

TOEFL対策授業の有効性に関する調査研究

高橋邦年、田島祐規子¹、カールマクガリー²、山ノ内哲也³、
山品みゆき⁴、加藤千博⁵、今村紅子⁶

A survey on the effectiveness of English classes focusing on TOEFL

Kunitoshi TAKAHASHI, Yukiko TASHIMA, Carl MCGARY,
Tetsuya YAMANOUCHI, Miyuki YAMASHINA, Chihiro KATO,
Beniko IMAMURA

0. はじめに

横浜国立大学（以下、本学）では、日本の他の多くの大学と同様に、英語科目は必修あるいは必修に近い科目として設定されている。実質的には、1年次生のほとんどが英語を履修する。日本の大学レベルでの語学の授業については様々な批判がある。中でも、授業の内容、レベル、および成績評価に関して、担当教員間でばらつきがあるということは以前から指摘され続けている。本学では、このような批判・指摘を反省材料とし、数年におよぶ議論の末、その改善方法の一環として成績の平準化を実現するために統一的なテストの導入が検討され、TOEFL Level 2 (ITP⁷) を利用することになった。

平成13年度から、1年次に英語 I（平成18年度より、英語実習 1）を履修する学生は、Listening/Reading（前期）、Listening/Reading（後期）、Speaking（前期あるいは後期）、Writing（後期あるいは前期）の4科目を受講するようになっている。そして、その中のListening/Reading（後期）（以下、後期LR）では、授業の一環として、後期末試験期間中に設定される「英語統一テスト」実施日に全学的規模で行われる統一テストの受験が義務付けられるようになった。その統一テストとしてTOEFL Level 2 (ITP) が実施されはじめた。⁸

後期LRは学期末のTOEFL受験を目的としているので、TOEFL対策の授業を通してより高得点の獲得を目指している。教科書については、統一教科書や指定教科書の使用には至っていないが、原則的に、推奨テキストのリストの中から担当教員が選択した教科書を使用している。

本調査では、後期LRの授業が、果たしてTOEFLの得点の向上につながっているのかどうかを調

¹ 横浜国立大学大学教育総合センター教員

² 横浜国立大学大学教育総合センター教員

³ 横浜国立大学非常勤講師

⁴ 横浜国立大学非常勤講師

⁵ 横浜国立大学非常勤講師

⁶ 福岡女学院大学短期大学部英語科教員（元横浜国立大学非常勤講師）

⁷ Institutional Testing Program

⁸ 統一テスト実施までの経緯についての詳細は『平成13年度英語 I 統一テスト実施報告書』に記されている。

べることにする。⁹

1. 本調査研究の目的

本調査研究は、下記の計画表にあるとおり、以下の2つを目的とする。

- ① 平成17（2005）年度入学者における、統一テストとしてのTOEFL受験を目的とする後期LRに関する授業効果の有無の分析。
- ② リスニング問題のないセンター試験を受験して入学した平成17年度入学者と、リスニング問題入りのセンター試験を受験して入学した平成18（2006）年度入学者における、リスニング問題の得点の比較と分析。

（表1）後期LRに関する調査研究プロジェクト実施計画（2カ年計画）

実施時期	調査内容
平成17年10月	平成17年度後期授業————① (予備テスト実施)
平成17年11月～12月	結果集計
平成18年2月	(統一テスト)
平成18年度	①の統計処理結果と分析(研究報告書作成)
平成18年10月	平成18年度後期授業————② (平成17年度と同じ予備テスト実施)
平成18年11月～12月	結果集計
平成19年2月	(統一テスト)
平成19年度	上記①②の統計処理結果と両年度の統一テスト結果の比較・分析(研究報告書作成)

①については、後期LRがTOEFL受験に効果を及ぼしているか、また、効果があるのであれば、どの程度のものであるかを検討する。また、②については、①と同じ検討に加えて、センター試験にリスニング問題が導入されたことが、リスニング能力の向上につながったのか、また、リスニング問題の導入が英語力自体の向上に影響を与えたかを検討する。

2. 調査に使用したテスト

理想的には、TOEFL Level 2を2度受験させるのが望ましいが、予算の都合上、TOEFLを同一学年に2回実施することは不可能である。そこで前年度に使用したTOEFL Level 2を活用できれば便利であるが、国際教育交換協議会(CIEE¹⁰)によって徹底したマテリアル管理がなされており、コピーをとることも、現物のテスト冊子を1冊たりともテスト会場に残すことができない。そのため、CIEEから入手できるTOEFL Sample Test (4th edition)を「予備テスト」として本調査に使用することにした。

⁹ 本稿は、執筆者全員による2カ年計画のプロジェクトの1年目の研究報告である。主として、高橋と田島が執筆にあたった。

¹⁰ Council on International Educational Exchange

3. 調査の方法

本調査は以下の手順による。

- 1) 後期授業開始前に、対象クラス担当教員に予備テスト実施の協力依頼。
- 2) 対象クラスの第1回目、もしくは第2回目の授業に予備テストを実施。
 - ①予備テスト実施後、学生による自己採点。
 - ②自己採点後に、問題用紙、解答用紙（自己採点済み）の両方を回収。
- 3) 回収された解答用紙（自己採点済み）の結果を集計。
- 4) 集計結果を分析。

4. 対象クラスについて

本調査の実施にあたり、対象クラスを選定する際に、英語力のレベルの差異を考慮する必要があると考えた。各学部のLRクラスは（旧RLクラス）は、習熟度別にクラス分けが行なわれている。そのため、上位クラスは、英語力が下位クラスよりも平均的には高いと判断される。この理由から、各学部の上位、中位、下位クラスから対象クラスを選出した。結果的には、帰国学生の多いクラスを対象からはせずなどの複数の条件整備をし、工学部、教育人間科学部、経営学部の3学部から抽出した合計10クラスが調査対象クラスとなった。498名の受講者が予備テストを受験したが、そのうち35名は統一試験を受験していないので調査対象とはできなかった。¹¹ また、予備テストで高得点を獲得した受講者の中には帰国学生が含まれている。極めて高い得点を調査対象に含めてしまうと、本学の学生の英語力の「平均的」な姿を見損なう恐れがあるので、外れ値を示した5名（593、610、630、637、640点獲得）も調査対象からはずし、残りの458名を調査対象の被験者とした。¹²

5. 得点の集計と分析

本節では、予備テストの集計方法について説明したのち、予備テストの得点と統一テストの得点を比較・分析する。

5.1 予備テストの得点の集計

上述のように予備テストにはTOEFL Sample Testを利用した。3つのセクションから成り立っており、セクション1がListening Comprehension、セクション2がStructure and Written Expression、セクション3がVocabulary and Reading Comprehensionである。サンプルテストでは、それぞれのセクションが、25問、20問、30問となっており、正式なTOEFLより問題数が少ない。その分だけ各問題の比重を重くして、素点（正答数）から各セクションの得点を算出した。使用した得点換算表は表2に挙げてある。

¹¹ この35名は、主として出席不足により統一テストの受験資格を認定されなかった者であるが、統一テスト当日に病気等の理由で欠席した者も若干含まれている可能性がある。

¹² 例年、全学で約1550名の一年次生が統一テストを受験する。本調査の被験者の463名はその約30%に相当するので、サンプルとしては十分な数であると言えよう。

(表2) TOEFLサンプルテスト用得点換算表

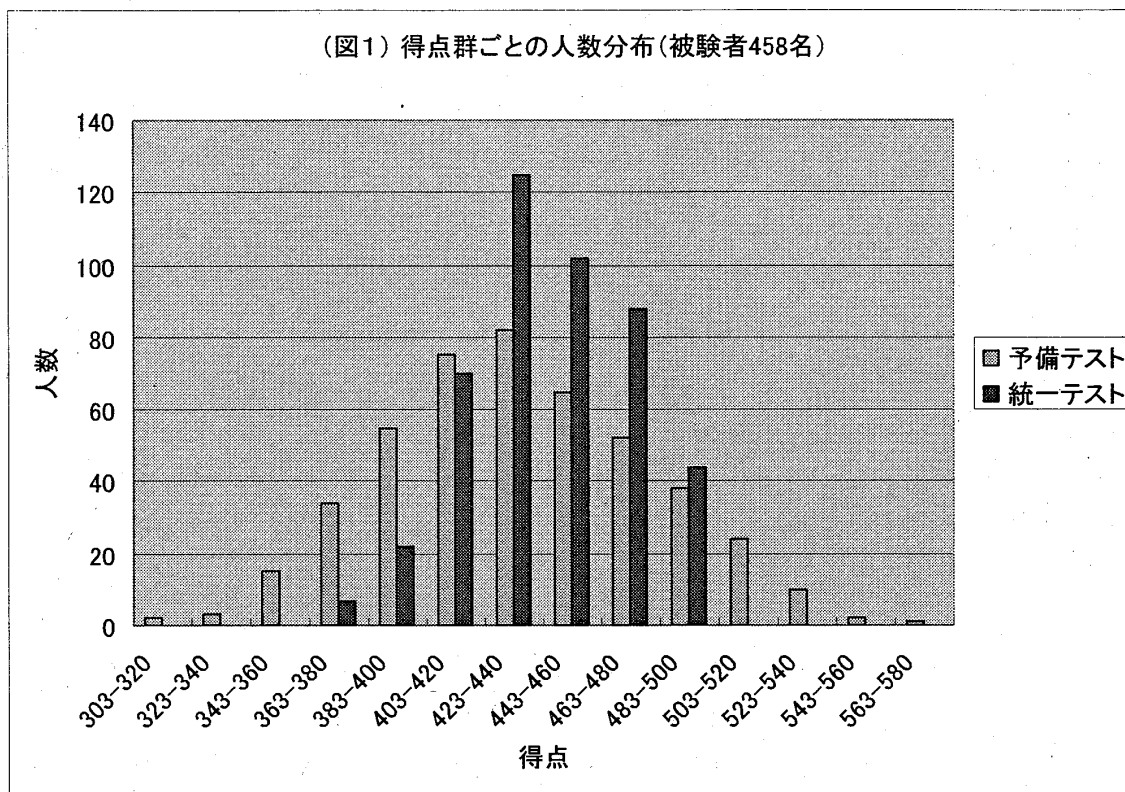
正答数	Section 1	Section 2	Section 3	正答数	Section 1	Section 2	Section 3
0	24	20	20	16	52	58	45
1	26	22	22	17	54	61	47
2	28	25	23	18	56	63	48
3	29	27	25	19	57	66	50
4	31	30	26	20	59	68	51
5	33	32	28	21	61		53
6	35	34	29	22	63		54
7	36	37	31	23	64		56
8	38	39	33	24	66		58
9	40	42	34	25	68		59
10	42	44	36	26			61
11	43	46	37	27			62
12	45	49	39	28			64
13	47	51	40	29			65
14	49	54	42	30			67
15	50	56	44				

5.2 予備テストと統一テストの得点比較

本節では、有効被験者458名の予備テストの得点と統一テストの得点を比較する。予備テストは平成17年10月（後期）の1週目（ないしは2週目）に行われ、統一テストは翌年2月の初旬に行われた。この4ヶ月間にわたり後期LRの授業が行われたことになるが、果たしてこの間に英語の実力が向上したかどうかを、このテストの得点を比較することで探してみる。

(表3) 得点群ごとの人数分布

得点群	予備テスト(人数)	統一テスト(人数)
303-320	2	0
323-340	3	0
343-360	15	0
363-380	34	7
383-400	55	22
403-420	75	70
423-440	82	125
443-460	65	102
463-480	52	88
483-500	38	44
503-520	24	0
523-540	10	0
543-560	2	0
563-580	1	0



予備テストと統一テストの得点を、20点ごとに得点群に分けて、それぞれの得点群に何人の被験者がいるかを見れば、それぞれのテストの点数のばらつきがわかりやすい。これを表にしたものが、表3であり、図示したものが図1である。どちらのテストにおいても、得点ごとの人数分布は、いわゆるベルカーブを描いていると言えよう。

予備テストの平均は434.7点で、統一テストの平均は444点である(表5参照)。平均点を比べるまでもなく、ベルカーブを比べれば、統一テストのほうが中心が右にずれており、全般的によりよい得点を得たことが見て取れる。

統一テストのほうが、500点を超える部分に分布がないのは、TOEFL Level 2は500点が満点で、それを超える能力は計測できないからである。TOEFL Level 1を利用していたら、Level 2で500点を獲得した被験者のうちのある割合の者が500点を超える得点を得ていたものと予想される(表6参照)。そうであれば、統一テストの人数分布の右端は500点以上の範囲に向かって裾野を広げていたはずであるし、統一テストの平均点も数点高かったはずである。

本調査では、予備テストと統一テストは、英語の同じ能力を測定し、難易度が同じであるということを前提としている(これについては、第6節で考察する)ので、図1の比較から、後期LRがTOEFLの得点獲得に何らかのよい影響を与えた可能性があることになる。

図1からは、同じ被験者、あるいは、同じグループの被験者らが、統一テストでよりよい得点を獲得したかどうかはわからない。そこでつぎに、予備テストの得点を20点刻みで10グループに分け、¹³それぞれのグループが統一テストでどのような得点を得たかを表4に示した。対応する図が図2である。

¹³ このグループ分けは、処理上のグループ、つまりバーチャルなグループであり、授業でのクラス分けとは異なる。

(表4) 得点グループ別得点推移

得点グループ	予備テスト (得点)	統一テスト (得点)	人数
303-360	345	419	20
363-380	373	426	34
383-400	393	426	55
403-420	412	439	75
423-440	431	441	82
443-460	452	452	65
463-480	470	455	52
483-500	490	460	38
503-520	510	462	24
523-580	538	489	13

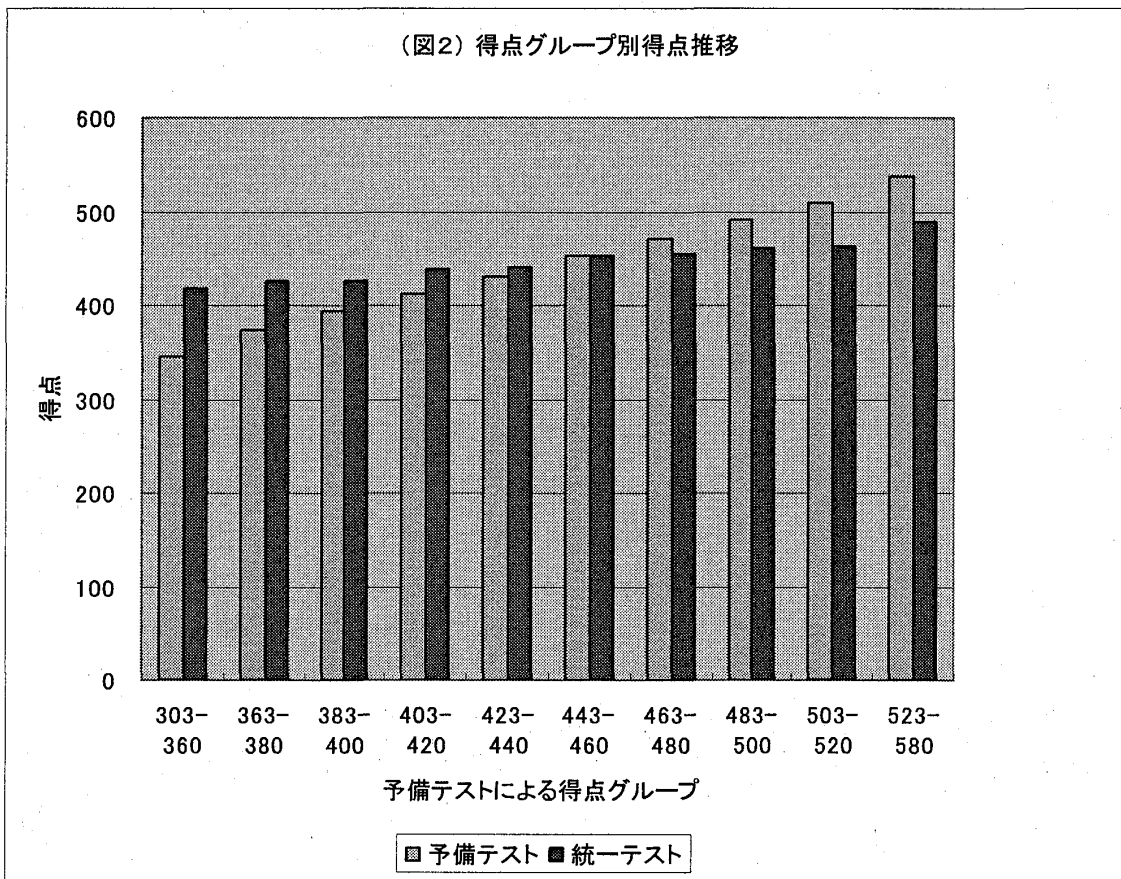


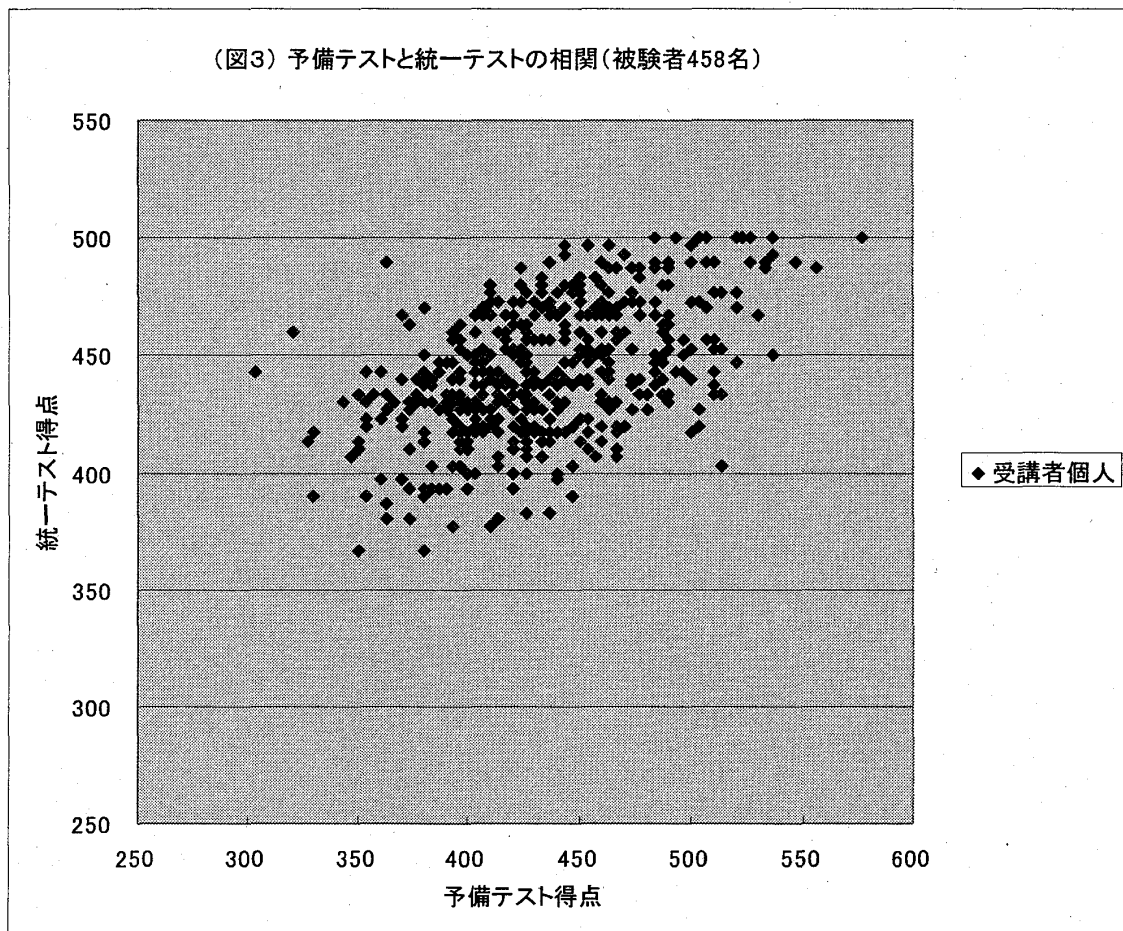
図1で示唆したように、もし後期LRが何らかのよい影響を与えているとすれば、どのグループもいくらかの得点の向上を示しても不思議はない。より高得点のグループですら得点向上の傾向があってもよいはずである。しかし、図2では、右半分の高得点のグループにおいて、予備テストのほうが統一テストよりも点が高いという逆転現象が起きている。図1についての考察で述べたように、図2でもTOEFL Level 2の500点頭打ちの影響が多少でている可能性はある。つまり、特に高得点グループについては、Level 1を使用していれば、平均点がこの結果よりは上がっていたであろうと推定できる。そうであれば、統一テストのグラフについては、実力的には図2よりも右上がりになっていると考えられる。この推定が正しいとすると、少なくとも低得点グループについては、後期LR

の効果がいくらか出ている可能性がある。¹⁴

ここで、予備テストと統一テストに見られた得点の差が有意であるかどうか、また、両テスト間の相関がどの程度であるかを検討してみる。

(表5) 有意水準1%でのt検定

	予備テスト	統一テスト
平均	434.7	444
分散	2010.7	772.7
観測数	458	458
ピアソン相関	0.5	
仮説平均との差異	0	
自由度	457	
t	-5.07	
P(T<=t) 片側	2.90E-07	
t 境界値 片側	2.33	
P(T<=t) 両側	5.81E-07	
t 境界値 両側	2.59	



¹⁴ 図2に見られる逆転現象の原因のひとつとして、表2で用いた得点換算方法が十分に適切ではないということが考えられるが、この点に関しては、今後検討を続けたい。

「2つのテストの母集団の平均値の差がない」という帰無仮説を、有意水準1% ($p < 0.01$) で t 検定した。結果は表5のとおりである。¹⁵ t 値の絶対値5.07は片側検定1%水準のときの t 値2.33よりも大きいので帰無仮説を棄却して、¹⁶「統一テストの平均点のほうがよい」と1%水準で言える。¹⁷

また、予備テストと統一テストの相関はピアソンの相関係数で $r=0.5$ であるので、そこそこの相関があることが明らかである。図に表したものが図3である。プロットのパターンから統一テストは500点で頭打ちになっているのがよくわかる。¹⁸

5.3 予備テストの信頼度

ここでは、予備テストの信頼度を間接的に調べてみる。後期LRの受講者の中から過去にTOEFLやTOEICの受験経験のある者に、任意にスコアの提示を求め、そのデータと予備テストの得点を比較してみた。結果的に11名のTOEICのスコアが集まった。このスコアはすべて予備テスト以前のTOEIC受験によるものである。したがって、予備テストと同等かあるいは、若干予備テストより低い得点であろうと予測できる。TOEICからTOEFLへの換算はつぎの式を使った。

$$\text{TOEFL} = (\text{TOEIC} + 865) \div 2.9$$

換算結果は表6のとおりである。図4には、TOEFL換算点と予備テストの得点の対応をグラフにしてある。

(表6) 予備テストの信頼度調査表

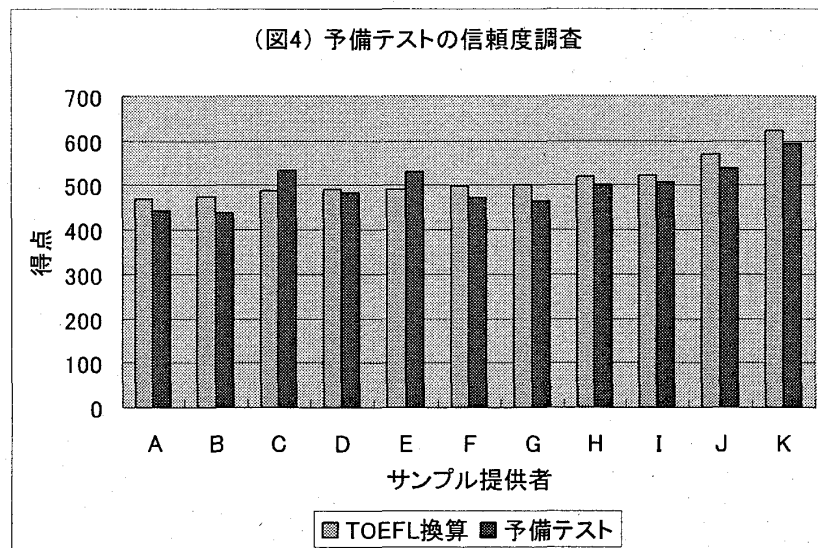
サンプル提供者	TOEIC	TOEFL換算点	予備テスト
A	495	469	443
B	510	474	437
C	550	488	533
D	555	490	483
E	560	491	530
F	580	498	473
G	590	502	463
H	645	521	500
I	650	522	507
J	785	569	537
K	940	622	593

¹⁵ 表中の $2.90E-07$ は 2.90 の小数点を左に7桁ずらした数値を表す。

¹⁶ 帰無仮説を棄却するには、設定した有意水準である 0.01 ($=1\%$) よりも片側検定の $P(T < t)$ の値である $2.90E-07$ のほうが小さいことを理由にすることもできる。

¹⁷ つまり、あることを 1% の有意水準 (つまり、 1% の危険率) で言えるということは、 99% の確信で言えるということである。

¹⁸ ちなみに、センター試験との相関は予備テストに対しては 0.28 、統一テストに対しては 0.33 であり、あまり相関はない。TOEFLで測定している英語力とセンター試験の英語で計測している英語力は別物であると言ってよいであろう。これ以上この問題については、ここでは探求しないが、実際に役に立つ英語力、あるいは、一般に言う英語の実力というものを測るには、いずれのタイプの試験のほうがより適切であるかを真剣に考える必要がありそうである。



一見してかなりの相関がありそうであるが、相関係数は $r = 0.8$ と高い。¹⁹ また、表7の数値から5%水準でも「両テストの平均値に差がない」という帰無仮説を棄却できない。²⁰ つまり、両テストは同じ能力を表している可能性がある。

(表7) 有意水準5%でのt検定

	TOEFL換算	予備テスト
平均	513.3	499.9
分散	2060.2	2166.1
観測数	11	11
ピアソン相関	0.8	
仮説平均との差異	0	
自由度	10	
t	1.53	
P(T<t) 片側	0.078	
t 境界値 片側	1.81	
P(T<t) 両側	0.156	
t 境界値 両側	2.23	

両テストの平均得点と図4を見るかぎりには、若干TOEFL換算点のほうが高得点である。もし換算式が適切であれば、予備テストはいくぶんか低めの測定値を提供する可能性があることになる。このあたりの数値の調整のためにも、表2のサンプルテスト用換算表をさらに精緻なものにする必要がある。

6. 過去の報告との関連 (まとめとして)

平成15年に報告された『平成13年度横浜国立大学英語 I 統一テスト実施報告書』では、教員アンケートの「リーディングおよびリスニングの技能の向上に効果があるか」という問いに対して、肯定的な回答が得られた一方で、効果に対しての高い期待がもたれていない点、技能向上に必ずしも

¹⁹ TOEIC と TOEFL は一般的にかなりの相関があるとされている。

²⁰ t 値 1.53 は片側検定の t 境界値 1.81 よりも小さい (大きくない)

好適ではない授業形態への不満などが報告されている (p. 16)。

また、平成16年度『英語統一テスト実施報告書』でも、統一テスト実施に対しての肯定的な意見が報告されているが、実施理念、(授業および統一テストの) 方法論についての課題も提示されている (p. 5)。

今回のように、「授業効果」を具体的なデータで示す試みは、上記2点のような報告・提示というものに、より具体的・客観的な視点を加え、今後の統一テストに対し有益な情報を提供するという点で、非常に意味を持つと思われる。さらに、今後の課題として、もし後期LRの授業効果があるとすると、なぜ、また、どのようにして効果が生まれるのかを調査する必要があると思われる。²¹

参考資料

平成15年3月、横浜国立大学教養教育運営委員会英語 I 統一テスト報告書作成委員会編『平成13年度横浜国立大学英語 I 統一テスト実施報告書』

平成16年3月、横浜国立大学教養教育運営委員会英語部会編『英語統一テスト実施報告書』

²¹ 想定できる理由には、つぎのようなものが含まれるであろう。「授業が刺激になった」、「教員によるTOEFLの説明が役に立った」、「練習問題等によりTOEFLの問題形式に慣れた」、「ひたすら勉強した」など。