

プレースメントテストとしての TOEIC スコアの検証

上 田 恒 雄

はじめに

本学グローバル英語学科では2年次及び3年次の習熟度別クラス編制を行うための基礎資料として TOEIC 試験を利用している。TOEIC (Test of English for International Communication) はリスニングとリーディングの二つのセクションから成る「英語によるコミュニケーション能力を総合的に評価することを目的」(ETS, 2006) とした集団基準準拠テストの言語テストである。またこの言語テストは受験者の一般的な英語力を測定するための習熟度テストでもある。このような特性をもった言語テストを使用目的としてプレースメントテストに用いる際の検証をおこない、テスト結果がその目的にあった機能を果たしているかどうかを本稿では検討する。過去に TOEIC スコアを使って、クラスを分ける境界の線引きをする得点を決めることが困難な状況に何度も遭遇したこともあり、TOEIC は本学科の英語プログラムに適切なクラス分けが行なえるほど受験生を分布させることができるのか、具体的に2009年度以降過去5年間の本学科の学生が受験した TOEIC スコアを使って確かめる。

研究対象

愛知学院大学文学部グローバル英語学科に在籍する学生が2009年度から2013年度までに受験した5年間分の TOEIC のスコアを用いて検証する。なお、国際コミュニケーション英語能力テストとして知られる TOEIC 試験の具体的な問題構成の特徴は、下記の通りである。

2時間で200問(990点) マークシートのテスト

問題は、問題形式の説明を含めて英語

リスニングは45分で100問(495点) Section I (Part I~IV)

写真描写、応答、会話、説明文問題(リスニングの設問文は、一度しか流れない)

リーディングは75分で100問(495点) Section II (Part V~VII)

短文穴埋め、長文穴埋め、読解問題

検証・分析方法

TOEIC は多様な受験者集団の幅広い英語能力を測定することを目的とした集団基準準拠テストであり、受験者を英語プログラムの適切なレベルに配置するクラス分けの目的に使うには、測定対象となる本学科の受験生のスコア(英語習熟度)がクラス分けを適切に行えるほど幅広く分布していないといけな。プレースメントテストの目的である「受験生をテスト結果により幅広く分布させる」ことができなければ、プレースメントテストとして不適切となり、本来測定すべき内容を測定していないことになる。そこで本学科の学生が実際に受験した TOEIC スコアの得点分布を調べて、プレースメントテストとして機能するほど十分な正規分布をなしているかどうかを検証する。理想的な正規分布とは度数分布図(ヒストグラム)を見ると、平均値を頂点として左右対称に釣鐘状(ベルカーブ)にケースが分布している状態である。グラフで視覚的に調べるだけでは客観的な正確さを欠くので、TOEIC スコアの正規性を仮説検定する。正規性検定として「シャピロ・ウィルク検定(Shapiro-Wilk test)」を用いて、統計処理ソフトを使って出力した結果を基に正規分布

にしたがうかどうかを検証することで判断する。統計的仮説検定における帰無仮説は「変数は正規分布に一致する」、対立仮説は「変数の分布は正規分布に一致しない」である。有意確率が $(P) < 0.05$ となり、帰無仮説が棄却されれば正規分布に一致するとはいえない (=正規分布にしたがわない) と判断され、逆に有意確率が $(P) \geq 0.05$ となり、帰無仮説が棄却されなければ正規分布に一致するとはいえない (=正規分布にしたがう) となる。定まった基準で正規分布かどうかを判定できることもあり、一般的にケース数が2000を超えない場合は特にシャピロ・ウィルク検定は正

規分布との違いを見つけ出し易くよく利用されており、ヒストグラムもあわせて確認することでより正確に正規性の判断ができる。

結 果

2009年度から2013年度までのデータを順に1年生から3年生まで各学年ごとのTOEICスコアの統計値を含めた正規性の検定結果を表にしたものを示す。続いてその右にヒストグラムにしたものも併せて提示する。なお、4年生は受験者数が少ないのでデータの信頼性の観点から今回の検証からは除外する。

2009年度 1年生

表 1

平均値	374.7200	
平均値の95% 信頼区間	下限	358.4763
	上限	390.9637
5%トリム平均	370.3000	
中央値	365.0000	
分散	8419.074	
標準偏差	91.75551	
最小値	185.00	
最大値	650.00	
範囲	465.00	
4分位範囲	110.00	
歪度	.710	
尖度	.538	
Shapiro-Wilk		
統計	自由度	有意確率
.965	125	.002

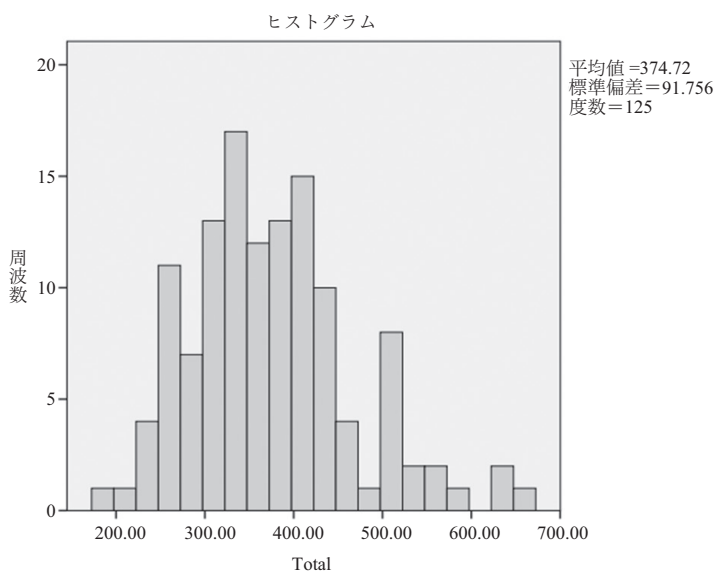


図 1

プレースメントテストとしての TOEIC スコアの検証 (上田)

2009年度 2年生

表 2

平均値	434.1964	
平均値の95%	下限	412.9707
信頼区間	上限	455.4222
5%トリム平均	429.4048	
中央値	412.5000	
分散	12850.700	
標準偏差	113.36093	
最小値	165.00	
最大値	765.00	
範囲	600.00	
4分位範囲	143.75	
歪度	.680	
尖度	.317	
Shapiro-Wilk		
統計	自由度	有意確率
.963	112	.003

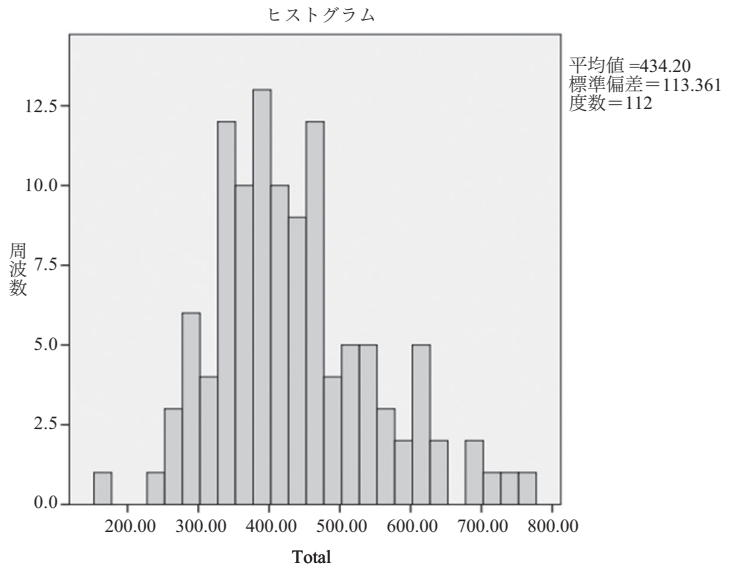


図 2

2009年度 3年生

表 3

平均値	470.2655	
平均値の95%	下限	448.3645
信頼区間	上限	492.1665
5%トリム平均	464.6657	
中央値	455.0000	
分散	13806.179	
標準偏差	117.49970	
最小値	260.00	
最大値	840.00	
範囲	580.00	
4分位範囲	160.00	
歪度	.726	
尖度	.342	
Shapiro-Wilk		
統計	自由度	有意確率
.963	113	.003

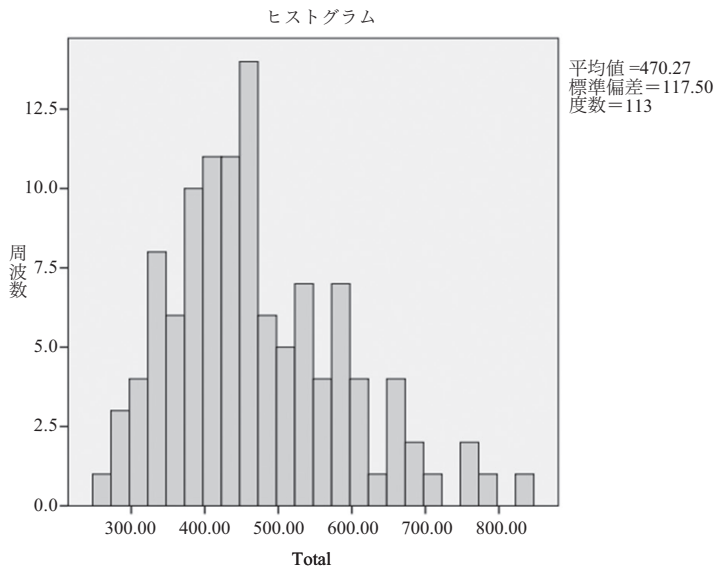


図 3

2010年度 1年生

表 4

平均値	392.4590	
平均値の95%	下限	373.4338
信頼区間	上限	411.4843
5%トリム平均	390.1184	
中央値	385.0000	
分散	11266.631	
標準偏差	106.14439	
最小値	155.00	
最大値	705.00	
範囲	550.00	
4分位範囲	136.25	
歪度	.353	
尖度	.076	
Shapiro-Wilk		
統計	自由度	有意確率
.989	122	.405

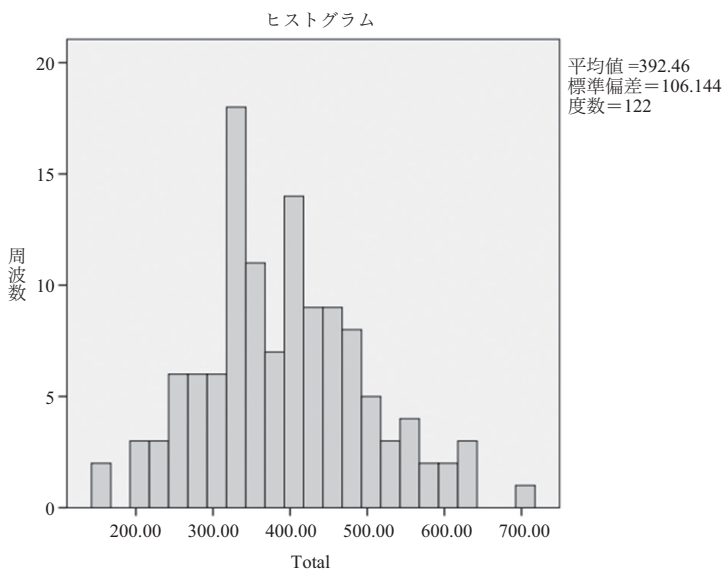


図 4

2010年度 2年生

表 5

平均値	444.9138	
平均値の95%	下限	424.5865
信頼区間	上限	465.2411
5%トリム平均	439.5977	
中央値	432.5000	
分散	12216.079	
標準偏差	110.52637	
最小値	195.00	
最大値	790.00	
範囲	595.00	
4分位範囲	130.00	
歪度	.781	
尖度	1.017	
Shapiro-Wilk		
統計	自由度	有意確率
.961	116	.002

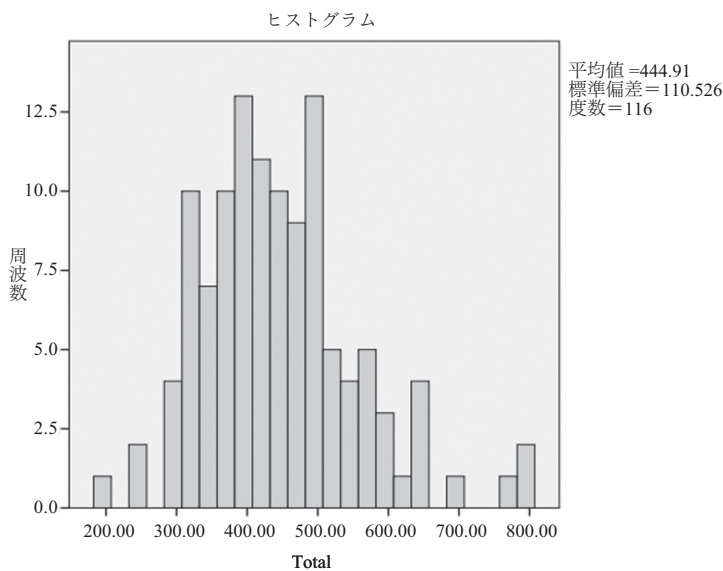


図 5

プレースメントテストとしての TOEIC スコアの検証（上田）

2010年度 3年生

表 6

平均値	498.9806	
平均値の95%	下限	474.6207
信頼区間	上限	523.3404
5%トリム平均	495.5906	
中央値	475.0000	
分散	15535.470	
標準偏差	124.64137	
最小値	240.00	
最大値	855.00	
範囲	615.00	
4分位範囲	175.00	
歪度	.475	
尖度	-.179	
Shapiro-Wilk		
統計	自由度	有意確率
.975	103	.052

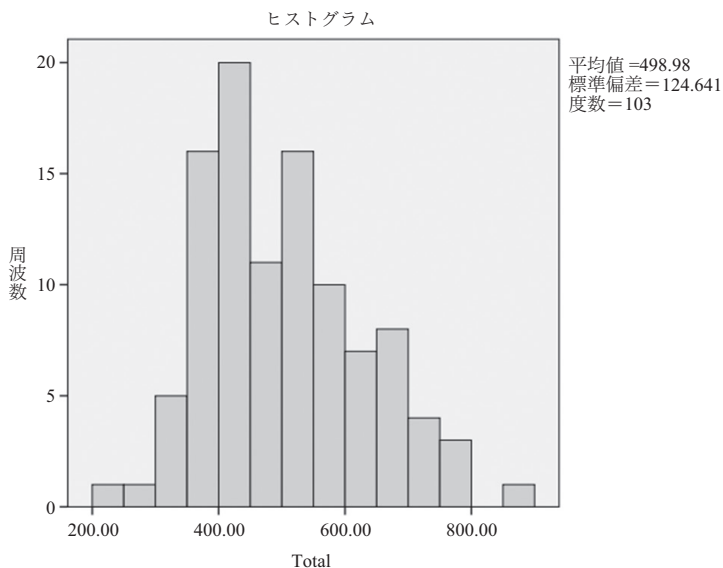


図 6

2011年度 1年生

表 7

平均値	402.2973	
平均値の95%	下限	375.2794
信頼区間	上限	429.3152
5%トリム平均	400.8183	
中央値	397.5000	
分散	13599.445	
標準偏差	116.61666	
最小値	155.00	
最大値	780.00	
範囲	625.00	
4分位範囲	156.25	
歪度	.240	
尖度	.672	
Shapiro-Wilk		
統計	自由度	有意確率
.983	74	.432

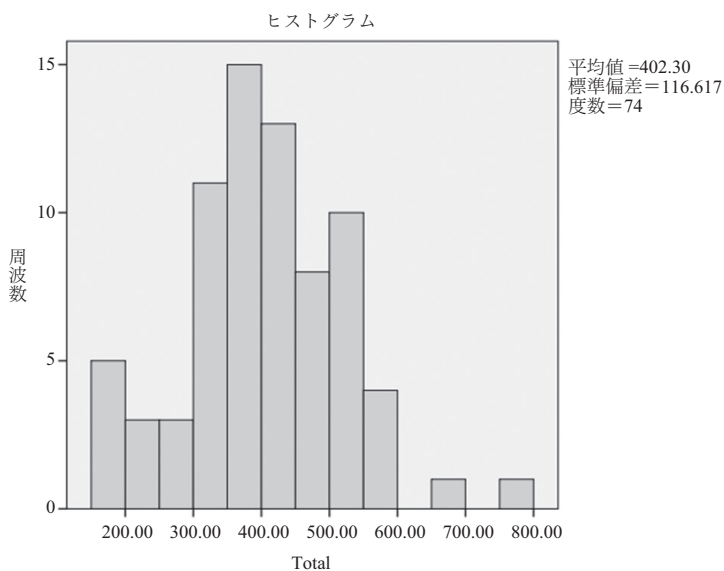


図 7

2011年度 2年生

表 8

平均値	492.4648	
平均値の95%	下限	464.6637
信頼区間	上限	520.2659
5%トリム平均	489.3701	
中央値	485.0000	
分散	13795.624	
標準偏差	117.45477	
最小値	275.00	
最大値	760.00	
範囲	485.00	
4分位範囲	155.00	
歪度	.477	
尖度	-.336	
Shapiro-Wilk		
統計	自由度	有意確率
.961	71	.025

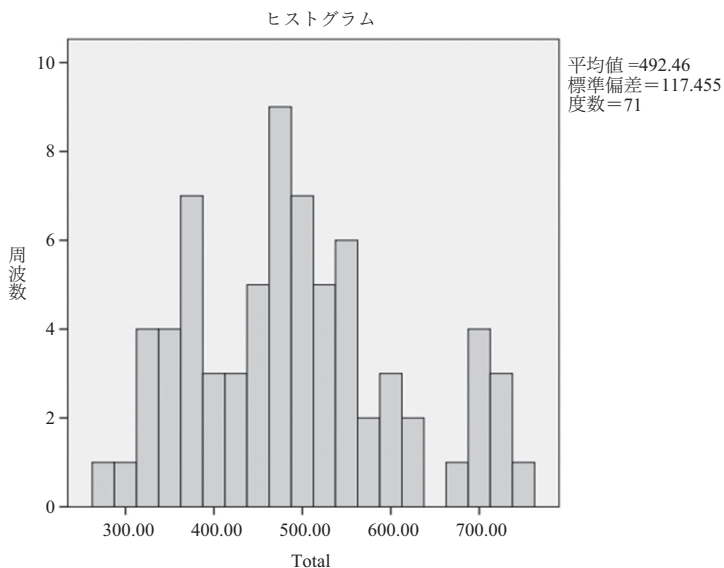


図 8

2011年度 3年生

表 9

平均値	500.3289	
平均値の95%	下限	468.8209
信頼区間	上限	531.8370
5%トリム平均	492.9971	
中央値	465.0000	
分散	19012.224	
標準偏差	137.88482	
最小値	230.00	
最大値	900.00	
範囲	670.00	
4分位範囲	155.00	
歪度	.969	
尖度	.843	
Shapiro-Wilk		
統計	自由度	有意確率
.929	76	.000

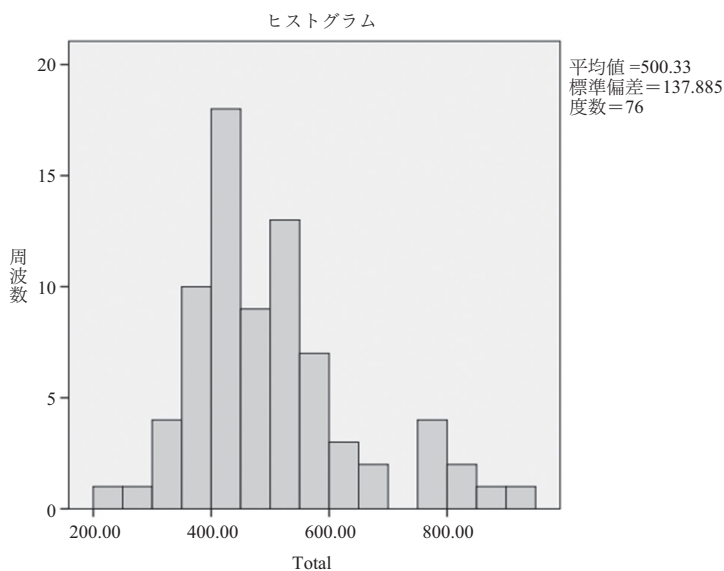


図 9

プレースメントテストとしての TOEIC スコアの検証 (上田)

2012年度 1年生

表10

平均値	398.9904	
平均値の95%	下限	375.7889
信頼区間	上限	422.1918
5%トリム平均	392.9808	
中央値	390.0000	
分散	14233.194	
標準偏差	119.30295	
最小値	205.00	
最大値	770.00	
範囲	565.00	
4分位範囲	173.75	
歪度	.651	
尖度	.198	
Shapiro-Wilk		
統計	自由度	有意確率
.964	104	.006

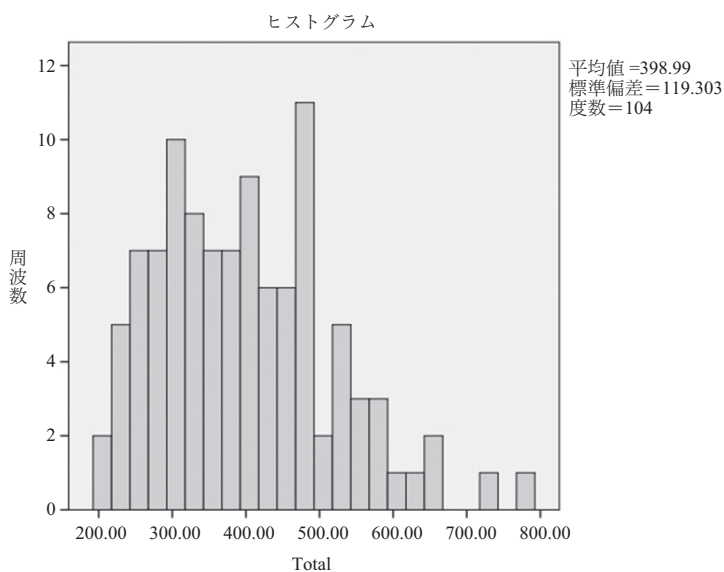


図10

2012年度 2年生

表11

平均値	452.8889	
平均値の95%	下限	427.0948
信頼区間	上限	478.6829
5%トリム平均	451.6975	
中央値	442.5000	
分散	15166.841	
標準偏差	123.15373	
最小値	210.00	
最大値	735.00	
範囲	525.00	
4分位範囲	176.25	
歪度	.105	
尖度	-.673	
Shapiro-Wilk		
統計	自由度	有意確率
.984	90	.320

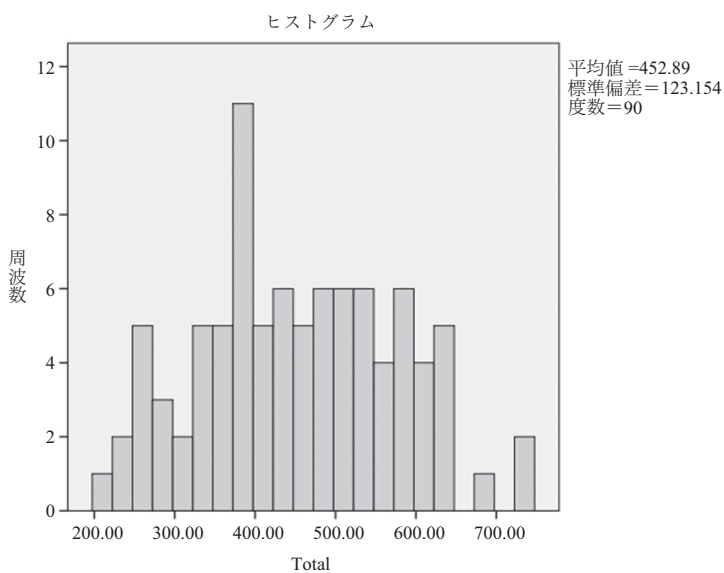


図11

2012年度 3年生

表12

平均値	486.3725	
平均値の95% 信頼区間	下限	452.4361
	上限	520.3090
5%トリム平均	480.8442	
中央値	460.0000	
分散	14559.078	
標準偏差	120.66101	
最小値	220.00	
最大値	880.00	
範囲	660.00	
4分位範囲	155.00	
歪度	.817	
尖度	1.456	
Shapiro-Wilk		
統計	自由度	有意確率
.955	51	.051

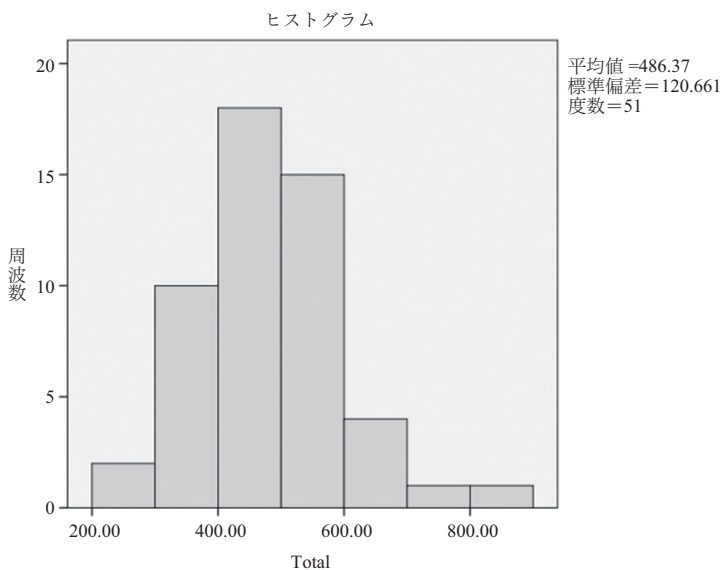


図12

2013年度 1年生

表13

平均値	437.2936	
平均値の95% 信頼区間	下限	417.2306
	上限	457.3565
5%トリム平均	435.8384	
中央値	425.0000	
分散	11166.913	
標準偏差	105.67362	
最小値	215.00	
最大値	750.00	
範囲	535.00	
4分位範囲	152.50	
歪度	.250	
尖度	-.103	
Shapiro-Wilk		
統計	自由度	有意確率
.991	109	.713

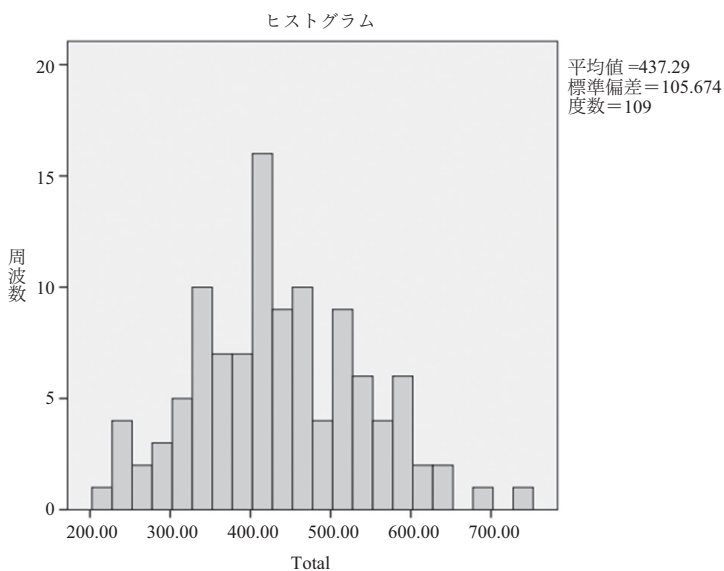


図13

プレースメントテストとしての TOEIC スコアの検証（上田）

2013年度 2年生

表 14

平均値	458.9691	
平均値の95%	下限	433.5524
信頼区間	上限	484.3858
5%トリム平均	459.9599	
中央値	440.0000	
分散	15903.614	
標準偏差	126.10953	
最小値	55.00	
最大値	790.00	
範囲	735.00	
4分位範囲	155.00	
歪度	-.041	
尖度	.427	
Shapiro-Wilk		
統計	自由度	有意確率
.988	97	.504

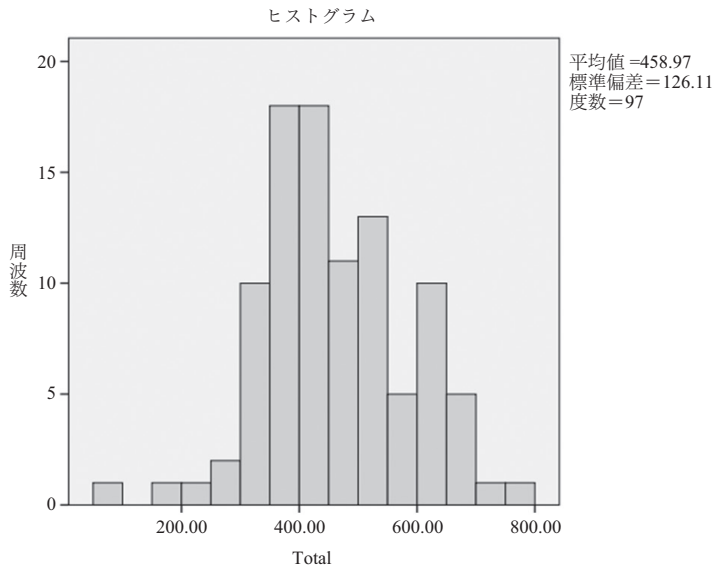


図 14

2013年度 3年生

表 15

平均値	585.6818	
平均値の95%	下限	535.9823
信頼区間	上限	635.3814
5%トリム平均	581.6162	
中央値	562.5000	
分散	12564.989	
標準偏差	112.09366	
最小値	415.00	
最大値	830.00	
範囲	415.00	
4分位範囲	156.25	
歪度	.635	
尖度	-.326	
Shapiro-Wilk		
統計	自由度	有意確率
.947	22	.271

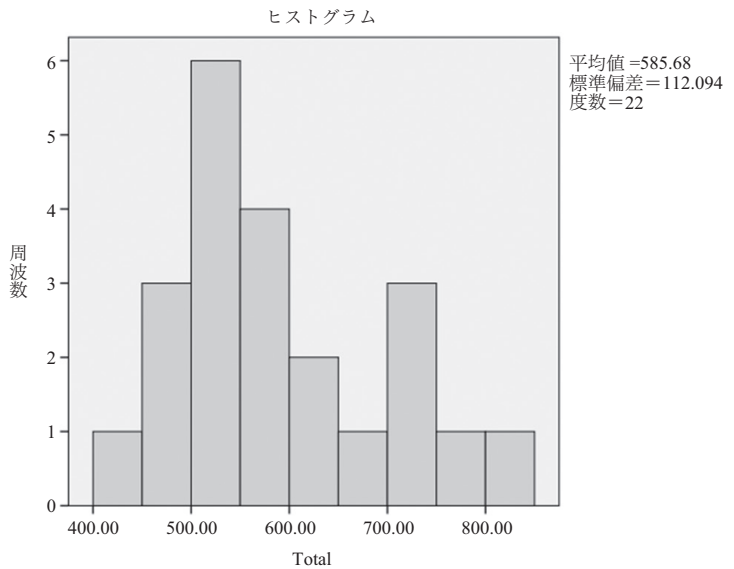


図 15

ま と め

2009年度から2013年度までの各学年毎のTOEICスコアを仮説検定した結果、帰無仮説が棄却されずに正規分布に従うと言えるのは、下記表16の7つのTOEICスコアのみである。残り8つのスコアは帰無仮説が棄却されて正規分布に従っているとは言えないことになる。

表16

年度	学年	$P \geq 0.05$
2010	1年生	0.405
2010	3年生	0.052
2011	1年生	0.432
2012	2年生	0.32
2012	3年生	0.051
2013	1年生	0.713
2013	2年生	0.504

実際の度数分布を表示したヒストグラムを確認すると仮説検定の結果を裏付ける様な形状が見られる。ちなみに統計上の天井効果、床効果は何れのスコア結果において見られないが、ヒストグラムのみで結論づけることは見るものの主観にも左右されるかもしれないので確実とは言えない。今回の結果が示すようにTOEICスコアを習熟度クラ

スのプレースメントに利用することは、必ずしも毎年すべての学年で正規分布が保証されているわけではないことを考えれば、学生の中には自分の英語力に応じた適切なクラスに配置されていないものもいる可能性がある。そこで今後はTOEICスコアのみで2年次以降の習熟度クラス編制することの適切性を再考する必要があるかもしれない。但し、TOEICの試験としての信頼性や、特に実用性を考慮すれば今まで本学科がプレースメントテストとしてTOEICを利用してきたことは妥当であるとも言えるので、TOEICスコアの結果のみに頼るのではなく、他の要素を加味してクラス分けをすることも重要であろう。

参考文献

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge, Cambridge University Press.
- Brown, J. D. (2005). *Testing in language programs: A comprehensible guide to English language assessment*. New York, Mcgraw-Hill College.
- ETS. (2006). *Test of English for International Communication*. Princeton, NJ: Educational Testing Service.
- 中村洋一 (2002). テストで言語能力は測れるか—言語テストデータ入門. 東京, 桐原書店
- 和田稔 (1999). 言語テストの基礎知識 (訳) (*Testing in language program*). 東京, 大修館書店