

一般化線形モデル (GLM) による G-TELP スコアから
TOEIC スコアの推定モデルの構築：
長崎大学学生の 2011 年から 2016 年のデータから

丸山 真純^{*1}・小笠原 真司^{*2}・宇都宮 譲^{*3}

^{*1}長崎大学 経済学部・^{*2}長崎大学 言語教育研究センター

^{*3}長崎大学 経済学部

**Estimating the TOEIC Scores from the G-TELP Scores
by the Generalized Linear Model:
From the Data Obtained from Nagasaki University Students
from 2011 to 2016**

Masazumi MARUYAMA^{*1}, Shinji OGASAWARA^{*2},
Yuzuru UTSUNOMIYA^{*3}

^{*1}Faculty of Economics, Nagasaki University

^{*2}Center for Language Studies, Nagasaki University

^{*3}Faculty of Economics, Nagasaki University

Abstract

This article aims to estimate the TOEIC scores from the G-TELP (Level 3) scores with the data obtained from Nagasaki University students from 2011 to 2016. The problems with the previous estimation research lie in the inadequate fit and the use of the linear regression model, which assumes residuals being normally distributed. This study uses both the linear regression model and the Generalized Linear Model (GLM), which can handle categorical variables and is more flexible in the assumption on error structure. Both departments and entrance years are included as factors in the GLM to predict the TOEIC scores from the G-TELP scores. The results indicate that the estimation by the GLM is better overall to predict the TOEIC scores than the linear regression model, suggesting (a) departments and entrance years should be included in the model in estimating the TOEIC scores, (b) as

is the case with the previous research, the estimated scores in the lowest or the highest score ranges are not so precise, and (c) the GLM is more appropriate in estimating the scores than the traditional linear regression model. Further research should be necessary that will take into account individual differences and/or time lag between the two tests.

キーワード：一般化線形モデル (GLM), 線形回帰モデル (LM),
G-TELP, TOEIC

はじめに

近年、英語客観テストを実施する大学が増えている。その理由のひとつは、同一科目間における評価の平準化への要請であり、評価の一部に英語客観テストのスコアを利用する大学が増えている（廣森・山西、2009；前田、2009；磯田、2009；山森、2009）。また、TOEIC スコアを獲得することは、大学教育においても今日重要な位置づけのひとつとなってきた。TOEIC は習熟度別クラス編成のプレイスメント・テストとして利用されることもある。さらには、留学時に必要となる TOEFL や IELTS も、近年、実施される機会が増えている。

長崎大学では、必須教養英語一部科目の評価の平準化を目的として、学期末に実施される G-TELP（レベル 3）スコアを評価の 20%とし、さらに、次学期の英語科目の習熟度別クラス編成のために利用される。評価の平準化を目的として導入された同試験は、はたして、すべての学部学生にとって、その難易度などが適当なものであるかを、試験の信頼性ととも、検証することが重要な課題の一つであった。

一方、TOEIC の大学教育ならびに社会における外部英語能力テストとしての重要性の高まりを受け、G-TELP 導入と同時期に、同大学では、1 年次と 3 年次の学生全員に TOEIC を大学の費用負担にて提供することとなった。そのため、1 年次に 2 つの英語客観テストのデータを利用し、相互にテストとしての信頼性と妥当性を検討しつつ、片方の試験からもう片方の試験のスコアを推定することが可能になった。

複数の英語能力試験において取得されたスコアを互換する要請は、入学試験、英語能力試験スコアによる英語科目の単位認定、留学時等における英語能力認定、さらには大学院の入試など、必要とされる機会が増えている（安間、2011）。そのため、複数の英語試験スコアの互換を可能とする推定式は、実用上必要とされる。近年、文教予算が逼迫し、今後、さまざまな英語試験を大学の費用負担にて提供する機会が減ることが予想される。そのため、一方のスコアから他方のスコアの推定が可能となるこ

とは、望ましいことである。もちろん、両テストは目的等が異なるため、完全な互換や代替は不可能であるのは言うまでもない。

TOEIC はテストとしての信頼性が担保されており、また社会的に頻用されることが知られる。一方、TOEIC は費用が高価であること、120 分という受験時間により 90 分授業内での実施が困難であるなど、実施における問題も多い。費用や時間という観点からより優位な外部テストから TOEIC スコア を推定できることは、それゆえ、実用上有意義である。

ある試験結果を用いて TOEIC スコア を推定しようとする試みは、数多くなされてきた。こうした先行研究は、そのほとんどが線形回帰モデルを用いて、TOEIC スコア を推定してきた (小笠原、2014; 小笠原・丸山・宇都宮, 2015; Ogasawara, Maruyama, Utsunomiya & Collins, 2016)。しかしながら、線形回帰モデルをスコア推定に用いるには問題も多い。より精緻なモデルによる説明、推定が可能であれば、一層有意義である。

したがって、本研究の目的は、これまで線形回帰による推定モデルに依存しがちであった 2 つの相異なる英語能力試験の換算式を、より精緻に説明・推定が可能な一般化線形モデル (Generalized Linear Model; 以後、GLM) によりつつ、2011 年から 2016 年の 6 年間に長崎大学にて収集されたデータを用いて、G-TELP (国際英検) レベル 3 のスコアから TOEIC スコア を説明、推定するモデルを構築することである。

先行研究

ある英語能力試験のスコアを、他の英語能力試験から得られるスコアから推定する試みは、これまでもいくつかなされてきた。たとえば、土肥・張 (2014) は、千葉大学学生を対象とした TOEIC から TOEFL (ITP) のスコアを、単回帰モデルによって求めている。同研究は、どうしても生じてしまう 2 つの試験の受験時期の間隔を、上昇量を算出し、それを用いた TOEIC の補正データを使うことによって、より相関の高い換算式を作成している。そして、実際それまで使われていた 2 つのテストの換算表が、TOEFL (ITP) の基準が TOEIC の基準よりも厳しくなっていると報告している。また、Hemingway (1999) は、360 名の TOEFL (ITP) ならびに TOEIC のデータから、TOEIC の 400-450 点は TOEFL (ITP) の 410-430 点に対応するということのように幅をもたせた両テストのスコア対照表を提示している。この研究では、両テスト間の相関は $r=.75$ であった。一方、Tannenbaum & Wylie (2005) は、CEFR の B1 から C1 レベルという限定ながら、TOEIC と TOEFL (ITP) の換算式を線形回帰モデルによって、算出している。

それ以外にも、江口 (2011) は、北星学園大学の学生 114 名から収集した TOEIC

および TOEFL (ITP) データを用いて、相関係数を求めている ($r = .65$)。この研究では、線形回帰せずに、両テストを 0-100 点のスコアに変換している。

TOEFL との換算以外にも、その他の英語能力試験との換算を試みた研究もある。斉田・野口・小林 (2009) は、茨城大学学生の 200 名強のデータを用いて、TOEIC や TOEFL (ITP) のスコアを Objective Placement Test (Cambridge University Press) のスコアから算出する単回帰モデルによる換算式を求めている。同試験は、茨城大学共通テストとして使用されており、外部試験ではなく、Cambridge University Press より出版されている *Interchange* というレベル別教科書の使用レベルを判断するためのテストである。同テストからの回帰式の決定係数は、TOEIC で $R^2 = .41$ 、TOEFL で $R^2 = .38$ であった。

Yoshida (2012) は TOEIC のスコアを TOEIC Bridge から予測するモデルを線形回帰モデルによって行った研究である。この研究では、TOEIC Bridge を 4 月に、その 2 か月後に TOEIC を受験した 292 名のデータを用い、単回帰分析により換算式を導き出している。そして、とくに TOEIC スコアが低いと予測される受験生にとって、この換算式が有効であると報告している。その一方、2 つのテストの実施期間に 2 か月間タイムラグがあることに由来する影響が存在する可能性にもふれつつ、理想的には、換算式作成のためには、2 つのテストが同時期に行われるべきであると主張している。

新しく開発したテストを用いて、TOEIC のスコアを予測する試みもある。静・吉成 (2012) は、時間と費用のかからない日本人大学生向けの英語テストとして、VELC Test¹を開発した。さらに、静 (2012) では、VELC Test を用いて TOEIC スコアを予測するモデルを提案し、かなり高い相関があることを報告している。モデルの作成には、単回帰、重回帰分析を用いているが、重回帰分析を用いたモデルの決定係数は、 $R^2 = .58$ であった。ただし、この研究では、VELC Test 受験者の TOEIC テスト受験時期がまちまちであり、TOEIC 得点も自己申請された得点に依拠する。

G-TELP テストを用いて、TOEIC のスコアを予測する研究としては、まず丸山 (2012) と小笠原 (2014) がある。G-TELP テストは、正式には General Tests of English Language Proficiency といい、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るテストであり、Grammar, Listening, Reading & Vocabulary の 3 つのセクションから成っている。G-TELP は、5 つの難易度に分類されており、これら一連の研究で用いられているのは、レベル 3 (TOEIC 400 点から 600 点の受験者対象) である。なお、本論文で紹介する一連の研究は、断りのない限り、このレベル 3 を用いたものである。

丸山 (2012) では、2012 年に G-TELP と TOEIC の両方を受験した 163 名の長崎大学経済学部学生のスコアを用いて、重回帰分析より換算式を作成している。その決

定係数は、 $R^2 = .48$ であった。一方、小笠原 (2014) では、2012 年に G-TELP (7 月実施) と TOEIC (4 月実施) の両方を受験した長崎大学全学部 1 年生、1,352 名のスコアを用い、単回帰分析により換算式を作成している。その決定係数は、 $R^2 = .57$ であった。両研究とも、G-TELP レベル 3 から TOEIC スコアを予測する換算式の限界を述べるとともに、換算式は TOEIC スコア 400 点から 600 点の受験生に限定して使用すべきであると主張している。

これらの研究を受けて、小笠原・丸山 (2014) は、2012 年に G-TELP (7 月実施) と TOEIC (4 月実施) の両方を受験した勤務校の全学部 1 年生のデータに、小笠原 (2014) では除外されていた追試験受験生も加えて、1,365 名のスコア分析を行った。この研究では、単回帰モデルに加えて、G-TELP の各セクションから TOEIC スコアを求めるため重回帰分析も行っている。さらに、外れ値を特定し、データクリーニングを行うとともに、2 次曲線を用いた換算式を作成し、 $R^2 = .59$ という小笠原 (2014) よりも高い決定係数を得ている。

さらに、小笠原・丸山・宇都宮 (2015) は、同大学の 2013 年に G-TELP (7 月実施) と TOEIC (4 月実施) の両方を受験した全学部 1 年生のデータを 1,389 名のデータを用い同様の分析を試みている。2013 年度は、単回帰モデルの決定係数が、 $R^2 = .52$ と下がったが、これはこの年から実施された e-Learning 課外学習必修化の効果であると報告し、4 か月間の英語学習効果が大きかったためであるとしている (小笠原・廣江・奥田, 2016)。さらに、この研究では、モデルに当てはまりのよい学部とそうでない学部の存在が決定係数に影響を与えていることを指摘し、学部間の影響力 (TOEIC スコアが低く、学部人数が他学部に比べて多いいくつかの学部の存在) を排除したモデルとして、多段抽出法を用いた換算式を提案している。この多段抽出法を用いた換算式の決定係数は、単回帰モデルで $R^2 = .59$ 、重回帰モデルで $R^2 = .60$ と高い数値を得ることができた。そして、小笠原・丸山 (2014) も小笠原・丸山・宇都宮 (2015) も、G-TELP (レベル 3) を用いて TOEIC スコアを予測する際、有効な G-TELP のスコアレンジは、180 点から 240 点あたりであろうと報告している。

先行研究の問題点

先行研究をいくつか概観してきたが、これらの先行研究は、多くの場合、線形回帰分析を用いて、さまざまな英語能力試験から TOEIC スコアを推定しているものである。これらの研究は、2 つの試験の実施時期が数か月以上開いていることが多く、必ずしも、推定精度が高いとは言えなかった。また、統計的観点からは、一般に、単回帰分析には、誤差 (残差) が等分散な正規分布であると仮定される。こうした分布に関する仮定は、しばしば現実にそぐわない。スコアは連続量として扱ってもよいとは考えられるものの、原理的にはカウントデータと呼ばれる離散量である。離散量に対

して正規分布を適用することには、躊躇せざるを得ない。残差が示す分布が正規分布するとも限らない。むしろ、正規分布しないことのほうが自然であろう。そこで、われわれはカウントデータを扱うことができ、かつ正規分布以外の誤差構造を取り扱うことができる手法を用いて、TOEIC スコアを説明、推定する必要があると考えられる。

さらに、スコアに影響するであろう要因も加味しなければならない。筆者たちは、これまで各年度に収集された G-TELP と TOEIC スコアから、換算式を回帰モデルによって年度別に求めてきた（丸山、2012; 小笠原、2013, 2014; 小笠原・丸山、2014）。しかし、年度によるデータから、年度ごとに異なる換算式が算出されるため、G-TELP から TOEIC スコアを推定するという目的に照らすと、実用上、どの換算式を用いればよいのかという問題が生じる。また、学部や入学年次（これは、G-TELP のフォームとも関わってくる。詳しくは、後述）といった要因は、両試験にシステムティックに影響を与えていると考えられるため、これらを統計的に統制する必要も生じる。言い換えるならば、これらの要因は、説明・推定モデルに組み込まれることが望ましい。

研究方法

参加者およびデータ

本研究は、長崎大学 8 学部 9 学科の全 1 年次学生である（本研究では、便宜上、医学部は、医学科と保健学科の 2 学科に分けて分析する）。2011 年度より、1・2 年次学生が履修する教養英語科目群の一部において、G-TELP（国際英検）のレベル 3 を全学部学生に入学時から 2 年間で 3 度の受験を課してきた。

G-TELP は、TOEIC のような Norm Referenced Test（集団基準準拠テスト）とは異なり、ある一定期間の語学学習の成果を測る Criterion Referenced Test（目標基準準拠テスト）として位置付けることができ（小笠原、2013）、そのため、G-TELP では、レベル 1（高い）からレベル 5（低い）まで難易度が 5 つのレベルに設定されている。G-TELP（レベル 3）は、TOEIC スコア 400 から 600 点程度を対象とし、文法、リーディング、リスニングからなる試験である（各セクションは、100 点で、合計 300 点満点）。

本研究では、2011 年度入学者から 2016 年度までの全学部入学者の 6 年間にわたるデータを用いることとする。G-TELP（レベル 3）には、同じ難易度の並行テストとして複数の異なるフォームが用意されている。本研究の国立大学では、年度比較が可能なように、1 年前期、1 年後期、2 年前期または後期には、それぞれ、2011 年から 2013 年まで同じフォームを用いてきた。ただし、2014 年度からはフォームを変更して実施してきている²。

各学期の教養英語科目における成績の 20% は、G-TELP スコア で決まり、習熟度別クラスを実施する学部においては、次学期習熟度別クラス編成基準として、G-TELP スコア を用いる。

データ収集方法

データ収集手順は、以下に示す通りである。G-TELP は 1 月に、「総合英語 II」の授業時間内に、実施に関するオリエンテーションを事前に受けた授業担当日本人英語教員が実施した。この授業は 1 年次必修科目である。また、再履修の学生のスコアは、分析から除いた。

TOEIC スコア は 1 年次のものである。G-TELP 同様、2011 年度入学者から 2016 年度入学者までのスコアをデータとして用いているが、実施月は 2013 年度入学者までは 4 月（つまり、入学直後）が基本であり、2014 年度からは 7 月（前期末）となっている。実施に関しては、各学部 に委ねられているが、年度による違いはあるものの、相当数の 1 年次生が受験してきた。受験者数およびスコアについては、補遺を参照されたい。なお、本研究では、TOEIC ないし G-TELP のいずれか片方を受験しなかった学生および双方を受験しなかった学生については、考察対象としなかった。

過去 6 年間に実施された両試験は、異なる傾向を示す。補遺は、学部・年次別に TOEIC および G-TELP 総合点について、平均点および標準誤差がどう推移するかを示している。2013 年と 2014 年との間に、G-TELP スコア が学部を問わず下降している。2014 年度から、異なるフォームが用いられ、より水準が高い問題となったことによる影響であろう³。

より子細に検討すると、補遺によれば、TOEIC は全学部において平均点が経時的に変動し、結果として学部別平均点順位が入れ替わる傾向を示すことがわかる。たとえば、医学部医学科は、他学部が微増か横ばいを保つなか、平均点が下落する傾向にある。医学部医学科学生平均点は、2012 年に 558.9 点を記録以降、年々下降している。2015 年には首位を薬学部 に譲った。医学部における TOEIC スコア 下降と薬学部におけるスコア 上昇が相互作用した結果と考えられる。医学部医学科および薬学部を除く 7 学部において、年毎に順位が入れ替わる事態が確認される。たとえば、経済学部は、2011 年においては医学部保健学科に次いで、平均点が高かった。しかし 2015 年には歯学部 に追い越されてしまった。単年度におけるスコアをもって学部学生の英語能力を論じることは、危険であることを示していると言えるかもしれない。

一方、G-TELP は比較的経時的変動は少ない。順位は下位学部 3 学部（工学部、教育学部、水産学部）においてのみ頻繁に入れ替わり、上位学部においてはさほど入れ替わらない。本学の入試制度や学部の入学難易度がさほど変わらない状況を考慮にいと、得点が安定していることから、英語能力試験としては、G-TELP がより適切

であるかもしれない。あるいは、G-TELP が成績評価に組み込まれるのに対し、TOEIC スコアによる明示的インセンティブがないことによって生じた結果かもしれない。

分析方法

本研究は、線形回帰モデル (LM; Linear Regression Model) および一般化線形モデル (GLM; Generalized Linear Model) を用いて、G-TELP スコアおよび諸要因から、TOEIC スコアを推定するものである。GLM は、統計モデリング手法のひとつである。いわゆる回帰分析や分散分析を内包するとされる (粕谷、2012; 馬場、2015)。GLM には、いくつかの特徴が存在する。

第 1 に、LM 同様、線形予測子を用いる。つまり、目的変数を推定する説明変数が線形結合することを意味する。線形モデルを名乗るゆえんである。ただし、推定される値が線形性を有するとは限らない。説明変数には連続量だけではなく、順位尺度で構成されたデータや名義尺度で構成されたデータ (カテゴリカルデータ) も活用できる。交互作用やオフセット (計数が必ず 1 である説明変数)、高次項 (説明変数の n 乗) を含むことも可能である。

第 2 に、リンク関数を用いる。リンク関数を用いることで、データが有する性質を反映させつつ、目的変数を推定できる。カウントデータや明らかに上限が存在するデータ、割合を表現するデータなど、LM が取扱困難なデータも取扱可能である。ただし、リンク関数を使うことで、両者の関係を常に精確に推定できるとは限らない。

第 3 に、誤差構造を選択可能である。残差 (実測値とモデルによる推計値との乖離。誤差) を表現する分布を適宜選択できる。実データにおいて、誤差が LM において仮定されるような等分散な正規分布を示すことは、まれである。むしろ、正規分布しない場合が自然であろう。GLM は、正規分布やポワソン分布、二項分布など、さまざまな分布を選択可能である。本研究は、得点がカウントデータであること、学部別に英語能力が大きく異なるという先験的知見から、負の二項分布を採用した。負の二項分布はカウントデータに用いる事ができるうえに、ポワソン分布よりも過分散に対して頑健である。

近年、階層ベイズモデルや状態空間モデルなど、より複雑で現実味がある統計モデリングを可能とする手法も提案され、用いられるようになりつつある。しかしながら、われわれが GLM を用いる理由は、その利便性からである。GLM は、コンピュータを用いれば計算はほぼ一瞬で終わる。階層ベイズモデルや状態空間モデルはしばしば計算に半日を超える膨大な計算時間を必要とする。モデルとしては現実的かもしれないが、クラス分けなど速報性を要求される場面においては、GLM が有利である。

本研究は、上記特徴をふまえて、以下に示すモデルを構築・分析した。

$$\log y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

y : TOEIC スコア

x_1 : G-TELP スコア

x_2 : 学部 (9 水準 ; 歯学部、経済学部、教育学部、工学部、環境科学部、水産学部、医学部保健学科、医学部医学科、薬学部)

x_3 : 年次 (6 水準 ; 2011 年度、2012 年度、2013 年度、2014 年度、2015 年度、2016 年度)

α : 切片

β_1 : G-TELP スコア の係数

β_2 : 学部 の係数

β_3 : 年次 の係数

ε : 残差

先行研究が提示した LM と比較するため、先行研究とおなじモデルを用いて分析した。

$$y = \alpha + \beta_1 x_1 + \varepsilon$$

LM による結果と GLM による結果とは、AIC (赤池情報量基準) を用いて統計モデリングとしていずれが望ましいか比較・検討する。本研究は、分析にあたって、統合統計解析環境 R (Ver. 3.4.0) を用いた。

結果

線形回帰モデル (LM) を用いた分析

表 1 は、線形回帰分析結果である。決定係数から G-TELP スコアにて、TOEIC スコアを 4 割強、推計可能であることが読み取れる。決定係数は 5% 水準において統計的に有意であった。本研究による推定結果は、先行研究が得た結果を支持するものと考えられる。

推定結果は、成績上位者および成績下位者において当てはまりが悪い。特に成績下位者における残差をそのまま適用すると、あり得ないことに、スコアが負になることがある。これら結果は、残差を表現する分布に正規分布を採用することに由来する齟齬と考えられる。予想通り、LM は TOEIC スコア推定に用いるには限界があると考えられよう。

表1 回帰分析結果要旨

残差	最小値	第一四分位	中央値	第三四分位	最大値
	-425.06	-58.71	-4.41	55.9	429.96
	推定値	標準誤差	t 統計量	p 値	
切片	78.72	4.55	17.29	<2.2e-16	
G-TELP	1.87	0.00	73.43	<2.2e-16	
残差標準誤差	87.46				
決定係数:	0.42				
調整済決定係数	0.42				
F 統計量	5,392				
p 値	2.2e-16				
AIC	89,381.1				

一般化線形モデル（GLM）による推定

表2は、一般化線形モデルを用いた TOEIC スコア推定結果要約である。図1は表2による結果を用いて、G-TELP スコアから TOEIC スコアを推定、学部別に得点散布図に重ね描きした。また、図2は表2による結果を用いて、G-TELP スコアから TOEIC スコアを推定、各年における得点散布図に重ね描きした。

われわれは、以下に示す事実を確認した。第1に、G-TELP スコアが伸びると、TOEIC スコアも伸びる。推定された傾きは1より大きな値である。G-TELP にて測定可能な英語能力が伸長するにつれて、TOEIC スコアが上昇すると考えられる。

第2に、学部と年次は TOEIC スコアに影響する。表2によれば、いずれの年次も5%水準において統計的に有意に作用すると確認される。推定値から、学部と年次はほぼ同程度、TOEIC スコア推定値に影響すると考えられる。2014年度および2015年度には、推定値が各々0.14および0.17となった。各学部による影響を示す推定値よりも数倍高い。同年以降、年次が学部よりも強く TOEIC スコアに影響するようになったことが確認される。ただし、2014年度と2015年度との間にある推定値差は、2011-2013年度における各推定値差とさほどかわらない。2014年度以降は7月に TOEIC 試験を実施するようになったことが、推定値に影響を及ぼしたと考えられる。

第3として、下位者における残差が、上位者における残差よりも大きいことが言える。最低点を鑑みるに、試験終了時間まで真剣に解答を続けたとは考えにくい受験者がいることが想定される。多肢選択式の990点満点の試験にて200点とは、いかにも不思議な得点である。試験監督を担当する折、しばしば居眠りする受験者が観察されるという事実は、こうした見解を支持するだろう。

AIC は、線形回帰モデル（LM）よりも小さい。一般化線形モデル（GLM）にお

ける AIC は 85,582 である。LM における AIC は、89,381.13 である。GLM を用いて、より精確な TOEIC スコア推計モデルを構築できたことがわかる。

表 2 一般線形モデルを用いた TOEIC スコア推定結果

Family	Negative binomial				
Link	log				
	Estimate	SE	z value	Pr(> z)	
(Intercept)	5.10	0.02	250.88	0.00	
gtelp.second	0.00	0.00	63.79	0.00	
facultyEconomics	0.04	0.01	3.02	0.00	
facultyEducation	-0.03	0.01	-1.91	0.06	
facultyEngineering	-0.02	0.01	-1.61	0.11	
facultyEnvironmental_Studies	0.01	0.01	0.93	0.35	
facultyFisheries	0.01	0.01	0.76	0.45	
facultyHealth_Sciences	0.04	0.01	2.47	0.01	
facultyMedical_Sciences	0.13	0.01	8.48	0.00	
facultyPhamaceutical_Sciences	0.09	0.02	6.00	0.00	
yearFY2012	-0.03	0.01	-3.88	0.00	
yearFY2013	-0.02	0.01	-3.11	0.00	
yearFY2014	0.14	0.01	16.52	0.00	
yearFY2015	0.17	0.01	19.93	0.00	
yearFY2016	0.15	0.01	17.45	0.00	
Null deviance	19,979 (Df: 7,438)				
Residual deviance	7,525.7 (Df: 7,428)				
AIC	85,582				

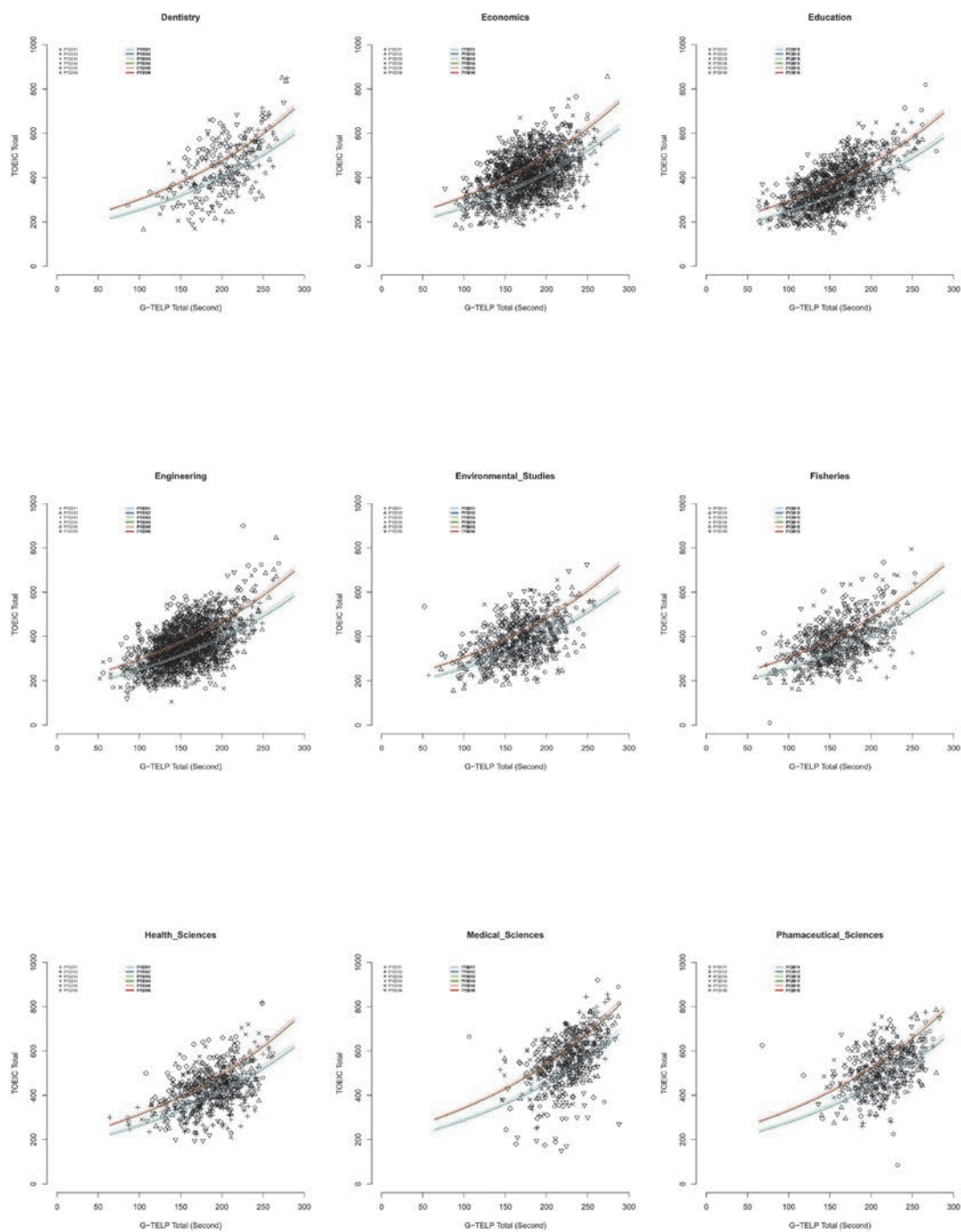


図1 学部別得点散布図得点推計曲線

一般化線形モデル (GLM) によるG-TELPスコアからTOEICスコアの推定モデルの構築:長崎大学学生の2011年から2016年のデータから

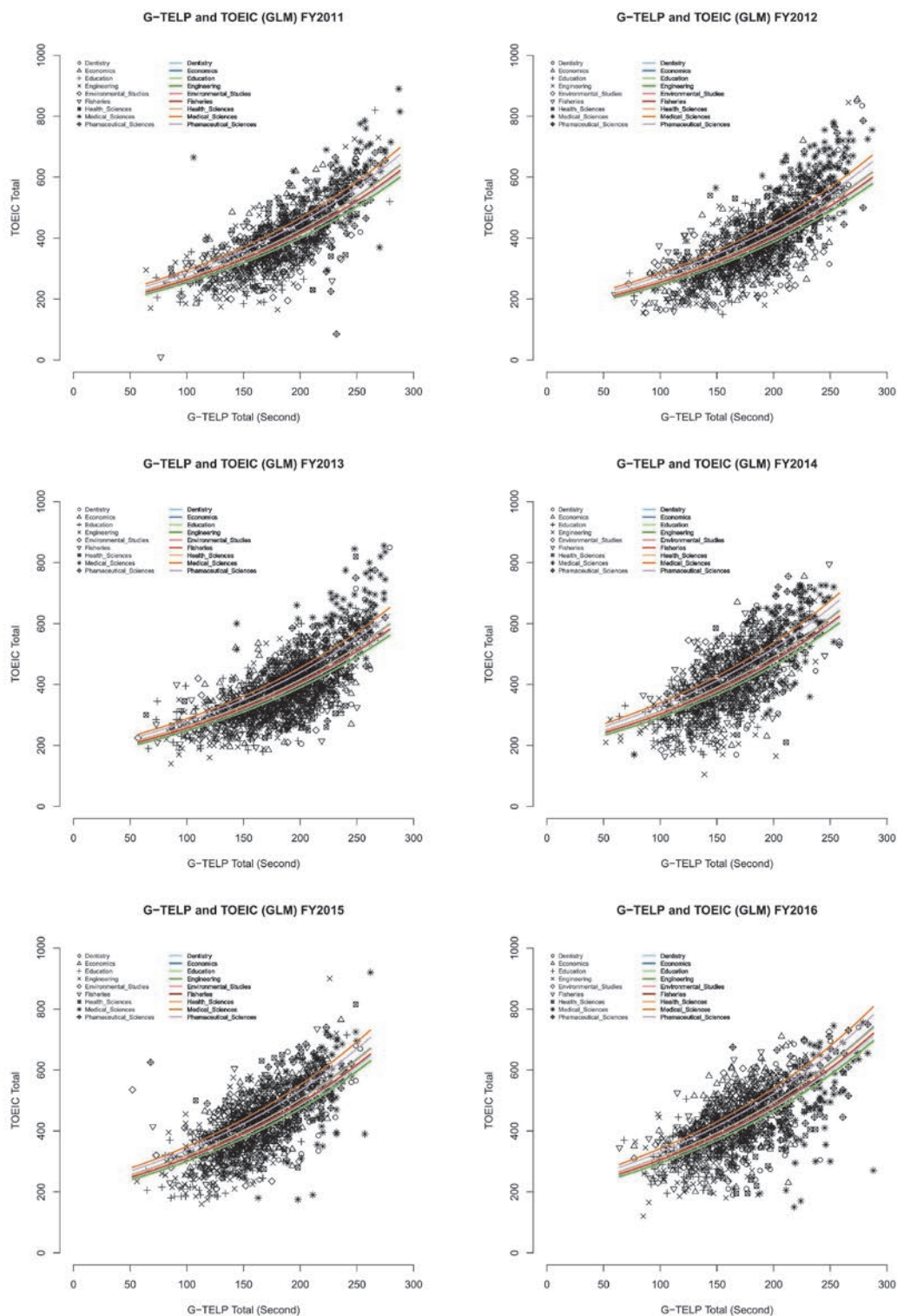


図2 年次別得点散布図得点推計曲線

考察

本研究は、G-TELP スコアを用いて、TOEIC スコアを推定するモデルを精緻化することを目的とした。長崎大学に通う全1年次学生を対象とした。モデリングには、一般化線形モデル（GLM）を用いて検討した。結果、入学年次と学部とは同じ程度に TOEIC スコアに影響すること、および GLM を用いることでより精確なモデリングが可能であることが明らかになった。安価な G-TELP のスコアから、TOEIC のより精度の高いモデリングができたことで、TOEIC を頻繁に受けることができない環境にある学生が、TOEIC スコアを推定する際に役立てることができであろう。また、GLM を用いることで、従来よりも精確にかつ現実的な計算速度で TOEIC スコアを推定することに成功した。

以上の結果から、われわれは以下に示す3点の結論を得た。第1に、年次と学部がもたらす影響を加味しつつ、英語能力を推定することが望まれるということである。両要因は統計的に有意に、TOEIC スコアに対して影響を与える。要因として検討しないという選択肢は、存在しないだろう。ただし、要因として有意に作用しない学部も存在することから、注意も必要である。年次は確実に作用することから、教育期間に、内容、方法を編成する際に検討する必要があるだろう。可能ならば、入学からできるだけ早期に、英語能力試験を実施することが望ましい。また、毎年データを収集して TOEIC スコアを推定する必要もあろう。ただし、過去数年分蓄積したデータを用いて TOEIC スコアを推定することについては、慎重に取り組むことが望ましい。

第2として、G-TELP 成績上位者および下位者においては、TOEIC スコア推定について、個別に検討を要するといえる。本研究が明らかにしたモデルによれば、先行研究による線形回帰モデルよりも残差は小さい。とはいえ、G-TELP 成績下位者については TOEIC スコアが過大に、G-TELP 成績上位者については TOEIC スコアが過小に各々推定される。事実上、G-TELP 成績が極端に高い受験者や低い受験者については先行研究同様、本研究によるモデルも役をなさないであろう。これは、G-TELP がレベル分けされたテストであり、レベル3が TOEIC スコア400から600点あたりの受験生がターゲットであると実施団体が公表している事実からして G-TELP レベル3を用いて TOEIC スコアを推定するには限界があること、非線形性を有し得る反応変数を推定するには GLM を用いても対処し得ないことがあることに由来する。GAM（一般化加法モデル）を用いることも検討するべきであるかもしれない。ただし GAM は一般に解釈が容易ではない。解釈や計算時間など、種々の制約条件を勘案して、よりふさわしい手法が採用されるべきである。

実務的には、スコアが過大に推定されることにはさしたる問題はないように考えられる。問題があるとすれば、成績上位者に期待される TOEIC スコアが過小に推定されることである。本来期待できるよりも低いスコアを目指すための英語教育が提供さ

れるかもしれないからである。幸いなことに、残差が大きくなるほどに TOEIC スコアが高い学生は、本学の場合、さほど多くない。しかしながら、対象学生を早期に発見、同学生が有する英語能力をさらに把握することが必要であろう。

第 3 に、スコアを推定する際、一般化線形モデルは線形回帰モデルよりもふさわしいといえる。本研究が採用したような、残差が負の二項分布に基づくとする仮定や対数を用いたリンク関数は、推定モデルを改善することに貢献している。すなわち、手頃な計算時間内に線形性が必ずしも担保されないデータを取り扱うことを可能にしつつ、線形モデルが仮定する等分散な正規分布を示す残差という条件を備えないデータを取り扱い可能にした。

最後に、今後の課題を述べる。今後、個体差を加味した形での研究をぜひすすめたい。おそらく、学生間に個体差があり、また個体差に由来する G-TELP スコアや TOEIC スコア差が存在すると考えられる。学生毎に英語能力試験や英語能力開発にかかる意気込みは異なるであろう。また、本大学の場合、成績評価の一部となる G-TELP に対して、成績評価に組み込まれておらず、インセンティブもない TOEIC の受験に対しても、個体差が大きく関係してくるものと思われる。これらの点を加味した研究は、ひいては、学習者個人のオートノミー育成の面からも大きな示唆をあたえるものと思われる。

受験時期ラグがもたらす影響も、検討に値するであろう。学事暦や学校行事等に制約されて、両試験受験時期が数ヶ月程度ずれることは、容易に想定される。往々にして、ずれた数ヶ月間に英語能力が伸長する。夏休み前に TOEIC を、夏休み明けに G-TELP を各々受験したケースを考えれば、わかりやすい。能力が上昇したか下降したかはとにかく、両受験日に英語能力が異なるかもしれない。こうした差異がもたらす影響は、スコア換算について検討する際、検討されるべきであろう。

註

1. VELC Test とは、英語能力測定・評価研究会による開発され、大学生に求められる英語力の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。
2. 2011 年度～2013 年度まで、1 年前期 (Form 312)、1 年後期 (Form 319)、2 年前期または後期 (Form 314) を使用した。2014 年度からは、1 年前期 (Form 310)、1 年後期 (Form 313)、2 年前期または後期 (Form 308、ただし 2014 年は Form314) を使用している。なお、本研究で

使用するの、1年生後期の G-TELP のスコアデータである。

3. 2014年7月、Form 310を受験した学生のうち、1週間後2クラス82名の学生に前年まで使用していた Form 312を受験させ、Form間の難易度を確認した。その結果、リスニング、リーディング（ボキャブラリー）のセクション間に有意な差はなかったが、グラマーでは10点程度の開きがあり、Form 310の方が、やや難しい傾向がみられた。

引用文献

- 安間一雄. (2011). 「英語能力テストとその利用」. 石川祥一・西田正・斉田智里（編）『テストニングと評価：4技能の測定から大学入試まで』. 144-172.
- 馬場真哉. (2015). 『平均・分散から始める一般化線形モデル入門』. プレアデス出版.
- 土肥充・張智君. (2014). 「千葉大学における TOEIC IP と TOEFL ITP のスコア分析と経年調査」. 『言語文化論叢』No.8. 15-32. 千葉大学言語教育センター.
- Eguchi, H. (2011). An examination of correlations between TOEFL and TOEIC scores among Hokusei Gakuen students: An attempt to find a score conversion *Hokusei Review, the School of Humanities*, 48(2), 35-44.
- Hemingway, M. A. (1999). *English proficiency tests—A comparative study*. Princeton, NJ: The Chauncey Group International.
- 廣森友人・山西博之. (2009). 「大学英語教育における成績評価と外部試験」第24回大学英語教育学会 (JACET) 中国・四国支部大会シンポジウムハンドアウト.
- 磯田貴道. (2009). 「広島大学における外部試験の活用例：授業に基づく評価と TOEIC」第24回大学英語教育学会 (JACET) 中国・四国支部大会シンポジウムハンドアウト.
- 粕谷英一. (2012). 『一般化線形モデル』共立出版.
- 前田啓朗. (2009). 「大学英語教育における成績評価と外部試験：広島大学における評価の現状と課題」第24回大学英語教育学会 (JACET) 中国・四国支部大会シンポジウムハンドアウト.
- 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』, 92(3), 71-91. 長崎大学経済学会.
- 小笠原真司. (2013). 「長崎大学学生の英語力伸長に関する研究—1年間の G-TELP のデータから」. 『長崎大学言語教育研究センター紀要』No.1. 47-66.
- 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』147-160. 東京：開隆堂.

- 小笠原真司・廣江顕・奥田阿子. (2016). 「2 種類の e-learning 教材による課外学習効果について—G-TELP のデータおよびアンケート結果からの考察—」、『長崎大学言語教育研究センター論集』 No.4. 139-161.
- 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・丸山真純・宇都宮譲. (2015). 「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—」 *Annual Review of English Learning and Teaching*, 20, 63-82.
- Ogasawara, S., Maruyama, M., Utsunomiya, Y. & Collins, W. (2016). Developing a prediction equation for the G-TELP scores from the TOEIC scores using linear regression model: A comparison of four data from 2014 to 2015. *Annual Review of English Learning and Teaching*, 21, 13-30.
- 斉田智里・小林邦彦・野口裕之. (2009). 「外部試験を活用した大学英語カリキュラム改革: 茨城大学共通テストと外部試験との関連」『日本テスト学会誌』 5(1), 96-105.
- 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・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*, 272-277.
- 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.
- Tannenbaum, R. J. & Wylie, E. C. (2005). *Mapping English language proficiency test scores onto the Common European Framework*. TOEFL Research Reports, RR-80.
- 山森光陽. (2009). 「大学英語教育における成績評価と外部試験—指定討議—」第 24 回大学英語教育学会 (JACET) 中国・四国支部大会シンポジウムハンドアウト.
- 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.

補遺 試験種別・年次・学部別得点分布記述統計量

TOEIC 2011	Faculty	N	Min.	Median	Mean	Max.	G-TELP					
							SD	Min.	Median	Mean	Max.	SD
2012	Dentistry	34	270	437.5	457.79	660	104.90	86	195.25	208.79	230.25	36.23
	Economics	119	185	445	438.99	685	100.70	108	173	190.44	214	31.14
	Education	131	185	335	353.17	820	108.85	73	146	168.32	188.5	40.41
	Engineering	239	165	335	347.36	730	88.68	64	143	164.19	184	33.00
	Environmental Studies	122	195	357.5	354.75	530	78.11	94	154.25	172.82	197	33.33
	Fisheries	111	10	330	337.21	685	81.38	77	142.5	163.43	183	34.61
	Health Sciences	89	230	420	414.66	570	74.96	86	172	190.22	210	29.64
	Medical Sciences	89	295	560	558.88	890	112.39	106	222	235.17	253	28.04
	Pharmaceutical Sciences	78	85	477.5	487.44	690	105.80	153	202.25	220.59	241	26.37
	Dentistry	48	165	407.5	436.88	850	141.16	105	191.25	210.56	237	37.61
	Economics	332	170	385	380.41	855	89.49	105	191.25	210.56	237	33.88
	Education	116	150	342.5	352.41	610	104.40	105	191.25	210.56	237	40.83
	Engineering	292	155	320	328.63	845	91.52	105	191.25	210.56	237	35.38
	Environmental Studies	125	155	335	344.76	620	87.13	105	191.25	210.56	237	36.48
	Fisheries	110	160	332.5	330.77	600	82.43	105	191.25	210.56	237	38.56
	Health Sciences	98	250	442.5	440.56	655	83.00	105	191.25	210.56	237	31.93
2013	Medical Sciences	103	385	600	600.97	780	85.66	105	191.25	210.56	237	23.25
	Pharmaceutical Sciences	76	275	495	502.11	785	103.45	105	191.25	210.56	237	29.06
	Dentistry	43	205	455	462.56	850	129.44	134	197.5	218.28	241	32.15
	Economics	348	185	380	382.63	620	81.90	100	166.5	187.30	207.5	31.49
	Education	234	180	335	347.01	650	90.39	66	141	165.44	193	38.58
	Engineering	354	140	342.5	351.74	655	87.55	82	145.5	172.11	197.75	35.05
	Environmental Studies	128	210	370	373.55	580	85.89	57	150.75	175.24	204.25	36.94
	Fisheries	106	195	330	338.92	640	80.63	73	145	168.51	196.75	39.23
	Health Sciences	84	225	395	401.18	820	90.23	64	174	195.08	218	33.83
	Medical Sciences	93	250	595	591.02	855	121.59	144	223	235.09	258	27.85
	Pharmaceutical Sciences	76	260	482.5	491.45	775	105.71	164	211	226.79	246.75	24.75

(Continued)

2014

Dentistry	40	170	455	454.63	660	122.43	125	169.5	190.88	209.25	31.22
Economics	121	200	430	431.28	755	100.17	104	147	165.02	181	25.44
Education	214	170	350	355.54	650	96.27	64	125	145.87	166.5	29.63
Engineering	336	105	360	360.36	675	89.64	52	129	147.80	168	30.63
Environmental Studies	116	185	415	415.99	655	93.37	99	145	161.56	177	27.15
Fisheries	105	165	395	400.14	795	105.56	85	138	157.68	176	32.59
Health Sciences	76	210	465	451.45	720	111.58	103	153	172.32	194	29.85
Medical Sciences	101	305	580	564.60	725	100.34	146	194	209.47	225	23.87
Pharmaceutical Sciences	78	375	530	540.71	755	84.12	143	185	198.83	211.75	21.53
2015											
Dentistry	41	240	490	481.83	680	106.87	113	173	192.44	211	31.38
Economics	252	215	427.5	428.75	765	89.17	99	145	161.50	179	27.29
Education	209	180	360	362.34	715	95.94	64	127	147.34	168	30.36
Engineering	335	160	380	384.01	900	100.63	56	130	148.01	166.5	29.07
Environmental Studies	120	210	402.5	398.92	620	87.05	52	130.75	149.54	168.25	29.47
Fisheries	107	225	420	423.22	735	92.74	70	136	154.67	174.5	30.49
Health Sciences	90	245	445	454.94	815	97.97	108	154	172.24	189	27.59
Medical Sciences	100	175	537.5	531.00	920	121.89	151	198	211.43	227	20.91
Pharmaceutical Sciences	65	295	545	541.00	740	91.48	68	185	197.26	215	29.41
2016											
Dentistry	46	190	385	412.39	740	141.34	128	172.25	192.63	214.75	30.35
Economics	249	200	460	455.96	710	92.69	77	149	168.69	189	28.40
Education	218	195	375	378.37	630	83.12	68	128.25	147.29	167	27.70
Engineering	298	120	387.5	388.31	690	88.99	78	134	153.72	172	28.02
Environmental Studies	87	230	420	428.56	725	103.12	77	148.5	162.64	176	29.23
Fisheries	90	210	410	421.44	660	87.11	64	133	152.06	171.75	30.53
Health Sciences	90	195	415	410.78	660	98.24	109	171.25	187.47	211	31.57
Medical Sciences	74	150	510	492.36	755	137.26	147	212.5	228.12	245	28.24
Pharmaceutical Sciences	66	315	522.5	526.06	750	99.50	136	196.5	215.70	240.75	31.83

