

VELC Test[®] スコアで
新 TOEIC[®] L&R Test スコアを予測する
— 式をアップデートする必要があるのか —

静 哲人

Using VELC Test[®] Scores
to Predict New TOEIC[®] L&R Test Scores:
Do We Need to Update Our Prediction Formula?

SHIZUKA Tetsuhito

Abstract

This study examined the current formula for predicting TOEIC[®] L&R scores from the VELC[®] Test scores, using a post-2016 data set, i.e., scores on the revised TOEIC[®] Test. A total of 112 L1-Japanese university students sat for the VELC Test[®] administered sometime between December 2019 and March 2020. They either had a known post-2016 TOEIC[®] score or sat for the TOEIC[®] Test on the same day that they took the VELC Test[®]. Their predicted TOEIC[®] scores computed by the formula from a regression analysis of a pre-2016 data set (i.e. based on the “old” TOEIC[®]) were compared against their actual TOEIC scores. The results indicate that the VELC Test[®] continues to predict post-2016 TOEIC[®] scores to the same satisfactory extent as it did pre-2016 TOEIC[®] scores, meaning that there is no need to update the current prediction formula. Meanwhile, the discrepancies between the actual and the predicted TOEIC[®] scores turn out to correlate weakly but positively with the predicted TOEIC[®] scores, indicating that the predicted scores tend to be lower than the actual scores toward the high end of the score spectrum and higher than the actual scores toward the lower end of the score spectrum.

キーワード：VELC Test[®] 新 TOEIC[®] L&R 予測式 重回帰分析 決定係数

1. 背景

1.1 VELC Test[®] とは

VELC Test[®] は英語能力測定・評価研究会が開発し、2012 年度より実施・運用を行ってきた、主として日本語を母語とする大学生を対象とした英語熟達度テストである（静・吉成，2012；静・望月，2014；静，2020）。リスニングセクション 60 問とリーディングセクション 60 問からなり、各セクションがそれぞれ L1・L2・L3 と、R1・R2・R3 の 3 つのパートに分かれている。各パートの具体的な問題形式については静（2020）を参照されたい。

受験者には 120 項目全体にもとづく VELC スコア、セクションごとのリスニングスコアとリーディングスコア、さらに各パートに基づく L1, L2, L3, R1, R2, R3 スコアが返される。これらのスコアはいずれも素点ではなく、ラッシュモデル（Rasch 1960）で推定された受験者能力値を、理論的には平均 500、標準偏差 100 として分布するように変換された値である。開発段階においてのべ 5,000 名を超える日本人大学生に施行した結果より大学生全体の平均的熟達度を推定し、それが VELC スコア 500 になるよう設定されたものである。

2012 年度の本実施開始以来、VELC Test[®] の信頼性、妥当性、項目特性などについては様々な報告がなされてきた（静，2012a；2012b；2013；2014；2015a；2015b；2017；2020；Shizuka，2016；静・望月，2014；Kumazawa *et al.* 2016）。静（2020）は、VELC Test[®] の本実施データが 8 年分蓄積したタイミングで、改めてテストの性能や特性を確認するため、複数年度データに関して、(1) 受験者スコアの分布状況、(2) 項目難度の安定性、(3) 項目のラッシュモデルへの適合度、(4) 信頼性と受験者分離度、を調査した。その結果、本実施受験者スコアの分布は、開発時の「全体平均が 500 で標準偏差が 100」という想定を幾分下回っている傾向はあったものの、全体として VELC Test[®] は日本の大学で学ぶ学習者の英語熟達度を測定するツールとして有効的に機能していることが再確認された。

1.2 VELC スコアによる TOEIC[®] L&R スコアの予測

VELC Test[®] の主たる用途のひとつは大学におけるプレイスメントであるが、その用途で現在最も多くの受験者に利用されているものは TOEIC[®] L&R であろう。国際ビジネスコミュニケーション協会（The Institute for International Business Communication [IIBC]）（2020）によれば、2019 年度 TOEIC[®] L&R の IP テスト受験者数は延べ 116 万 7 千人である。

ここで指摘しておかねばならないことは、VELC Test[®] の設計思想は TOEIC とはそ

もそも異なるという点である。VELC スコアは日本全体の大学生母集団を想定したときの集団基準準拠的な解釈と、現実のタスクに照らしたときの目標基準準拠的な解釈が可能なものである。すなわち、VELC リスニングスコア、VELC リーディングスコアの特定の数値に関して、それは日本の大学生全体のなかでおおよそどのあたりに位置するもので、かつ状況別の Can Do レベル(特定の活動において何%くらいが理解できるか)はどの程度であるか、についての情報をフィードバックする。すなわち VELC Test[®] のスコアはそれ単体で受験者の英語力についての多様な視点からの情報を提供する。

しかしながら TOEIC[®] L&R のシェアが圧倒的に大きい状況下において、テストユーザー(実施する大学および受験する学生)の側が、「VELC スコアは TOEIC スコアとどういう関係があるのか」「VELC スコアを TOEIC スコアに換算するとどうなるのか」という点に関心を寄せるのは、その是非はともかく、避けられないひとつの現実である。このような観点から VELC Test[®] は本実施初年度から結果レポートの一部として、TOEIC 換算点を提供してきている。

1.3 VELC スコアから TOEIC[®] L&R スコアへの換算式と精度

VELC Test[®] が TOEIC 換算点を算出する式は以下のようにして求めたものである(静・望月, 2014)。開発段階(2011年)のトライアル協力者(すべて日本語を母語とする大学生)のなかで、比較的近い時日(1年以内程度)に受験した TOEIC(当時は L&R という名称はなかった)スコアが明らかであった 375 名に関して、重回帰分析を行った。目標変数は TOEIC リスニングスコア(以下、TOEIC_L)および TOEIC リーディングスコア(以下、TOEIC_R)、予測変数は VELC の L1, L2, L3, R1, R2, R3 の各パートスコアである。まず TOEIC_L と TOEIC_R をそれぞれ別々に予測し、ふたつの予測値の単純合計を TOEIC トータルスコア(以下、TOEIC_T)の予測値とすることとした。

ステップワイズ回帰分析で得られたモデルの中から Cp 統計量を目安にして最終的に選択した予測式と、その決定係数(R^2)を表1に示す。

TOEIC_L に関しては R1 と R2 を除いた予測変数を、TOEIC_R に関しては 6 つの予測変数をすべて利用するモデルが選択された。これらのモデルによる予測値と実際の TOEIC スコアの相関係数は、TOEIC_L、TOEIC_R、TOEIC_T に関してそれぞれ .763、.799、.825 であった。これを決定係数になおすとそれぞれ .583、.639、.680 なので、VELC[®] Test は TOEIC_T の分散のおおよそ 68% を説明できたことになる。2012 年の本実施開始以来現在までこの式によってユーザーに TOEIC 予測値を提供してきている。70 分で終了する 120 項目のテストが、120 分かかる 200 項目のテストの分散の 7 割近くを説明するのは興味深い。テストである限り必ず存在する TOEIC 側

**表 1 VELC スコアから TOEIC スコアを予測する式とその決定係数
および重相関係数**

	TOEIC_L	TOEIC_R	TOEIC_T
予測式	TOEIC_L = - 74.88592 + 0.0753525 * L1 + 0.1986157 * L2 + 0.2481834 * L3 + 0.1187556 * R3	TOEIC_R = - 199.5994 + 0.0746046 * L1 + 0.0792655 * L2 + 0.1483208 * L3 + 0.1086506 * R1 + 0.1735139 * R2 + 0.2118343 * R3	TOEIC_T = TOEIC_L + TOEIC_R
R ²	0.583	0.639	0.680
重相関係数	0.763	0.799	0.825

注：ユーザーに示す予測スコアは、5 点刻みの数値になるように四捨五入

のスコアの測定誤差の存在を考えたとき、重相関係数 .825 という数値は、別形式のテストのスコアが TOEIC スコアとの間に取りうる数値のなかでおおよそ上限に近いのではないかと考えられる。

1.4 TOEIC[®] L&R Test の形式変更

その一方、2016 年に TOEIC[®] L&R の問題形式の一部とセクションごとの項目数のバランスに変更があった。実施団体である国際ビジネスコミュニケーション協会が 2016 年 1 月に発行した『IIBC Newsletter 特別号』（IIBC, 2016 とする）によれば、「よりオーセンティック（实际的）な出題形式を採用」とし、具体的な出題形式変更の概要としては表 2 を掲げている。

このような変更に関して、開発元である Educational Testing Service (ETS) のエグゼクティブ・ディレクターは、これによるテストの難易度に変化は生じない、としている（IIBC, 2016）。Q&A の回答内の該当箇所を以下に引用する（p.11）。

Q5：変更後の TOEIC テストの 600 点と現行の TOEIC テストの 600 点は同じレベルを意味しますか。

A：変更後の TOEIC テストの 600 点と現行の TOEIC テストの 600 点は同じレベルを意味します。変更後の TOEIC テストは従来のテストと難易度、テストの長さ、出題数、評価スケール（10 点～990 点）、そしてクオリティに変化が出ないよう、ETS のテスト開発の専門家により設計されました。変更後のスコアは現行のスコアと比較できます。

表2 2016年のTOEICテストの出題形式の変更の概要（IIBC, 2016より抜粋）

リスニングセクションの変更点	リーディングセクションの変更点
<ul style="list-style-type: none"> ・写真描写問題（Part 1）と応答問題（Part 2）の設問数が減る。 ・会話問題（Part 3）の設問、数が増える。 ・会話問題の中に、発言が短くやりとりの多いものが加わる。 ・3名で会話する設問がある。 ・Elision（省略形：going to が gonna など）、Fragments（文の一部：Yes, in a minute: Down the gall: Could you? など）を含む会話が流れる。 ・会話やトークの中で聞いたことと、問題用紙に印刷された図などで見た情報を関連づけて解答する設問が加わる。 ・会話やトークの中で話し手が暗示している意図を問う設問が加わる。 	<ul style="list-style-type: none"> ・短文穴埋め問題（Part 5）の設問数が減る。 ・長文穴埋め問題（Part 6）の1つの文章に含まれる設問は3問から4問に増える。 ・文書の全体的な構成を理解しているか問う設問が加わる。 ・テキストメッセージやインスタントメッセージ（チャット）、オンラインチャット形式で複数名がやり取りを行う設問が加わる。 ・Part 7（1つの文書、複数の文書）の設問数が増える。 ・Part 7（複数の文書）で3つの関連する文書を読んで理解する設問が加わる。 ・文書中で書き手が暗示している意図を問う設問が加わる。

Q6：変更後のTOEICテストは現行のTOEICテストより難しくなってしまうのではないですか。

A：変更後のテストと現行のテストの難易度は変わりません。テストの難易度に変化が生じないよう、ETSのテスト開発専門家が調査、検証を重ね設計しています。

しかしこのような実施団体側の公式見解にもかかわらず、少なくともインターネット上には「問題が難化したようだ」という感想が散見される。またVELC Test[®]を運営している英語能力測定・評価研究会の事務局には、「VELCによるTOEIC予測スコアは、TOEICの形式変更に対応しているのか？」といった問い合わせが寄せられるようになった。TOEIC実施団体の発表をそのまま受け入れるならばVELCの予測式をアップデートする必要はないはずである。しかしTOEIC受験者側の素朴な感想や、VELC Test[®]ユーザー側の懸念は無視できない。以上のことから、新しくなった後のTOEICスコアを用いて、改めてVELCテストによる予測の精度を確かめることとした。

2. 研究目的

本研究の目的は、形式の新しくなった現在のTOEIC[®] L&Rのデータを目標変数と

して、(1) VELC Test[®] の現在の予測式による予測スコアの精度を確認し、(2) その結果に基づき、果たして新たな予測式を創出する必要があるのか、について検討することである。

3. 方法

3.1 協力者

関東圏の4つの私立大学に通う学生に、研究目的を明らかにした上で英語能力測定・評価研究会メンバーが協力を呼びかけた。その結果、A大学の67名、B大学の26名、C大学の16名、D大学の5名より、文書による研究協力の意思表示を得た。そのうち直近のTOEICスコアが得られなかったC大学の2名を除いた合計112名が最終的なデータ提供者となった。

3.2 データ収集

それぞれの大学において以下のいずれかの方法で同一協力者からVELC、TOEICという2種類のテストスコアを入手した。VELC Test[®] はすべて授業外の時間帯に実施され、試験監督には英語能力測定・評価研究会メンバーが当たった。

【方法1】新たにVELC Test[®] を実施し、かつ同日にTOEIC[®] L&R IPを実施する。

【方法2】新たにVELC Test[®] を実施し、その受験学生から直近のTOEICスコア（公開あるいはIP）の開示を受ける。

VELC Test[®] はA大学では1月と3月、B大学では1月、C大学では2019年12月、D大学は3月に実施された。TOEICスコアの「直近」の範囲は、VELC Test[®] の実施日からおおよそ1ヶ月から12ヶ月の間である。

4. 結果

4.1 VELCスコアとTOEICスコアの記述統計

VELCスコアとTOEICスコアの記述統計を表3に示す。(VELCスコアはLとRのセクションスコアだけでなくさらに細分化したパートスコアL1, L2, L3, R1, R2, R3も算出されるがここではスペースの関係で割愛する。)

VELCスコアはトータル、L、Rともに日本人大学生の平均が500になるように設定している。実際に2012~2019年度にVELC Test[®] を受けている受験者の平均はそれよりもやや低く、毎年475~485程度である(静, 2020)。今回の協力者グループの平均は558.4 (VELC_T), 577.2 (VELC_L), 547.5 (VELC_R)なので、通常の実験者平均よりもかなり英語力が高いことがわかる。標準偏差を見ると86.8 (VELC_T), 94.8 (VELC_L), 102.3 (VELC_R)であり、これは2012~2019年度の実験者の値(静, 2020)である73.3 (VELC_T), 78.2 (VELC_L), 84.1 (VELC_R)よりもいずれもかな

表3 VELC Test と TOEIC L&R の記述統計

	VELC_T	VELC_L	VELC_R	TOEIC_T	TOEIC_L	TOEIC_R
平均値	558.4	577.2	547.5	526.6	290.6	235.9
標準偏差	86.8	94.8	102.3	165.5	84.6	88.1
最大値	819	877	890	990	495	495
最小値	350	396	306	195	105	90
歪度	0.76	0.57	1.15	0.74	0.35	0.89
尖度	1.16	0.56	2.24	0.45	-0.20	0.73

注：n = 112

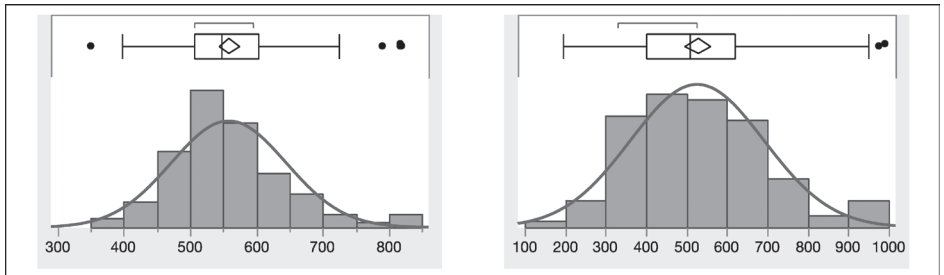


図1 VELC Test (左) と TOEIC L&R のスコア分布

り大きい。ここからは英語力が幅広いグループであることが示唆される。

一方 TOEIC の平均点が 526.6 であるが、これは 2019 年度の日本全国の IP テストの平均スコア 470 (IIBC, 2020) よりも、やはりかなり高い。TOEIC の最低が 195 で最高が 990 だということを見ても、多様な英語力をもった協力者グループであったと言える。

歪度と尖度は正規分布からの逸脱を示す指標であり、正規分布であればいずれも 0.0 であるが、母集団が正規分布であってもサンプルの歪度と尖度は 0.0 にはならない。おおむね ± 1.5 までは許容範囲とされる (芳賀・奥原・野澤, 2009)。唯一 VELC_R の尖度が 2.24 であるほかはいずれも許容範囲に収まっている。

視覚的に確認するために、VELC_T と TOEIC_T のヒストグラムに正規曲線を重ねてみたのが図 1 である。厳密な正規分布ではないにせよ極端な逸脱はないことが確認できたので、次の相関分析に進む。

4.2 現在の予測スコアの精度

表1に示した現在の予測式を使用して VELC スコアから算出した TOEIC 予測スコアと実際の TOEIC スコアの相関を確認してゆく。

まず TOEIC_L について確認する。図2にプロットと直線の当てはめを示す。X軸の「TOEIC L 相当点」とは TOEIC 予測スコアのことである。相関は $r = .769$ で決定係数は $R^2 = .591$ である。

次に TOEIC_R について、図3にプロットと直線の当てはめを示す。リスニングと同様、「TOEIC R 相当点」が TOEIC 予測スコアのことである。相関は $r = .844$ 、決定係数は $R^2 = .712$ である。

最後に、TOEIC_T について図4にプロットと直線の当てはめを示す。TOEIC_T の予測スコアは単純に TOEIC_L 予測スコアと TOEIC_R 予測スコアの和である。相関は $r = .852$ 、決定係数は $R^2 = .726$ である。

表4に、VELC Test[®] 本実施開始前のトライアル段階に、旧問題形式の TOEIC を予測したときの相関係数と、今回の新形式の TOEIC を予測した相関係数を改めてまとめた。

数値は $0.763 \rightarrow 0.769$ (TOEIC_L)、 $0.799 \rightarrow 0.844$ (TOEIC_R)、 $0.825 \rightarrow 0.852$ (TOEIC_T) と、いずれもわずかながら上昇している。この数値を見る限り、目標変数である TOEIC 側に問題形式の変更があったことにより VELC スコアで TOEIC スコアを予測する精度が下がったのではないかという懸念はまったく当たっていないと言えるだろう。

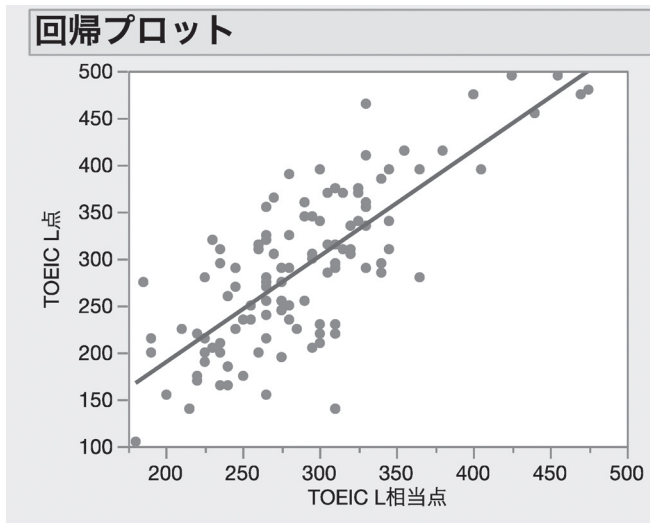


図2 VELC スコアによる TOEIC_L の予測スコアと実際の TOEIC_L スコアの関係

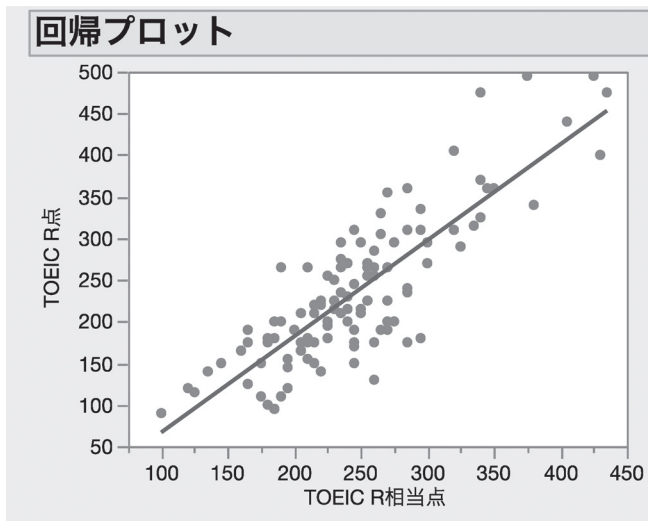


図3 VELC スコアによる TOEIC_R の予測スコアと
実際の TOEIC_R スコアの関係

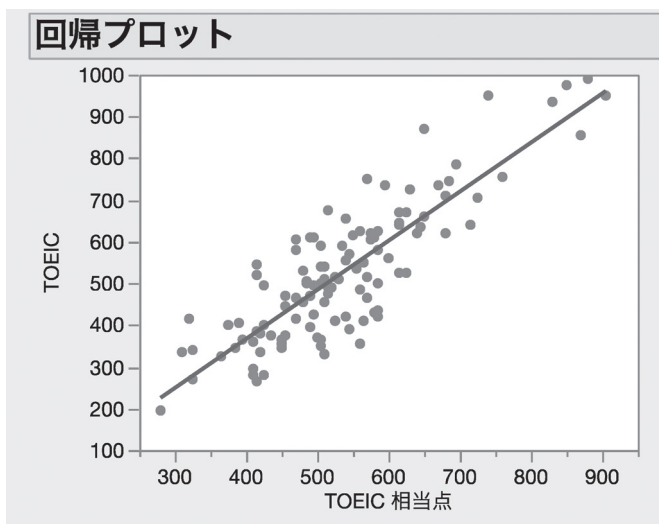


図4 VELC スコアによる TOEIC_T の予測スコアと
実際の TOEIC_T スコアの関係

表 4 2011 および 2020 データに基づく TOEIC 予測スコアと
実際のスコアの相関係数

	TOEIC 問題形式	<i>n</i>	TOEIC_L	TOEIC_R	TOEIC T
前回 2011	旧形式	375	0.763	0.799	0.825
今回 2020	新形式	112	0.769	0.844	0.852

4.3 予測のずれ幅の確認

次に個々の受験者についての予測のずれの度合いに関して分析を行った。「実際の TOEIC スコア - TOEIC 予測スコア」を「差異」と呼ぶこととし、TOEIC_L に関する差異を「L 差異」、TOEIC_R に関する差異を「R 差異」とする。まず記述統計と分布を確認したのが図 5 と図 6 である。

差異は 0 を平均値として正規分布するのが理想である。まずグラフの形状や歪度・尖度の値から概ね正規分布に近いとみなしてよいだろう。平均値に関しては、L 差異は想定どおりだが R 差異はわずかに負の値のようである。つまり TOEIC_R の予測スコアは実際よりもわずかに（10 点ほど）高かった。

次に X 軸にデータ番号、Y 軸に差異をとったプロットを図 7 に示す。当然、想定平均値である $y=0$ の直線の上下にランダムに分布しているはずである。その想定は概ね正しいが、図 6 でも確認したように R 差異の平均が僅かに $y=0$ よりも低いのが視認できる。

最後に、予測スコアと差異のプロットを吟味する。まず図 8 は、TOEIC_L 予測スコアと L 差異をプロットし、平均に加えて直線を当てはめたものである。当てはめた直線がわずかに右上がりなのが視認できる。この直線は

$$L \text{ 差異} = -35.28966 + 0.1271404 * \text{TOEIC_L 相当点}$$

であり、勾配の推定値 0.1271404 の p 値は $p = .1582$ であり、すなわち有意ではない。つまり TOEIC_L 予測スコアが高くなるほど差異の値が上昇する傾向がわずかにあるが、有意であるほど強くない、という意味である。

同様に図 9 として TOEIC_R 予測スコアと R 差異のプロットおよび直線の当てはめを示す。L 差異と同様、当てはめた直線は右肩上がりであるが、勾配が L 差異の場合よりもわずかに大きい。

この直線は

$$R \text{ 差異} = -48.52113 + 0.1532803 * \text{TOEIC_R 相当点}$$

であるが、勾配推定値の 0.1532803 は、 p 値が $p = .0303$ で有意である。つまり得点レベルが上がるに従ってマイナスの差異よりもプラスの差異が多くなる有意な傾向がある、という意味である。

式の上で差異が 0 になる、すなわち差異の期待値が 0 であるのは、L 差異に関して

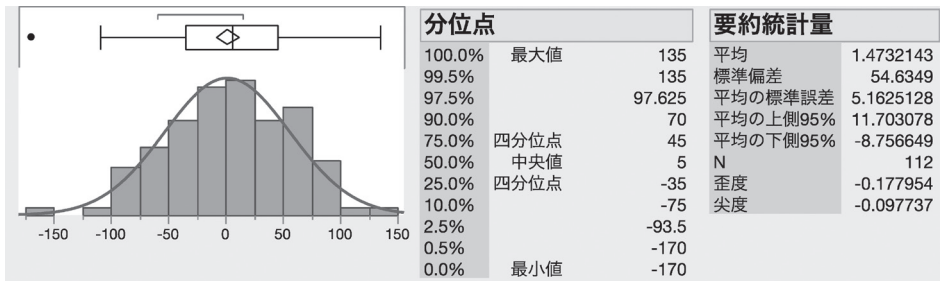


図5 L 差異の分布状況

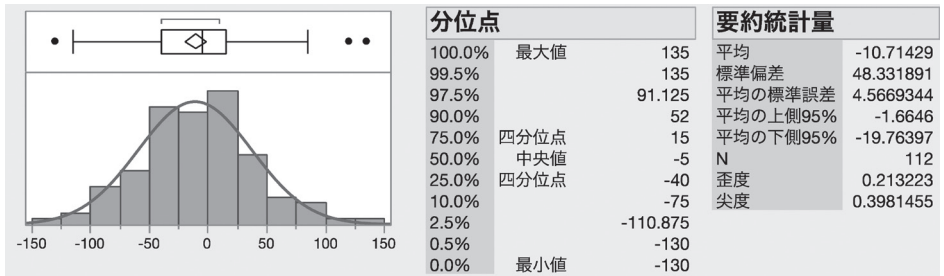


図6 R 差異の分布状況

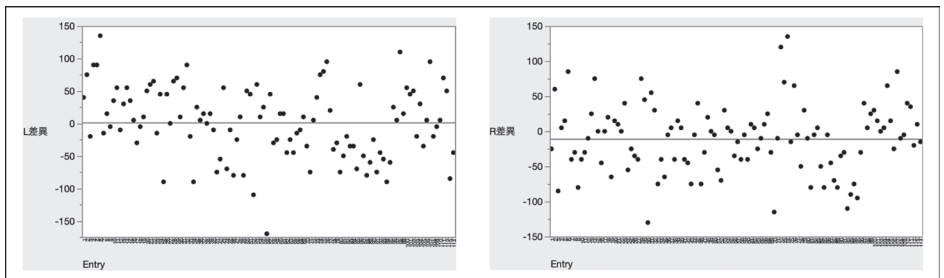


図7 データ入力番号とL 差異(左)、データ入力番号とR 差異(右)のプロット

は、TOEIC_L 予測スコア=277.6 の時、R 差異に関しては、TOEIC_R 予測スコア=316.6 の時である。TOEIC 予測スコアは TOEIC_L 予測スコアと TOEIC_R 予測スコアの合計なので、TOEIC 予測スコアが 594.1 の時、差異の期待値がゼロである。すなわちこれより予測スコアが低いと差異はプラスよりもマイナスが多い (= 予測スコアが高すぎる場合のほうが多い)、これより高いレベルだと差異はマイナスよりもプラスが多い (= 予測スコアが低すぎる場合のほうが多い) 傾向が、今回のデータからは観

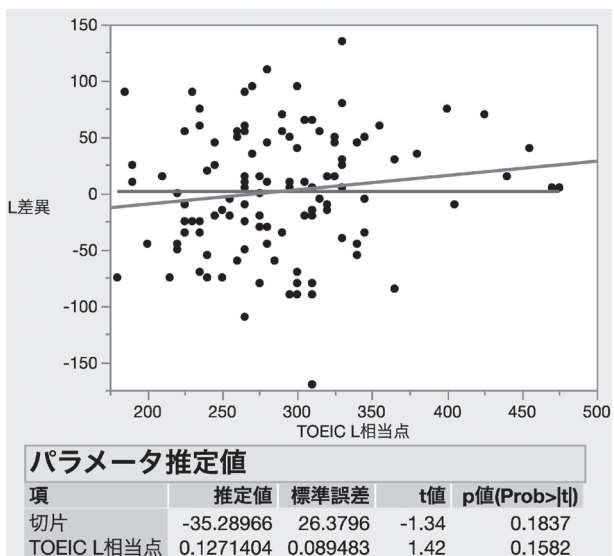


図 8 TOEIC_L 予測スコアと L 差異のプロットへの直線のあてはめ

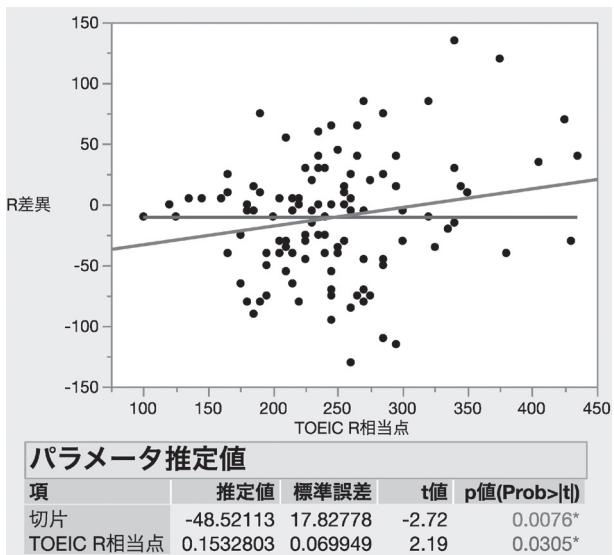


図 9 TOEIC_R 予測スコアと R 差異のプロットへの直線のあてはめ

察されたということである。

5. 考察

以上の結果は大きくふたつの点にまとめることができる。ひとつは、2016年度以前の旧形式のTOEICデータにもとづく重回帰分析によって求めたVELCスコアからTOEICスコアを予測する式は、2016年度以降の新形式によるTOEICスコアを、旧形式の場合と少なくとも同程度以上の精度(=決定係数)を持って予測していることが確認された。重相関係数 $r=0.852$ 、決定係数 $R^2=0.726$ は実用上十分に高い数値であり、さらに言えばひとつの言語テストAが、形式を異にするもうひとつの言語テストBを予測する時の現実的な上限に近い精度ではないかと考える。

なお、今回VELC Test[®]による予測精度が、目標変数が旧形式のTOEIC問題によるスコアでも新形式のTOEIC問題によるスコアでもほぼ同一だったということは、新旧のTOEIC間で難易度の互換性が保たれていることの間接的なエビデンスになっているとも言えよう。

その一方で差異分析からはVELCスコアによるTOEIC予測スコアの解釈に際しての注意点も明らかになった。すなわち予測スコアと実際のスコアのずれは、当該受験者の英語力レベルとわずかながら関係がありそうだという可能性が示唆された。具体的にはおおそ予測スコア590点あたりを境にして、それよりスコアが低い場合は「甘い」予測になっているケースのほうが「辛い」予測になっているケースよりも多く、逆にそれよりスコアが高い場合はやや「辛い」予測になっているケースのほうが多い、という傾向がわずかに見られるということである。

ただし図8、図9からも明らかのように、これは直線を当てはめてみればそうなるということに過ぎないのであって、比較的高い能力バンドに位置する受験者の予測スコアが「甘い」場合もあるし、比較的低い能力バンドにいる受験者の予測スコアが「辛い」場合もあることは言うまでもない。

以上のふたつを考え合わせるとVELC Test[®]として現時点で直ちに予測式をアップデートする必要は認められない。新形式のデータを用いた今回の分析でも決定係数がほとんど変わらず、十分に高いものであった以上、VELC Test[®]の提供するTOEIC相当スコアの精度はTOEIC側の問題形式が変更された現在でも、変更前と同じ程度であると考えるのが適当である。

なお本研究に参加してくれたのはすべて、英語能力測定・評価研究会メンバーの担当する授業をその時点でまたは過年度において履修しており、研究目的を理解した上で授業外の貴重な時間を提供してくれた、意欲の高い学生たちである。2011年度の開発時の協力者と同程度かまたはそれ以上にVELC Testを真剣に受験してくれたと推測され、そのことがTOEICテストとの一層高い相関につながったのかも知れない。

最後に、今回のデータに見られたTOEICの実際のスコアと予測スコアの間の差異の傾向に関しては、潜在的にはさらに追究する価値があると思われる。しかしながら

今回利用できたデータは $n=112$ と十分とは言えないため、このデータセットについてこれ以上の分析を進めても得られるものは多くないだろうと判断している。今後少なくとも予測式を算出した時の $n=375$ 以上のサンプルを集めた上で、改めて回帰分析を行って探索することを考えている。

6. まとめ

新形式の TOEIC によるデータで VELC Test[®] の予測精度を改めて検証してみた結果、旧形式 TOEIC の場合と同程度の予測が出来ていることを確認した。また予測スコアの解釈に際しての注意点も明らかになった。当面は現在の予測式の使用を続けながら、新形式による TOEIC データがより多く蓄積された段階で再度分析を行い、アップデートの必要性について改めて検討してみたい。

謝辞

本研究のために貴重な時間を割いて VELC Test[®] を受験してくださった協力者の学生諸君に感謝します。またデータを管理してくださっている英語能力測定・評価研究会の事務スタッフの方々と、データ収集に協力して下さりかつ本論文の草稿に有益なコメントを下された同研究会の望月正道先生と熊澤孝昭先生に、心より御礼申し上げます。分析結果の解釈に際しては、大東文化大学教職課程センターの児玉佳一先生に貴重なコメントを頂戴いたしました。ここに記して感謝の意を表します。

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：確認の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」

日本語テスト学会第21回研究大会(2017.9.10)会津大学.

静哲人(2020)「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・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・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.

芳賀敏郎・奥原正夫・野澤昌弘(2009).『エクセルによる多変量解析実務講座テキストI』(財)実務教育研究所.

IIBC(2016).『IIBC Newsletter 特別号』.

IIBC(2020). TOEIC Program: DATA & ANALYSIS 2020.

https://www.iibc-global.org/library/default/toEIC/official_data/pdf/DAA.pdf

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>

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.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago.)