の大学入試政策に生かす必要がある。
4. 以上のことが可能になるまでは、現行のまま進めることを旨とする。国家規模の教育改革は高いリスクを覚悟で性急に進めることは戒めるべきである。
5. 当然のことながら、その間公教育で行われる英語教育を充実させるべく努力を重ねることは当然である。特に大学に進学しない生徒、センター試験を受けない生徒は埒外となっている。
6. 大学における教員養成を改善しなければならない。大学の英語教育は徹底的に改善する必要があることは言うまでもない。
7. **本委員会は今後予定されている内容を終えた後には、今後の対策を練る機会とするのが相応しくかつ建設的である。昨年までの成果は一旦白紙に戻した上で、理想的な試験内容および制度について継続して審議するのが妥当である。**
8. 以上は全てセンター試験・共通テストとして民間試験を導入することに関する案であり、すでに行われている各大学での民間試験の使用に関するものではない。



## おわりに

### 日本語テスト学会で提出した提言（2016）について

日本語テスト学会（JLTA）という、言語能力測定にかかわるさまざまな課題を研究している研究者の集団があり、現在小職が会長を務めている。当学会は 2016 年に「大学入学希望者学力評価テスト（仮称）における 英語テストの扱いに対する提言」（以下「JLTA 提言」）を日英 2 カ国語で公にしている。（参考資料 G 参照。）当時の文科省の担当者の方をお願いをして当時の文部科学大臣にもお渡ししている。JLTA 提言公表当時すでに入試改革が端緒についていたが、それまでに公開された情報を見る限り、研究の成果を参考にした形跡が全く見えず会員は懸念を共有していた。そこで 2015 年の会合で同意を得た上で有志の協力の下提言を作成し、公表することとなった。

ところが、JLTA 提言は改革会議において全く話題になった形跡がなく、そのまま進みそして今回の突然の方針大転換となった。改めて昨年までの会議の報告を見る限り、残念ながら JLTA 提言が考慮されたほとんど形跡はなく、したがってそこで述べた事項はほぼそのまま当てはまる。さらに、他国の事例を研究した形跡が見受けられない。例えば、韓国の 4 技能テスト NEAT（National English Ability Test）は膨大な国家予算をかけ長年の準備期間を経てとして大きな期待のうちに始まったが、資格試験として機能せず、さらにその後は予算を削減されるなどの経緯で今では実施されていない。このような近隣の国の試みがあるにもかかわらず、十分に検証されていないという印象を私は受けている。

このようなことはすべて、民間試験導入という結論が先にあり、そのために都合のよい情報を収集し、都合の悪い研究成果は考慮しようとしなかった結果であると疑われてもやむを得ないところがあるのではないだろうか。このようなことは委員の見識を疑われることとなり、信頼を損なう結果となる。このようなことが繰り返されてはならない。

本日は JLTA 提言で指摘した課題のうち 4 つの論点に絞って実証研究の成果を紹介し、必要に応じてデータを示し委員の皆様と検討することを目的とした。ただし、今回の発表に際して作成委員全員の下承を受けているわけではない。作成にかかわった会員はいずれも超一流の研究者である。本発表に納得していただける点があるとすればそれは作成委員の貢献であり、瑕疵があるとすればそれは小職の責任であることをご了承願いたい。

**大学入学希望者学力評価テスト（仮称）における英語テストの扱いに対する提言**

[http://jlta2016.sakura.ne.jp/?page\\_id=865](http://jlta2016.sakura.ne.jp/?page_id=865)

**1. 序**

平成 28 年 8 月 31 日に、高大接続改革について文部科学省より発表された文書において、「大学入学希望者学力評価テスト（仮称）」の進捗状況が発表され平成 29 年度初頭に「実施方針」が策定されるという見通しが明らかになった。日本言語テスト学会は、高大接続改革において大学入試に 4 技能テストを導入する方向性についてここに支持を表明し、その中心的な改革である大学入学希望者学力評価テスト（仮称）における英語テストとそれに関連する諸問題に対して以下のような提言をしたい。

**2. 大学入学希望者学力評価テストにおける英語テスト**

**(1) 透明性の高い手続き** 民間の資格・検定試験を活用する際に、国（センター）が認定を行うとあるが、その認定の手順を明確化するべきである。その際、以下の 2 点の実施を提案する。

- i. テストの実行可能性だけでなく、テストの質も十分検討した上で決定を行うこと。
- ii. 認定の手順を策定する際には、言語テスト研究者を複数名配置し、最新の言語テスト研究の成果を反映した、透明性が高いものにする。

**(2) 情報公開**

民間の資格・検定試験に関して、以下の 3 点の実施を提案する。

- i. 民間の資格・検定試験はそれぞれ、テストの目的、測定しようとする能力及び受験対象者が異なるケースが多い。このことをテスト使用者（スコアを利用する各大学、受験者、その他大学入試に関わる者）は広く認識すること。
- ii. テスト機関は、テストの目的と測る能力、採点基準や採点方法、また作成・実施方法やテストの適切な使用法など、テストを選ぶ際に欠かせない詳細な情報を公表すること。特に、学習指導要領にどのように準拠しているテストなのかを明示すること。
- iii. 文部科学省主導で、テスト使用者とテスト機関に上記 i および ii の徹底を明確に要請すること。

**(3) 大学入試における選抜に関する考慮点**

- i. **スコア換算表の扱い**：複数の民間の資格・検定試験を認定する場合、現行案ではスコアを換算表で比較することが前提とされている。しかし、各テストは測る能力が異なるため、テスト間でスコアを直接比較したり、そこに現れる細かな差を一般的な英語力の差と解釈したりすることは適切ではない。文部科学省においては、このような制約について明示し、適切なテスト結果の解釈と使用をテスト使用者側に要請すること。
- ii. **各大学のニーズやアドミッション・ポリシーとテストの関連づけ**：各大学は、それぞれのニーズやアドミッション・ポリシーに対して体系的な分析を十分行った上で、それをふまえたテストを選定し、テスト結果を適切に活用すること。文部科学省は、適切なテストの選定と結果の活用のための具体的な手順を、言語テスト研究の知見を十分にふまえながら提案すること。

**(4) 公正な機会**

受験者に関して、経済格差や地域格差、障害等による受験機会の不平等が生じないよう、引き続き具体的な対処策を検討・実行すること。特に、地域格差については、地方によって受験できるテストが限られ、たとえ受験が可能であっても、そのためかなりの時間と費用が必要になることを理解した上で、公平な機会を提供すべく支援や改善すること。（現筆者下線）

**(5) 波及効果だけに頼らない方策** これまでの波及効果の実証研究を概観すると、4 技能テストを導入するだけでは、高校の英語教育改善につながるとは限らないことが予測される。これらの研究の知見を十分鑑み、より良い波及効果を起こすような方策を策定し、実行すること。同時に、英語教育の改善をテスト改革のみに頼るのではなく、教員の指導力向上を目指す教員養成・教員研修の強化を実現するため、以下の 2 点を行うこと。

- i. 言語テスト研究における成果を十分反映させつつ、教員養成課程や教員研修の内容を精査すること。
- ii. その内容に、英語教育におけるテスト本来の役割と教師が果たすべき役割、テストについての基礎知識と適切な使用方法を含めること。

**(6) 学習につながるテスト結果の活用** 技能テストの結果を、高校と大学入学後の学習に十分生かすために、全てのテスト機関は、指導や学習により役立つようなスコアレポートと関連資料を提供すること。さらに、テスト使用者が指導や学習に活かすことができるような手順や方策を提案すること。

**3. 結び** 日本言語テスト学会は、教員やテスト関係者に対するアセスメント・リタラシー（評価・テストについての知識・スキル）習得を促進する啓蒙活動を過去 20 年間行ってきた。当学会では、このような活動を引き続き行うとともに、他の学会や組織とも協力しながら、大学入試改革に向けて積極的に協力する準備がある。また、学会構成員である我々も、教員や評価者、テスト関係者の一人として、それぞれの専門分野・知識を生かし、大学入試や教室における評価、教育実践の改善に全面的に協力をしていく所存である。

平成 29 年（2017 年）1 月 4 日  
日本言語テスト学会

**作成委員** 渡部良典（上智大学）〔会長、文責〕他 9 名（原本には明記、本資料作成に当たって割愛）

## **参考：複数回受験について**

### **期待されている効果**

- 受験の機会を増やすことは生徒のやる気を増す。
- 受験の機会を増やすと、生徒は最も出来の良い得点を大学に提出することができる。
- 受験の機会を増やすことは「受験戦争」の緩和につながる。

### **留保事項**

- 最高点が取れるまで試験を受け続けることはかえって心理的圧迫感を増すかもしれない。
- ある一定のレベルに到達しない生徒は入学が許可されないことになる。

### **複数受験機会得点の使用方法**

- 平均点を採用する。
- 受験応募期間内において最も近い日程に受験した結果を採用する
- 最高点を採用する
- 最高点と最低点は排除して、残りの点数のうち最高点を採用する

### **実証研究が少ない**

- 平均点を採用した場合が入学後の成績（GPA）と相関—予測妥当性—が高い（Boldt, et al、1986）。
- 複数回受験者は一回だけ受験する学生よりも平均点が低い傾向がある（Wightman、1990）
- 回数に関係なく平均点をもっとも妥当である（Zhao, et al、2010）。