

プレースメント・テストとしての G-TELP についての一考察

—大阪商業大学での信頼性・妥当性に焦点をあてて—

前 田 和 彦

1. はじめに
2. 調査方法
 2. 1. G-TELP (L4) の本学における信頼性
 2. 2. G-TELP (L4) の本学における妥当性
 2. 3. G-TELP (L4) の文法問題の出題方法の妥当性
3. 課題と展望

1. はじめに

言語テストはその目的・機能・特性に応じて様々な形が存在する。言語テストのカテゴリーには、検定試験、プレースメント・テスト、学力テスト、診断テスト (Alderson, Clapham, & Wall, 1995; Brown, 2005)、そして進捗テスト (Alderson, Clapham, & Wall, 1995) があげられる。本研究では、プレースメント・テストとしての国際英検 G-TELP を取り上げ、大阪商業大学 (以下、本学) における信頼性と妥当性に焦点をあて、調査する。

「4年制大学における英語プレースメント・テスト実施の現状」(清水、2001)によると、1999年度から2000年度の間で回答のあった4年制大学の学部単位200件のうち、96件(48.00%)が英語統一テストを実施し、そのうち61件(63.54%)がプレースメント・テストとして利用していたことが報告されている。また「英語統一テスト・習熟度別クラス編成・到達目標の設定および測定に関する実態調査の報告」(杉森、2003)では、習熟度別クラス編成が加速度的に増加していることを報告し、プレースメント・テストの必要性が述べられている。これは、年々広がる学力格差が原因となっている。この問題は他大学でも年々深刻化しており、同一クラス内の学生の英語学力格差のため授業進行が非常に困難となり、習熟度別クラスを余儀なくされている大学が増加してきているのである。

学力格差は小・中学校や高等学校でもすでに問題になっており、2007年の文部科学省の公立小・中学校、及び公立高等学校の「教育課程編成・実施状況調査」によれば、「理解や習熟の程度に応じた指導を実施」している小学校は21,860校中18,581校(85.0%)、中学校は10,030校中7,412校(73.9%)、全日制普通科の高等学校は2,485校中1,627校(65.5%)であった。習熟度別授業の有効性については、梅原(2005)が学力向上フロンティアスクール

の報告書の多くが、標準学力テストの偏差値が上昇したことを紹介しており、国立教育政策研究所(2003)が実施した調査研究「指導方法の工夫改善による教育効果に関する比較調査研究」でも、小・中学校において成績向上に最も効果的であったことを報告している。そして岸田(2005)は、成績分布が二極化した事実を記していることから、特に下位グループへの対処が課題となっていることが分かる。

本学では、1回生820名を対象とした先行研究(加賀田、小磯、前田、2007)により、英語に関しては学習初期にすでにつまずいている者が多いことが判明しており、英語の基礎からのやり直しを余儀なくされている学生が多いことが現状である。そこで、本学では2003-2005年度に学生の自己申告による「英語基礎クラス」を開講し、スローラーナーの学修支援を試みた。だが、年々増加する英語クラスの脱落者の数から、自己申告でのクラス編成は信頼性に欠けると判断し、2006年度より本学で作成したプレースメント・テストを使用し、習熟度別クラス編成を行い、上位・中位・下位グループに分け、それぞれをA(Advanced)クラス、S(Standard)クラス、B(Basic)クラスとした。しかし、プレースメント・テスト実施時、及び終了直後に発生するクラス分けの事務処理において膨大な人手・時間・手間を要するため、2008年度より国際英検G-TELP(General Test of English Language Proficiency)日本事務局にテスト実施と習熟度別クラス分け業務を委託することとなった。

本研究では、過去2年間の本学でのG-TELPの結果を古典的テスト理論に基づいて分析し、本学でのプレースメント・テストとしての信頼性・妥当性を検証する。

2. 調査方法

本学では習熟度別クラス編成に向けて2008年(1240名対象)、2009年(1270名対象)共にG-TELPを使用した。G-TELPは、「ネイティブスピーカーでない人が、実際の状況下でどの程度英語をコミュニケーション手段として駆使できるかを測定する」という新しいコンセプトによる国際的な英語検定試験であり、生活やビジネスに関する簡単なコミュニケーションから複雑なものまでを問題としている。

本学が利用しているG-TELPのレベルは、難易度が最も低いとされるLevel 4(形式的な表現方法を用いて、ネイティブと簡単なコミュニケーションができるレベル)である。2009年度は、本学学生の英語力を比較するため、2008年度と同一問題を使用した。本研究では、G-TELP(L4)が本学において、適切なプレースメント・テストと言えるかどうかを、特に信頼性・妥当性の面から調査していきたい。

2.1. G-TELP(L4)の本学における信頼性

まず、G-TELP(L4)の本学における信頼性について調査する。

2008年度、2009年度それぞれのG-TELP(L4)の結果から、クロンバックの α 信頼係数、最小値、最大値、平均値、標準偏差(SD)を表1に示す。

クロンバックの α 信頼性係数、最大値、平均、標準偏差(SD)には2008年度と2009年度

表1 過去2年間のG-TELP(L4)の結果

	テストの種類	問題数	受験者数	時間(分)	α	最小値	最大値	平均	SD
2008	G-TELP(L4)	40	1240	50	0.60	0	35	14.99	4.55
2009	G-TELP(L4)	40	1270	50	0.63	0	32	14.53	4.71

との間には大きな差は見られない。クロンバックの α 信頼性係数は一般に0.80以上であれば良いとされているが、2008年度は0.60、2009年度は0.63とやや低い。これらの数字から、標準誤差(標準偏差 \times SQR(1-信頼性係数))を求めると、それぞれ2.88、2.86となり、大友(2002)によると、95%の受験者の真の得点(得点 \pm 標準誤差 \times 2)は、2008年度は \pm 5.76、2009年度は \pm 5.72の誤差が生じている可能性を示していることとなる。つまり、例えば10点を取った受験者の95%の真の得点が2008年度では4.24から15.76点、2009年度では4.28から15.72点までの開きが生じている可能性があるのである。本学では2008年度・2009年度は14点、あるいは13点を境に下位と中位のクラス分けをしていることから、G-TELP(L4)によるクラス編成には疑問が残る。

尚、G-TELP ホームページでは、目安として表2の通り相関を表している。

表2 他テストとG-TELPの相関表

G-TELP	TOEIC	TOEFL	英検
Level 1	990点	650点	1級
	900点		
Level 2	800点	550点	準1級
	700点		
Level 3	600点	500点	2級
	500点		
Level 4	400点	400点	3級
	300点		
	200点		

この表から判断すると、2008年度、2009年度共に本学受験生の平均値がそれぞれ40点中、14.99点、14.53点と半分にも到達していないことから、先述の誤差を考慮したとしても、本学の多くの学生の英語力は比較的低い方であることが分かる。

次に、過去2年間のG-TELP(L4)の結果より、出題形式別に見たクロンバックの α 信

信頼係数、正答率の平均を表3に示す。回答形式は全て四者択一である。

表3 過去2年間の出題形式別の結果

	出題形式	N	α	正答率の平均
2008年 G-TELP	文法問題 (I=20)	1240	0.47	0.38
	読解問題 (I=20)	1240	0.48	0.34
2009年 G-TELP	文法問題 (I=20)	1270	0.52	0.37
	読解問題 (I=20)	1270	0.48	0.35

出題形式別に見ると、クロンバックの α 信頼性係数は全体よりさらに下がり、0.8を大きく下回っている。また、正答率の平均に関しては、それぞれ0.38、0.34、0.37、0.35であった。大友(2002)は、四者択一形式の問題でプレースメント・テストとして使用する場合の最適困難度(正答率)は0.625(0.5+0.5×1/4)が望ましいと述べていることから、G-TELP(L4)は本学学生にとっては過去2年共に、比較的難題と言える問題が多く出題されていたと考えられる。また、2008年度入学の新生と、2009年度入学の新生との英語力の差はほとんどない。

以上のことから、本学でのプレースメント・テストとしてのG-TELP(L4)の信頼性は低いと考えられる。

2.2. G-TELP(L4)の本学における妥当性

次に、G-TELP(L4)の本学での妥当性について調査する。

G-TELP(L4)は本来、文法問題・リスニング問題・読解問題で構成されているが、本学ではオリエンテーション時の実施ということで、時間的に制限されていたため、リスニング問題は省き、文法問題と読解問題を使用することとなった。

図1・2に文法得点と読解問題の相関を示す。過去2年間のG-TELP(L4)の結果から

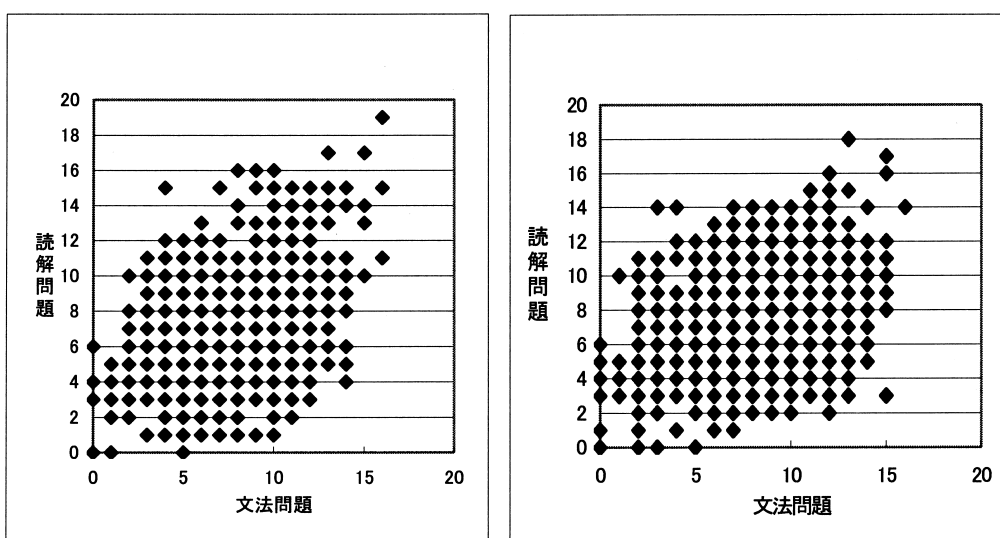


図1 文法問題と読解問題の相関図(2008年度) 図2 文法問題と読解問題の相関図(2009年度)

は、文法問題の得点が高い者が必ずしも読解問題の得点も高いとは言えないことが分かる。つまり文法問題と読解問題との相関は低い ($r=.324$ 、 $.344$)。

従って、文法問題と読解問題の合計点を参考にクラス分けをした場合、Sクラスはもちろん、Aクラスに割り当てられた学生の中にも英語基礎文法力の養成を必要とする学生がいる可能性がある。同様に、Sクラス・Aクラスに割り当てられた学生の中にも基礎の読解力の養成を必要とする学生がいる可能性も否定できない。

次に、どのような割合で本学が新生を上位・中位・下位のレベルに分けているかについて述べる。本学は経済学科、経営学科、商学科、公共経営学科に分かれており、学科別に見ると、経済学科はA (Advanced) S (Standard) B (Basic) がそれぞれ2クラス、4クラス、6クラス、経営学科ではそれぞれ2クラス、4クラス、4クラス、商学科・公共経営学科では1クラス、2クラス、2クラスという編成になっている。A・S・Bの割合では、経済学科ではA、S、Bはそれぞれ16.67%、33.33%、50%、経営学科・商学科・公共経営学科では20%、40%、40%となる。しかし、学科別の英語力の差に関しては表4に示す通り、2008年度、2009年度の結果からは大きな差が見られない。

表4 2008年度、2009年度の新入生の学科別比較

年 度	学 科	受験者数	最小値	最大値	平 均	SD
2008年	経済学科	456	0	28	14.34	4.42
	経営学科	405	0	31	14.88	4.77
	商学科	188	3	35	14.22	4.55
	公共経営学科	187	4	27	14.29	4.33
2009年	経済学科	471	0	31	15.10	4.81
	経営学科	432	0	32	14.52	4.77
	商学科	212	0	26	13.90	4.18
	公共経営学科	155	1	26	13.65	4.66

そこで、これらの2種類の割合(16.67%・33.33%・50%、20%・40%・40%)でそれぞれのクラスに振り分けた場合を想定し、上位・中位・下位について考えてみたい。

まず、上位・中位・下位を総受験生を16.67%・33.33%・50%で分類した場合の過去2年間のクラス別点数・度数分布を表5に示す。

表5 過去2年間のクラス別点数・度数分布(16.67%・33.33%・50%)

クラス	2008年 G-TELP (N=1240)				2009年 G-TELP (N=1270)			
	点数	度数	累積度数	累積相対度数	点数	度数	累積度数	累積相対度数
A	20~35	151	151	0.12	20~32	180	180	0.14
S	15~19	440	591	0.48	15~19	425	605	0.48
B	0~14	649	1240	1.00	0~14	665	1270	1.00

この割合で分類したと仮定した場合のB・S・Aに割り当てられた学生の文法・読解の得点の度数を図3・4・5・6に示す。

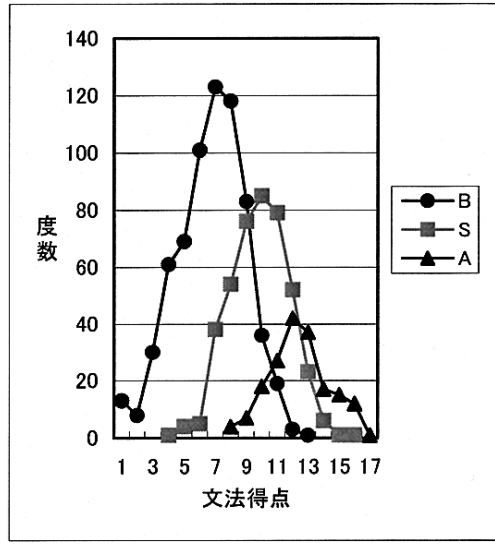
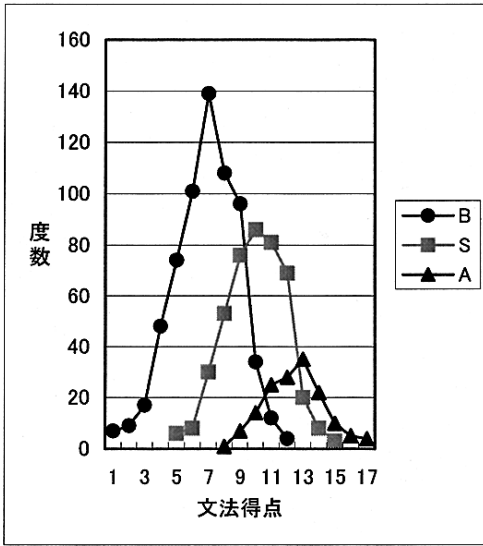


図3 上・中・下クラスの文法得点・度数(2008) 図4 上・中・下クラスの文法得点・度数(2009)

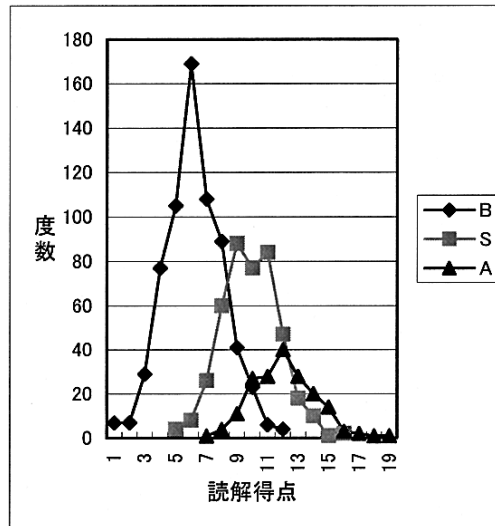
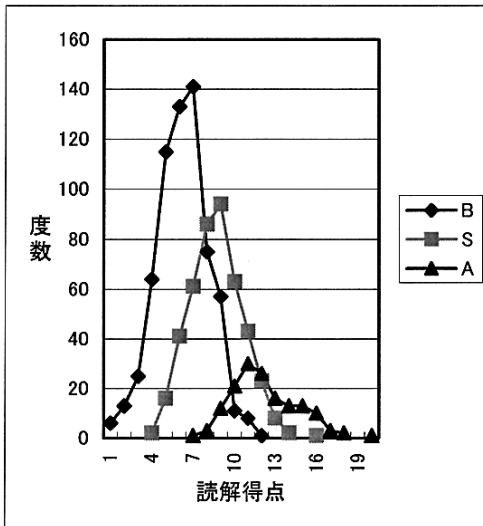


図5 上・中・下の読解得点・度数(2008) 図6 上・中・下の読解得点・度数(2009)

図3・4・5・6が示すように、文法と読解の得点を合計してクラス分けを行った場合、それぞれの技能に適応したクラスに振り分けられていないことが分かる。そこで、2009年度の文法得点を例にとると、Sクラスに振り分けられた学生425名中417名(98.00%)、Aクラスに振り分けられた学生180名中135名(75.00%)がBクラスの学生の一部と文法得点と同じ、あるいは、より低い点を取っていた。また、2009年度の読解得点を例にとると、Sクラスに振り分けられた学生425名中412名(96.94%)、Aクラスに振り分けられた学生180名中111名(61.67%)がBクラスの学生の一部と読解得点と同じ、あるいは、より低い点を取っていた。つまり、文法得点と読解得点の合計によるクラス編成は適切ではないことが分かる。

次に、上位・中位・下位を総受験生の20%・40%・40%で分類した場合の過去2年間のクラス別点数・度数分布を表6に示す。

表6 過去2年間のクラス別点数・度数分布(20%・40%・40%)

クラス	2008年 G-TELP (N=1240)				2009年 G-TELP (N=1270)			
	点数	度数	累積度数	累積相対度数	点数	度数	累積度数	累積相対度数
A	19~35	209	209	0.17	19~32	244	244	0.19
S	14~18	485	694	0.56	14~18	469	713	0.56
B	0~13	546	1240	1.00	0~13	557	1270	1.00

さらに、この割合で分類したと仮定した場合のB・S・Aクラスに割り当てられた学生の文法・読解の得点の度数を図7・8・9・10に示す。

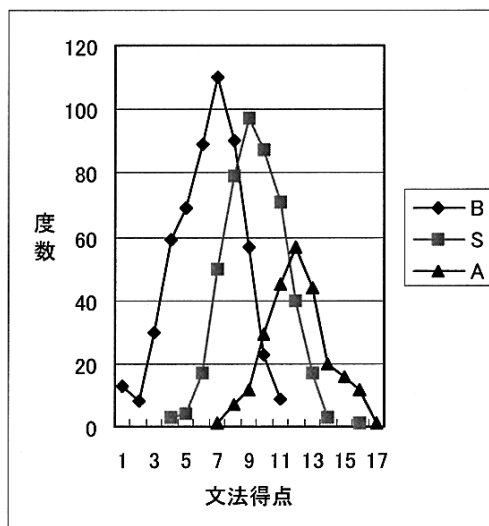
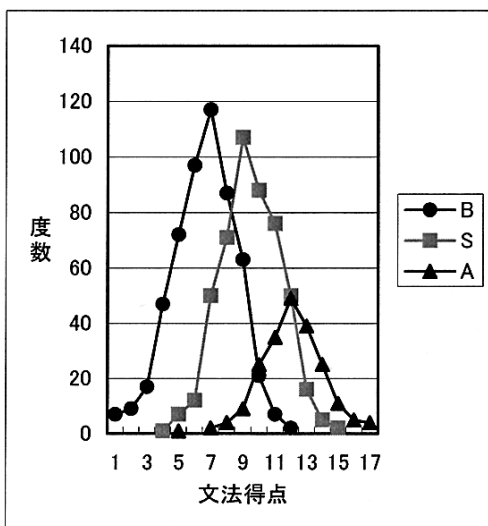


図7 上・中・下クラスの文法得点・度数(2008) 図8 上・中・下クラスの文法得点・度数(2009)

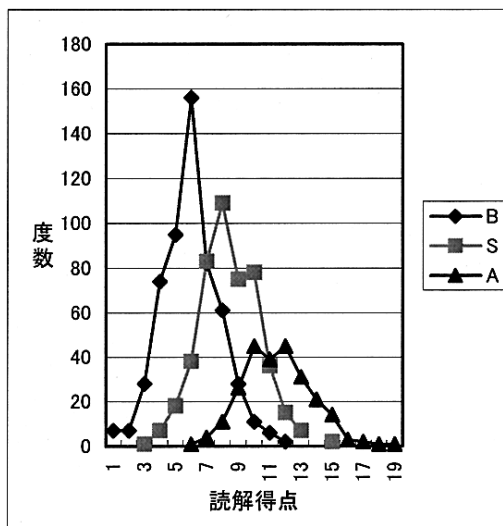
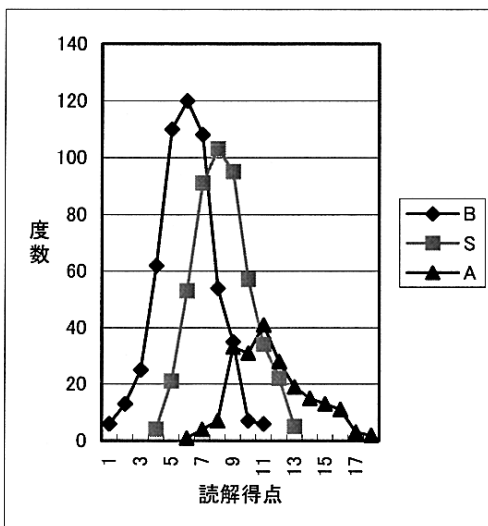


図9 上・中・下の読解得点・度数(2008) 図10 上・中・下の読解得点・度数(2009)

図7・8・9・10が示すように、20%・40%・40%の割合でクラス分けをしても16.67%・33.33%・50%の割合でクラス分けをした場合と同様に、学生の文法・読解技能に合わせた振り分けにはなっていないことが分かる。

本学の1年次の英語授業内容は英語Ⅰと英語Ⅱを履修することになっており、英語Ⅰでは主に読解、英語Ⅱでは主に文法を指導している。英語ⅠのBクラスでは、主に現在時制の短文から構成される長文読解能力の習得を目標とし、英語ⅡのBクラスでは、基礎文法項目であるBe動詞・一般動詞の現在時制の習得を目標としている。しかしながら、上記の結果から、S・Aクラスにも、Bクラスと同じレベルの文法・読解指導を必要としている学生がいることが判明した。

このように、G-TELP (L4) の文法と読解の相関が低いことから、本学ではG-TELP (L4) の文法と読解の得点合計による習熟度別クラス編成の妥当性は低いといえる。

2.3. G-TELP (L4) の文法問題の出題方法の妥当性

次に、本学におけるG-TELP (L4) の文法問題の出題方法の妥当性について考察する。G-TELP (L4) の文法問題は20問であり、「現在進行形」「There 構文」「未来形」「WH 疑問文」にターゲットを絞った問題がそれぞれ5問ずつ割り当てられている。これらの文法項目は中学1年から中学2年の間に学ぶ項目である。2008年、2009年におけるそれぞれの問題の正解率は図11・12・13・14の通りである。

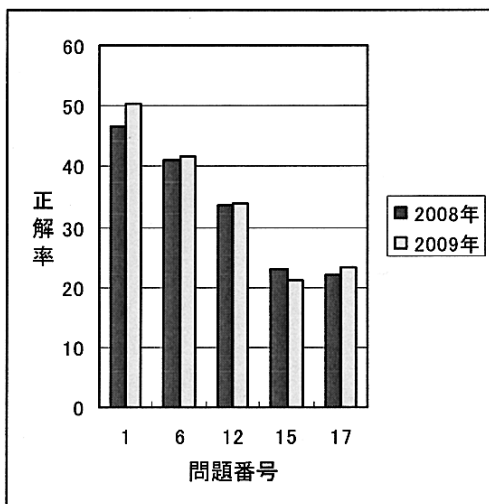


図11 「現在進行形」の出題順別正解率

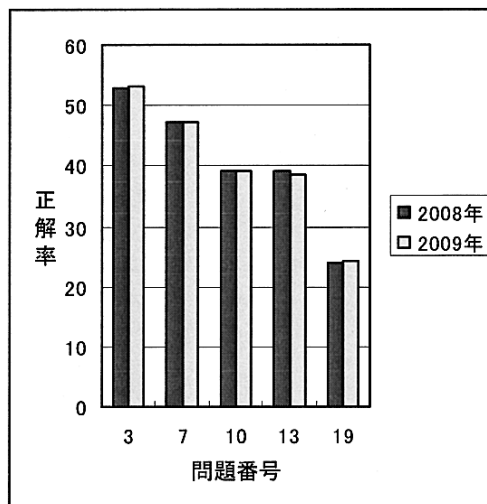


図12 「There 構文」の出題順別正解率

図11・12・13・14が示す通り、2008年度と2009年度の受験者の問題別の正解率にはほとんど差は見られない。興味深いのは、同じ文法項目のターゲット内での問題の正解率に大きなばらつきが観察される点である。その原因として、G-TELP (L4) の文法問題の出題方法の特徴が考えられる(前田、2008)。

G-TELP (L4) の文法問題はすべて2文で構成されており、問題によっては基礎の英文法項目をターゲットとしていたとしても、文全体の意味を把握していないと正答を選ぶこと

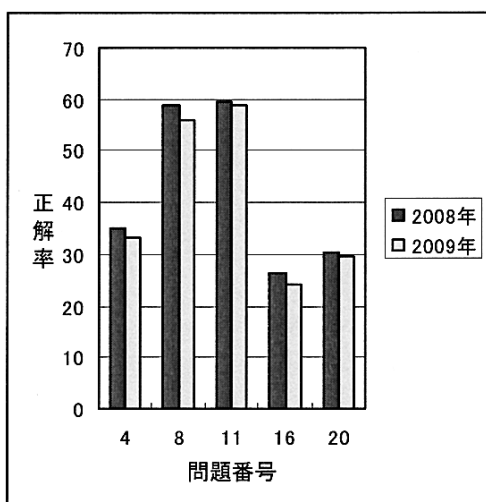


図13 「未来形」の出題順別正解率

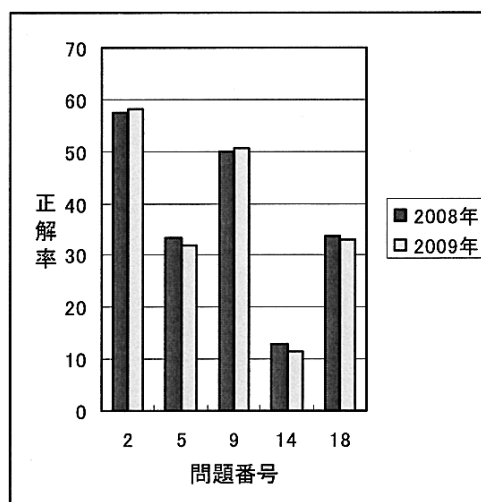


図14 「WH 疑問文」の出題順別正解率

ができない問題が存在するのである。従って語彙力や、読解力も必要とされる場合があるのである。例えば、「現在進行形」をターゲットとした問題では、問題番号15の “ Bill needs some help with the project. He () difficulty with it. ” が該当し、正答率が他の問題と比べて極端に低い(2008年度23.06%、2009年度21.02%)。この問題の過去2年間の点双列相関係数による項目弁別力指数と実質選択肢数、また選択肢別回答率を表7に示す。尚、選択肢Aが正解の「現在進行形」“is having”で、Bは「過去形」“had”、Cは「現在形」“have”、Dは「仮定法過去」“would have”である。

表7 問題番号15「現在進行形」の年度別の弁別力指数・実質選択肢数・選択肢別回答率

実施年度	問題番号	弁別力	実質選択肢数	NA	(A)	B	C	D
2008年度	15	0.23	3.78	9.68	23.06	27.90	22.74	16.61
2009年度	15	0.20	3.73	12.44	21.02	27.40	22.91	16.22

点双列相関係数による項目弁別力指数は大友(2002)によると、一般的には0.30以上であればよい項目とされているが、表7が示すように2008年度・2009年度共に「現在進行形」をターゲットとしている問題番号15の項目弁別力指数は低く、本学のプレースメント・テストとして妥当であるとは言い難い。

実質選択肢数は0.00から選択肢の数までの数値となり、表7からは四者択一の問題としては回答に偏りがなことが分かり、「現在進行形」という中学1年のレベルの文法ターゲットであるにもかかわらず、2008年度・2009年度共に80%近くの受験者が錯乱肢に惑わされていたことが分かる。このような結果となった原因として、次のことが考えられる。前文の意味をよく把握しないまま、“He () difficulty with it.”だけに注目したため、「現在進行形」以外の選択肢が可能と判断されたと思われる。

このように、G-TELP(L4)の文法問題には、問題とされている文の直前の文の意味を正確に把握していないと正答を選ぶことができない問題が含まれているのである。つまり、

上記のような問題の場合、「現在進行形」と「未来形」の文法項目を習熟した上で、英文の内容を正確に把握できる能力を習得した者が正答を選ぶことができる問題となっており、文法項目をターゲットとするだけでなく、英文の理解力も求められていると考えられる。

尚、前文の意味を理解しなくても正答を選ぶことができる問題の例を上げる。問題番号11の“Nancy has to mail a letter. She () to the post office tomorrow.”が該当し、選択肢Aは「未来形」“will go”、Bは「現在形」“go”、Cは「仮定法過去」“would go”、Dは「現在完了形」“has gone”であった。前文を理解しなくても、後の文の最後の“tomorrow”の意味が分かれば「未来形」が正答であることが分かる。この問題の正答率・弁別力は2008年度0.59・0.43、2009年度0.59・0.46であり、難易度も適切で弁別力も高い。このようなタイプの問題は、本学のプレースメント・テストとして有効な問題であると言える。

これまでの調査の結果から、同じ文法項目をターゲットとしている問題の中で全体の正答率が極端に下がり、また、点双列相関係数による項目弁別力指数が0.30を下回っている問題が存在している以上、G-TELP (L4)の文法問題の中でも、英文の理解を必要とするタイプの問題は、本学のプレースメント・テストとしては妥当性が低いと言えよう。

3. 課題と展望

英語の学力格差は今後も広がっていくことが容易に予想される中、習熟度別クラスによる指導は急務である。しかしながら現実問題、それぞれのレベルでどのような指導をするのか、また、何を目標とするのか、に合わせて適切なクラス編成ができていないと、指導も困難となる。

G-TELPは、英語を母国語としない人々が、実際の場面でどれだけ英語でコミュニケーションがとれるか、という力を測るテストである。つまり、英語の文法能力については、どれだけ定着しているのかを正確に測るテストではないことが、今回の調査で判明した。

本学では、教師全員にそれぞれのレベルにあった指導を周知徹底しているが、学生側からの不満が後を断たないのが現状である。原因の1つとしては、今回の調査からも判明したように、本学でのG-TELP (L4)の信頼性・妥当性が共に低い故、学生が適切な習熟度別クラスに配置されるような分類の方法ではなかったことがあげられよう。

今後の課題として、本学でのプレースメント・テストの信頼性・妥当性を高めるには、過去2年間のG-TELP (L4)の結果から判断すると次のことが考えられる。まず、妥当性を高めるには、今回の調査で本学学生の文法力と読解力の相関が低いことが判明したため、英語I・英語IIの習熟度別クラスについては、文法問題・読解問題の合計得点ではなく、各々の得点を基に編成する必要がある。次に、信頼性を高める方法として、文法問題については問題数を増やすことが考えられる。Test Data Analysis Program (TDAP) Ver.2.0 (『テストで言語能力は測れるか』(大友賢二監修/中村洋一著)添付プログラム)を使用し、項目数と信頼性係数の予測値を測定したところ、2008年度・2009年度の文法問題の信頼性係数を0.80以上に上げるためには、それぞれ92問・74問以上の問題数が必要だということが分

かった。しかし、G-TELP (L4) 文法問題の中には、文の意味を理解しないと解けないという本学学生には妥当性が低いと考えられる問題が存在することが今回の調査で判明した。そこで、弁別力が0.30を下回る問題を削除し、さらに予測値を測定すると39問・37問以上でも信頼性係数が0.80以上になることが分かった。このことから、G-TELP (L4) の問題でも、上記のような問題を削除することによって、問題数が40問程度であれば、英語Ⅱの習熟度別クラス編成は可能だと考えられる。

G-TELP (L4) の読解問題に関しては、前回の調査(前田, 2009)からスキニングの技能の問題以外(語彙・推量問題等)のほとんどの問題の弁別力が本学では0.30を下回っていることが判明している。そこで、スキニング問題のみで同ソフトにより項目数と信頼係数の予測値を測定したところ、2008年度・2009年度それぞれ61問・62問以上の問題数にすれば信頼性係数が0.80以上になるという結果となった。このことから、本学学生に対してはスキニング技能を中心とした問題を62問程度使用することで、英語Ⅰの習熟度別クラス編成は可能だと考えられよう。

また、B・S・Aクラスの配分率についても検討する必要があるだろう。本学学生の英語力が低いことや、今後も広がると予想される学力格差から、Bクラスの割合を増やすことも考慮すべきであろう。

このように改善すべき点は多々存在するが、今後も、本学での英語教育をよりいっそう充実したものにするために、本学の習熟度別クラス編成の在り方についての研究をさらに進めていきたい。

参考文献

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press
- Brown, J. D. (2005). *Testing in language programs*. New York, NY: McGraw-Hill
- Corrigan, A., Dobson, B., Kellman, E., Spaan, M., & Tyma, S. (1993). *English placement test*. The Testing and Certification Division, English Language Institute, University of Michigan
- G-TELP ホームページ <http://www.g-telp.jp/>
- 加賀田哲也、小磯かをる、前田和彦(2007)「英語学習についての調査研究」『大阪商業大学論集』第151・152号合併号、pp.597-608
- 岸田幸雄(2005)「習熟度別クラス編成の現場でおこっていること」『教育』712号5月号、2005年、p.44-51
- 国立教育政策研究所(2003)「指導方法の工夫改善による教育効果に関する比較調査研究」
<http://www.nier.go.jp/shochu/seika/sidouhouhou/sidou-gaiyo.htm>
- 前田和彦(2009)「本学におけるブレースメント・テストについての一考察」『大阪商業大学論集』第144号、pp.13-32
- 文部科学省(2007)「教育課程編成・実施状況調査」
<http://www.mext.go.jp/a-menu/shotou/new-cs/1263169.htm>
- 西部真由美(2003)「英語カリキュラムの作成と実践 習熟度別クラス編成の実施と経過報告」『東邦

学誌』第35巻、第2号、2006年12月、pp.107-114

大友賢二(2002)『テストで言語能力は測れるか』桐原書店

佐藤学(2004)『習熟度別指導の何が問題か』岩波ブックレット、2004年

清水裕子(2001)『4年生大学における英語ブレースメント・テスト実施の現状』立命館大学経済学部

杉森幹彦(2003)「英語統一テスト・習熟度別クラス編成・到達目標の設定および測定に関する実態調査の報告」『政策科学』第10巻3号、立命館大学政策科学学会、2003年3月、pp.3-25