

# 「RLG テスト」の信頼性と妥当性の検討および形成的利用法に関する研究

## A Study on the Use of RLG Test: Reliability, Validity, and Formative Use

キーワード：語彙・英文法、各種テスト、大学生用英語力測定テスト

加藤 千博、田島 祐規子、村上 嘉代子、前川 浩子

KATO Chihiro, TASHIMA Yukiko, MURAKAMI Kayoko, MAEKAWA Hiroko

### 1. はじめに

現在、TOEIC、TOEFL、CASEC、GTEC、G-TELP といった外部試験を英語カリキュラムに導入する大学が増加傾向にある。その利用目的は「習熟度別クラス編成」や「国際基準での英語力の把握」が主たるものであるが、他にも「単位認定」や「動機付け」としての利用が報告されている(斉田他, 2009)。しかしそれぞれのテストにはそれぞれの特性があり、目的に応じて相応しいテストを選択しなければならない。Messick (1989) は、テストの得点の解釈と使い道、そして社会への影響を考慮することがテストの妥当性を確立するためには必要であると提言している。それは、テストの妥当性はテストそのものに内在する絶対的なものではなく、テストの使用目的や状況によって変わりうる相対的なものであることを意味している(小林, 2006)。

TOEIC-IP を用いて新入生対象のプレースメント・テストを行い習熟度別のクラス編成を行う大学が少なくないが、これには問題がある。センター試験が受験生を序列化するために英語知識を測定するテストである一方で、TOEIC は実際の場面でのコミュニケーション力、つまり英語の運用力を測るためのテストである。そうすると、入学したばかりの新入生は、これまで受験勉強として学習してきた英語力とは異なったタイプのテストを受けてレベル判定を受けることになる。また TOEIC では 1 回目と 2 回目の受験の間に *practice effect* という現象が生じて得点が上昇しやすいなど、スコアの揺れの問題がある(Childs, 1995; 竹蓋他, 2007)。実際に個々の学生レベルでは標準誤差を超える範囲で得点が上下する例がしばしば見られる。よって、TOEIC 受験経験の少ない新入生にとっては実力通りのスコアを獲得することは困難であることが予想され、正確な習熟度別のクラス分けが行えない可能性が生じ得る。同様のことが TOEFL-ITP にも言えよう。

このような現状の中、簡易に実施ができて使い勝手のよいテストを提供しようと大場(2009)が「RLG テスト(大学標準英語学力テスト)」を開発し、サイバー化を行って無料で提供している。このテストは、「R(読んでわかる単語)テスト」、「L(聞いてわかる単語)テスト」、「G(文法の基礎知識)テスト」からなっており、R テストは 15 分 50 問、L テストは 15 分 40 問、G テストは 30 分 60 問の四者択一式で、トータル 1 時間を要し、各問 1 点で 150 点満点となっている。R テストは JACET8000 の頻度順 5000 語から、L テストは 4000 語をもとに作成されており、G テストは大場文法シラバスの「基本品詞」「文」「動詞の拡充」「文の拡充」「動詞の転換」「文の転換」の 6 領域から各 10 問ずつ 60 題、初級、中級、上級レベルの問題が 20 問ずつになるよう配置されている。

この RLG テストは TOEIC や TOEFL のような英語の運用力を測るテストではなく英語学習者の基

礎力、つまり潜在能力・基礎体力を測るテストと言える。本稿では、最初に 1) このテスト及び作成方法が信頼のおけるものであるかどうかを検証し、次に 2) 外部テストに代わる指針となりうるのかを検討した上で、3) 実際に授業内でどのように活用することが可能かを考察することにする。

## 2. 等価性と信頼性

テストを作成する上でその作成方法の信頼性を示すことは必須である。本研究では、信頼性の指標となる等価性を測り、RLG テストの作成方法の信頼性を検証した。等価性とは、同一個人に同じ作成方法によるテストを実施して得られる結果が安定しているかどうかということである。等価性を証明することで、この RLG テストの作成方法によって作られたテストは代替可能であると言える。等価性を検証するためにはさまざまな方法があるが、本研究では平行テスト法を使用した。RLG それぞれのテストにおいて等質と仮定された 2 種類のテスト A、B を用意し、その二つのテストを同一集団に行い、テスト A、B の一致度を見た。テストの一致度の指標は二つのテストの得点の相関係数である。したがって、テスト A、B の相関係数の値が高ければ、テストの等価性が高いと言える。二つのテストが等価であるということを検証するためには、1) テスト A、B の平均点に統計的に有意な差がない、2) テスト A、B 間に有意な正の相関がある、という 2 点を示す必要がある。

以上のことから RLG テストの等価性の検証を以下のように実施した。

ステップ 1：RLG それぞれのテストにおいて等質と仮定された 2 種類のテスト A、B を用意する。

ステップ 2：5 つの大学の学生に両テストを 10 日以内に受験してもらう。それぞれのテストには制限時間が設けられている。受験者数は、R テスト 340 名、L テスト 257 名、G テスト 431 名である。

ステップ 3：得られた結果を分析する。分析方法として、テスト A とテスト B の平均値、標準偏差、相関係数、 $t$  検定による検討を行う。

### 2.1 データ分析結果

上述の方法で実施されたテスト結果と分析結果を表 1 にまとめた。RLG テストそれぞれ 2 種類のテスト A、B の最大値や最小値にあまり開きはなく、平均値を見てもほぼ同じ数値である。G テストの平均値に 4 点近い開きはあるが、標準偏差はほぼ同じでばらつきはないと言える。R、L テストも同様である。さらに、二つのテスト A、B の平均値に関して、対応のある  $t$  検定を行ったところ、L テストにおいてはテスト A、B の平均点に統計的に有意な差がないことが示され ( $t(256) = 0.48, n.s.$ )、R テストと G テストでは、平均値間に有意な差があることが示された (R テスト： $t(339) = -2.85, p < .01$ , G テスト： $t(430) = -15.21, p < .01$ )。このことは R、G テストにおいては、テスト B のほうが平均値がわずかではあるが有意に高いことを意味している。また、それぞれのテストについて、テスト A、B との相関係数を算出したところ、0.71~0.85 といずれも有意な正の強い相関が見られた。RLG いずれのテストもテスト A、B 間に強い相関があると言える。

この相関を実際に図で示した。図 1 は R テストの散布図を示している。x 軸をテスト A の得点、y 軸をテスト B の得点とし 340 名のデータを表示した。図からわかるように、右上がりの強い正の相関

があることがうかがえる。データは近似曲線上に並び、「外れ値」と呼ばれるようなばらつきのあるデータは少ない。ここでの相関係数は  $r=0.85$  ( $p<.01$ ) と高く、1%水準で有意であった。つまり、二つのテストには統計的に意味のある相関関係があるといえよう。これはL、Gテスト共に同様である。

表1 RLGテスト分析結果

	Rテスト (50問)		Lテスト (40問)		Gテスト (60問)	
	RA	RB	LA	LB	GA	GB
<i>n</i>	340		257		431	
<i>Max</i>	50	49	39	37	52	56
<i>Min</i>	13	11	8	7	4	0
<i>M</i>	33.1	33.8	22.9	22.8	23.1	27.5
<i>SD</i>	7.6	7.6	5.5	5	10.1	10.6
<i>t</i>	-2.85**		0.48 <i>n.s.</i>		-15.21**	
<i>r</i>	0.85**		0.71**		0.83**	

\*:  $p<.05$ , \*\* :  $p<.01$

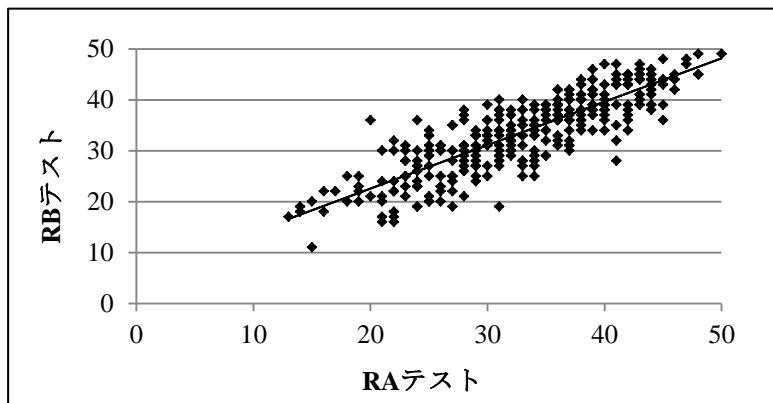


図1 RA/RBテストの散布図

以上のことから、まず、Lテストについては、テストA、Bの平均値に有意差がなく、強い相関係数が得られたことから、等価性は示されたと言える。次にR、Gテストについては、テストA、Bの平均値に有意差が見られたが相関係数は十分な値が得られている。対応のある *t* 検定では、2つの変数間の相関係数が正で大きいほど *t* 値が大きくなる傾向があり、有意差が出やすくなると言われている(吉田, 1998)。このことから、今回のRテストとGテストにおいてテストA、B間の平均値差が有意になったことは、等価性がないことを示すものであるということではなく、相関係数の高さが反映している可能性がある。従って、R、GテストにおいてもテストA、Bの等価性はほぼ示されたと考えられ、RLGテストの作成方法には信頼性があると結論付けられる。

### 3. TOEFL・TOEIC との相関

RLG テストの作成方法が信頼のおけるものであり、テスト自体の信頼性が証明されれば、今度は TOEFL や TOEIC などの外部テストとの相関を示すことにより、この RLG テストの妥当性を示し、活用方法を広げていくことが可能となる。

TOEFL・TOEIC との相関を見るために、3 大学 4 グループ、計 228 名を対象に RLG テストを実施し、その前後 1 か月以内に受験した TOEFL-ITP、TOEFL-Level2、TOEIC のスコアを分析した。以下の表 2 はグループごとの人数、RLG テストにおける平均値、標準偏差、そして TOEFL もしくは TOEIC の平均値、標準偏差を示している。

表 3 はグループごとの RLG テストと TOEFL および TOEIC との相関係数を示したものである。いずれにおいても有意な相関が得られたが、特に RLG テストの合計点との相関係数を見ると、0.55 から 0.87 までという中程度からかなり強い相関を示していることがわかる。するとこの RLG 合計点との高い相関から、グループごとで TOEFL や TOEIC の得点予測ができるのではないかと推測される。

表 2 RLG テストおよび TOEFL・TOEIC スコアの平均・標準偏差

	n	R テスト		L テスト		G テスト		RLG 合計		TOEFL・TOEIC	
		M	SD	M	SD	M	SD	M	SD	M	SD
A 大(09)	25	39.2	4.4	27.0	2.9	44.3	4.6	110.4	8.8	470.9	21.1
A 大(08)	57	37.9	4.0	25.9	2.9	43.5	5.8	107.3	9.3	483.2	21.9
B 大(08)	114	39.0	3.7	27.5	3.4	46.9	4.8	113.3	8.7	466.9	21.7
C 大(08)	32	32.2	9.2	30.7	6.7	21.7	5.1	84.6	19.2	448.1	167.8

A 大 : TOEFL-ITP B 大 : TOEFL-Level2 C 大 : TOEIC

表 3 RLG テストと TOEFL・TOEIC との相関係数

		R テスト	L テスト	G テスト	RLG 合計
A 大(09)	TOEFL-ITP	0.73**	0.50*	0.58**	<b>0.83**</b>
A 大(08)	TOEFL-ITP	0.31*	0.27*	0.52**	<b>0.55**</b>
B 大(08)	TOEFL-Level2	0.32**	0.47**	0.50**	<b>0.60**</b>
C 大(08)	TOEIC	0.84**	0.73**	0.80**	<b>0.87**</b>

\*:  $p < .05$ , \*\*:  $p < .01$

図 2、3 は A 大 (09) における RLG テストと TOEFL-ITP との相関を示した散布図である。RLG テストで合計 100 点以上の者が 23 名いるが (図 2)、そのうちの 2 名を除いた 21 名 (点線楕円部分) が TOEFL-ITP で 450 点以上のスコアを獲得できている。すなわち、RLG テストで 100 点以上獲得したうちの 9 割強の学生が TOEFL-ITP で 450 点以上を獲得できている。RLG テストで 110 点以上獲得した者に関しては (図 3) 9 割強の学生 (14 名中 13 名) が TOEFL-ITP で 470 点以上獲得することができている。

同様に B 大 (08) の学生の TOEFL-Level2 の得点を分析すると、RLG テストで 100 点以上獲得

した学生の8割が TOEFL-Level2 で 450 点以上のスコアを獲得し、RLG テストで 120 点以上の者の9割弱が TOEFL-Level2 で 470 点以上を獲得していることがわかった。よって、似たようなレベルと特性を備えた A 大 (09) と B 大 (08) の学習者集団に関しては、仮に TOEFL-ITP と TOEFL-Level2 のスコアがともに TOEFL-PBT と等質だとすれば、RLG テストで 100 点以上獲得すれば TOEFL で 450 点以上を、RLG テストで 110~120 点以上獲得すれば TOEFL で 470 点以上獲得できる可能性が高いと言える。

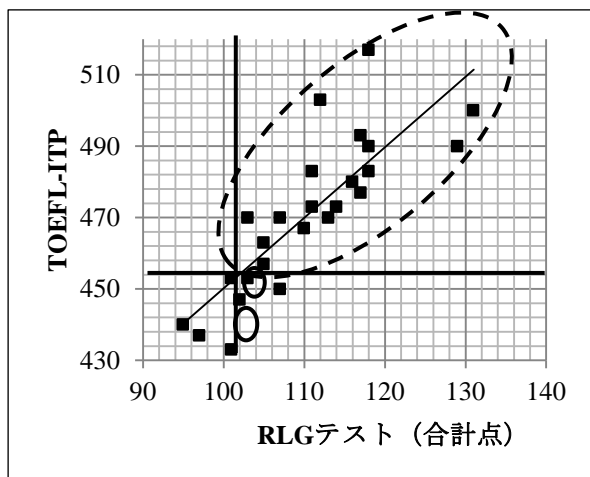


図2 RLG100点以上のTOEFL得点予測

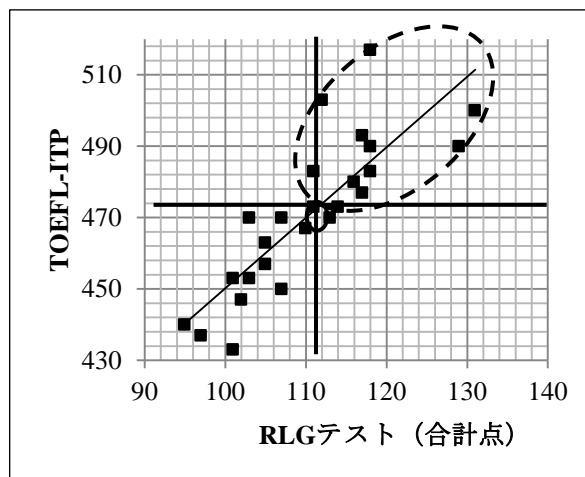


図3 RLG110点以上のTOEFL得点予測

C 大 (08) からは、RLG テスト 80 点以上で TOEIC 420 点以上を、RLG テスト 110~120 点以上で TOEIC 650 点以上を獲得できる可能性が高いことがわかった。しかしこれらはいずれもそれぞれの学習者集団に限った得点予測値であり、一般化するにはデータ数が少なすぎ、各グループのレベル、特性、学習方法によってこの予測値は異なってくるのが予測される。しかしながらこの RLG テストを用いることにより、各学習者集団がどのような特性を備え、どのようなことが学習課題となるかが見えてくる。その上でこの TOEFL や TOEIC の得点予測を行えば、その予測値に満たなかった学生の学習課題を見つけ出すことが可能となる。

この RLG テストは受験者の英語における基礎力を測定するものであり、TOEFL や TOEIC のようにある程度練習しないと実力通りのスコアが出せないという類のものではない。いつ受けてもその時点での基礎力が正確に判定でき、いわば英語の潜在能力を測ることができる。そして先に示したように TOEFL や TOEIC とも高い相関がある。よって、RLG テストは TOEFL や TOEIC に不慣れな大学新入生対象のプレースメント・テストや TOEFL・TOEIC や英検などの外部テストを受ける前のプレテストとして活用が可能である。更には、R、L、G テストのそれぞれの得点を分析することにより、外部テストからだけでは見えてこないその学習者や学習者集団の特性を顕在化させることが可能となる。そして、この数値によって可視化された学習者の特性を自らのクラス指導にどのように反映させることができるかが次の課題となりうる。以下ではクラスルームにおける活用法を考察することにする。

#### 4. 形式的利用—R 語彙サイズ

RLG テストの等価性および信頼性が確認され、TOEFL や TOEIC との高い相関も示された。また、RLG テストが英語の基礎力を測るという点も指摘された。これらの点を前提にした時、RLG テストを授業指導に活用することが可能であると考えられる。語彙研究で著名な Nation (1990) など多くの語彙研究者が指摘するように、語彙力は英語力を示す一つの指標として捉えられている。そのことから、今回は、RLG テストのうち読んでわかる単語を試す R テストを実施し、その結果について語彙指導を含めた授業指導の一助にするべく形式的利用を試みた。

R テスト実施に先立って、中学・高校までの語彙と JACET8000 の関係を確認した。これに際しては、教科書会社が文科省検定教科書を作成する際に語彙数調整に使用する中学・高校 1 年までの語彙リストを活用した。このリストには、中学必修基本語の 100 語を含む中学終了時点での既習語 920 語、加えて高校 1 年で学習する英語 I 終了時点での既習語 405 語、の合計 1325 語が記載されている。その 1325 語と JACET8000 の両方に含まれない語を対象から外し、最終的に両リストに含まれる 1268 語についての重なり具合を確認した。

結果として、その 1268 語は JACET 2000 語までにおいて 85%、JACET 3000 語までならば 96%含まれていることが判明した。R テストは、JACET 8000 の頻度順 100 語ごとに 1 語を任意に抽出して作成されている。そのため、R テストの間 1~10 番までは JACET の 1000 語レベルまでの語彙、問 11~20 番までは JACET の 2000 語レベルまでの語彙、問 41~50 番までは JACET の 5000 語レベルまでの語彙という非常に分かりやすい構成になっている。つまり、上記の結果からするならば、R テストの間 20 番までは英語 I を学習する高校 1 年生までの基本語彙の定着状況を示すことになる。また、大場(2009)は高校修了時での平均受容語彙サイズを約 3000 語、難関大学の平均ラインを 4000 語と設定しているため、R テストの間 31~40 番、問 41~50 番の結果からも学生の語彙力に関する特徴が観察できると考えられる。

上記の考え方にに基づき、B 大教養英語 (Reading /Listening) 上位、中位、下位 3 クラスの初回授業において、R テストを実施した。結果は表 4 および図 4 の通りである。

この観察結果から担当する上位、中位クラスのレベルが非常に近接しており、この 2 クラスにおいては 4000 語レベル以降での語彙増強が課題であると判断した。また、下位クラスについては予想以上に高校修了時点での語彙力に不安があることが判明した。今学期は、ニュース記事を扱ったやや難易度の高いテキストを上位クラス用、初級レベルのものを中位・下位クラス用に用意していたが、この結果から 3 クラスの語彙指導を含めた授業方針について大まかに以下のように設定を変更した。

- 1) 上位クラスについては、JACET 3000-5000 レベルの語彙学習をする。英英辞典、類義語辞典の活用・多用を促し、authentic な (本物の) 英語に触れさせるためインターネット上の英語ニュースを授業内で活用する。また、テキスト以外の英語に積極的に触れていくように様々な情報 (インターネット上の英語ニュース、英字新聞、英文多読用リーダーズなど) を提供する。
- 2) 中位クラスについては、上位クラスとテキストの難易度が異なるものの、上位クラス同様の語彙指導を積極的に行う。
- 3) 下位クラスについては、JACET 2000-3000 語レベルの語彙学習を目標にする。テキストから語彙・

構文・文法問題を含む宿題プリントを作成し、毎授業の解答作業を通してテキスト文中での語彙の正確な意味や語法を丁寧に学習する。

表4 B大 Rテスト結果

JACET 8000 のレベル	1000	2000	3000	4000	5000
R テスト	1~10	11~20	21~30	31~40	41~50
上位クラス(35)	9.31	9.57	9.09	8.11	6.23
中位クラス(52)	9.22	9.61	8.69	7.45	5.69
下位クラス(52)	8.84	8.49	7.27	5.16	3.76

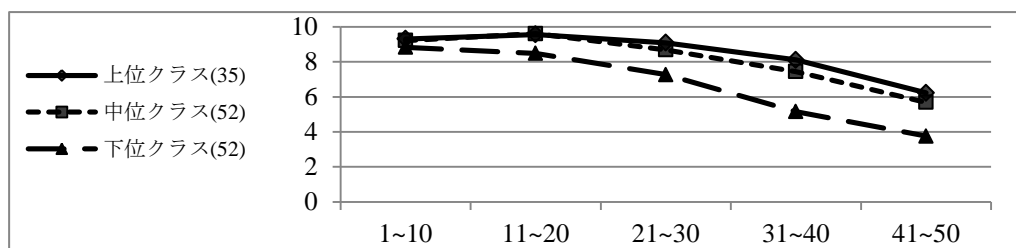


図4 B大 Rテスト結果

B大では、1年次の教養英語クラスを、センター試験の得点結果に基づき上位、中位、下位の習熟度別の3レベルに分ける。ただし、各々の学部におけるレベル分けのため、各学部の上位、中位、下位を横軸に見たときに、全く等しい英語力であることは保証されていない。たとえば、A学部の下位クラスがB学部の中位クラスとほぼ同等レベルにあることも起きうる。しかし、テキストは学生の英語力を確認できない学期前の発注となるため、最終的には担当教員が上位、中位、下位という大まかなレベル情報をもとに、経験的予測を働かせながら授業を行うことになる。ところが、大学入試方法の多様化に伴う学生の英語力の多様化が指摘されている通り、このような経験的予測が外れてしまうことが多くある。しかし、授業開始前に大がかりなプレースメントを実施する教育環境に恵まれていない場合、英語力の多様化を問題視しつつも実際に学生の英語力を検証できないのが現状である。この点において、Rテストは簡易でありながら信頼性の高い語彙の基礎力が測れる。今回のように、授業開始直後に実施したRテストの場合でも、その結果から学生の語彙力を認識し、その後の授業指導に活かすことができた。これは、RLGテストの価値ある形成的利用の一つであると言える。

## 5. おわりに

以上本稿ではRLGテストが大学英語教育の現場でどのように利用可能かを考察してきた。初めに、RLGテスト作成方法の信頼性を検証するために、テストA版とB版の等価性を示した。Lテストにおいては、A版とB版の平均値に有意差がなく強い相関関係が得られたことから、問題なく等価性が示された。R、Gテストにおいては、A版とB版の平均値に有意差が見られたものの十分な相関係数が得られたことから、等価性がないものではないと結論付けられた。次に、TOEFLやTOEICとの相

関を示し各学習者グループにおける得点予測を行った。4 グループにおいて、RLG テストと TOEFL もしくは TOEIC との相関はいずれも有意な中程度から強い相関を示した。A 大学と B 大学のグループに関しては、RLG テストで 110~120 点以上獲得すると TOEFL 470 点以上を、C 大学のグループに関しては、RLG テスト 110~120 点以上で TOEIC 650 点以上獲得できる可能性が高いことが示された。そして更に、R テストに焦点を絞り、R テストの結果から得られた語彙サイズを分析することにより、各クラスにおいてよりの確な指導を行うことが可能となることを実証した。B 大学において、教養英語の上位・中位クラスでは JACET 3000-5000 レベルが、下位クラスでは JACET 2000-3000 レベルの語彙学習が適切な目標設定であることが判明し、当初用意されていたテキストのレベルと実際の学生のレベルの差を補完することが可能となった。

本稿冒頭で述べたようにテストの妥当性はテスト自体の信頼度だけではなく、それをどう使用し、どう利用するかによってくる。この RLG テストの信頼性を保持するためには更なるデータ収集と分析が必要である。それをクリアした後に、このテストの適切な使用法を更に追究し、テスト結果の解釈を適切に行って指導に還元していくことで、このテストの真の妥当性が獲得できよう。我々大学英語教員の目的はテストによって学習者を評価することではなく、テストを利用しながら学習者の英語力を向上させることにある。RLG テストを到達度評価の尺度とすることは容易であるが、それよりも形成的評価のための一要素として活用法を見出していくことのほうに価値があるのではないだろうか。大学英語教育の役割は入学前の英語学習者 (learner) から卒業後の使用者 (user) への架け橋となることである。その役割を果たすような RLG テストの利用法を追究していくことが今後の大きな課題と言えよう。

(防衛大学校、横浜国立大学、芝浦工業大学、金沢学院大学)

## 引用文献

- Childs, M. (1995). Good and bad uses of TOEIC by Japanese companies. In J.D. Brown & S.O. Yamashita (Eds), *Language Testing in Japan* (pp.66-75). Tokyo: The Japan Association for Language Teaching.
- 小林美代子 (2006) 「日本の高等教育機関における英語能力評価についての一考察：結果的妥当性の視点から」『言語科学研究：神田外語大学大学院紀要』第 12 号 93-109 頁
- Messick, S. (1989). Meaning and value in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5-11.
- Nation, I.S.P. (1990). *Teaching & learning vocabulary*. Boston, MA: Heinle & Heinle.
- 大場昌也 (2009) 「大場昌也ホームページ」2009 年 4 月 1 日検索 <http://www9.ocn.ne.jp/~bigarden/>
- 斉田智里・小林邦彦・野口浩之 (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』Vol.5 95-105 頁
- 竹蓋幸生・与那覇信恵 (2007) 「英語教育カリキュラムの中の評価、その方法と問題点」『文教学院大学外国語学部文教学院短期大学紀要』第 7 号 143-158 頁
- 吉田寿夫 (1998) 『本当にわかりやすいすごく大切なことが書いてあるごく初歩の統計の本』京都：北大路書房