

— Article —

G-TELP から TOEIC スコアを予測する回帰モデルの検証
—2 年間のデータから示唆されること—

**Regression Model Predicting the Scores of TOEIC from the G-TELP:
Suggestions and Implications Obtained from the Data for the Two Years**

小笠原 真司 (長崎大学)

Shinji OGASAWARA (Nagasaki University)

E-mail: ogasa-s@nagasaki-u.ac.jp

丸山 真純 (長崎大学)

Masazumi MARUYAMA (Nagasaki University)

E-mail: masazumi@nagasaki-u.ac.jp

宇都宮 譲 (長崎大学)

Yuzuru UTSUNOMIYA (Nagasaki University)

E-mail: yuzuru@nagasaki-u.ac.jp

Abstract

The purpose of this study is to investigate the possibility of predicting the scores of TOEIC tests from the scores of G-TELP Level 3 (General Tests of English Language Proficiency), by using data collected from April 2013 to July 2013. To obtain the predicting model, the data collected from all first-year students of Nagasaki University in 2013 ($N=1,388$) were submitted for simple regression analyses. Then, to predict the TOEIC scores more precisely, the scores of each section of G-TELP test (Grammar, Listening, Reading & Vocabulary sections) were analyzed using the multiple regression analyses. These models obtained from the 2013 data were compared with the ones obtained from the 2012 data (Ogasawara & Maruyama, 2014). Although R^2 is slightly lower than in 2012 ($R^2=.56$ vs. $R^2=.52$), the model obtained from the 2013 data also indicates that the overall fit of G-TELP Level 3 scores to TOEIC scores is sufficient but the fit among the higher scores is not adequate due to the fact that G-TELP Level 3 does not differentiate well enough TOEIC scores of 600 and higher. To ascertain the higher R^2 , a different kind of technique was used. Forty students from the nine faculties were randomly sampled and their scores of the TOEIC test and G-TELP test were analyzed by the multiple regression analyses as well as the simple regression analyses. The model obtained from this random sampling technique also indicates the overall fit of G-TELP Level 3 scores to TOEIC scores is sufficient, except for the TOEIC scores of 600 or higher and of 400 or lower. Moreover, R^2 is higher than one of the first model ($R^2=.59$ vs. $R^2=.52$).

Keywords: G-TELP, TOEIC, language testing, regression analysis, sampling

はじめに

長崎大学では、2011 年度より、全学部の 1・2 年生全員に対して G-TELP（国際英検）レベル 3¹ と TOEIC-IP テスト² を大学が受験費用を全額負担した上で受験させる制度を開始した。本学学生は、G-TELP レベル 3 を入学後 2 年間で 3 度受験する。一方、授業担当教員は、そのスコアを教養教育の英語科目（総合英語 I, II, III）の成績評価の一部（20%）に利用することとなっている。また、TOEIC-IP は、英語の成績評価とは直接関係はないが、1 年次と 3 年次に 2 度受験が義務付けられている。1 年次の TOEIC-IP の受験は、ほぼすべての学部で、入学時か 4 月中に実施されている。³

G-TELP の正式名称は、General Tests of English Language Proficiency であり、英語母語話者以外の英語学習者がどの程度英語をコミュニケーション手段として駆使する能力を有しているかを測定するテストである。G-TELP は、Grammar、Listening、Reading & Vocabulary⁴ の 3 つのセクションからなり、各セクションは 100 点満点で合計 300 点である。G-TELP は、Listening と Reading の二つのセクションからなる TOEIC とは違い、さらに Grammar のセクションが独立しているのが、特徴である。

本学では、1 年生のほぼ全員が、2011 年度より、毎年、TOEIC-IP と G-TELP レベル 3 を受験している。このような状況の中、教育的側面からも、G-TELP の得点から、社会的に知名度が高く、とくに就職活動において広く活用されている TOEIC の得点を推定できることは、重要であると思われる。

本研究では、2013 年の本学のデータを利用して、まず 2012 年のデータを分析した小笠原・丸山（2014）と同様に単回帰、重回帰によるモデルを作成し、2012 年データとの比較検討を行う。さらに、より信頼性の高いモデルの作成として、多段抽出単回帰、多段抽出重回帰モデルを作成し、本学によりあてはまりのよいモデルの考察を行うこととする。

研究背景・目的

小笠原・丸山（2014）は、2012 年前期実施（4 月～5 月）の TOEIC-IP と前期 7 月実施の G-TELP レベル 3 の両方の試験を受けた本学学生のスコアを用いて、単回帰および重回帰分析により、G-TELP のスコアから TOEIC のスコアを予測する換算式を作成した。換算式の作成においては、まず学部別の記述統計量を求め、相関分析を行い、さらに、相関分析等を用いていくつかのデータ・クリーニングを行った。その結果、単回帰分析により、以下のような換算式を得ることができた。

$$\text{TOEIC スコア} = -10.08 + 2.29 \times \text{G-TELP スコア} \quad (R^2 = .56) \quad \dots\dots\dots (1)$$

この回帰モデルは統計的に有意であり（ $F=1750.1, p<.0001$ ）、G-TELP のパラメータ推定値も有意であった（ $t=41.83, p<.0001$ ）。さらに、小笠原・丸山（2014）では、

以下のような 2 次曲線による回帰式を求めている。

$$\text{TOEIC スコア} = -27.59 + 2.29 \times \text{G-TELP スコア} + 0.01 \times (\text{G-TELP スコア} - 174.59)^2$$

$$(R^2 = .59) \dots\dots\dots (2)$$

この回帰モデルも有意であり ($F=992.55$, $p<.0001$)、パラメータ推定値も有意であった。2 次曲線モデルの方が、直線モデルよりもやや当てはまりがよく、決定係数が、0.03 高い結果となった。これは低得点層と高得点層での当てはまりが、直線モデルよりも、2 次曲線による回帰式のほうがややよいことに起因すると思われる。

受験者内での位置を測ることを目的としている（相対評価）集団基準準拠テスト（Norm Referenced Test）である TOEIC と異なり、もともと G-TELP は、あるレベルの学習内容をどれだけ習得したかを測る視点（絶対評価）のテストとしての目標基準準拠テスト（Criterion Referenced Test）といえる（小笠原，2011）。したがって、G-TELP は、難易度がレベル 1（高い）～レベル 5（低い）の 5 段階に分けられており、理想的には、学生の英語力や授業での学習内容に応じて、テストレベルを選択すべきである。しかしながら、本学では教養課程の英語の授業の統一評価基準という観点から、学部やクラスにかかわらず、すべての学生がレベル 3 を受験している⁵。

レベル 3 は、G-TELP の実施団体である ITSC (International Testing Services Center) が提示した対応表によると、TOEIC の 400 点～600 点を対象としているテストである。したがって、2012 年のデータを用いた小笠原・丸山（2014）で作成された回帰式において、低得点層と高得点層で当てはまりが悪くなったのも当然といえよう。表 1 は、ITSC が提示したレベル 3 における TOEIC との対応表である。

表 1 G-TELP（レベル 3 Form 310-319、300 点満点）と TOEIC の対応表

G-TELP 得点	100 点以下	150 点	200 点	250 点	300 点
TOEIC 得点	400 点未満	400 点前後	450 点前後	500 点前後	600 点以上

(国際英検、2014)

丸山（2012）、小笠原（2014）、小笠原・丸山（2014）のいずれの研究でも、G-TELP の 200 点前後で TOEIC スコアは ITSC の公式換算（対応）点とほぼ一致するが、それを下回れば下回るほど換算値が過小に、それを上回れば上回るほど過大になることが、報告されている。

小笠原・丸山（2014）では、単回帰分析に加えて、さらに重回帰分析を用いて、G-TELP の各セクションのスコアから TOEIC のスコアを推測する回帰式を求めている。その結果、最も当てはまりがよかったのは、TOEIC の総得点を G-TELP の 3 つのセクション (Grammar, Listening, Reading) から求めるものであり、 $R^2=.57$ であった。ちなみに、最も当てはまりの悪いのは、TOEIC のリスニングスコアを G-TELP の 3 つのセクションから求めるものであり、 $R^2=.43$ であった。この際、外れ値を除

外するモデルも作成されたが、全受験生を対象とするモデルと比較して、決定係数 (R^2) は、ほとんど変わらなかった。

本研究では、小笠原・丸山 (2014) で得られた知見の再検証を、2013 年度のデータで行う。小笠原・丸山 (2014) の結果と異なる結果が得られた場合は、その原因を考察することとする。また、決定係数の高い回帰式を得る方法はどうにすればよいのか、他のサンプリングも試みしてみる。

Yoshida (2012) が述べているように、信頼性の高い換算式を推定するには、2 つのテストの間の期間があいていないことが一つの条件となる。しかし、実際には二つの試験をほぼ同時期に実施することは、困難である。様々な試験を利用して、TOEIC との換算式を求める試みが行われているが、その多くが二つの試験の間に数か月から 1 年程度の期間が存在している (静, 2014; 丸山, 2012; 小笠原, 2014; 小笠原・丸山, 2014)。

2013 年度のサンプリングは、小笠原・丸山 (2014) で使用された 2012 年度のデータとほぼ同じ条件で行われている。すなわち、2012 年度のデータの場合、G-TELP の実施時期は、2012 年 7 月の 1 週目であり、TOEIC-IP の場合は、ほとんどの学部が、入学時の 4 月に実施した。それは、2013 年度もかわらず、実施時期等はまったく同じであった。なお、G-TELP のレベル 3 には、並行テストとして多くの Form が存在し、その等価性は保証されているが、2012 年度および 2013 年度は条件を同じにするために、同じ Form である Form 312 を使用した。

本研究は、2013 年度の G-TELP と TOEIC のスコアを用いて、双方の総点を用いた単回帰モデル、また、G-TELP の各セクションから TOEIC 総点を予測する重回帰モデルを求め、それらのモデルを 2012 年度のもの (小笠原・丸山, 2014) と比較検討する。次に、学部定員の違いによる特定学部の影響を取り除くために、多段抽出法によって、各学部から 40 名を抽出し、同様の分析を試みる。本研究のデータである 2013 年度のデータを用いて G-TELP レベル 3 から TOEIC のスコアを推測する換算式が、どの程度まで信頼性の高いものであるかを再検証し、同時にその限界や問題点を明らかにする。

全受験生のスコアから TOEIC スコアを予測する換算式

単回帰モデルの作成手順

表 2 は、学部ごと⁶の 2013 年 7 月実施の G-TELP レベル 3 Form 312 の合計点 (300 点満点) と 2013 年 4 月実施の TOEIC-IP の総合点 (990 点満点)、リスニングセクション、リーディングセクション別の記述統計量と相関係数を表している。両試験を受けた 1 年生の総数は、1,389 名である。TOEIC-IP の平均点は 390.6 ($SD=112.2$) であり、G-TELP は 180.0 ($SD=37.7$) であった。全学部の両テスト間の相関係数は、 $r=.72$ ($p<.0001$) であり、両テスト間には強い相関があった。G-TELP および TOEIC の最低点は、それぞれ、69 と 140 であり、最高点は 275 と 855 であった (表 3 参照)。

一方、小笠原・丸山 (2014) で使用された 2012 年度のデータも、両試験を受けた 1 年生の数は、1,365 名とほぼ同数である。4 月実施の TOEIC-IP の全体の平均点

は、389.4 とほぼ同じ結果である一方、7月実施の G-TELP レベル 3 Form 312 の平均点は 174.6 であった。2012 年度も 2013 年度も入学時の成績はほぼ同じであるが、7月実施の G-TELP の成績では、2013 年度の方の成績が向上していることがわかる。また、G-TELP と TOEIC-IP との相関では、小笠原・丸山（2014）の 2012 年のデータと同様に、TOEIC の合計点との相関が高く、TOEIC リスニングセクションとの相関が低くなっている。

表 2 学部ごとの G-TELP（総合点）と TOEIC スコア：記述統計量と相関係数

	G-TELP スコアの平均点 (標準偏差)	TOEIC スコア総合点の平均点 (標準偏差)	G-TELP と TOEIC 総合点との相関	TOEIC リスニングの平均点 (標準偏差)	G-TELP と TOEIC リスニングとの相関	TOEIC リーディングの平均点 (標準偏差)	G-TELP と TOEIC リーディングとの相関
教育学部 (n=229)	162.6 (36.0)	347.4 (91.1)	0.63	197.8 (52.6)	0.51	149.6 (48.3)	0.63
経済学部 (n=346)	183.3 (31.1)	383.4 (83.4)	0.57	212.9 (47.7)	0.40	170.5 (50.1)	0.56
医学部医学科 (n=94)	233.0 (23.0)	588.7 (118.7)	0.52	300.1 (66.8)	0.44	288.6 (65.0)	0.51
医学部保健学科 (n=66)	198.1 (23.7)	421.7 (78.4)	0.58	229.2 (44.2)	0.44	192.4 (50.2)	0.51
歯学部 (n=45)	205.4 (28.5)	464.7 (127.2)	0.72	250.0 (66.9)	0.62	214.7 (71.4)	0.70
薬学部 (n=73)	218.7 (25.7)	492.2 (107.3)	0.62	262.9 (57.3)	0.50	229.3 (60.6)	0.62
工学部 (n=295)	167.1 (31.4)	353.8 (89.4)	0.67	201.5 (52.0)	0.54	152.3 (50.0)	0.64
環境学部 (n=132)	168.1 (33.8)	373.1 (84.3)	0.63	209.6 (48.4)	0.57	163.5 (45.9)	0.55
水産学部 (n=109)	162.4 (35.5)	337.2 (80.1)	0.60	189.8 (46.7)	0.44	147.4 (46.1)	0.59
全 体 N=1,389	180.0 (37.7)	390.6 (112.2)	0.72	216.4 (59.0)	0.61	174.3 (63.6)	0.70

単回帰分析結果

次に、2012 年度と同様に、G-TELP のスコア（総合点）から TOEIC のスコア（総合点）を予測するために、TOEIC のスコアを従属変数、G-TELP のスコアを独立変数として単回帰分析を行った。これらの変数の記述統計量および相関係数は、表 3、単回帰分析の結果は表 4 の通りであった。

表 3 記述統計量及び相関係数

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>r</i>
G-TELP	1,389	180.0	37.7	69	275	0.72
TOEIC-IP	1,389	390.6	112.2	140	855	

表 4 TOEIC-IP を従属変数とした単回帰分析の結果

項	推定値	標準誤差	<i>t</i> 値	<i>p</i> 値
G-TELP	2.14	0.05	38.55	0.0001
切片	5.62	10.20	0.55	0.58

式 (3) は、G-TELP スコアを独立変数とし、TOEIC-IP スコアを従属変数とした単回帰式である。この回帰モデルは、統計的に有意であり ($F=1485.9, p<.0001$)、

G-TELP のパラメータ値も有意であった ($t=38.55, p<.0001$)。なお図 1 は、回帰分析から得られた散布図と回帰直線である。

$$\text{TOEIC スコア} = 5.62 + 2.14 \times \text{G-TELP スコア} \quad (R^2=.52) \quad \dots\dots\dots (3)$$

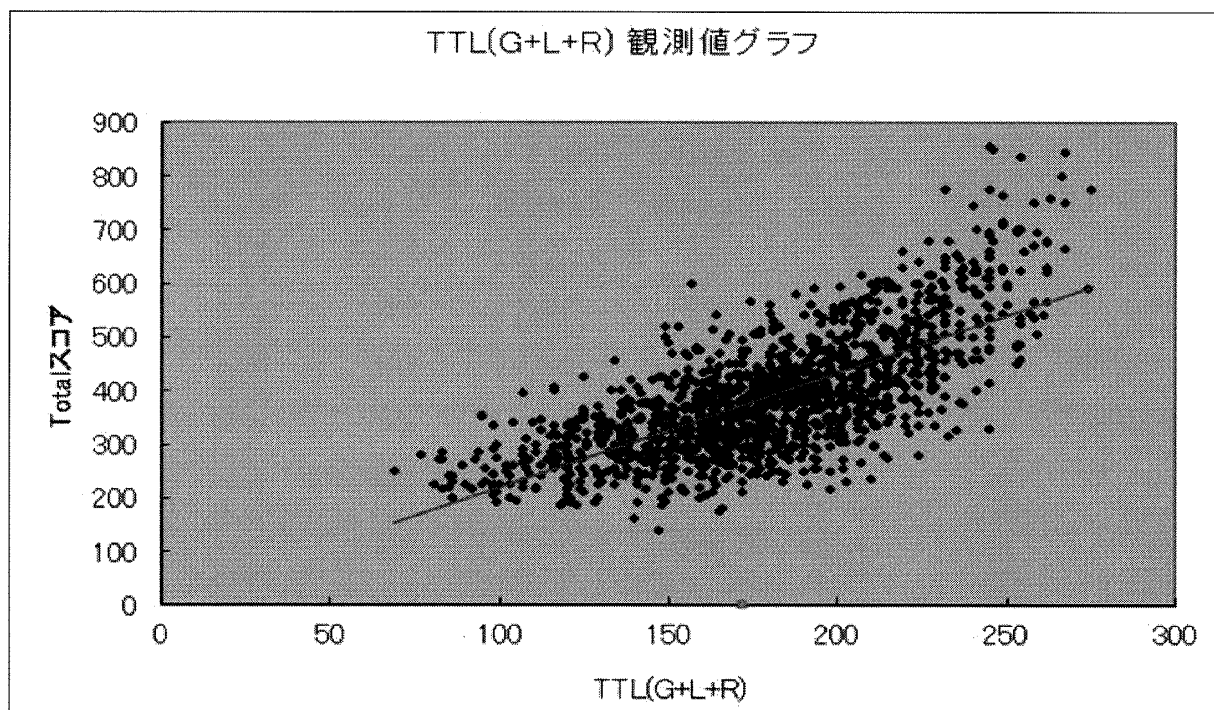


図 1 直線によるあてはめと散布図

式 (3) の決定係数は $R^2=.52$ であり、小笠原・丸山 (2014) の 2012 年度のデータ $R^2=.56$ (直線単回帰) より、0.04 悪い結果となった。

小笠原・丸山 (2014) では、マルハノビスの距離 (95%) を用いて、1,365 名中 68 名のスコアを外れ値とみなしている。外れ値は、おおよそ、G-TELP スコアの 240 点以上か 90 点以下で、TOEIC スコアでは 650 点以上または 400 点以下であることを報告している。さらに、それらを以下の 3 つのグループに分類している：

- ①両テストとも高いスコアであるが、TOEIC スコアが G-TELP スコアと比して高いグループ
- ②G-TELP スコアに比して TOEIC スコアが低いグループ
- ③G-TELP スコア 120 点以下で、TOEIC スコアが相対的に高いグループ

図 1 からみて、2013 年度のデータでも、2012 年の結果と同様に 3 つグループの存在が認められるが、特に①のグループが目立っているのが特徴であろう。

重回帰モデルの作成手順

次に G-TELP の各スコアから、TOEIC スコアを予測する重回帰分析を行った。

単回帰によるモデルと比較するために、外れ値も入れた全受験生を対象 ($N=1,389$) として、データ解析を行った。表 5 は、最小値と最大値を表記した記述統計である。また、表 6、表 7、表 8 は、それぞれ G-TELP の Grammar, Listening, Reading セクションと TOEIC 総合点の記述統計量と相関係数を学部別に表している。

表 5 最小値、最大値を含んだ記述統計

テスト (セクション)	<i>N</i>	<i>df</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max</i>
G-TELP (Grammar)	1,389	1,388	67.8	17.8	14	100
G-TELP (Listening)	1,389	1,388	51.6	12.6	17	92
G-TELP (Reading)	1,389	1,388	60.5	16.7	8	100
G-TELP (TOTAL)	1,389	1,388	180.0	37.7	69	275
TOEIC (Listening)	1,389	1,388	216.4	59.0	75	455
TOEIC (Reading)	1,389	1,388	174.3	63.6	55	455
TOEIC (TOTAL)	1,389	1,388	390.6	112.2	140	855

表 6 学部ごとの G-TELP (Grammar) と TOEIC スコア：記述統計量と相関係数

	TOEIC スコアの平均点 (標準偏差)	G-TELP Grammar Section の平均点 (標準偏差)	相関
教育学部 ($n=229$)	347.4 (91.1)	60.7 (17.9)	0.53
経済学部 ($n=346$)	383.4 (83.4)	69.5 (16.2)	0.42
医学部医学科 ($n=94$)	588.7 (118.7)	89.1 (9.7)	0.31
医学部保健学科 ($n=66$)	421.7 (78.4)	74.0 (14.5)	0.30
歯学部 ($n=45$)	464.7 (127.2)	77.6 (13.3)	0.34
薬学部 ($n=73$)	492.2 (107.3)	82.4 (11.9)	0.61
工学部 ($n=295$)	353.8 (89.4)	63.3 (16.3)	0.50
環境科学部 ($n=132$)	373.1 (84.3)	63.5 (16.2)	0.51
水産学部 ($n=109$)	337.2 (80.1)	59.0 (17.1)	0.48
全体 $N=1,389$	390.6 (112.2)	67.8 (17.8)	0.58

表 7 学部ごとの G-TELP (Listening) と TOEIC スコア：記述統計量と相関係数

	TOEIC スコアの平均点 (標準偏差)	G-TELP Listening Section の平均点 (標準偏差)	相関
教育学部 ($n=229$)	347.4 (91.1)	48.2 (12.2)	0.39
経済学部 ($n=346$)	383.4 (83.4)	51.5 (11.4)	0.38
医学部医学科 ($n=94$)	588.7 (118.7)	64.1 (12.1)	0.39
医学部保健学科 ($n=66$)	421.7 (78.4)	56.6 (11.6)	0.37
歯学部 ($n=45$)	464.7 (127.2)	55.8 (12.7)	0.61
薬学部 ($n=73$)	492.2 (107.3)	59.3 (11.5)	0.51
工学部 ($n=295$)	353.8 (89.4)	48.7 (11.6)	0.45
環境科学部 ($n=132$)	373.1 (84.3)	48.9 (11.8)	0.43
水産学部 ($n=109$)	337.2 (80.1)	49.4 (12.6)	0.45
全体 $N=1,389$	390.6 (112.2)	51.6 (12.6)	0.52

表 8 学部ごとの G-TELP (Reading) と TOEIC スコア : 記述統計量と相関係数

	TOEIC スコアの平均点 (標準偏差)	G-TELP Reading Section の平均点 (標準偏差)	相関
教育学部 ($n=229$)	347.4 (91.1)	53.8 (15.9)	0.52
経済学部 ($n=346$)	383.4 (83.4)	62.0 (14.4)	0.46
医学部医学科 ($n=94$)	588.7 (118.7)	79.9 (11.6)	0.38
医学部保健学科 ($n=66$)	421.7 (78.4)	67.6 (11.4)	0.44
歯学部 ($n=45$)	464.7 (127.2)	72.0 (13.4)	0.62
薬学部 ($n=73$)	492.2 (107.3)	76.9 (11.6)	0.45
工学部 ($n=295$)	353.8 (89.4)	55.2 (14.9)	0.51
環境科学部 ($n=132$)	373.1 (84.3)	55.7 (15.9)	0.49
水産学部 ($n=109$)	337.2 (80.1)	54.0 (16.0)	0.46
全体 $N=1,389$	390.6 (112.2)	60.5 (16.7)	0.61

重回帰分析結果

式 (4) は、重回帰分析によって求めた G-TELP の各セクション、Grammar、Listening、Reading のスコアを独立変数、TOEIC のスコアを従属変数とした換算式である。表 9 は、G-TELP 各セクションを独立変数とし、TOEIC スコアを従属変数とした重回帰式のパラメータの推定値を表している。この回帰モデルは、統計的に有意であり ($F=499.1, p<.0001$)、G-TELP のパラメータ推定値も、有意であった (表 9 参照)。

$$\text{TOEIC スコア} = 0.75 + 1.89 \times \text{G-TELP Grammar スコア} + 2.53 \times \text{G-TELP Listening スコア} + 2.16 \times \text{G-TELP Reading スコア} \quad (R^2=.52) \quad (4)$$

表 9 TOEIC-IP を従属変数とした重回帰分析の結果

項	推定値	標準誤差	t 値	p 値
切片	0.75	10.48	0.07	0.94
Grammar	1.89	0.14	13.22	<0.0001*
Listening	2.53	0.18	13.74	<0.0001*
Reading	2.16	0.15	13.65	<0.0001*

表 10 は、回帰統計表である。単回帰に加えて、重回帰も行なった結果、表 10 にもあるように、決定係数は、 $R^2=.52$ と同じであった。同様の結果は、小笠原・丸山 (2014) でも報告されており、G-TELP の得点から TOEIC スコアを推定する場合、単回帰のモデルでも、重回帰のモデルでも大きな差がないといえる。

表 10 回帰統計結果

回帰統計	
重相関 R	0.720745
重決定 R^2	0.519473
補正 R^2	0.518432
標準誤差	77.84498
観測数 (N)	1,389

2012 年度と 2013 年度のデータによる回帰モデルの比較検討

2013 年度のデータでは、単回帰も重回帰のモデルも $R^2=.52$ であり、同じ条件の 2012 年度の単回帰モデル $R^2=.56$ 、重回帰モデル $R^2=.57$ と比較して、決定係数自体が下がる結果となった。言い換えれば、相関が下がったのであるが、その原因はどこにあるのであろうか。

小笠原・丸山（2014）の G-TELP スコアから TOEIC 総得点を予測する 2012 年度の重回帰モデルにおいて、G-TELP リーディングの影響力が強いことが報告されている。表 9 のパラメータ推定値より⁷、2013 年度の TOEIC 総得点を予測するモデルには、G-TELP のリスニングの影響力が一番強いことがわかる。これは、G-TELP リーディングの影響力が TOEIC 総得点を予測するモデルに一番影響力が強いという 2012 年度の結果とは異なっている。

相関や決定係数が下がった主な原因として、2013 年度に導入にした学期中の e-learning の課題による学生の学習効果が考えられる。本研究では、先に述べたように、G-TELP（入学時 4 月実施）と TOEIC-IP（前期末 7 月実施）は、受験時期が異なっており、これは 2012 年度も 2013 年度も同様であった。本学では、学生の英語力を飛躍的に向上させたいとの願いから、2012 年度より、e-learning 教材の充実を図り、千葉大学作成の「3 ラウンド・システム」⁸ およびアルク社作成の「パワーワーズ」を教養教育の英語に導入した。2012 年度は、一部の専任教員の授業における部分的な使用にとどまっていたが、2013 年度より、非常勤講師担当のクラスも含め、教養課程の 1・2 年生全員に必修の課外学習として課されることとなった。特に、リスニング強化に大きな実績のある「3 ラウンド・システム」に関しては、1 学期の間に 15 時間以上の学習をすべての学生に課し、その学習を英語の授業評価に加えることで、学生のモチベーションを高め、2013 年度から徹底的にリスニング力を鍛えはじめた。

その成果に関しては、小笠原・廣江・奥田（2014）で詳しく報告がなされている。小笠原・廣江・奥田（2014）は 2013 年度の前期の G-TELP の成績を、2012 年度、2011 年度の同時期のものと比較している。図 2 は、G-TELP レベル 3、Form 312 の前期 7 月の 3 年間の総合点（300 点満点）の変化を学部別、および全体の成績を比較したものである。2013 年度は、2011 年度、2012 年度と比較して、総合点で 10 点程度全体の平均点が高いことがわかる。

G-TELP の成績に関して、小笠原・廣江・奥田（2014）は、2011 年度から毎年 4 月に行われた TOEIC-IP のテストをプリテストとし、7 月実施の G-TELP レベル 3 Form 312 をポストテストとして、年次間に差があるかどうか 2 元配置の分散分析を行っている。その結果、4 月入学時の TOEIC-IP のスコアには 2011、2012、2013 年度間で有意差がない一方、7 月の G-TELP の成績には有意差があることがわかった。さらに、2013 年度は、2011 年度、2012 年度と比較して、統計上リスニングセクションの平均点の伸びが大きいことが、ほとんどの学部で確認されたことを報告している。

Yoshida（2012）は、二つの試験を利用して一方の試験の換算式を作成する際に、

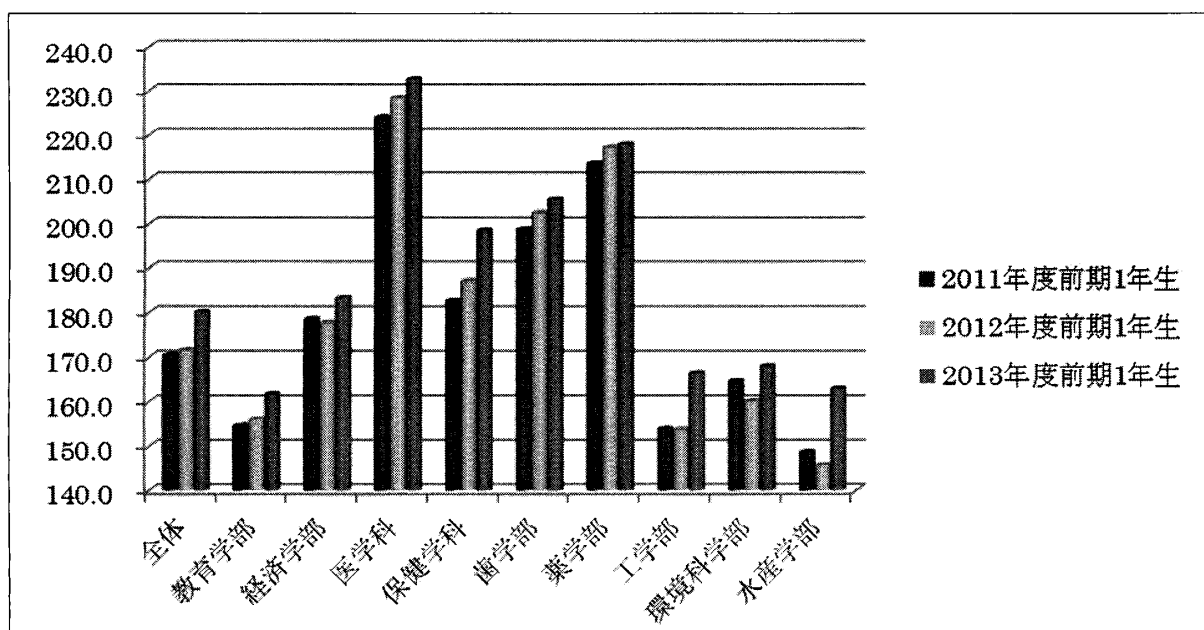


図2 2011年度、2012年度、2013年度、G-TELP レベル3 Form 312 の総合点

(小笠原・廣江・奥田, 2014)

二つの試験の間に生じる時間差の問題について指摘している。理想的には換算式作成のためには、ほぼ同時期に二つの試験を実施する必要がある。実施時期が大きく違えば、その間に学習効果などの影響が入り込む可能性が大であろう。今回の2013年度のデータを用いたモデルの決定係数が、2012年度のデータによるモデルよりも低く算出された理由のひとつは、すでに述べた2013年度のe-learningによる学習成果により、英語力、とくにリスニングスキルを入学時から飛躍的に伸ばした学生が多く存在するからであろうと思われる。

では、2013年度のデータから、もう少し決定係数の高い回帰分析モデルを、作成することはできないのであろうか。外れ値を省く方法が考えられるが、小笠原・丸山(2014)で報告されているように、マルハノビスの距離(95%)を用いて外れ値を除外しても、決定係数は大きく変化しなかった。二次曲線を用いた場合、多少あてはまりがよくなったが、大きく決定係数が変わることはなかった。

小笠原(2011)が指摘しているように、G-TELPは、あくまで目標基準準拠テストであり、レベル分けされている以上、カバーできる範囲が限定されていると言えよう。G-TELPを実施しているITSCによると、レベル3は、TOEICスコアの400点～600点を対象とすべきテストである。小笠原・丸山(2014)では、G-TELPレベル3では、260点を超えたり、逆に120点以下の場合は、学生の英語力を正確に測定できないと述べている。

以上の先行研究を検討するに、特定集団が推定式正確度に影響したことが疑われる。小笠原・丸山(2014)や今回TOEICスコア推定式を構成するために用いたサンプルは、経済学部や工学部といった学部から、相対的に多くサンプリングしている。今回、歯学部学生は45人であるが、経済学部346人であり、工学部は295人が所属する。明らかに、歯学部学生による結果は、相対的に少なく反映されるとい

えよう。言い換えれば、経済学部や工学部に所属する学生が獲得したスコアに、推定結果が左右されることを意味する。

そこで、次章では、各学部の影響が等質になるように、多段抽出法にて抽出したサンプルを用いて、換算式を推定する。多段抽出法は、各サブグループから同数、サンプルを抽出する方法である。極端に標本数が多いサブグループによる影響を排除することに向いている抽出法である。多段抽出法を用いると、相対的にサンプルサイズが少ないサブグループによる成果が強調されることになる。当該サブグループが極端により成果を出している場合、バイアスがかかることが懸念される。とはいえ、本研究が対象とする母集団は、さほど極端な差を内包するとは考えられない。したがって、多段抽出法を用いる合理性があらうと考えられる。

多段抽出法を用いた TOEIC スコアを予測する換算式

単回帰モデル作成手順

前章で試みたように、2013 年度のデータを用いた回帰分析では、2012 年度のデータよりも R^2 の値が 0.05 低くなった。小笠原・丸山（2014）では、2012 年度データを用いて、さまざまな条件での R^2 の値を求めている。単回帰による直線モデルが $R^2=.56$ であったが、2 次曲線によるモデルは $R^2=.59$ と当てはまりがややよくなった。さらに、外れ値を除外した単回帰によるモデルも求めたが、直線モデルが $R^2=.57$ 、2 次曲線によるモデルが $R^2=.58$ であった。G-TELP の 3 つのセクションのスコアから TOEIC のスコアを予測する重回帰分析では、 R^2 は全受験生でも外れ値を除いた形でも大きな差はなく、決定係数はともに $R^2=.57$ であった。どのような分析をしても、 R^2 は 0.56 から 0.59 の程度である。

小笠原・丸山（2014）では、回帰モデルにおいて当てはまりのよい学部と当てはまりの悪い学部があることが指摘されている。本節では、各学部（9 学部）から 40 名を無作為に抽出する多段抽出法による回帰モデルを試みることにする。40 名という値は、本学では最も学生定員が少ない歯学部（定員が 50 名（実際両試験を受けた学生数は、45 名））であることから決定した。全体のサンプルサイズは、360 となった。

表 11 は、2013 年度 4 月実施の TOEIC-IP と 7 月実施の G-TELP レベル 3 Form 312 の両方を受験した 1 年生 1,389 名から、各学部から無作為に 40 名を抽出した学生の TOEIC-IP および G-TELP の総合点の記述統計量である。

多段抽出法を採用すると、TOEIC のスコアが低い、しかも受験生の多い学部が 40 名となるので、本学の TOEIC の平均点は、全受験生の 390.6 から 417.0 に上がる結果となった。また、同様に G-TELP の平均点も、全受験生の 180.0 から 186.4 と 6.4 点上がった。多段抽出法による分析は、各学部の影響が均等となる分析法であることから、受験生の多いいくつかの学部が全体の平均点を下げていることが容易に推測できる。したがって、この平均点の変化は、当然と言えよう。なお、G-TELP と TOEIC 間の相関係数は、ピアソン (0.771)、スピアマン (0.779) であり、ともに統計的に有意であった ($p<.0001$)。

表 11 G-TELP および TOEIC-IP 総合点の記述統計量

	G-TELP				TOEIC-IP			
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
教育学部 (n=40)	154.4	33.4	89	209	338.8	88.1	180	505
経済学部 (n=40)	184.8	30.3	85	242	375.6	74.3	220	535
医学科 (n=40)	232.8	23.5	156	267	593.0	115.1	335	855
保健学科 (n=40)	189.3	29.7	107	237	428.1	97.0	235	640
歯学部 (n=40)	206.6	30.0	126	262	466.0	132.6	205	850
薬学部 (n=40)	216.6	26.3	165	274	493.0	103.5	260	775
工学部 (n=40)	161.6	31.3	104	244	359.2	78.5	215	560
環境学部 (n=40)	168.6	34.0	83	228	365.8	79.3	225	560
水産学部 (n=40)	162.3	33.9	99	228	339.4	82.3	210	640
全体 N=360	186.4	39.7	83	274	417.0	125.1	180	855

単回帰分析結果

次に、多段抽出法によるデータを用いて、G-TELP のスコア（総合点）から TOEIC のスコア（総合点）を予測するために、TOEIC-IP のスコアを従属変数、G-TELP のスコアを独立変数として単回帰分析を行った。結果は、表 12 の通りであり、式 (5) はその回帰式である。なお、この回帰モデルも、統計的に有意であり、($F=524.2, p<.0001$)、G-TELP のパラメータ値も有意であった ($t=22.90, p<.0001$)。

表 12 TOEIC を従属変数とした単回帰分析の結果

項	推定値	標準誤差	t 値	p 値
G-TELP	2.43	0.11	22.90	0.0001
切片	-35.13	20.22	-1.73	0.0832

$$\text{TOEIC スコア} = -35.13 + 2.43 \times \text{G-TELP スコア} \quad (R^2=.59) \quad \dots\dots\dots (5)$$

多段抽出法による単回帰モデルの決定係数は、 $R^2=.59$ となり、全学生 1,389 名を用いた単回帰モデルの決定係数 $R^2=.52$ より、より精度が高い回帰式を得ることができた。図 3 は、回帰分析から得られた、散布図と回帰直線である。

多段抽出法による図 3 の散布図を見ると、小笠原・丸山 (2014) で指摘された①のグループ、すなわち両テストとも高いスコアであるが、TOEIC スコアが G-TELP スコアと比して高いグループの存在が、相変わらず目立っている。①のグループで、直線からかなり外れている学生の多くが、医学部医学科の学生であり、小笠原・丸山 (2014) の指摘のように、G-TELP スコア 240 点以上は、順位づけのテストとしての機能が極度に下がることわかる。図 4 は、TOEIC-IP の学部別得点の分布図である。

図 4 より、薬学部、歯学部、医学部保健学科の学生の分布が、TOEIC スコア 400

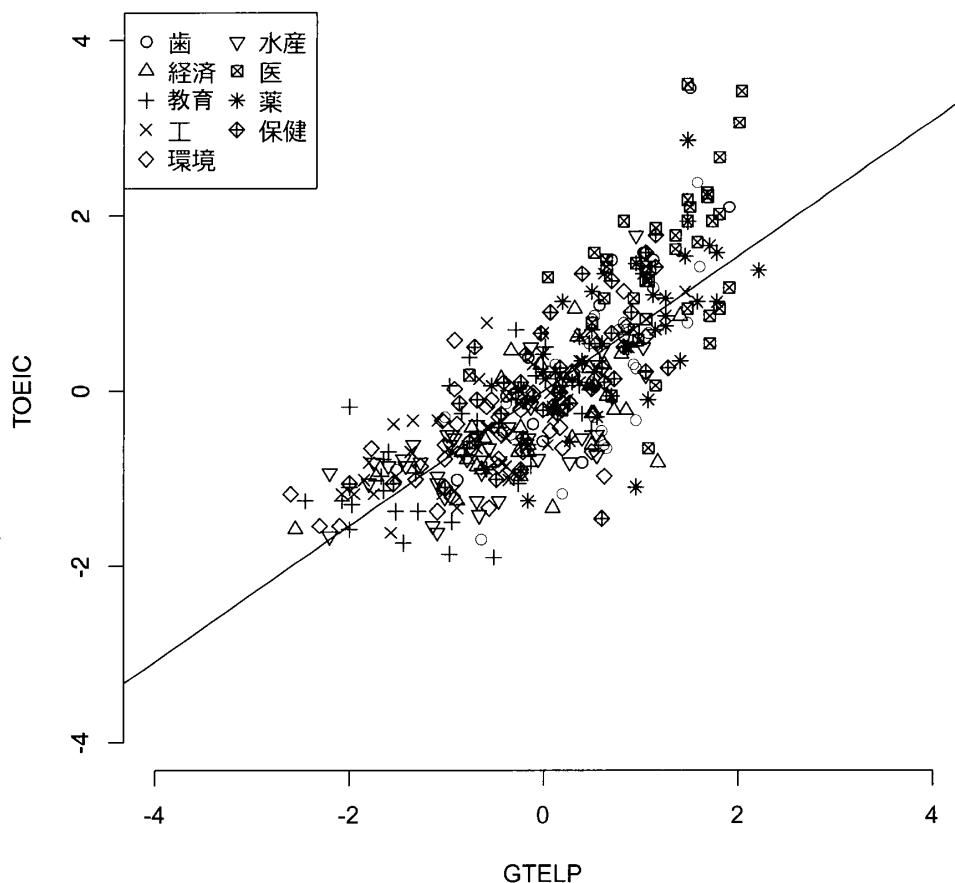


図3 直線によるあてはめと散布図

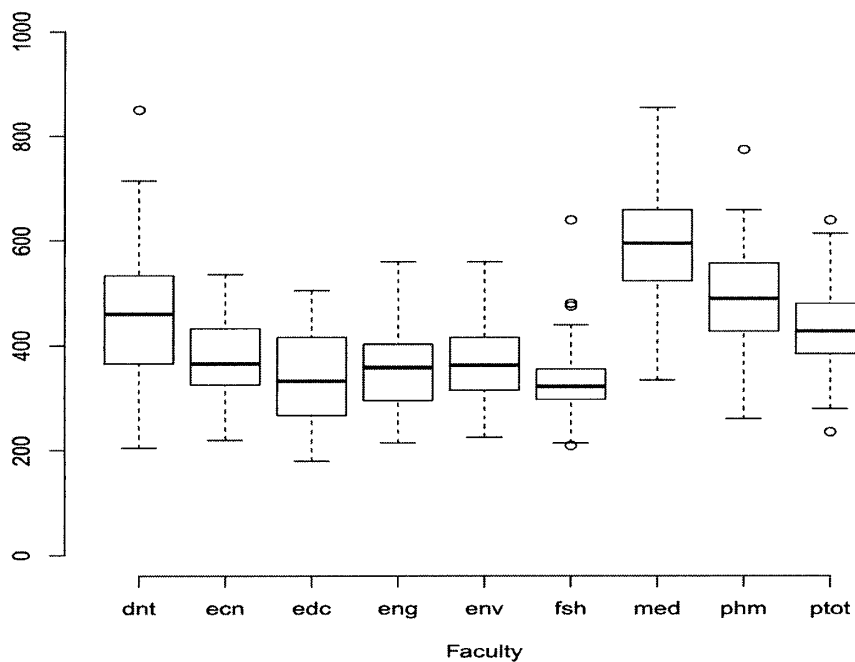


図4 多段抽出法による学部別分布表 (TOEIC 総合点)

(図中で使用されている文字は、それぞれ、dnt (歯学部)、ecn (経済学部)、edc (教育学部)、eng (工学部)、env (環境科学部)、fsh (水産学部)、med (医学部医学科)、phm (薬学部)、ptot (医学部保健学科) を示している。)

～600点程度にほどよく分布しており、G-TELP レベル3の理想範囲に比較的あてはまっていることがよくわかる。

換算式と公式換算点の比較考察

本節では、今回の結果得られた換算式⁹とITSCが発表しているG-TELP レベル3とTOEICとの対応表との比較検討を行うこととする。小笠原・丸山(2014)では、回帰分析の方法を変えて、外れ値も含めた直線モデルの換算式、外れ値を除外した直線モデルの換算式、外れ値も含めた2次曲線モデルの換算式の3つの式を、ITSCの対応表との比較を行っている。その結果も参照しながら、本研究で得られた2つの換算表を用いて、ITSCの対応表の検証を行う。

小笠原(2012)が指摘しているように、ITSCの発表した表は、G-TELPの得点が50点刻みであり、こまかな得点変換が不可能である。さらに、対応表自体も信頼性に疑問が残っている。今回研究で得られた換算式が、どの程度ITSCの対応表と一致するか、検証してみることとする。使用する式は、全受験生のデータによる単回帰直線モデルの換算式(3)と多段式抽出法による単回帰直線モデル換算式(5)である。

$$\text{TOEIC スコア} = 5.62 + 2.14 \times \text{G-TELP スコア} \quad (R^2=.52) \quad \dots\dots\dots (3)$$

$$\text{TOEIC スコア} = -35.13 + 2.43 \times \text{G-TELP スコア} \quad (R^2=.59) \quad \dots\dots\dots (5)$$

表13は、本研究から求められた換算式から換算点を求め、対応点数を比較し、そのずれを表したものである。表の最初の2行がITSC発表のG-TELPとTOEICの対応点数である。なお公式表は、G-TELPスコア100点以下という表記となっているので、100点以下に関しては、便宜上100点で換算式の点数を計算している。

表13 本研究による換算式による換算点と公式対応点のずれ

G-TELP スコア	100点以下	150点	200点	250点	300点
TOEIC スコア	400点未満	400点前後	450点前後	500点前後	600点以上
換算式(3)	219.62	326.62	433.62	540.62	647.62
換算式(5)	207.87	329.37	450.87	572.37	693.80
差					
換算式(3)	-180.38	-73.38	-16.38	+40.62	+47.62
換算式(5)	-192.13	-70.63	+0.87	+72.37	+93.87

本研究においても、丸山(2012)、小笠原(2014)、小笠原・丸山(2014)の研究と同様、G-TELPの200点前後でITSCの公式換算(対応)点とほぼ一致した。

表14は、小笠原・丸山(2014)で報告された同様な換算式と公式対応点とのずれを示した表である。

表 14 換算式による換算点と公式対応点のずれ (小笠原・丸山, 2014, p. 57 より)

G-TELP スコア TOEIC スコア	100 点以下 400 点未満	150 点 400 点前後	200 点 450 点前後	250 点 500 点前後	300 点 600 点以上
換算式 (直線) 外れ値除外した	218.92	333.42	447.92	562.42	676.92
換算式 (直線)	221.18	331.18	441.18	551.18	661.18
換算式 (2 次式)	257.05	321.96	436.87	601.78	816.69
差					
直線	-181.08	-66.58	+2.08	+62.42	+76.92
外れ値除外	-178.82	-68.82	+8.82	+51.18	+61.18
2 次式	-142.95	-78.04	+13.13	+101.78	+216.69

表 13 と表 14 から、今回の多段抽出法による単回帰直線モデルによる式 (5) が、200 点あたりでのあてはまりが一番よく、ITSC の対応と G-TELP スコア 200 点でほぼ一致している。また、すべての換算式にこの傾向がみられるとともに、G-TELP スコア 200 点を境に、それを下回れば下回るほど換算値が過小に、それを上回れば上回るほど過大になることが明らかである。

式 (5) を使用すると、TOEIC の 400 点は、G-TELP180 点である (G-TELP スコア 180 点で、TOEIC スコア 402 点と換算される)。本学のこの 2 年間のデータを見る限り、G-TELP スコア 150 点で TOEIC スコア 400 点前後という ITSC の対応表は、再検証する必要があるように思われる。また、同時に G-TELP レベル 3 を用いて、TOEIC スコア 500 点以上を換算することも不可能と言えよう。TOEIC スコア 500 点以上の学力にあてはまりのいいのは、G-TELP のレベル 2 以上であり、TOEIC スコア 500 点以上を G-TELP で換算するモデルを作成するには、レベル 2 の成績データが必要となってくるであろう。TOEIC スコアも 400 点以下は、あまり弁別力をもたず、得点そのものに意味がない (Min, 2012) ことから、G-TELP レベル 3 のスコアを用いて TOEIC スコアを換算するのは、G-TELP180 点から 240 点前後に限定するべきであろう。

ただし、本学の場合、この得点範囲の学生は多く、本研究で得られた換算式そのものは、彼 (女) らの TOEIC スコアを予測するのに、適切なものであり、有効なものと言える。

重回帰モデル作成手順

本節では、多段抽出法によるデータで、G-TELP 各セクションから TOEIC スコア (総合点) を予測する重回帰モデルを作成し、前節の単回帰モデルの結果と比較してみる。表 15 は、多段抽出法 (各学部から無作為に 40 名を抽出) による基本記述統計量である。また、表 16 は、G-TELP 各セクションと TOEIC-IP (総合点) の相関行列である。TOEIC (総合点) との相関は、これまでの研究では、セクション

別では、G-TELP の Reading が一番高かった（小笠原，2014；小笠原・丸山，2014）。

一方、今回は G-TELP Listening が、G-TELP Reading とほぼ同じ相関を示していることがわかる。2013 年度の G-TELP のデータは、すでに指摘したように e-learning によるリスニング学習の効果が大きく反映しているからであろう。特に、多段抽出法では、Listening の平均点が Reading より高くでており（表 15）、英語学力が上位である学生（歯学部、薬学部、医学医学科、医学部保健学科）が、大きくリスニングの力を伸ばした可能性が読み取れる。

表 15 最小値、最大値を含んだ記述統計

テスト（セクション）	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Median</i>	<i>Max.</i>
G-TELP (Grammar)	360	61.0	0.7	17	59	100
G-TELP (Listening)	360	65.9	17.7	8	68	100
G-TELP (Reading)	360	60.8	17.0	4	62	96
TOEIC-IP (TOTAL)	360	417.0	125.1	180	400	855

表 16 G-TELP、TOEIC-IP（総合点）の相関行列（ピアソン）

	TOEIC (TOTAL)	G-TELP Grammar	G-TELP Listening	G-TELP Reading
TOEIC (TOTAL)	1			
G-TELP Grammar	0.52	1		
G-TELP Listening	0.62	0.31	1	
G-TELP Reading	0.64	0.28	0.58	1

重回帰分析結果

式 (6) は、重回帰分析によって求めた G-TELP の各セクション、Grammar、Listening、Reading のスコアを独立変数、TOEIC のスコアを従属変数とした換算式である。表 17 は、G-TELP 各セクションを独立変数とし、TOEIC スコアを従属変数とした重回帰式のパラメータの推定値を表している。この回帰モデルは、統計的に有意であり ($F=332.3, p<.0001$)、G-TELP のパラメータ推定値も有意であった。

$$\text{TOEIC スコア} = -36.02 + 2.20 \times \text{G-TELP Grammar スコア} + 2.13 \times \text{G-TELP Listening スコア} + 2.85 \times \text{G-TELP Reading スコア} \quad (R^2=.60) \quad (6)$$

決定係数は $R^2=.60$ と算出され、これまで求めた式の中で一番高い数値となった。また、この TOEIC 総得点の換算式に関しては、これまでの結果と同様、表 17 のパラメータ推定値より、G-TELP Reading の影響力が大きかった。一方、相関係数では、相関が比較的高かった G-TELP Listening であるが、換算式では、影響力はそれほど強くなかった（表 17）。

表 17 TOEIC-IP を従属変数とした重回帰分析の結果

項	推定値	標準誤差	<i>t</i> 値	<i>p</i> 値
切片	-36.02	14.60	-2.47	0.014
Grammar	2.20	0.18	12.19	< 0.0001*
Listening	2.13	0.21	9.72	< 0.0001*
Reading	2.85	0.22	12.66	< 0.0001*

考 察

G-TELP のデータから、どの程度まで TOEIC スコアを予測することが可能かという課題に基づき、分析方法やサンプリングの方法を変えて、換算式の作成を試みてみた。静・望月 (2014) は、幅広い能力の受験生に適した短時間のテストがないことを指摘している。TOEIC は、幅広い受験生の能力を測るには適した試験ではあるが、スコア 400 点以下ではそのテストとしての信頼性が下がり、何よりも時間と費用がかかる。G-TELP はレベルに応じた受験が可能であり、費用や時間があまりかからない長所がある。その G-TELP レベル 3 を利用して、ある程度 TOEIC の得点を予測できることを明らかにすることができた。

時間と費用のかからない試験として、静・吉成 (2012) は、VELC Test¹⁰ を作成し、静 (2012) および静・望月 (2014) では、回帰分析から TOEIC を予測するモデルを提案している。本研究では、それに近い R^2 を G-TELP レベル 3 を用いて、得ることができた。すなわち、全受験生対象とした分析ではなく多段抽出法というサンプリング法を用いることで、より決定係数の高いモデル式を作成した ($R^2=.60$)。

ただし、本研究による全受験生を対象とした回帰モデルの決定係数の数値は、2012 年度のデータを用いた小笠原・丸山 (2014) よりも大幅に下がったのも事実である。本研究のデータも、静 (2012)、Yoshida (2012)、丸山 (2012)、小笠原 (2014)、小笠原・丸山 (2014) などの先行研究と同様に、TOEIC 受験時と換算表作成用のためのテストの実施時期に、大きな時間差が存在している。特に本研究のデータにおいては、明らかに 4 月実施の TOEIC から 3 か月の間に、e-learning 学習によるリスニング強化の成果がみられ (小笠原・廣江・奥田, 2014)、本研究の結果に影響を及ぼした。

理想的には、2 つの試験は、ほぼ同時期に実施される必要がある (Yoshida, 2012)。2014 年度からは、本学では TOEIC も G-TELP と同時期の 7 月に実施することとなった。したがって、2014 年度のデータからは、より信頼性の高い換算式が得られるものと思われる。今後は、2014 年度のデータの分析を行い、2012 年度以降 3 年間の比較をしてみることで、換算式とともに、学習効果の影響を考察することが可能になる。

結 語

本研究により、G-TELP レベル 3 でカバーできる範囲がかなり明らかになった。TOEIC の得点予測ができそうなのは、G-TELP 180 点から 240 点あたりが限界であ

り、TOEIC の 500 点以上を予測するのは困難である。言い換えれば、G-TELP レベル 3 は、TOEIC400 から 500 点の得点予測用に限定的に使用すべきであろう。もちろん、G-TELP レベル 3 が、本学の多くの学生にあてはまりのいい試験であることも事実である。

また、成績評価の平準化という点ではいたしかたないが、すべての学部で G-TELP レベル 3 を実施することの問題点もあきらかになった。G-TELP レベル 3 ではやさしすぎるという学生層（主に医学部医学科）が存在しており、またその一方、TOEIC の点数から判断して、G-TELP レベル 3 でも難しい学生層が一部の学部（工学部、水産学部、教育学部）にいる。今後、レベル 2 など異なる G-TELP 試験実施も必要となろう。同時に、評価の平準化を重視するのであれば、レベル 2 とレベル 3 を対応させる研究も、今後すすめる必要があるであろう。

註

- 1 G-TELP レベル 3 の難易度は、TOEIC スコア 400 点から 600 点までの学習者用に設定されており、TOEIC スコア 600 点以上の実力の学習者は、G-TELP レベル 2（TOEIC スコア 600 点以上）や G-TELP レベル 1（TOEIC スコア 800 点以上）を受験することが理想とされている（国際英検）。
- 2 本論では、TOEIC-IP と TOEIC は同じ意味合いで使用している。
- 3 TOEIC-IP は、医学部医学科のみ、毎年学科の事情で 4 月より遅い時期に実施しているが、7 月実施の G-TELP よりも前に受験が完了するので、他の学部と同様に、4 月実施のデータとして本論では扱うこととする。
- 4 G-TELP の Reading & Vocabulary セクションは、本論では Reading セクション、あるいは、Reading と表記することとする。
- 5 本学では、英語の授業クラスは、学部単位で構成されており、再履修の学生を除き、他学部の学生が混じることはない。なお、本研究において、G-TELP の試験は成績評価に利用するため再履修の学生も受験するが、データ分析においてはすべて除外している。
- 6 本論では、医学部医学科と医学部保健学科のデータを、独立したものとして扱うこととする。
- 7 独立変数の影響力の強さの判断には、通常標準偏回帰係数 β を用いるが、本研究は換算式作成を目的としており、強さの判断にはパラメータ推定値を利用した。
- 8 「3 ラウンド・システム」は、千葉大学から使用契約をかわし、本学に導入してからは、本学では「3 STEP」と呼んでいる。また、e-learning 教材は、CALL 教室や大学内のみならず、自学自習用として、すべての学生が自宅でも学習できる環境を用意している。このシステムは、千葉大学名誉教授竹蓋幸生氏の理論に基づき作成された聴解力と語彙力を軸において構築された英語教育の総合システムである。教材は、初級から上級まであり、authentic な音声を用いた繰り返し学習により、リスニング力を向上させるところに特徴がある。

- 9 全受験生のデータによる重回帰直線モデル換算式 (4) に関しては、単回帰のモデルと R^2 が同じであり、ITC の発表した対応表が G-TELP も TOEIC も総合点を対象としているため、重回帰による式は検討しないこととした。
- 10 VELC Test とは、英語能力測定・評価研究会により開発され、大学生に求められる英語力のスコアを直接測定するテストである。VELC とは、Visualizing English Language Competency (英語能力の可視化) の略である。VELC Test は現在金星堂から提供されている。リーディングとリスニングの 2 つのセクションから構成されている。リスニングセクションは 3 つのパートからなり、パート 1 (リスニング語彙力)、パート 2 (音声解析力)、パート 3 (内容把握力) となっている。

参考文献

- 国際英検. Retrieved July 28, 2014, from <http://www.g-telp.jp/about/a003.html>
- 丸山真純. (2012). 「長崎大学経済学部生の G-TELP/ (レベル 3) と TOEIC スコア—記述統計量と換算式からの検討—」. 『経営と経済』, 第 92 号 (3), 71-91、長崎大学経済学会.
- Min, S. (2012). *English Education and TOEIC Speaking and Writing Tests in Korea*. 2012 年度 TOEIC 研究会基調講演ハンドアウト.
- 小笠原真司. (2011). 「英語習熟度別クラスの効果と G-TELP による成績分析—工学部総合英語 III のデータを中心に—」. 『長崎大学大学教育機能開発センター紀要』, 第 2 号, 9-19.
- 小笠原真司. (2012). 「英語習熟度別クラスの効果的運用について—工学部総合英語 III の G-TELP データによる分析—」. 『長崎大学大学教育機能開発センター紀要』, 第 3 号, 9-20.
- 小笠原真司. (2014). 「G-TELP レベル 3 による TOEIC スコアの予測—回帰分析による予測式の作成と考察—」. 『第 2 言語習得研究と英語教育の実践研究—山岡俊比古先生追悼論文集』 (pp. 147-160). 東京: 開隆堂.
- 小笠原真司・丸山真純. (2014). 「G-TELP レベル 3 は、どの程度 TOEIC スコアを予測できるか」 *Annual Review of English Learning and Teaching*, 19, 45-63.
- 小笠原真司・廣江顕・奥田阿子. (2014). 「二種類の e-learning 教材の必修化による英語教育改革とその成果—G-TELP のデータおよびアンケート結果からの考察」. 第 27 回 JACET 九州沖縄支部大会 (鹿児島)、ハンドアウト.
- 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」. *The JACET International Convention Proceedings-The JACET 51st International Convention*, 272-77.
- 静哲人. (2012). 「VELC テストによる 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」. 第 16 回日本言語テスト学会 (JLTA) 全国研究大会発表要綱、ハンドアウト.
- 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」 英語力「可視化」の試み: 熟達度診断のための VELC

Test の開発」、*JACET Journal No. 58*, 121-141.

Yoshida, H. (2012). Can TOEIC Bridge Test scores predict TOEIC Test scores? An investigation of the relationship between TOEIC Bridge and TOEIC Tests. *JLTA Journal*, 15, 101-114.