

VELC Test® の

PP 120 問版、Online 120 問版、Online 90 問版

実施データの信頼性と受験者分離指数の比較

静哲人 (大東文化大学)

[口頭発表原稿]

1. VELC Test とは

1.1 120 問版

VELC Test®とは英語能力測定・評価研究会[VELC 研究会]によって開発され、2012 年度から実施されてきている日本語を母語とする大学生のための英語力熟達度テストである(静・吉成, 2012)。リスニングとリーディングの 120 項目によって受容面からみた熟達度を測定するテストであり、主としてプレイスメントや授業効果の測定のために利用されている。その信頼性、妥当性、項目特性などについては繰り返し検証がなされてきた(静, 2012a; 2012b; 2013; 2014; 2015a; 2015b; 2017; Shizuka, 2016; 静・望月, 2014; Kumazawa et al. 2016)。実施開始から 8 年間のデータについては静(2020a)が総括的に分析し、VELC Test®が日本の大学生の英語力測定のために有効に機能してきていることを改めて確認している。2020 年には新型コロナウイルス感染症拡大によってテストの対面実施が困難となったため、オンライン版(以下 OL 版)の運用がスタートした。OL 版と従来のペーパー版(以下 PP 版)との等価性については確認されている(静, 2020b; 静・望月, 2020)。

1.2 90 問版

OL 版と同時に運用開始されたのが 90 問の短縮版である。すなわち 2021 年現在、VELC Test®には、120 問 PP 版、120 問 OL 版、90 問 OL 版が存在する。

90 問版は 120 問版をもとにしながら、6 つのパート各 20 問から難易度が異なる各 5 問を削除することにより作成した。図 1 にイメージを示す。

Listening Part 1			Listening Part 2			Listening Part 3			Reading Part 1			Reading Part 2			Reading Part 3		
Item L1-1	+	+	Item L2-1	+	+	Item L3-1	+	+	Item R1-1	+	+	Item R2-1	+	+	Item R3-1	+	+
Item L1-2	+	+	Item L2-2	+		Item L3-2	+	+	Item R1-2	+		Item R2-2	+	+	Item R3-2	+	
Item L1-3	+		Item L2-3	+	+	Item L3-3	+		Item R1-3	+	+	Item R2-3	+		Item R3-3	+	+
Item L1-4	+	+	Item L2-4	+	+	Item L3-4	+	+	Item R1-4	+	+	Item R2-4	+	+	Item R3-4	+	+
Item L1-5	+	+	Item L2-5	+	+	Item L3-5	+	+	Item R1-5	+	+	Item R2-5	+	+	Item R3-5	+	+
Item L1-6	+	+	Item L2-6	+		Item L3-6	+		Item R1-6	+		Item R2-6	+	+	Item R3-6	+	
Item L1-7	+		Item L2-7	+	+	Item L3-7	+	+	Item R1-7	+	+	Item R2-7	+		Item R3-7	+	+
Item L1-8	+	+	Item L2-8	+	+	Item L3-8	+	+	Item R1-8	+	+	Item R2-8	+	+	Item R3-8	+	+
Item L1-9	+	+	Item L2-9	+	+	Item L3-9	+	+	Item R1-9	+		Item R2-9	+	+	Item R3-9	+	+
Item L1-10	+	+	Item L2-10	+		Item L3-10	+		Item R1-10	+	+	Item R2-10	+	+	Item R3-10	+	
Item L1-11	+		Item L2-11	+	+	Item L3-11	+	+	Item R1-11	+	+	Item R2-11	+		Item R3-11	+	+
Item L1-12	+	+	Item L2-12	+	+	Item L3-12	+	+	Item R1-12	+	+	Item R2-12	+	+	Item R3-12	+	+
Item L1-13	+	+	Item L2-13	+	+	Item L3-13	+		Item R1-13	+		Item R2-13	+	+	Item R3-13	+	
Item L1-14	+		Item L2-14	+	+	Item L3-14	+	+	Item R1-14	+	+	Item R2-14	+	+	Item R3-14	+	+
Item L1-15	+	+	Item L2-15	+		Item L3-15	+	+	Item R1-15	+	+	Item R2-15	+		Item R3-15	+	+
Item L1-16	+	+	Item L2-16	+	+	Item L3-16	+	+	Item R1-16	+	+	Item R2-16	+	+	Item R3-16	+	+
Item L1-17	+		Item L2-17	+	+	Item L3-17	+		Item R1-17	+	+	Item R2-17	+	+	Item R3-17	+	
Item L1-18	+	+	Item L2-18	+	+	Item L3-18	+	+	Item R1-18	+		Item R2-18	+	+	Item R3-18	+	+
Item L1-19	+	+	Item L2-19	+		Item L3-19	+	+	Item R1-19	+	+	Item R2-19	+		Item R3-19	+	+
Item L1-20	+	+	Item L2-20	+	+	Item L3-20	+	+	Item R1-20	+	+	Item R2-20	+	+	Item R3-20	+	+

図1 120問版と90問版の使用項目の比較イメージ図。各パート内のアイテムは1～20まで項目難度で昇順にソートされている。120問版は“+”を付した全項目を使用し、90問版は“+”のない網掛け部分の項目を使用しない。

この方法により、90問版と120問版のパート毎および全体としてのテスト難易度はほぼ等しくなるよう調整されている。

理論上、90問版は120問版よりも項目数が30問少ないぶんだけ測定誤差が大きくなる。具体的にVELCスコアにどの程度の測定誤差があるかはそのスコアレベルによって異なるが、受験者が多いVELCスコア400～600程度の能力帯では、120問版で18～21程度、90問版では21～24程度が測定標準誤差であると想定されている（表1）。

表1 能力帯別の120問版および90問版のVELCスコアの測定標準誤差

VELCスコア	200	300	400	500	600	700	800
120問版の誤差	30	23	18	18	21	30	46
90問版の誤差	37	26	22	21	24	32	53

このような根拠により、90問版には120問版とまったく同じ精度はないものの、それでも主たる用途である学内プレイスメントや授業効果測定などの目的にとっては十分な精度を保持するはずであると考えられる。ただし、これはあくまで理論的な想定であり、実際の受験者データによって得点の信頼性などを検証したものではない。

2 静(2021): 模擬データでの検証

静(2021)は、120問版の受験者データから、それぞれのフォームに対応する90問版のデ

ータを模擬的に作り出し、妥当性と信頼性を検証した。この場合の妥当性とは TOEIC L&R との相関であり、信頼性とはクロンバック α である。

2.1 データ提供者

データ提供者は、研究の目的について説明を受けた上でスコアの提供に同意した首都圏の大学生 ($n = 135$) である。そのうち 112 名は、2019 年末から 2020 年春にかけて VELC Test® 120 問 PP 版を受験してかつ TOEIC® L&R のスコアを提供してくれた学生、23 名は 2021 年 1 月～2 月に VELC Test® 120 問 OL 版を受験してかつ TOEIC® L & R スコアを開示してくれた学生である。詳細を表 2 に示す。

表 2 データセット別 VELC Test 実施時期、フォーム、モード、受験者属性と人数

	実施時期	フォーム	モード	受験者所属大学	n
セット 2051	2019 年 12 月	2051	PP	A 大学	14
セット 2052	2020 年 1 月	2052	PP	B 大学	53
	2021 年 1～2 月		OL	B 大学	23
セット 2055	2020 年 3 月	2055	PP	B 大学 (14)、C 大学 (26)、 D 大学 (5)	45

前述したように OL 版については PP 版との等価性が確認されている (静, 2020b; 静・望月, 2020) ため、今回の分析では PP 版と区別せず扱った。そこで、フォーム別に整理し、フォーム 2051 ($n = 14$) の解答データを「セット 2051」、フォーム 2052 ($n = 76$) の解答データを「セット 2052」、フォーム 2055 ($n = 45$) の解答データを「セット 2055」と呼ぶこととする。

2.2 記述統計

まずセット毎に 90 問版と 120 問版の記述統計を表 3 に示す。平均値としてはセット 2052 と 2055 がおおよそ同じで、セット 2051 がそれよりも 90 問版で 15 点、120 問版で 20 点ほど低い。一方、標準偏差に関してはセット 2051 と 2052 には大きな差がなく、セット 2055 のみ明らかに大きい。

表 3 90 問版と 120 問版の記述統計

	n	90 問版		120 問版	
		平均	標準偏差	平均	標準偏差
セット 2051	14	50.14	11.33	66.64	14.40
セット 2052	76	65.82	10.70	85.87	14.27
セット 2055	45	64.51	14.39	87.04	18.47

標準偏差の傾向についてはセット 2051 とセット 2052 はそれぞれ単一の大学に所属する受験者によるデータであるのに対し、セット 2055 は 3 つの異なる大学に所属する学生のデータを合わせたものであることの反映だと考えられる。

2.3 信頼性

次に信頼性の指標として、それぞれのセットにおける 90 問版と 120 問版のクロンバック α を表 4 に示す。

表 4 90 問版と 120 問版の信頼性係数（クロンバック α ）の比較

セット	n	90 問での α	120 問での α
セット 2051	14	0.8798	0.8958
セット 2052	76	0.8870	0.9104
セット 2055	45	0.9359	0.9492
3 セット合体	135	0.9181	0.9333

まず 120 問版の数値を見ると、セット 2051 は、サンプルが $n = 14$ と非常に少ないにもかかわらずほとんど .90 に近い値である。セット 2055 は .9492 と最も高い。誤解を避けるため述べておくと、これはテストフォームとして 2055 が他のフォームよりも「信頼性」の高い得点を得やすい、ということではない。

クロンバック α はその算出式からも明らかなように、テスト項目の性能とともに受験者能力の散らばりの関数でもある。セット 2055 の α の値の高さは、異なる 3 つの大学に所属する受験者の標準偏差が他のセットよりも明らかに大きかったことが原因であると考えられる。90 問版の値を見ると最も低いセット 2051 であっても .8798 と十二分に高い。最も高いセット 2055 では .9359 である。セット毎に 90 問版と 120 問版の値の差を比較してみると、0.01~0.02 程度と小さい。

2.4 基準関連妥当性その 1: 120 問版との相関

3 つのセットそれぞれに関しての 90 問版と 120 問版の相関を表 5 に、ヒストグラムおよびプロットを図 2~図 4 に示す。

表 5 90 問版と 120 問版の相関係数および上下 95% 区間

	n	相関係数	下側 95%	上側 95%
セット 2051	14	0.987903	0.961089	0.996274
セット 2052	76	0.983780	0.974458	0.989718
セット 2055	45	0.989232	0.980372	0.994105

すべてのセットにおいて相関係数は $r = .98$ を上回っており、極めて高い。すなわち、プレイスメント目的であれば 90 問版は 120 問版の代替として十分に機能すると言えよう。

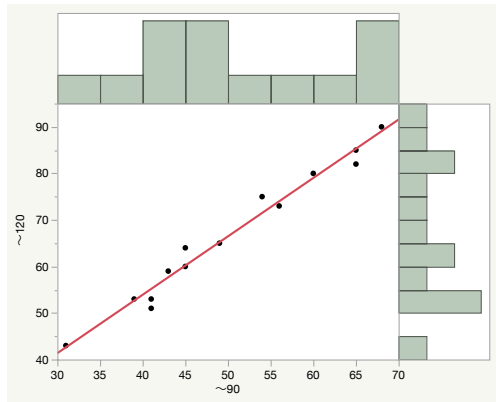


図2 90問版と120問版のヒストグラムおよび散布図（セット2051）

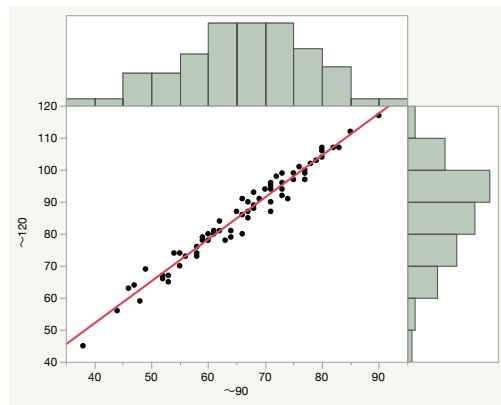


図3 90問版と120問版のヒストグラムおよび散布図（セット2052）

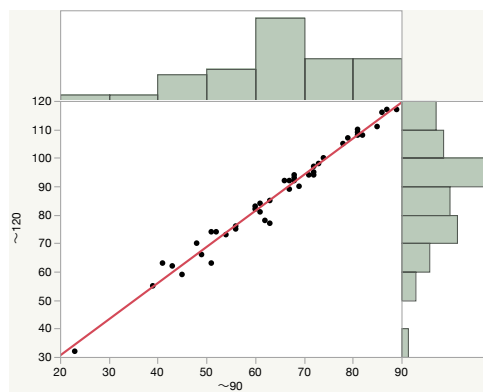


図4 90問版と120問版のヒストグラムおよび散布図（セット2055）

2.5 基準関連妥当性その2: TOEIC との相関

次に TOEIC® L&R のトータルスコアを基準としたときの、基準関連妥当性を吟味するが、数値を示す前に指摘しておきたいことがある。それは、VELC Test®は設計思想からして、TOEIC® L&R を模そうという意図がそもそもないということである。その意味で TOEIC® L&R

との相関の高低は、VELC Test®の日本語母語大学生の英語熟達度テストとしての妥当性の高低を必ずしも意味しない。VELC Test®はそれ自体として個別受験者が日本の大学生全体を基準としてスキル別にどのような位置にいるのかという集団基準準拠的情報と、スキル別にどの程度のタスクがどの程度の正確さでできるのかという目標基準準拠的情報の両方をもたらすことのできる測定ツールである、と開発者として考えている。

そのことを確認した上で、TOEIC® L&R との相関を表 6 に示す。3つのセットは素点ベースでは厳密には等価されているとは言えないが、事実上はおおむね等しいとみなせると思われるため、3セットを合わせた場合の相関係数も求めてみた。

表 6 TOEIC L&R を基準とした場合の 90 問版と 120 問版の妥当性係数

セット	<i>n</i>	90 問版	120 問版
セット 2051	14	.7168	.7016
セット 2052	76	.7361	.7592
セット 2055	45	.8283	.8394
3 セット合体	135	.7962	.8080

セット別に比較すると、やはり α の高低と同じように、セット 2051 が一番低く、セット 2055 が一番高い。セット 2051 に関しては 90 問版のほうが 120 問版より値がわずかながら高いが、これはサンプルサイズの小ささに起因するノイズであると解釈するのが妥当だと思われる。セット 2055 の値の高さは、やはりこのセットを受けた受験者グループの能力幅 (=標準偏差) の広さの表れであると考えられる。3セット合体した場合の相関係数は 90 問版で $r = .7962$ 、120 問版で $r = .8080$ である。

なお、再度指摘しておく、この相関係数は素点との数値である。Rasch モデリングを経た VELC スコア、さらに 6 つのパート別 VELC スコアのすべてを使用した回帰式で求める予測点 (TOEIC 相当点) を用いたときの数値は、より高い。具体的には 3セット合体した場合、120 問版の VELC スコアトータルと TOEIC® L&R の相関は $r = .8373$ 、予測点 (TOEIC 相当点) との相関は $r = .8437$ である (表 7)。

表 7 3セット合体での素点、VELC スコア、予測点の TOEIC L&R との相関係数

	素点	VELC スコア Total	予測点 (相当点)
TOEIC L&R Total	.8080	.8373	.8437

すなわち、今回は簡易的に素点ベースで算出した $r = .8080$ (120 問版) という数値に関しては、絶対値に着目するのではなく、同じく素点ベースで算出した $r = .7962$ (90 問版) とそれほど値に差がない、という点にのみ着目するのが適当である。

以上、静(2021) は、120 問版 3 フォームと、そこから模擬的に作り出した 90 問版 3 フォームを、信頼性係数と TOEIC® L&R を基準にした場合の基準関連妥当性を調査した。その結果、想定どおり、いずれのフォームにおいても 90 問版は 120 問版よりも信頼性・基準

関連妥当性ともにやや減ずるものの、テスト目的に照らして十分に高い信頼性と基準関連妥当性を保持していることを確認した。

3 本研究: 実データでの検証

静(2021)は、模擬データをもとに90問版の信頼性と妥当性を探ったわけだが、実際の90問版の受験データを検証したものではない。そこで、本研究では実際のOL90のデータをOL120、PP120と比較することとした。

3.1 目的

本研究はPP120、OL120、OL90の受験データを信頼性と受験者分離の観点から検証する。基準関連妥当性の検証にはTOEIC L&Rなどの別の基準スコアが必要となるが、今回調べるVELC Testのデータには付随するTOEICデータがない。またVELC Testが利用される主たる目的のひとつが同一大学内でのプレイスメントである。これらの2点に鑑み、PP120、OL120、OL90を実施した各大学内のスコア分布を中心に検証することに一定の意義があると考えた。

3.2 データ

テストデータを管理しているVELC研究会事務局(金星堂)より、受験者の個人情報削除された形式で、2021年3月から6月に実施されたPP120、OL120、OL90の全実施データの提供を受けた。同一大学で同一日に実施した場合であっても、複数フォームを実施した場合には別々のデータセットとして扱ったところ、データセットの受験人数の最小は $n = 3$ 、最大は $n = 4659$ 、平均が $n = 631.8$ であった。VELCスコアの平均値は、最低で376.6、最高で623.1、平均が487.1であった。(この数値は静(2020)が総括した、2012~2019年度の120PPの平均480.0とほぼ一致する)。標準偏差が最小のセットは、 $SD = 38.5$ 、最大のセットは、178.6、標準偏差の平均は $SD = 62.2$ であった(この数値は2012~2019年度の120PPデータの平均73.3よりやや低い)。ベルクスコアの幅(最低スコアと最高スコアの差)は、最小で95、最大で688、平均で391.2であった。同一の大学内でもかなりのスコア差があったことがわかる。

これらの全データの中から分析対象を絞ってPP120、OL120、OL90を比較対照するため、それぞれのテスト形式(注:フォームは問わない)の受験によって得られた各4セット(4つの大学)、合計12セット(12の大学)を抽出した。抽出に当たっては以下の手順によった。

- (1) 各テスト形式の中で、VELCスコアでソートする。
- (2) VELCスコア平均が449未満(レベル4)、450~499(レベル3)、500~549(レベル2)、550以上(レベル1)という4つのレベルに分ける。

(3) それぞれのレベルのなかで最も受験者数が少ない（ただし 50 名を下回らない）セットを選ぶ。

他の条件が等しければ受験者数が多くなるほど英語力の散らばりも大きくなる傾向があるので、弁別・分離の観点からは最も条件の悪い集団を選んだといえる。こうして抽出されたのは最小が $n = 55$ 、最大が $n = 258$ の 12 大学のデータである。

3.3 結果

テスト形式 (PP120、OL120、OL90 に)、1~4 の数値のいずれかを付して PP120_1、PP120_2、... OL90_3、OL90_4 と呼ぶこととする。1~4 の数値は、スコアレベル (4: 449 未満、3: 450~499、2: 500~549、1: 550 以上) を表す。すなわち OL120_3 は、OL120 を受験し、平均 VELC スコアが 450~499 であったセットである。

3.3.1 スコアの分布状況

VELC スコア (トータルスコア) の分布状況を表 8 に、リスニングスコアの分布状況を表 9 に、リーディングスコアの分布状況を表 10 にまとめた。リスニング、リーディングについてはトータルスコアの傾向とよく似ているため、トータルスコアに絞って傾向を記す。

(1) 平均値については、OL120、OL90、PP120 いずれのグループにおいてもレベル 1 > レベル 2 > レベル 3 > レベル 4 である。これはスコアバンド別にそのようになるように大学を抽出したものである。

(2) 同じく平均値について、同じレベルに属するセットであっても、テストの種類によってある程度スコアが異なる。例えばレベル 1 を比べると、OL120_1 は 591.8、OL90_1 は 623.1、PP120_1 は 572.3 である。言うまでもなくこれはテストの種類が異なることの効果ではなく、それぞれのテストを受けた大学の受験者集団の英語力の違いが現れたものである。本研究の目的のためには、大学ごとに英語力が異なることはまったく問題にはならない。

(3) 大学ごとに英語力が多少あるいはかなり異なるのは既知の現実であって意外性はない。本研究で着目すべきはそのような平均値の相対的高低ではなく、スコアの散らばり (dispersion) である。もっとも英語力の高いレベル 1 では天井効果が起こっていないか、もっとも英語力の低いレベル 4 では床効果が起こっていないか、いずれでないにしても、同一集団内で「団子状態」になっているケースがないか、などがチェックすべき点である。

標準偏差の欄を見るともちろんそれぞれのセットで異なるが、OL のほうが PP よりも小さい、あるいは 90 問版のほうが 120 問版よりも小さい、などの傾向は読み取れない。テスト形式が同じ 4 つの標準偏差の平均は、OL120 が 63.9、OL90 が 59.8、PP120 が 60.25 である。現実に見られる違いは今回抽出した特定大学集団の特性だと解釈したほうが適切だ

と思われる。また最低点と最高点の差（スコア幅）を見てみても、天井効果も床効果も見られないと言ってよいだろう。

表8 VELC（トータル）スコアの状況

	<i>n</i>	平均値	標準偏差	標準偏差平均	最低	最高	スコア幅
OL120_1	101	591.8	70.3	63.9	445	916	471
OL120_2	258	519.2	76.5		325	737	412
OL120_3	55	491.8	67.7		358	641	283
OL120_4	63	424.0	41.1		325	546	221
OL90_1	126	623.1	64.6	59.8	444	839	395
OL90_2	223	519.9	58.4		334	673	339
OL90_3	212	485.2	59.4		304	662	358
OL90_4	190	441.0	57.2		345	635	290
PP120_1	220	572.3	59.9	60.28	376	774	398
PP120_2	203	505.7	65.2		304	772	468
PP120_3	125	469.3	65.5		350	740	390
PP120_4	97	437.0	50.4		311	580	269

表9 VELC（リスニング）スコアの状況

	<i>n</i>	平均値	標準偏差	標準偏差平均	最低	最高	スコア幅
OL120_1	101	613.4	87.1	73.6	423	943	520
OL120_2	258	541.1	91.2		310	787	477
OL120_3	55	506.9	69.8		366	652	286
OL120_4	63	438.8	46.3		322	572	250
OL90_1	126	618.9	76.2	71.0	448	841	393
OL90_2	223	516.1	65.9		294	707	413
OL90_3	212	492.9	77.5		274	796	522
OL90_4	190	461.1	64.3		312	687	375
PP120_1	220	575.8	73.7	69.4	399	831	432
PP120_2	203	509.1	69.9		290	826	536
PP120_3	125	494.3	75.7		369	770	401
PP120_4	97	447.6	58.4		291	652	361

表 10 VELC（リーディング）スコアの状況

	<i>n</i>	平均値	標準偏差	標準偏差平均	最低	最高	スコア幅
OL120_1	101	581.4	69.8	68.5	408	890	482
OL120_2	258	504.8	78.0		273	739	466
OL120_3	55	472.6	76.5		322	653	331
OL120_4	63	412.0	49.7		306	549	243
OL90_1	126	620.4	83.0	67.9	412	862	450
OL90_2	223	499.3	65.1		323	650	327
OL90_3	212	479.9	60.9		297	651	354
OL90_4	190	416.7	62.4		301	609	308
PP120_1	220	579.6	71.1	65.8	357	793	436
PP120_2	203	505.7	74.1		311	741	430
PP120_3	125	450.4	63.5		315	738	423
PP120_4	97	420.3	54.6		305	547	242

3.3.2 ヒストグラムによる視覚的確認

前項で数値から読み取った分布状況を視覚的に確認するために、それぞれの大学のトータルスコア、リスニングスコア、リーディングスコアの分布をヒストグラムにしたのが、図 5～図 16 である。いずれのレベルにおいてもおおむね正規分布に近い形でスコアが分布し、天井効果も床効果も見られないことが確認できる。

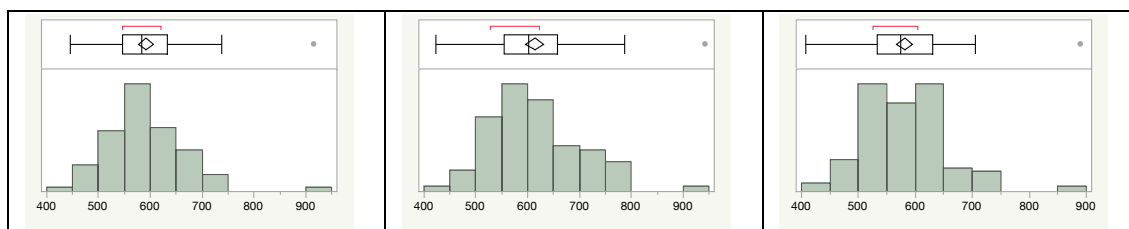


図 5 120OL_1 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

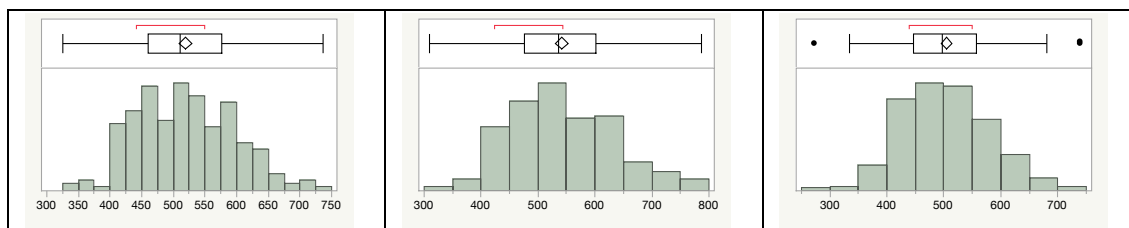


図 6 120OL_2 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

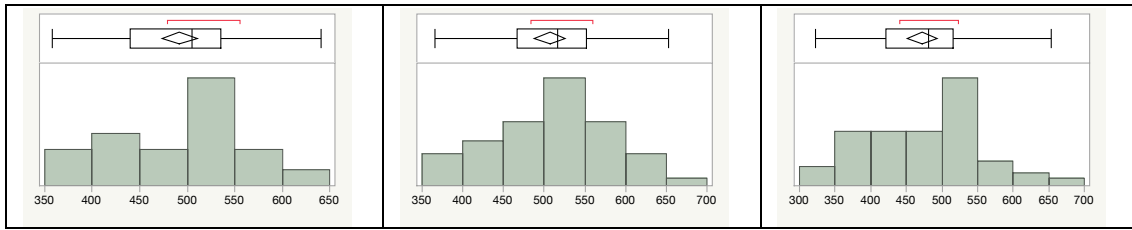


図7 120OL_3のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

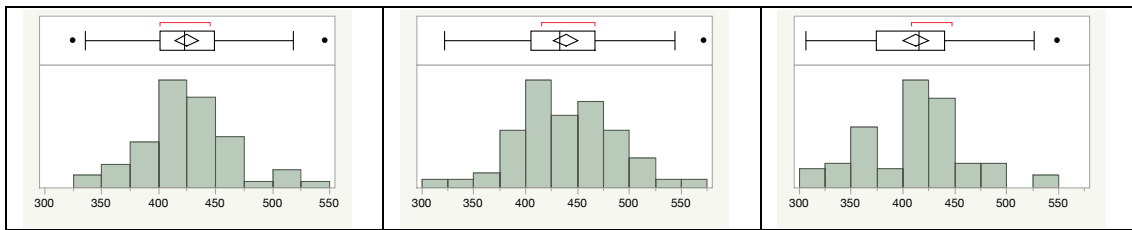


図8 120OL_4のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

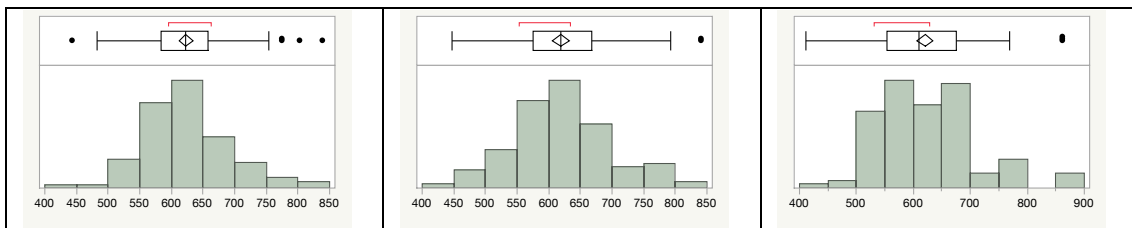


図9 90OL_1のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

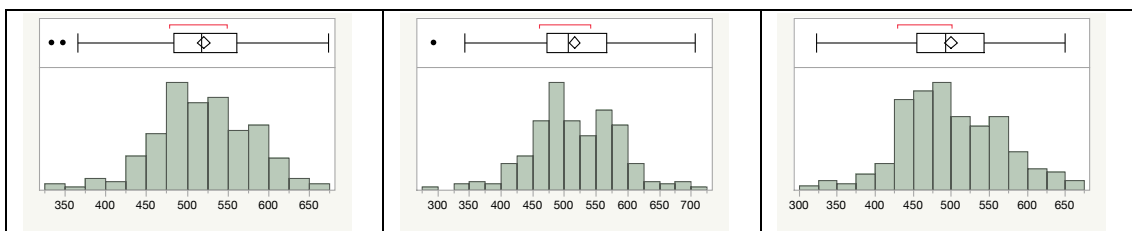


図10 90OL_2のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

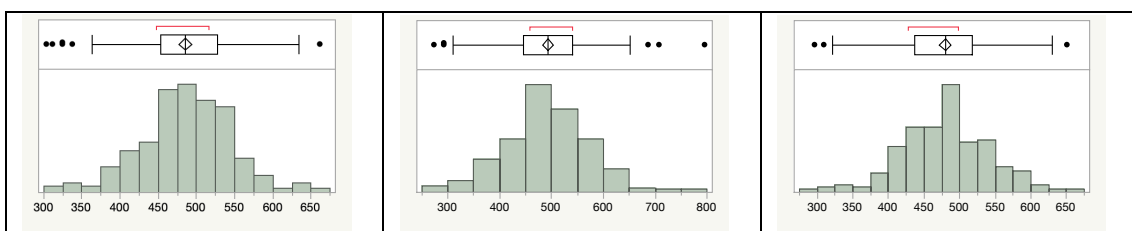


図11 90OL_3のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

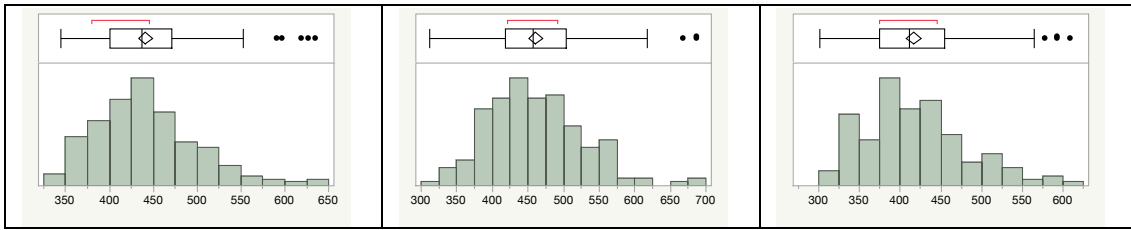


図 12 90OL_4 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

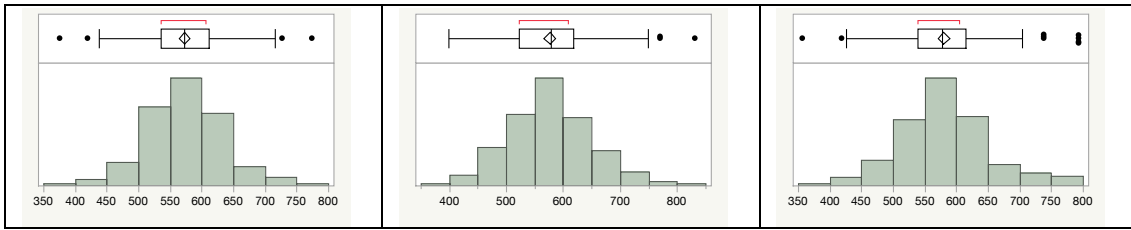


図 13 120PP_1 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

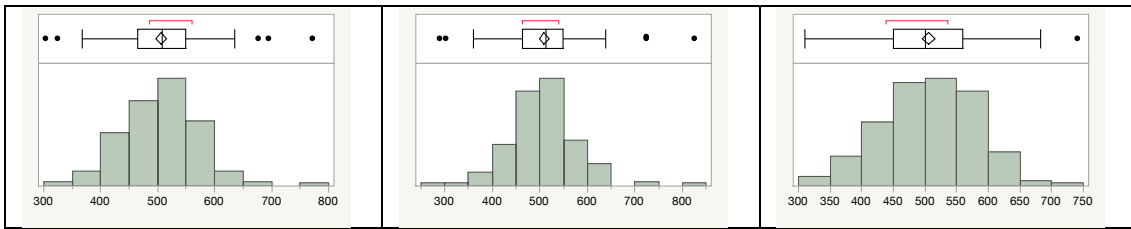


図 14 120PP_2 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

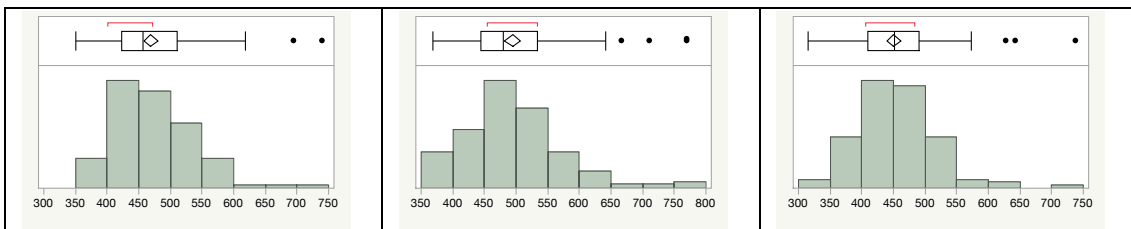


図 15 120PP_3 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

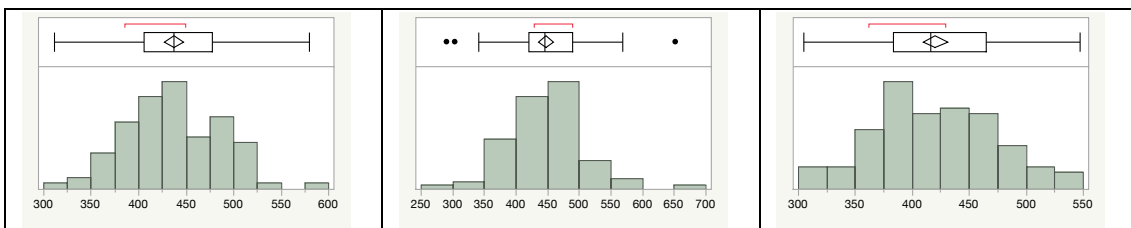


図 16 120PP_4 のトータルスコア (左)・リスニングスコア (中)・リーディングスコア

同じ情報を、同一のスケールで一覧したのが、図 17 ～図 19 である。トータルスコア、リスニングスコア、リーディングスコアのいずれに関しても、比較的英語力が高い方に分布している大学も、比較的低いほうに分布している大学も、それぞれの集団のなかで大きくスコアが散らばっていることがわかる。

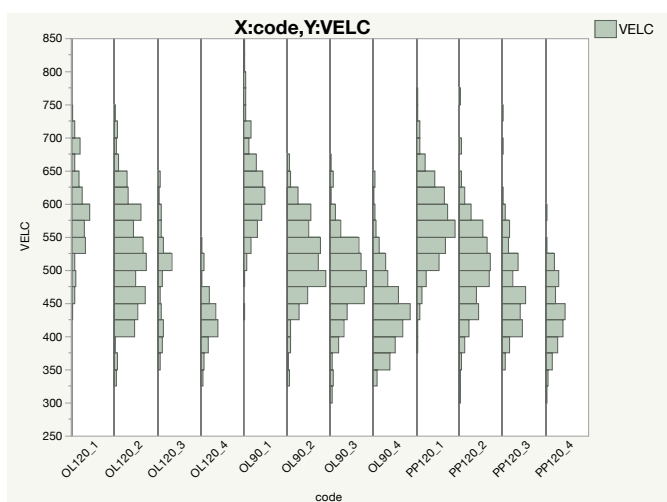


図 17 VELC スコア（トータル）の 12 大学の分布

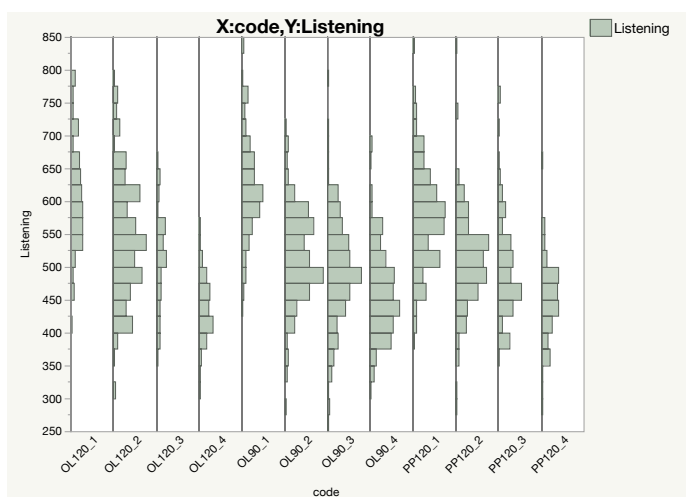


図 19 VELC スコア（リスニング）の 12 大学の分布

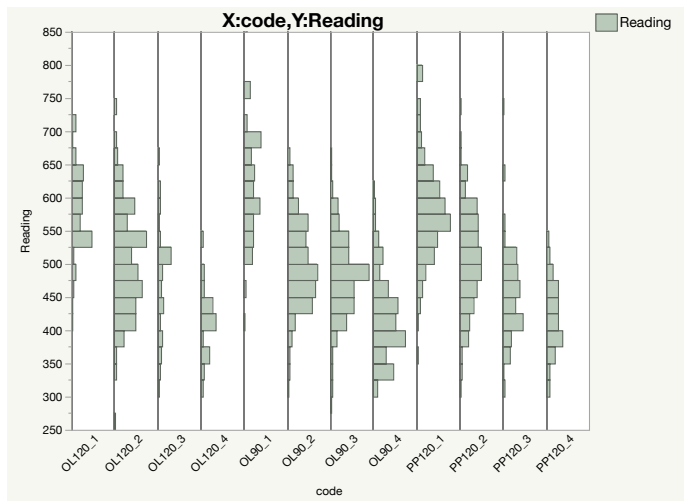


図 20 VELC スコア（リーディング）の 12 大学の分布

3.3.3 クロンバック α

信頼性の指標としてのクロンバック α はすべての項目に対する応答に基づいて計算した。すなわち 90 問版についてはリスニングとリーディングを合わせた 90 項目、120 問版についてはリスニングとリーディングを合わせた 120 項目を用いて求めた（表 11）。

表 11 12 セットのクロンバック α

	Level 1	Level 2	Level 3	Level 4	平均
OL120	.90	.94	.93	.79	.890
OL90	.84	.87	.87	.86	.860
PP120	.89	.92	.91	.87	.898

同一バージョン 4 大学の平均は、PP120 が.898、OL90 が.860、OL120 が.890、であった。信頼性は直接項目数に影響されるので、120 問版よりも 90 問版の数値が低いのは当然であるが、それでも平均で .860 は十分に高いとみなせる。

3.3.4 受験者分離度

受験者分離度（person separation）はラッシュモデリングのソフトウェア Winsteps のアウトプットのひとつで、受験者集団を統計的に異なるいくつかのレベルに分離可能であるかを示すものである。表 1 2 に示す。受験者分離指数の平均は、PP120 が 2.95、OL120 が 2.99、OL90 が 2.34 であった。

大学内でのプレイスメントは 3 レベル程度で行われることが多いが、そのためには 120 問版は十分に、90 問版はほぼ十分に機能していると言える。PP 版と OL 版には差がなく、120 問版と 90 問版には多少の差が認められたわけだが、これは信頼性が項目数の関数であ

るためである。

表 12 12 セットの受験者分離度

	Level 1	Level 2	Level 3	Level 4	平均
OL120	2.93	3.76	3.41	1.87	2.99
OL90	2.11	2.44	2.44	2.35	2.34
PP120	2.78	3.31	3.22	2.47	2.95

レベル別にみると中央の 2 レベルが上端と下端の 2 レベルよりも信頼性・分離の程度がわずかに高い傾向が観察された。これも理論上の想定どおりである。ただしレベル 4 の受験者分離指数の平均が 2.61、レベル 1 の平均が 2.23 と、いずれのレベルであっても実用上十分に高い。

4. 結論

本研究は、PP120、OL120、OL90 を、スコアの分布、信頼性の値、受験者分離度に関して吟味した。これらのテストの全データのなかから、それぞれ 4 つのレベルの大学を抽出して比較してみた。その結果、VELC Test® のオンライン版はペーパー版と同等の精度をもって、90 問版は 120 問版とは同等とは言えないまでも十分な測定精度をもって、平均受験者のレベルに関わらず、個別大学の受験者集団内の英語力の個人差を顕在化していることが確認できたと言える。

引用文献

- 静哲人(2012a)「VELC Test®による TOEIC スコアの予測：リスニングとリーディングについて示唆されるもの」日本言語テスト学会第 16 回全国研究大会(2012. 10. 27) 専修大学生田キャンパス.
- 静哲人(2012b)「VELC Test®の妥当性を検証する：2012 年度データにもとづいて」2012 年度 JACET 関西支部秋季大会(2012. 11. 24) 京都産業大学.
- 静哲人(2013)「VELC Test®の測る英語力構造：確認的因子分析がスコアレポート方式に示唆するもの」大学英語教育学会第 52 回国際大会(2013. 8. 30). 京都大学吉田キャンパス.
- 静哲人(2014)「VELC Test® フォーム A の選択肢分析から見える各アイテムの特性」大学英語教育学会第 53 回国際大会(2014. 8. 28). 横浜市立大学.
- 静哲人(2015a)「VELC Test® フォーム A の選択肢特性分析」大東文化大学語学教育研究所創設 30 周年記念フォーラム, 97-115.
- 静哲人(2015b)「VELC Test®の概要とよくある質問：Listening Section Part 2 の作問意図と項目特性」ベルク研究会第 4 回研究会基調講演(2015. 9. 12). 研究社英語センター.

- 静哲人 (2017) 「2017 年度実施 VELC Test® データからみる同一大学内での受験者分離の成功度」日本言語テスト学会第 21 回研究大会 (2017. 9. 10) 会津大学.
- 静哲人 (2020a) 「VELC Test® 2012-19 年度実施データの分析および総括」『語学教育研究論叢』第 37 号, 75-89.
- 静哲人 (2020b) 「VELC Test® Online と VELC Test® P&P の等価性を検証する (その 1)」言語教育 EXPO2021 (2020. 10. 25 オンラインにて開催).
- 静哲人 (2021). VELC Test® 短縮版の信頼性および基準関連妥当性の検証: 項目数の漸減はテスト特性にどの程度影響を与えるか? 第 6 回 VELC 研究会基調講演 (2021.9.3. オンラインにて開催)
- 静哲人・望月正道 (2014) 「日本人大学生のための標準プレイスメント・テスト開発と妥当性の検証」 *JACET Journal* 58, 121-141.
- 静哲人・望月正道 (2020) 「VELC Test® Online と VELC Test® P&P の等価性を検証する (その 2)」日本言語テスト学会 (2020. 12. 12) Zoom 上にて開催.
- 静哲人・吉成雄一郎 (2012) 「大学生の英語力『可視化』の試み: 熟達度診断のための VELC Test® の開発」 第 51 回大学英語教育学会研究大会 (2012. 9. 1) 愛知県立大学.
- Bond, T. G., & Fox, C. M. (2007) *Applying the Rasch model. (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kumazawa, T. Shizuka, T. Mochizuki, M., & Mizumoto, A. (2016). Validity argument for the VELC Test® score interpretations and uses. *Language Testing in Asia* 6:2 <https://doi.org/10.1186/s40468-015-0023-3>
- Linacre, J. M. (2005) Winsteps (Version 3.55) [Computer software]. <http://www.winsteps.com/>
- Shizuka, T. (2016) Modification of VELC Test® listening section part 2 type multiple-choice 1-blank partial "dictation" items: Effects on distractor discriminations and TOEIC®-relatedness. 大学英語教育学会第 55 回国際大会 (2016. 9. 3). 北星学園大学.