

IELTS と TOEIC を用いた CEFR 対照表の検証

神崎 正哉¹

要 旨

本研究では、IELTS（アカデミック・モジュール）と TOEIC（L&R と S&W）を用いて「各資格・検定試験とCEFRとの対照表」（文部科学省、2018）のCEFRレベルの対応関係のずれ具合を検証した。被験者は84名の大学生で、彼らに両方の試験を受けてもらい、両試験の得点のCEFRレベルの対応関係を調べた。両試験で等しいレベルになった者は49名（約58%）であったが、IELTSのCEFRレベルを0.5点分、下方に修正すると、その人数を67名（約80%）まで増やすことができた。これは対照表上の両試験間の対応関係にずれがあることを意味する。他の試験の間の対応関係もずれている可能性があるため、ずれの検証を行わずに対照表を大学入試で用いるのは、公平性と公正性の観点から好ましくない。

1. はじめに

文部科学省は2018年3月、「各資格・検定試験とCEFRとの対照表」（以下、対照表；文部科学省、2018）を発表した。その当時は、2020年度（2021年度入学者選抜）から始まる大学入学共通テスト（以下、共通テスト）の枠組みに、民間事業者が実施する英語の4技能試験が組み込まれることになっていた（その後、2019年11月に延期、2021年7月に中止が決まる；大学入試のあり方に関する検討会議、2021）。異なる民間試験の結果を比較するためには、共通の指標が必要である。その指標として、Common European Framework of Reference for Languages;

¹ 神田外語大学外国語学部国際コミュニケーション学科准教授。

Learning, teaching, assessment (CEFR; 外国語の学習・教授・評価のためのヨーロッパ共通参照枠; Council of Europe, 2001) の6段階のレベル分け(低い方から A1、A2、B1、B2、C1、C2)が使われることになり、2017年7月に発表された「大学入学共通テスト実施方針」(文部科学省、2017)では、「国は、活用の参考になるよう、CEFRの段階別成績表示による対照表を提示する」(p. 3)という方針が示された。対照表は、これを受けて作成された。

共通テストの枠組みで行われる民間試験の成績の集約と提供は、大学入試英語成績提供システムを介して行われることになっていた。対照表には、同システムへの参加申し込みを行い、要件を満たしていると判断された8種類の試験(ケンブリッジ英語検定、実用英語技能検定、GTEC、IELTS、TEAP、TEAP CBT、TOEFL iBT、TOEIC)が載っている。なお、TOEICは、運営団体の国際ビジネスコミュニケーション協会が2019年7月に大学入試英語成績提供システムへの参加申し込みを取り下げたため、改訂版の対照表では除外されている(文部科学省、2019)。

対照表が発表されると、関連する様々な問題点が指摘されるようになった。そのひとつに、対照表の作成にあたり、各試験運営団体の自己申告に基づくCEFRとの対応づけを検証せずにそのまま使っているというものがある(羽藤、2018)。本研究は、その指摘から着想を得て、IELTS(アカデミック・モジュール)とTOEIC(L&RとS&W)を用いて対照表で示されたCEFRレベルの対応関係を検証するものである。

CEFRを介して異なる試験を対応づけることに関して、目的、内容、対象者、難易度が異なる試験の成績を互いに比較することには無理があるという指摘がある(南風原、2018; 羽藤、2018、2020)。異なる言語試験の結果の比較に関して、Davies et al. (1999)は、各々の試験は異なる目的で異なる対象者を想定して設計されており、言語特性の捉え方や言語能力の測定・評価の方法も異なるので、Aという試験の何点がBという試験の何点に相当するという同値化は、正当性を欠

く、としている。しかし、多くの英語試験は、英語を外国語として学ぶ学習者の英語能力を測定する目的で作られているという点は共通している。また、専門知識や英語力以外の能力の影響を最小限にして、英語力のみを測定しようとしている点も共通である。試験の仕様に違いはあっても、測定している能力が近ければ、得点の比較は可能である。

異なる試験の得点比較が可能かという判断には、得点間の相関がひとつの目安になる。Kanzaki (2021) は、IELTS と TOEIC の得点間に 0.79 の相関があったと報告している（被験者数 84 名）。この値が得点比較の可能なレベルであるかという点に関しては、議論の余地がある。Dorans (2004) は、ある試験を別の試験の代替として用いるには、両試験の得点間に 0.866 以上の相関が必要であるとしている。しかし、英語試験には測定誤差がある。測定誤差があると、相関は本来あるべき値より低くなる (Murphy & Davidshofer, 2004)。よって、英語試験の場合、得点間の相関が 0.866 未満であっても代替として使われることが珍しくない。例えば、英語圏の大学で英語を母国語としない入学志願者の英語力の要件として IELTS の代わりに TOEFL iBT や PTE Academic の得点も認めているところがあるが、これまでに報告されている IELTS と TOEFL iBT の得点間の相関は 0.73 (Educational Testing Service, 2010)、IELTS と PTE Academic の得点間の相関は 0.74 (Clesham & Hughes, 2020) と 0.70 (Elliot et al., 2021) である。IELTS と TOEIC の得点間には、これらの値より高い相関があるので、本稿では両試験の得点比較は可能であると見なす。

また、CEFR は、言語試験の基準設定に適さないという指摘もある。Milanovic (2009) は、その理由として、CEFR は多様な言語および多様な用途に対応させるため、レベルの記述を意図的に非特定の不完全な形にしていることを挙げている。同様な指摘は他の研究者も行っている（例えば、Alderson et al., 2004; Fulcher, 2004; Weir, 2005）。言語試験の基準設定に適さない CEFR を介して異なる試験の得点を対応づけると、得点を直接比較して作った換算表との間にずれが生

じる。例えば、IELTS の 5.5 点は、CEFR レベルの B2 相当であるが、この得点は IELTS と TOEFL iBT の得点を直接比較して作った換算表では、TOEFL iBT の 46-59 点に相当し (Educational Testing Service, 2010)、IELTS と GTEC CBT の得点を直接比較して作った換算表では、GTEC CBT の 1001-1100 点に相当する (Kim et al., 2017)。しかし、TOEFL iBT の 46-59 点と GTEC CBT の 1001-1100 点の CEFR レベルは、共に B1 である。先行研究のこのような結果から、対照表で示された IELTS と TOEIC の CEFR レベルの対応関係にもずれがあることが予想される。

英語の民間試験が共通テストに組み込まれることはなくなったが、独自の方法で民間試験を入試に活用している大学は一定数ある。2021 年度の大学入学者のうち、民間試験を使った選抜区分で入学した者の割合は、国立大学で 14.5%、公立大学で 6.0%、私立大学で 19.8%であった (大学入試のあり方に関する検討会議、2021)。民間試験を活用している大学の中には、複数の試験の結果を文科省版の対照表に準ずる形で対応づけているところもある。よって、対照表の検証は、今日においても意義がある。

2. 方法

大学生から参加者を募り、IELTS (アカデミック・モジュール) と TOEIC (L&R と S&W) を受けてもらった。IELTS は 2019 年 9 月 7 日、TOEIC は 2019 年 9 月 10 日に大学の施設内で実施した。試験結果が出た後、得点の統計的な分析を行った。そして、対照表を使って、両試験の得点の CEFR レベルの対応関係を調べ、ずれの修正を試みた。

本研究は、当該大学の研究倫理審査委員会の承認を得て行った。

参加者

本研究の参加者は、関東地方にある外国語を専門とする大学に通う大学生 84 名 (女性 63 名、男性 21 名 ; 1 年生 24 名、2 年生 20 名、3 年生 23 名、4 年生 17

名)である。そのうち、主たる外国語として英語を専攻する者が76名、英語以外の外国語を専攻する者が8名であった。参加者は、IELTSとTOEICが無料で受けられ、研究協力終了時に謝金が支払われるという条件の下、自らの意思で参加した。参加者には事前に研究目的、研究内容、参加条件に関する説明を行い、参加同意書への署名により参加意思を確認した。

IELTS

IELTSはInternational English Language Testing Systemの略称で、1989年にEnglish Language Testing Serviceの後継として始まった(Charge & Taylor, 1997)。IELTSの運営は、British Council、IDP: IELTS Australia、Cambridge Assessment Englishの3団体から成るIELTS Partnersが行っている。

IELTSには、アカデミックとジェネラル・トレーニングの2つのモジュールがあり、前者は英語圏の大学や大学院への留学を目指している人向け、後者は英語圏への移住または英語圏での研修受講を目指している人向けの内容になっている(日本英語検定協会、2020)。本研究では、アカデミック・モジュールを用いた。また、IELTSには、紙と鉛筆で受ける試験とパソコン上で受ける試験があるが、本研究では前者を用いた。

IELTSは、ライティング(60分、2問)、リーディング(60分、40問)、リスニング(約30分、40問)、スピーキング(11~14分、試験官との1対1の面接)の順に行われる(詳細は、日本英語検定協会、2020を参照)。リスニングとスピーキングは、アカデミックとジェネラル・トレーニングの両モジュールで共通の問題が使われる。すなわち、アカデミック・モジュールであっても、リスニングとスピーキングに関しては、アカデミック色が薄い。

IELTSの結果は、4つの技能別得点と総合点がそれぞれ0点から9点の間で、0.5点刻みで表される。総合点は、4つの技能別得点の平均を少数第1位が0または5になるように調整した値である(例えば、6.125点は6.0点に、6.25点は6.5

点に、6.625 点は 6.5 点に、6.75 点は 7.0 点に調整される)。

IELTS Partners (2021) によると、IELTS の信頼度は、リスニングで 0.91、リーディング (アカデミック・モジュール) で 0.91、スピーキングで 0.83–0.86、ライティングで 0.81–0.89、総合点で 0.96、測定標準誤差は、リスニングで 0.39 点、リーディング (アカデミック・モジュール) で 0.37 点、総合点で 0.23 点である。スピーキングとライティングの測定標準誤差は公表されていない。

TOEIC

TOEIC は、Test of English for International Communication の略称で、職場環境における日常的な英語力を測定する試験である。アメリカのテスト開発機関である Educational Testing Service によって開発され、1979 年にリスニングとリーディングの問題から成る TOEIC L&R が始まり、2006 年にスピーキングとライティングの力を試す TOEIC S&W が加わった (Powers & Schmidgall, 2018)。

TOEIC L&R は、リスニングセクションとリーディングセクションから成る単独の試験で、正式な表記は TOEIC Listening and Reading test (test が単数形) となっている。それに対して、TOEIC S&W は、スピーキングとライティングが別の試験になっており、正式な表記は TOEIC Speaking and Writing tests (test が複数形) である。このため、TOEIC L&R の結果票にはリスニングとリーディングの合計点が表示されるが、TOEIC S&W の結果票にはスピーキングとライティングの合計点は表示されない。

TOEIC のリスニングとリーディングは、全問選択問題のマークシート方式の試験で、得点はそれぞれ 5 点から 495 点の間で、5 点刻みで表される。スピーキングとライティングは、パソコンを使って行われる試験で、得点はそれぞれ 0 点から 200 点の間で、10 点刻みで表される。リスニングは約 45 分で 100 問、リーディングは 75 分で 100 問、スピーキングは約 20 分で 11 問、ライティングは約 60 分で 8 問である (詳細は、国際ビジネスコミュニケーション協会、2021a、2021b

を参照)。

Educational Testing Service (2019a, 2019b) によると、TOEIC の信頼度は、リスニングで 0.90、リーディングで 0.90、スピーキングで 0.80、ライティングで 0.82、測定標準誤差は、リスニングで 25 点、リーディングで 25 点、スピーキングで 13 点、ライティングで 17 点である。

TOEIC の仕組み上、4 技能を測定するには、TOEIC L&R、TOEIC Speaking、TOEIC Writing の 3 種類の試験を受けなければならない。このため、4 技能の総合点が公式な形で示されることはなく、3 種類の試験の得点から算出する必要がある。総合点の計算方法に関して、対照表の欄外に「TOEIC L&R/ TOEIC S&W については、TOEIC S&W のスコアを 2.5 倍にして合算したスコアで判定する」と記されている。この「2.5 倍して合算」というのは、リスニングとリーディングの得点（各 495 点満点）とスピーキングとライティングの得点（各 200 点満点）をほぼ同じ比率にするための処置である。本研究でもこの計算方法を用いて、総合点を算出した。

対照表の CEFR レベルの細分化

CEFR のレベル分けは、各レベルの幅が広いので、対応関係のずれを細かく分析することが難しい。そこで本研究では、各レベルを細分化し、より詳細な分析を試みた。表 1 は、対照表の IELTS と TOEIC の部分と、各 CEFR レベルを上位 (H)、中位 (M)、下位 (L) の 3 層に分割した表を並べたものである (A1 と C2 は省略)。IELTS の得点は 0.5 点刻みで、C1 (7.0–8.0)、B2 (5.5–6.5)、B1 (4.0–5.0) には、それぞれ 3 個の得点が割り当てられているので、3 等分が容易である。TOEIC の得点は 5 点刻みで、C1 (1845–1990)、B2 (1560–1840)、A2 (625–1145) に割り当てられている得点の数は、それぞれ 30 個、57 個、105 個になっており、これらは 3 で割り切れるので、3 等分が可能である。B1 (1150–1555) に割り当てられている得点の数は 82 個で、これを 3 で割ると 1 余るので、中位の B1M に

得点をひとつ多く割り当てることで調整した。

表 1.

IELTS と TOEIC の CEFR 対照表と 3 分割対照表 (A2-C1)

CEFR 対照表			3 分割対照表		
CEFR	IELTS	TOEIC	CEFR	IELTS	TOEIC
C1	7.0–8.0	1845–1990	C1H	8.0	1945–1990
			C1M	7.5	1895–1940
			C1L	7.0	1845–1890
B2	5.5–6.5	1560–1840	B2H	6.5	1750–1840
			B2M	6.0	1655–1745
			B2L	5.5	1560–1650
B1	4.0–5.0	1150–1555	B1H	5.0	1425–1555
			B1M	4.5	1285–1420
			B1L	4.0	1150–1280
A2		625–1145	A2H		975–1145
			A2M		800–970
			A2L		625–795

注) H=上位、M=中位、L=下位

3. 結果

得点の記述統計と相関

表 2 は、IELTS と TOEIC の技能別得点および総合点の記述統計を示す。リスニングとリーディングの平均点を比較すると、IELTS ではリーディングの方が高いが、TOEIC ではリスニングの方が高い。スピーキングとライティングの平均点を

比較すると、IELTS ではスピーキングの方が高いが、TOEIC ではライティングの方が高い。

表 2.

IELTS と TOEIC の技能別得点および総合点の記述統計 ($N = 84$)

	可能得点	平均点	標準偏差	最低点	最高点	歪度	尖度
IELTS							
L	0-9	5.30	0.07	3.5	7.0	0.10	0.35
R	0-9	5.46	0.10	3.5	8.0	0.30	-0.09
S	0-9	5.54	0.09	3.5	8.0	0.20	0.26
W	0-9	5.33	0.06	3.5	6.0	-0.67	0.68
総合	0-9	5.45	0.07	4.0	7.0	0.13	-0.16
TOEIC							
L	5-495	383.15	8.32	145	495	-0.89	0.78
R	5-495	330.06	8.75	85	470	-0.82	0.69
S	0-200	137.74	2.11	90	190	0.21	0.34
W	0-200	144.76	2.45	70	190	-0.92	1.88
総合	10-1990	1419.46	24.27	720	1865	-0.76	1.22

注) L=リスニング、R=リーディング、S=スピーキング、W=ライティング
 TOEIC 総合=L+R+2.5(S+W)

表3は、IELTS と TOEIC の同じ技能得点間および総合点間の相関 (Pearson の r) を示す。リスニングとリーディングの相関を比べると、リーディングの方が高く、スピーキングとライティングの相関を比べるとライティングの方が高い。また、技能別得点間の相関が 0.50-0.69 であるのに対して、総合点間の相関は 0.79 になっ

ている。技能別得点間の相関より総合点間の相関の方が高いというのは、異なる試験の得点比較を行った先行研究の結果と一致する（例えば、Clesham & Hughes, 2020; Elliot et al., 2021; Educational Testing Service, 2010; Kim et al., 2017）。

表 3.

IELTS－TOEIC 間の技能別得点および総合点の相関（ $N = 84$ ）

	リスニング	リーディング	スピーキング	ライティング	総合
相関	0.64*	0.69*	0.50*	0.58*	0.79*

注) * $p < .001$

CEFR レベルの検証とずれの修正

表 4 は、IELTS と TOEIC の CEFR レベルの対応関係を示すものである。横軸が IELTS、縦軸が TOEIC の CEFR レベルになっており、各升の数字は、対応する両軸の CEFR レベルに該当する人数である（例えば、IELTS の B2 と TOEIC の B1 が交わる升の「26」は、IELTS が B2 で、かつ TOEIC が B1 になった者が 26 名という意味である）。網掛け部分は、両試験の CEFR レベルが一致する升である。84 名の参加者のうち、TOEIC の CEFR レベルの方が 1 段階高い者が 2 名、両試験で等しい者が 49 名、IELTS の方が 1 段階高い者が 33 名であった。2 段階以上ずれた者はいなかった。

表 4.

IELTS と TOEIC の CEFR レベルの対応関係 (N = 84)

		IELTS		
		C1	B2	B1
TOEIC	C1	0	1	0
	B2	1	22	1
	B1	0	26	27
	A2	0	0	6

表 5 は、表 1 の右側の 3 分割対照表を使った場合の両試験における対応関係を示したものである。横軸が IELTS、縦軸が TOEIC の 3 分割 CEFR レベルになっており、各升の数字は、対応する両軸の CEFR レベルに該当する人数である（例えば、IELTS の B2L と TOEIC の B1H が交わる升の「11」は、IELTS が B2L で、かつ TOEIC が B1H になった者が 11 名という意味である）。網掛け部分は、両試験の CEFR レベルが一致する升である。84 名の参加者のうち、TOEIC の CEFR レベルの方が 1 段階高い者が 4 名、両試験で等しい者が 14 名、IELTS の方が 1 段階高い者が 29 名、IELTS の方が 2 段階高い者が 26 名、IELTS の方が 3 段階高い者が 10 名、IELTS の方が 4 段階高い者が 1 名であった。

表 5.

IELTS と TOEIC の 3 分割 CEFR レベルの対応関係 (N = 84)

		IELTS						
		C1L	B2H	B2M	B2L	B1H	B1M	B1L
TOEIC	C1L	0	1	0	0	0	0	0
	B2H	0	3	1	0	0	0	0
	B2M	0	1	0	1	0	0	0
	B2L	1	3	9	4	1	0	0
	B1H	0	0	3	11	6	0	0
	B1M	0	0	3	6	6	1	0
	B1L	0	0	0	3	12	2	0
	A2H	0	0	0	0	0	2	0
	A2M	0	0	0	0	1	1	0
	A2L	0	0	0	0	0	0	2

注) H=上位、M=中位、L=下位

上記の対応関係の分類中、人数が一番多いのは、IELTS の方が 1 段階高い者 (29 名) である。そこで、この集団の 3 分割 CEFR レベルが両試験で等しくなるように IELTS の CEFR レベルを 0.5 点分、下方にずらすことで対照表の修正を行った。表 6 は、表 1 の CEFR 対照表と 3 分割対照表の修正版である。

表 6.

IELTS を 0.5 点下方修正した CEFR 対照表と 3 分割対照表 (A2–C1)

修正 CEFR 対照表			修正 3 分割対照表		
CEFR	IELTS	TOEIC	CEFR	IELTS	TOEIC
C1	7.5–8.5	1845–1990	C1H	8.5	1945–1990
			C1M	8.0	1895–1940
			C1L	7.5	1845–1890
B2	6.0–7.0	1560–1840	B2H	7.0	1750–1840
			B2M	6.5	1655–1745
			B2L	6.0	1560–1650
B1	4.5–5.5	1150–1555	B1H	5.5	1425–1555
			B1M	5.0	1285–1420
			B1L	4.5	1150–1280
A2	4.0	625–1145	A2H	4.0	975–1145
			A2M		800–970
			A2L		625–795

注) H=上位、M=中位、L=下位

表 7 は、表 6 の右側の修正 3 分割対照表を使った場合の両試験における対応関係を示したものである。この表では、TOEIC の方が 2 段階高い者が 4 名、TOEIC の方が 1 段階高い者が 14 名、両試験で等しい者が 29 名、IELTS の方が 1 段階高い者が 26 名、IELTS の方が 2 段階高い者が 10 名、IELTS の方が 3 段階高い者が 1 名になっている。

表 7.

IELTS を 0.5 点分下方修正した場合の 3 分割 CEFR レベルの対応関係 (N = 84)

		IELTS						
		B2H	B2M	B2L	B1H	B2M	B2L	A2H
TOEIC	C1L	0	1	0	0	0	0	0
	B2H	0	3	1	0	0	0	0
	B2M	0	1	0	1	0	0	0
	B2L	1	3	9	4	1	0	0
	B1H	0	0	3	11	6	0	0
	B1M	0	0	3	6	6	1	0
	B1L	0	0	0	3	12	2	0
	A2H	0	0	0	0	0	2	0
	A2M	0	0	0	0	1	1	0
	A2L	0	0	0	0	0	0	2

注) H=上位、M=中位、L=下位

表 8 は、修正前と修正後の 3 分割 CEFR レベルの対応関係を「TOEIC の方が高い」、「両試験で等しい」、「IELTS の方が高い」の 3 カテゴリーに分類して、人数を比較したものである。修正前は、IELTS の方が高い者が 66 名と全体の 8 割近くになっていたが、修正後は 37 名 (約 44%) にまで減り、「TOEIC の方が高い」と「両試験で等しい」に該当する人数が増えた。修正後も「IELTS の方が高い」の人数が一番多いが、修正前に比べると、偏りは小さくなっている。

表 8.

修正前と修正後の 3 分割 CEFR レベルの対応関係の比較 (N = 84)

	修正前	修正後
TOEIC の方が高い	4 (4.76%)	18 (21.43%)
両試験で等しい	14 (16.67%)	29 (34.52%)
IELTS の方が高い	66 (78.57%)	37 (44.05%)

表 9 は、表 6 の左側の修正 CEFR 対照表を使った場合の両試験における対応関係を示したものである。表 4 に示した修正前の対応関係に比べて、TOEIC の方が 1 段階高い者が 5 名増えて 7 名に、両試験で等しい者が 18 名増えて 67 名に、IELTS の方が 1 段階高い者が 23 名減って 10 名になっている。両試験のレベルが等しい者が 49 名（約 58%）から 67 名（約 80%）に増え、TOEIC の方が高い者と IELTS の方が高い者の人数の比が 2 対 33 から 7 対 10 になったので、対応関係のずれが減り、均衡のとれた形になったと言える。

表 9.

IELTS を 0.5 点分下方修正した場合の CEFR レベルの対応関係 (N = 84)

		IELTS		
		B2	B1	A2
TOEIC	C1	1	0	0
	B2	18	6	0
	B1	6	47	0
	A2	0	4	2

4. 考察

表2に示した記述統計では、技能別得点の平均点の高低関係がIELTSとTOEICで逆転している点が興味深い。リスニングとリーディングの平均点を比べると、IELTSでは、リーディングの方が高いが、TOEICでは、リスニングの方が高い。同様に、スピーキングとライティングの平均点を比べると、IELTSでは、スピーキングの方が高いが、TOEICでは、ライティングの方が高い。IELTSとTOEICは、同時期に実施したので、両試験の受験時における参加者の英語力は等しいと見なせる。それでもIELTSとTOEICの技能別得点の平均点の高低関係が逆転しているのは、Davies et al. (1999) が指摘しているように、言語特性の捉え方や言語能力の測定・評価の方法が試験によって異なっているためであろう。また、英語試験の得点は、英語力という物理的実体のない能力を任意の数字で表したものに過ぎず、数値化の方法や基準も試験ごとに異なる。

表3に示した技能別得点間の相関は、リーディングの方がリスニングより高く、ライティングの方がスピーキングより高くなっている。これは、IELTSの技能別問題のアカデミック色の濃淡を考えると、意外な結果であった。本研究では、IELTSのアカデミック・モジュールを用いたが、リスニングとスピーキングはジェネラル・トレーニング・モジュールと問題が共通で、アカデミック色が薄く、その点に関してはTOEICに近い。しかし、アカデミック・モジュール専用の問題を使っているリーディングとライティングの方がTOEICとの間により高い相関があった。得点の相関には、アカデミック色の差異より問題形式の差異の方が強く関係していると考えられる。例えば、技能別得点間の相関が一番低いのは、スピーキング ($r = 0.50$) であるが、面接官と会話のやり取りをするIELTSのスピーキングとパソコンに向かって一方的に話すTOEICのスピーキングでは、問題形式が大きく異なる。

両試験のCEFRレベルの対応関係を示した表4では、両試験のCEFRレベルが等しい者が49名(約58%)と多数派になっており、ずれ具合がわかりにくい

各 CEFR レベルを 3 層に分割して調べることで、明確にすることができた。3 分割 CEFR レベルの対応関係を示した表 5 では、IELTS の方が 1 段階高い者の人数が一番多くなっているため、IELTS の CEFR レベルを 0.5 点分、下方修正して、CEFR レベルのずれを調整した。それとは逆に、TOEIC の CEFR レベルを上げて修正することも可能である。TOEIC は、各 CEFR レベルに割り当てられている得点の数が多いので、TOEIC の得点を動かす方がより微細な調整が可能である。しかし、その分、調整方法が複雑になる。

本研究では、IELTS と TOEIC の得点間に高い相関 ($r = 0.79$) があったので、CEFR レベルのずれを比較的容易に修正することができた。得点間の相関が低い場合は、ずれの修正は困難になる。

5. おわりに

本研究では、対照表の IELTS と TOEIC の CEFR レベルの対応関係にずれがあることを示した。また、先行研究から IELTS と TOEFL iBT および IELTS と GTEC CBT の間の対応関係にもずれがあることがわかっている。よって、対照表上の他の試験間の対応関係もずれている可能性がある。対応関係がずれていると、英語力が等しい者が X という試験を受けると B2 であるが、Y という試験を受けると B1 になるというようなことが起きてしまう。対応関係がずれている対照表を大学入試に使うと、公平性と公正性が損なわれてしまうので、事前にずれの検証と修正を行うことが望ましい。しかし、大学によっては、7 種類の民間試験を使っているところもあり、それらすべての対応関係を横断的に検証することは不可能に近い。入試における公平性・公正性を重視するのであれば、複数の試験を対応づけて使うことはやめて、使う試験を 1 種類に限定するべきである。1 種類の試験のみにすると、試験結果を CEFR レベルに置き換える必要がなくなり、その試験の得点を直接合否判定に利用できるようになる。1 点刻みで合否が決まる大学入試には、そのような活用の仕方が適している。

謝辞

本研究は、JSPS 科研費 JP17K03018 の助成および神田外語大学研究助成の支援を受けて行った。ここに謝意を表す。

付記

本稿は Kanzaki (2021) と同じ得点データを使っており、同稿の内容と重複する部分がある。

データ公開

オープンサイエンスの一環として、本研究で用いた得点データを以下リンク先ページにて公開する。

ダウンロードページ短縮 URL : <https://bit.ly/MK2022>

参考文献

- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2004). *The development of specifications for item development and classification within the Common European Framework of Reference for Languages: Learning, teaching, assessment: Reading and listening: Final report of the Dutch CEF Construct Project*. Retrieved from https://eprints.lancs.ac.uk/id/eprint/44/1/final_report.pdf
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT Journal*, 51(4), 374–380. <https://doi.org/10.1093/elt/51.4.374>
- Clesham, R., & Hughes, S. R. (2020). *2020 concordance report: PTE Academic and IELTS Academic*. Retrieved from https://assets.ctfassets.net/yqwtwibiobs4/1hXHbkTLYCJly7JryACWjK/5a20dbe26d8ca2c36a3b0dd5a32868d7/2021_PTEA_2020_PTE_IELTS_Concordance_White_Paper.pdf

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. Retrieved from <https://rm.coe.int/1680459f97>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge University Press.
- 大学入試のあり方に関する検討会議 (2021)『大学入試のあり方に関する検討会議 提言』
https://www.mext.go.jp/content/20210707-mxt_daigakuc02-000016687_13.pdf
- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, 28(4), 227–246. <https://doi.org/10.1177/0146621604265031>
- Educational Testing Service. (2010). *Linking TOEFL iBT scores to IELTS scores—A research report*. Retrieved from https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf
- Educational Testing Service. (2019a). *Score user guide: TOEIC Listening & Reading test*. Retrieved from <https://www.ets.org/s/toeic/pdf/toeic-listening-reading-test-user-guide.pdf>
- Educational Testing Service. (2019b). *Score user guide: TOEIC Speaking & Writing tests*. Retrieved from <https://www.ets.org/s/toeic/pdf/toeic-sw-score-user-guide.pdf>
- Elliot, M., Blackhurst, A., O’Sullivan, B., Clark, T., Dunlea, J., & Saville, N. (2021). Aligning IELTS and PTE-Academic: A measurement study. In N. Saville, B. O’Sullivan & T. Clark (Eds.), *IELTS Partnership Research Papers: Studies in Test Comparability Series*, No. 2, (pp. 42–63). IELTS Partners. Retrieved from <https://www.ielts.org/-/media/research-reports/ielts-pte-comparisons.ashx>
- Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly*, 1(4), 253–266.
https://doi.org/10.1207/s15434311laq0104_4

- 南風原朝和 (2018) 「英語入試改革の現状と共通テストのゆくえ」 南風原朝和 (編) 『検証 迷走する英語入試 スピーキング導入と民間委託』 (pp. 5-25) 岩波書店
- 羽藤由美 (2018) 「民間試験の何が問題なのか—CEFR対照表と試験選定の検証より」 南風原朝和 (編) 『検証 迷走する英語入試 スピーキング導入と民間委託』 (pp. 41-68) 岩波書店
- 羽藤由美 (2020) 「英語入試改革の挫折から対案の可能性を探る」 宮本友弘 (編) 『変革期の大学入試』 (pp. 96-129) 金子書房
- IELTS Partners. (2021, October 22). *Test performance*.
<https://www.ielts.org/for-researchers/test-statistics/test-performance>
- Kanzaki, M. (2021). Correlations between IELTS and TOEIC scores. In P. Clements, R. Derrah, & P. Ferguson (Eds.), *Communities of teachers & learners* (pp. 397-406). JALT. <https://doi.org/10.37546/JALTPCP2020-49>
- Kim, M., Smith, W. Z., & Chin, T. Y. (2017). *Validation and linking scores for the Global Test of English Communication: White paper*. Retrieved from
<https://www.benesse.co.jp/gtec/schoolofficials/research/pdf/doc-2016-02.pdf>
- 国際ビジネスコミュニケーション協会 (2021年10月22日a) 『TOEIC Listening & Reading Testとは』 <https://www.iibc-global.org/toEIC/test/lr/about.html>
- 国際ビジネスコミュニケーション協会 (2021年10月22日b) 『TOEIC Speaking & Writing Testsとは』 <https://www.iibc-global.org/toEIC/test/sw/about.html>
- Milanovic, M. (2009). Cambridge ESOL and the CEFR. *Research Notes*, 37, 2-5. Retrieved from <https://www.cambridgeenglish.org/Images/23156-research-notes-37.pdf>
- 文部科学省 (2017) 『大学入学共通テスト実施方針』
https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afiel_dfile/2017/10/24/1397731_001.pdf

文部科学省 (2018) 『各資格・検定試験と CEFR との対照表』

https://www.mext.go.jp/b_menu/houdou/30/03/___icsFiles/afieldfile/2019/01/15/1402610_1.pdf

文部科学省 (2019) 『大学入試英語成績提供システム参加予定の資格・検定試験と CEFR との対照表』

https://www.mext.go.jp/a_menu/koutou/koudai/detail/___icsFiles/afieldfile/2019/07/09/1418843_004_1.pdf

Murphy, K. R., & Davidshofer, C. O. (2004). *Psychological testing: Principles and applications* (6th ed.). Pearson Education.

日本英語検定協会 (2020) 『IELTS 日本版受験者向け情報』

https://www.eiken.or.jp/ielts/test/pdf/information_for_candidates_jp.pdf

Powers, D. E., & Schmidgall, J. E. (2018). The TOEIC test: A brief history. In D. E. Powers & J. E. Schmidgall (Eds.), *The research foundation for the TOEIC tests: A compendium of studies* (Vol. 3, pp. 1.1–1.5). Retrieved from

<https://www.ets.org/s/toEIC/pdf/research-compendium.pdf>

Weir, C. (2005). Limitations of the Council of Europe's Framework of Reference in developing comparable examinations and tests. *Language Testing*, 22(3), 281–300. <https://doi.org/10.1191/0265532205lt309oa>