

大学入試センター試験の英語問題で測られる学力と リスニング・テストで測定される言語運用能力の 能力推定値を介した関係性の検討

試験問題の事後標準化とリスニング・テストでの
問題提示順序の影響の検討と共に

内田 照久*
中畝菜穂子**
石塚 智一***

要 約[†]

共通受験者法と共通項目法を組み合わせたデザインで、過去の大学入試センター試験（一部に共通第1次学力試験）英語試験問題の標準化を試みた。最も困難なフォームと最も容易なフォームの間での正答率にして6パーセントほどの困難度の差が、これを共通尺度上に等化することにより、両フォームを受験した集団間の θ の平均値の差は0.05ほどになり、困難度の問題は解消した。

同じ受験者集団に英語のリスニングテストも併せて実施した。英語筆記テストとリスニングテストの間の相関は0.65ほどであった。同時に、リスニング問題の出題順序や話者の連続性についても、幾つかの統制を行って実施したが、これらの要因と試験成績の間に関連は見られなかった。

1 問題と目的

1.1 英語試験で測られる学力とリスニング・テストで測定される言語運用能力の関係性

平成18年度の大学入試センター試験から、英語にリスニング・テストが課せられることになった（大学入試センター，2003）。その導入にあたっては、実施方法の検討や、問題作成体制の整備が求

められる一方で、従来型のペーパー・テストによる試験成績との関連性に係わる知見や情報が必要とされている。そこでは、異なるテストの結果を、単一の科目の試験成績として統合してよいかどうかに関する、実証的な基礎データとしての役割に加え聴覚障害者に対するリスニング・テストの代替措置を議論するための資料としても必要とされている。

本稿では、そのような情報の提供のために、研究開発部が、平成11-13年度（1999-2001年度）に

* 大学入試センター 研究開発部 試験環境研究部門

** 大学入試センター 研究開発部 試験作成支援研究部門（現所属：新潟大学 入学センター）

*** 大学入試センター 研究開発部 試験作成支援研究部門
2003年11月27日 受理

実施してきた共同研究、「大学入学者選抜における評価の標準化に関する研究」の下で行われたモニター試験のデータについて、再分析を行うこととする。このモニター試験は、過去の共通第1次学力試験やセンター試験の問題の事後標準化を主な目的として、研究期間の3年間、年に一回、継続的に行われたものである。この研究プロジェクトは大きく分けて、平成11-12年度（1999-2000年度）の第I期と、平成13年度（2001年度）の第II期に分かれる。第I期の結果については、前川・石塚・菊地・内田・中畝（2001）によって報告がなされている。また、第I期と第II期をまとめた結果に関しては、前川・菊地・内田・中畝・石塚（2003）の報告がある。

さて、この第II期には、従来型の英語の過去問題から構成されている試験に加え、英語リスニング・テストにおける問題提示順序の影響を検討する実験も同時に行われている。そこでは、同一の被験者が双方の試験を受験しているので、従来型の英語試験の成績とリスニング・テストでの成績を比較検討できる可能性がある。しかし、ここで行われた英語の試験やリスニング・テストでは、共に所与の実験目的に即して、異なる問題冊子が用いられていたり、問題提示順序が異なっていたりするので、素得点同士での比較はできない。そこで、素得点同士の関係を扱うのではなく、それぞれのテストが測定している潜在能力推定値（ θ ）をもとめ、その間の関係性について検討することとする。

またここでは、これまでに整理されていなかった第II期に実施された問題冊子間の事後標準化の評価、及び、リスニング・テストでの問題提示順序の影響についても検討を行う。

1.2 実施年度間での大学入試センター試験の成績比較の困難性

大学入試センター試験は、その前身である共通第1次学力試験まで含めると、1979年（昭和54年）の第1回目の実施から2004年（平成16年）までに26回が行われ、四半世紀あまりの歴史を刻んでいる。

この共通第1次学力試験や大学入試センター試験といった共通テストでは、年度間での試験成績の比較が困難である。これらの試験では、同一の教科・科目の成績であっても、前年度と当該年度の

試験の難易度に差がある場合には、年度間の得点は相互比較できない。さらに近年、平成15年度までは、18才人口が減少しているにも関わらず、センター試験受験者は増加していた。このことは、従来ならばセンター試験を利用しなかったと思われる生徒層の受験による現象であるとする、受験者集団は年々刻々と変化していることになる。その場合、試験成績の年度間での比較のために、集団内での相対的な位置に基づいた偏差値を用いたとしても、年度ごとに規準集団そのものが変容しているために、その比較は困難である。

また、教科内での選択科目間の得点差は、いわゆる得点調整の問題として、依然、理想的な解決策は見出されておらず、センター試験以降のア・ラ・カルト方式の導入も相まっていっそう複雑さを増し、今なお検討課題である（村上、1998）。

このような共通テストにおける試験成績の取り扱いの勝手の悪さは、その試験の設計時の思想、実施運営上の制約により、必然的に生ずるものである。この共通テストには、開発当初の大学入試状況が反映されており、年1回の試験実施が念頭におかれ、常に初出の問題から構成されることが前提になっている。またこの共通テストでは、受験者の便宜を計るために自己採点方式が導入されているため、全試験問題が即日公開される。使用された試験問題が全て公表されてしまうので、当然のことながら同じ問題は二度と使うことができず、同一条件での試験を再実施することが現実的に不可能となっている。この点は、アメリカの大学入試であるSAT（Scholastic Assessment Test）やAAP（ACT Assessment Program）などとは、テスト運営方式そのものが全く異なる。

しかしながら、今後は、試験成績の複数年度に亘る利用、センター試験の年複数回実施の可能性をにらんで、大学入試センター試験においても、異なるテストが用いられた場合でも成績の比較を可能にすることがもとめられる。そのためには、試験成績の標準化を視野に入れた実証的な研究の蓄積が必要とされる。

これまでに、試験成績の標準化に関連するものとして、鈴木・前川（1988）や前川・鈴木（1988）、また石塚・中畝・内田・前川（2001）が、共通第1次学力試験や大学入試センター試験のデータに対して項目反応理論（Item Response Theory, IRT）

を適用した研究を行っている。しかし、それらはいずれも単年度ごとでの試験結果の分析に留まっており、年度をまたいだ試験問題を総合的な対象とはしていない。なお、この項目反応理論については、芝（1991）や渡辺・野口（1999）、Lord & Novick（1968）やハンプルトン（1992）に詳しい。

さて最近になって、前川他（2003）は、過去に出題された問題群を同一の被験者に解答させるモニター試験形式の実験を行い、共通テストの試験成績の標準化を視野に入れた実証的な研究を試みている。

本稿では、その共通受験者法の手続きを用いて収集されたデータの中のサブセット、すなわち第II部の平成13年度（2001年度）に集められたデータの再分析を行うこととした。今回は、モニター試験受験者のみを対象として、異なる年度に実施された問題項目の困難度（*b*）や識別力（*a*）を推定し、同一の共通尺度上に整理する。その上で、それらの統計指標に基づいてテスト冊子間の等化を行った結果として、その補正機能が有効に働いているかどうかについての、その評価に主眼をおいて分析を行う。

1.3 英語リスニング・テストにおける問題提示順序の影響に関する検討

2004年（平成16年）現在、これまでのところ大学入試センター試験では、60万人規模の一斉試験での実施上の難しさから、外国語リスニング・テストは行われてこなかった。しかし、昨今の国際的なグローバル化の潮流の中で、センター試験にも英語リスニング・テストを導入すべきであるという意見は根強く（文部科学省 英語教育改革に関する懇談会，2002）、ついに平成18年度から、大学入試センター試験に英語リスニング・テストが導入されることとなった（大学入試センター，2003）。

一斉実施のリスニング・テストでは、実施にあたってCDやカセット・テープなどを利用することが多い。このような機器を使用する場合、音声は連続再生するのが基本となるので、必然的に問題項目の提示順序や解答順序は系時的に固定されたものになる。このことは、いわゆる筆記試験型のペーパー・テストのように、問題の解答順序が受験者の裁量に大きく任されているものとは大きく

異なる。なお、平成15年11月に、センター試験のリスニング・テストでの個別音源の使用が発表されたが、問題の提示と解答が系時的に固定されることには変わりはない。

この問題の提示順序や解答順序は、ペーパー・テストにおいては、その問題項目ごとの成績に影響を与えることが報告されている（石塚・中畝・内田・前川・斎藤・高橋・Ross・清水・根岸・野澤・杉野，2003；Sugino, Saito, Takahashi, Shimizu, Negishi, Nozawa, Ishizuka, Uchida, & Mayekawa, 2003）。そこでは、入試水準の英語の多肢選択型テストで、問題の提示順序が試験成績にどのような影響を与えるかを検討している。そこでの実験では、提示順序の異なる2種類の問題冊子を準備し、試験時の制限時間を遵守した条件で受験させている。その結果、テスト全体の平均点や標準偏差については冊子間に差が見られないものの、それぞれ後部に配置した問題項目の成績が低下する傾向があり、無答の増加が観測された。このことは、現実のテスト場面では、問題項目の項目通過率はその問題に固有の困難度によってのみ定まるのではなく、テスト全体の中での位置からも影響を受けることを如実に示している。

異なる試験間の成績比較を考える場合、それぞれのテストの中に共通の問題項目を一部分に配置し、そこで得られる統計的な情報を基に、複数のテスト間の等化を行って成績を比較する方法がある。このような共通項目法による試験成績の等化にあたっては、その共通のアンカー項目が、異なるテストの中にあっても同一の性質を備えていることが前提となる。しかし、出題順序が問題項目の困難度に影響を及ぼしている場合はその前提を満たせず、そのような状況下ではテストの等化を行うことはできない。このような場合には、共通アンカー項目の出題位置については十分な配慮が必要であることがわかる。

さて、リスニング・テストでは、テスト内での問題の出題順序は成績に影響を及ぼすのだろうか。先に触れたように、一般的なリスニング・テストでの問題の解答順序は、ペーパー・テストに比べてより固定的である。したがって、もし出題順序の影響があるとする、受験者の自由裁量部分が少ないため、その影響がいっそう顕在化することも考えられる。

このようなリスニング・テストにおいて、問題項目の出題順序の影響が存在するかを吟味するためには、実証的な実験を行う必要がある。テストを解き進める中で生じる練習効果や、逆に疲労などによって起こる影響などに関して、現実的なテスト場面での検討が必要となる。さらに、音声問題の進行と共に、受験者がその話者の声質や話し方に慣れていくことによって、同じ問題であっても、はじめの方で提示された場合より、後ろの方で提示された場合の方が解答し易いといった可能性についても検討しておく必要がある。このような問いに対して、経験的なデータを蓄積できれば、リスニング・テストにおいて、共通アンカー項目を用いたテストの等化を行う場合にも、その基礎資料として活用できると考えられる。

1.4 多面的情報を引き出すための複数の実験を系統的に組合せた研究デザイン

教育や学力の測定評価に関わる調査研究では、参加する被験者にかかなりの負担がかかる。そのため、研究目的ごとに個々に調査を計画し、やみくもに実験を繰り返すことは望ましくない。したがって、複数の目的を持つ、異なる実験であっても、それらを系統的に組合せることによって、実験実施の回数を最小限に抑えることに配慮することが研究計画の段階で重要となる。

また、教育に関わる研究では、その営みそのものにも教育的な貢献がもとめられる。したがって、研究活動の中では、研究者のための知見の確保に留まらず、被験者にとっても有用な情報を提供できるように配慮して研究をデザインすることが必要となる(内田・菊地・中畝・前川・石塚, 2002)。

本研究では、実験によって得られる、限られたデータから、できる限り多面的な情報を引き出すため、複数の異なる実験を系統的に組合せた研究デザインを用いる。具体的には、まず目的として、1) 大学入試センター試験の英語問題の事後標準化による補正機能の評価、2) 英語リスニング・テストにおける問題提示順序の影響の測定、3) 従来型のペーパー・テストで測られる英語学力とリスニング・テストで測定される英語運用能力の関係性の検討、の3点を据える。そして、1) 2) の目的達成のために、まずそれぞれの実験ごとの目的に沿った分析を実施する。その上で、実験ごとに測定値

の背後にある被験者の潜在能力特性値(θ)について推定を行う。そして、1) 2) の実験とも同一の被験者が参加している利点を生かすことにより、各被験者のそれぞれの能力推定値を資料として、3) の目的である能力間の関係性の検討を行うこととする。

2 実験1：英語試験問題の事後標準化のためのモニター実験

大学入試センターが実施するテストでは、年度ごとに初出の問題を使用しており、年度間ではテストに難易差が生じ得るため、年度を越えた成績の相互比較は困難である。この問題に対処するため、項目反応理論(IRT)に基づいた尺度化を試みる。共通受験者法と共通項目法を組合せたモニター実験を実施し、複数年度の試験問題の項目パラメタ推定と水平的等化を行う。その上で、事後標準化の補正機能について評価する。

なお、以降の実験実施のあらましは、前川他(2003)の研究中の第II部と共通である。

2.1 実験1の被験者

3つの国立大学(S大学・T大学・N大学)の文科系・理科系の1年生が参加した。男性444名、女性137名の計581名(18-23才)であった。文科系と理科系の受験者の割合はほぼ半数ずつである。実験参加者には、終了後に謝金が支払われた。実験は大学ごとに行なわれ、実施日時は、S大学が2001年12月1日、T大学が2001年12月8-9日、そしてN大学が2001年12月15-16日であった。なお、1日あたりの被験者数は約100名ずつであった。

2.2 実験1の実験用ペーパー・テスト冊子

大学入試センター試験で過去に実施したテストから、ある程度難易度が高く、識別力も高い問題を大問単位で2-3題ずつ選び出した。そして、複数年度の問題が混在した、異なる問題構成から成る10種類のテスト冊子を作成した。各冊子は、他の2つの冊子と相互に共通の問題が配置され、その共通問題部分が冊子間で循環的に重なるように、テスト冊子全体を設計した(図1)。また冊子中での大問の出題順序の効果を排除するため、各冊子

冊子 1	S63	H8	H5							
冊子 2		H8	H5	H2						
冊子 3			H5	H2	H9					
冊子 4				H2	H9	H6				
冊子 5					H9	H6	H3			
冊子 6						H6	H3	H10		
冊子 7							H3	H10	H7	
冊子 8								H10	H7	H4
冊子 9	S63								H7	H4
冊子 10	S63	H8								H4

冊子	版 1-1	S63-4,6,7	H8-3A,3B,5	H5-3A,3B,5
	版 1-2	H8-3A,3B,5	H5-3A,3B,5	S63-4,6,7
1	版 1-3	H5-3A,3B,5	S63-4,6,7	H8-3A,3B,5

注. S：昭和，H：平成，その後の数字が実施年度を示し，ハイフン以降は大問番号をあらわす

図 1 複数年度分の英語問題のテスト冊子への割り当て

から3つの版を作成し，それぞれで各年度に属する大問の出現場所を変化させた．したがって，各年度に属する大問は，問題冊子の前半，中盤，後半に同頻度で出現することとなった．なお，例として，冊子1の版1-1をAppendixに示す．

2.3 実験1の手続き

多肢選択形式の問題で，従来型のペーパー・テストの形式で実施した．なお，試験時間は90分であった．なお，この試験時間は，全ての問題を解答するのに十分であることを，あらかじめ事前の予備実験で確認している．

また被験者は，大学ごと，さらに文科・理科系別に層化した上で，各冊子をランダムに割り当てた．したがって，冊子ごとに，ほぼ等質と考えられる被験者群を配置したことになる．

2.4 実験1の分析と結果

各冊子の合計得点を大学別に標準化し，得点が平均から2標準偏差分より低い受験者について，問題用紙と答案用紙の状況を確認した．また，当日の試験監督者の報告書に記載された受験態度も考慮し，本来の実力通りに試験を受けていないと判断された受験者に関してはデータ解析の対象外とした．

次に，10種類のテスト冊子に含まれる全ての問題項目をまとめて，全冊子を統合した1つの大きなテストと見なし，IRTを用いて分析した．分析

にあたっては，大問の中に含まれる各小問を一つ一つの問題項目として扱った．各項目は，大問単位でそれぞれのテスト冊子に割り振られているので，被験者×項目のデータ行列としては，構造的に欠損値が生ずる形式である．したがって全体としては，共通受験者法と共通項目法を併用した形で項目間の等化がなされる．具体的な手順は，内田他（2002）で用いられている方法とほぼ同様である．

分析にあたっては，Zimowski, Muraki, Mislevy, & Bock (1996) によるBILOG-MGを用いた．被験者は各大学ごとに文科系，理科系に層化した上で各テスト冊子にランダムに割り当てられているので，全体では能力的に等質な1つの集団からのサンプルと考えることができる．そこで，分析では単一集団が受験したものとして扱った．ここでは2パラメタ・ロジスティック・モデルを採用した．項目困難度 (b) に正規分布，項目識別力 (a) はその対数値に正規分布，また能力推定値 (θ) には正規分布を，それぞれ事前分布として仮定した条件で，パラメタの推定を行った．尺度係数 D は1.7を使用し，周辺最尤推定によって項目困難度 (b)，項目識別力 (a) をもとめた．各パラメタの推定条件の詳細については，Zimowski et al. (1996) を参照されたい．

次に，分析を通じて，異なる冊子間での試験成績が比較できるようになったかどうかについて，被験者の成績を指標として吟味した．まず単純に冊

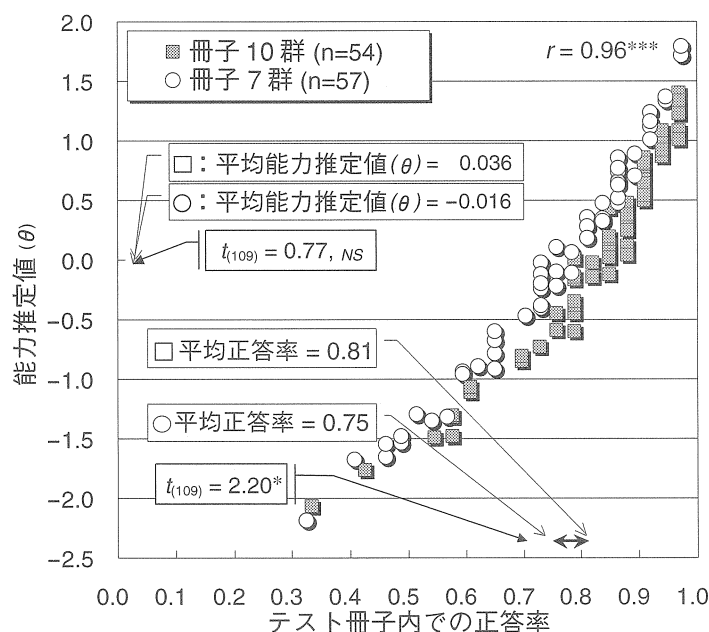


図2 正答率が最も異なる冊子間での被験者の正答率と能力推定値 (θ) の関係

子ごとでの正答率で比較した場合、冊子10を解いた受験者群の平均正答率が最も高く(81%)、冊子7の受験者群の平均正答率が最も低かった(75%)。被験者は各冊子にランダムに割り当てられているので、それぞれの冊子の受験者はほぼ等質なものと考えられる。したがって、この平均正答率の差(6ポイント)は、問題の難易度の相違に起因するものと想定される。実際、この両者の冊子の間には共通問題が存在せず、全く異なる問題構成から成るテストになっており、テストそのものの難易差によって平均正答率に違いが生じているという解釈が整合的であると考えられる。

ここで、各被験者の冊子ごとでの正答率とIRTによって推定した能力推定値(θ)の散布図を示す(図2)。図中のプロットは、各被験者の成績を示したものである。横軸は受験したテスト冊子の中での単純な正答率を表す。一方、縦軸はIRTを用いて推定された能力推定値(θ)が示されている。図から、単純な平均正答率の観点では低い値を示した冊子7を受験した者の成績が、縦軸の能力推定値(θ)では、たとえ同等の単純正答率であっても、冊子10を受験した者より高い値を与えられていることが読み取れる。群間で比較すると、冊子10群と冊子7群の平均正答率では6ポイント相当分の

難易度の違いが観測されており、この指標では両群に有意な差が見られる($t_{(109)} = 2.20, p < .05$)。しかし、能力推定値(θ)では、その平均差は解消し($t_{(109)} = 0.77, N.S.$)、両群がほぼ等質である様子が見て取れる。また、正答率と能力推定値(θ)は、線形関係と見なした場合でも相関係数は0.96と高く、能力推定値(θ)は、正答率とよい対応を保持した成績の指標としても十分な説得力を持つと考えられる。

このように、冊子間に全く共通の問題が配置されていないテストの間で、さらに平均正答率をもっとも離れたもの同士であるにも関わらず、能力推定値(θ)では、正答率との関係を保持しながら、その難易度の違いが解消されていることがわかった。したがって、今回の分析を通じて、共通受験者法と共通項目法を組合せて収集したデータに対して、IRTを用いた得点等化を行うことによって、異なる冊子間での良好な補正機能が提供できることが確認できたと考えてよい。

以上、共通受験者法と共通項目法を組合せたデザインモニター実験を通じてデータを収集し、IRTに基づいた分析を行うことにより、複数年度間の事後標準化を一部実現することができた。なお、項目パラメタをはじめ、各種統計量について

実験ブロック1 [問題提示位置効果の検討]

	提示位置①	提示位置②	提示位置③	提示位置④	提示位置⑤
冊子1	A	B	C	D	G
冊子2	B	C	D	A	G
冊子3	C	D	A	B	G
冊子4	D	A	B	C	G
小計1-4	A-D	A-D	A-D	A-D	G

注. A-D, 及び G は, いずれも異なる 7 つの小問からなる項目セット

実験ブロック2 [同一話者の連続効果の検討]

	提示位置①	提示位置②	提示位置③	提示位置④	提示位置⑤
冊子1	P1	P2	Q1	Q2	G
冊子2	Q1	Q2	P1	P2	G
冊子3	Q1	P2	P1	Q2	G
冊子4	P1	Q2	Q1	P2	G
小計1-2	P1-Q1: 初出	P2-Q2: 連続	P1-Q1: 不連続	P2-Q2: 連続	G
小計3-4	P1-Q1: 初出	P2-Q2: 不連続	P1-Q1: 連続	P2-Q2: 不連続	G

注. P1-2, Q1-2, 及び G は, いずれも異なる 7 つの小問からなる項目セット

P セットと Q セットは異なる話者セットによる出題

図3 リスニング問題項目の実験ブロックごとのテスト冊子への割り当て

は, 大学入試センター研究開発部 (2002) を参照されたい。

3 実験2: リスニング・テストの問題提示順序の影響を検討するための実験

多くのテストでは, 問題の提示順序が項目の正答率などに影響を与える。解答順序や解答時間が実施手続きの上で拘束されるリスニング・テストにおいて, 問題の提示順序, 同一話者による発話の連続が項目特性に与える影響を検討する。

3.1 実験2の被験者

実験1と共通である。

3.2 実験2の実験用リスニング・テスト冊子

英検準2級-準1級用のリスニング・テスト練習問題を使用し, 会話問題と説明文問題の2種類の問題形式を採用した (旺文社エディタ, 1997)。その利用にあたっては, 編集部の許諾を受けた。会話問題は, 2人の間で交わされる英会話のスク립トとそれに関する短い質問文が音声で提示された後, 答えとして適切なものを4つの選択肢文章から選ぶ形式である。説明文問題も同様に, 英語

の単独発話のスク립トとそれに関わる質問文の音声提示後に, 選択肢を選ぶ形式である。

問題の選定には, 英語の教員免許を持つ者があつた。会話文問題では, 難易度の低い問題では, 会話の自然な流れの中で, 会話の展開の理解に必要な 5W1H が的確に捉えられているかを確認するもの, 難易度の高い問題では, それに加えてその後の展開や背後関係を類推することができるかを確認するような問題を選定した。また, 説明文問題では, 描写されている内容に関して, 場面状況の理解に必要な 5W1H が的確に捉えられているかを確認するような問いが選ばれた。

実験内に2つの実験ブロックを配置した (図3)。実験ブロック1では, 問題の提示順序を循環的に操作した4つのテスト冊子を作成した。提示位置ごとに全ての冊子についての小計をまとめると, それぞれの位置ごとに A-D の問題項目セット (各7項目) が全て提示されるようになっている。各冊子に等質な被験者を割り当てれば, 提示位置の効果を各位置での平均点の違いの形で取り出すことができる。また, 被験者側から見ると, 出題順序が異なるだけで A-D セットの全ての問題項目に解答することになる。

次に実験ブロック2では, 同一話者の問題が続

く条件と、別の話者による問題が提示される条件が、それぞれ対比するように問題項目セットを配置した4つのテスト冊子を準備した。実験ブロック1と2をまとめた全8冊子について、冊子間、及びブロック間の比較を可能にするため、最終位置に共通アンカー項目（Gセット：7項目）を配置した。各冊子は計35項目で構成された。

実験条件に沿って作成した音声データを、上記の8種類のテスト冊子用に配置し、実験用のコンパクト・ディスク・レコーダブル（Compact Disk Recordable, CD-R）を作成した。CD-Rの作成にあたっては、音楽用CD-DAフォーマットで記録し、この8種のCD-Rを計128枚準備した。なお、原音声は、英検準2級・準1級用のリスニング・テスト練習問題CDから、デジタル・データのまま直接抽出したものである。

3.3 実験2の手続き

被験者に携帯型CDプレーヤ（AIWA, XP-V30）を配布して実験を行った。被験者は、イヤークラスから提示される問題を聞き、テスト冊子に解答を記入していった。

テストの各問題の音声提示は1回限りであった。解答時間は10秒で、テスト全体での試験時間は約30分であった。なお、大学入試センター試験では、ICプレーヤが用いられる予定であるので、本実験で使用した機器とは実際は異なるものになる。

3.4 実験2の分析と結果

実験ブロック1でのリスニング・テスト問題の提示位置の影響を整理するため、まず出題位置ごとに正答項目数の小計をまとめた。さらに、リスニング・テストの総得点で各被験者を5段階に群分けし、それぞれの群別に提示位置ごとの正答項目数の平均を示した（図4）。

提示位置ごとにA-Dセット全ての問題が提示されているので、その平均値の違いが提示位置の効果をあらわすのであるが、いずれの学力層の被験者群においても、問題項目の提示位置の影響は特に観測されなかった。

また、実験ブロック2の同一話者の連続効果についても、実験ブロック1同様に整理した。その結果、同一の話者セットが連続する場合と、そうでない場合との間には顕著な違いは見出されなかつ

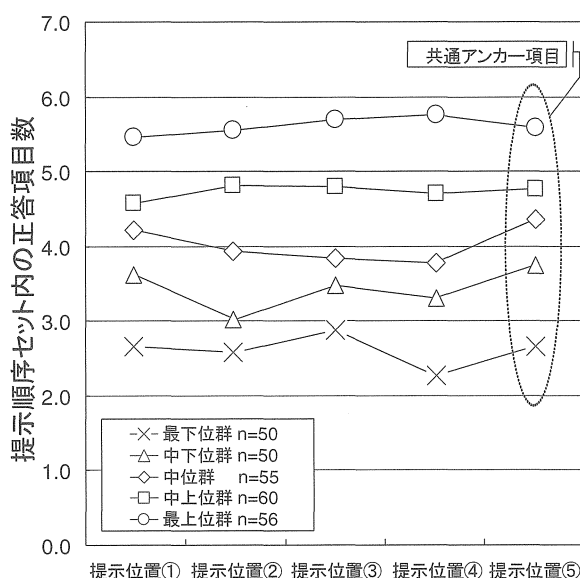


図4 問題提示位置と成績レベル別の平均正答項目数

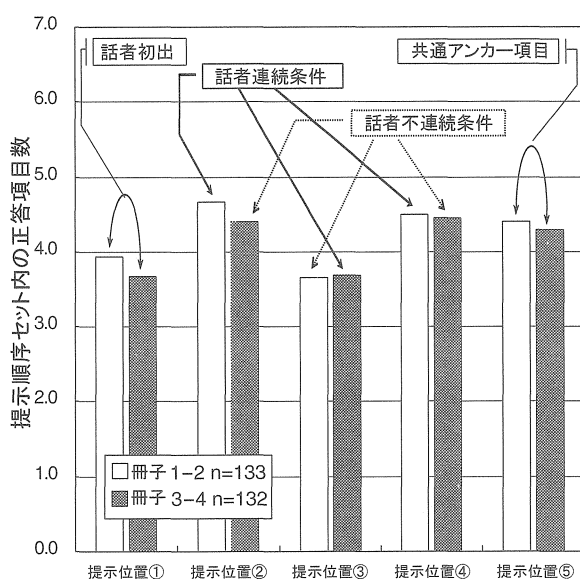


図5 問題提示時の話者の連続・不連続条件別の平均正答項目数

た（図5）。

これらの結果から、少なくとも本実験の環境下では、リスニング・テストでの問題項目の提示順序は、試験成績に対して必ずしも系統的影響は与えないものと判断された。

このリスニング・テストでは、問題項目ごとの解答時間に厳しい制約があり、ペーパー・テストなどと違って、特定の問題に集中的に時間を割くといったことができない。その一方、総試験時間が

表 1 筆記試験・リスニング・テスト間の正答率と θ を指標とした相関係数 ($n = 523$)

	筆記試験 正答率	筆記試験 能力推定値 (θ)	リスニング 正答率	リスニング 能力推定値 (θ)
筆記試験正答率	1.000			
筆記試験能力推定値 (θ)	0.960	1.000		
リスニング正答率	0.613	0.637	1.000	
リスニング能力推定値 (θ)	0.615	0.647	0.978	1.000

足りなくなつて、最後の方の問題には手をつけられないということは原理的に生じない。したがって、今回の実験で問題提示順序の影響が観測されなかった理由としては、このような各問題項目の解答時の時間面での独立性の高さを指摘できよう。

今回、リスニング・テストの実験の結果、問題提示順序の影響は見られなかった。今後のテスト設計において、共通アンカー項目の位置をどこにすべきかといった議論では、問題解答時の時間面での独立性が十分に高いのならば、その位置についてはそれほど神経質になる必要はないという示唆が得られたものと言えよう。

さて、実験ブロック 1 と実験ブロック 2 では、最後の共通アンカー項目を除いて、異なる問題構成になっている。このままでは、両実験ブロックの被験者の学力については相互比較や検討ができない。そこで、先の共通アンカー項目の情報を生かして、両実験ブロックの試験の等化を行い、全被験者の能力についての情報を得ることとした。そこで、先の実験 1 と同様に、内田他 (2002) の手続きに準じて IRT を用い、2 パラメタ・モデルで分析を行った。その分析を通じて、実験 2 に参加した全被験者のリスニング・テストにおける能力推定値 (θ) を算出した。

4 ペーパー・テストで測定された英語学力とリスニング・テストで測られた英語運用能力に関する分析結果

実験 1, 2 について、被験者の成績を整理した。冊子ごとでの正答率、及び能力推定値 (θ) を指標とした際の相互の相関係数を表 1 に示す。

筆記試験とリスニング・テストでの試験成績の関係は、共に能力推定値 (θ) を指標とした場合の相関が最も高い。この θ は、各冊子間の難易差を補正した推定値であるので、本研究のデザインの

下では、両者の能力の関係性はこの指標に基づいた解釈が適切であると考えられる。そこで、この θ を指標とした散布図を図 6 に示す。

この筆記試験での成績とリスニング・テストでの成績との相関 $r = 0.647$ の値について検討する。藤森 (2002) は、局所独立を満たさない項目に関する部分得点処理に関する研究の中で、英語のリーディング分野とリスニング分野の成績の相関は 0.65 程度であるとしている。今回の結果は、その報告とほぼ同等の値を示しており、いわゆる読解を中心としたペーパー・テストで測定する英語学力とリスニング・テストで測られる英語運用能力の関係は、相関で 0.65 程度を目安にすることができよう。厳密には、相関係数の比較は、比較する両群の学力の平均値と標準偏差が等しくないと、適切に比較できないのであるが、今回、一応の目安としては有効であるものと考ええる。

さて、大学入試センター試験の本試験と追試験で出題された両方の英語の筆記試験を、大学 1 年生の被験者に受験させたところ、その相関は 0.654 であった旨の報告がある (椎名・石塚, 2003)。この値は、被験者が既に国立大学に合格している者なので、全般的に平均点が高く、いずれの試験でも天井効果が生じている可能性もあり、いささか低めに見積もられているきらいがある。しかし、その調査自体は再テスト法の手続きに則った実測値として一定の意味があろう。また、現実の大学入試での英語の共通 1 次学力試験と個別大学の 2 次試験での成績の相関については、学部や年度によって必然的にばらつきがあるものの 0.41–0.86 の範囲であったとする報告がある (国立大学入学者選抜研究連絡協議会, 1983)。

このような具体例を勘案すると、英語のペーパー・テストでの成績とリスニング・テストでの成績の関係は、ペーパー・テストで再テストを行ったような場合や、入試における 1 次試験と 2 次試

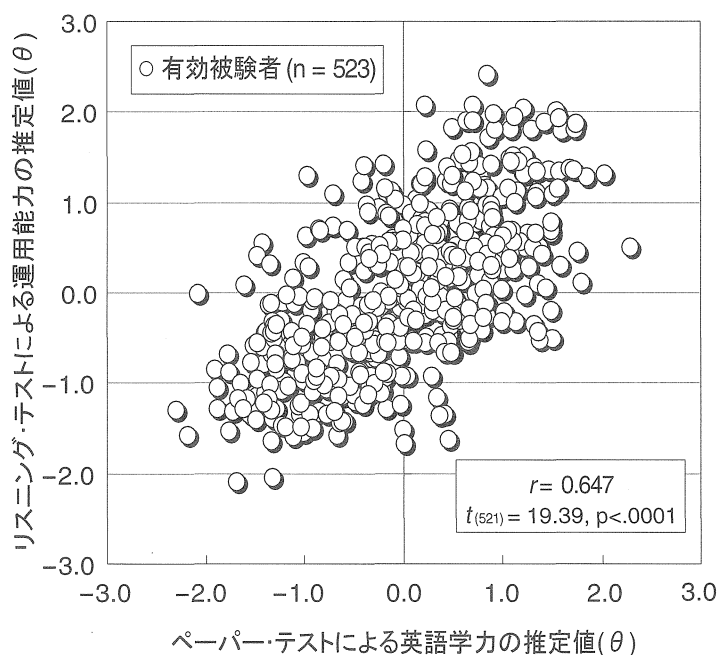


図6 ペーパー・テスト成績とリスニング・テスト成績の能力推定値(θ)を指標とした散布図

験のように反復測定に類する形で試験を行った場合での関係の強さと、ほぼ同等であるとも考えられる。このことは、聴覚障害者に対して、リスニング・テストの代わりにペーパー・テストを用いて試験を行うことを検討するような場合、その代替利用の精度面での目安となろう。

韓国の大学入試であるところの韓国教育課程評価院(KICE)が実施する大学修学能力試験の英語リスニング・テストでは、聴覚障害者にはリスニング・テストで用いられている内容を文字の形で提示して代替テストとして用いている(内田・林, 2004)。一方、大学入試センター試験においては、現在のところ、聴覚障害者についての特別措置の方法は検討中である。本研究の成果は、その議論を深める上での基礎資料の一つになるものと考えられる。

さて、今回のリスニング・テストの実験では英語検定試験の練習問題を用いているが、大学入試センター試験に導入されるリスニング・テストについては、その内容や難易度、問題形式や構成については、まだ明らかにされていない。実際に用いられる問題のあり方によっては、新しいリスニング・テストによって測定される能力が、本研究で扱ったものと、幾分か異なる側面の能力である

ことも想定される。今後、センター試験のリスニング・テストをめぐる動向に即して、それを考慮に入れた研究の継続がもてめられる。

なお、本実験の被験者には、試験の得点通知を行っている。そこでは、単純な正答率だけでなく、得点等化後の尺度得点を用いて成績を知らせている。このことは、内田他(2002)が提唱する、研究活動の中で、研究者のみならず、被験者にとっても有用な情報を提供できるよう配慮した研究デザインを採用し、それを実践した活動の一つとしても位置づけることができよう。

5 まとめ

1. 大学入試センター試験での年度間での試験成績の相互比較を目的として、項目反応理論に基づいた尺度化による事後標準化を試みた。共通受験者法と共通項目法を組合せたモニター実験を実施し、複数年度の英語問題の項目パラメータ推定と水平的等化を行った。実験に参加した被験者の能力推定値を算出して検討したところ、難易度の異なるテスト冊子間の結果についても良好に補正できることが示された。

2. 解答順序や解答時間が実施手続きの上で拘束

されるリスニング・テストにおいて、問題の提示位置、話者の連続性が試験成績に与える影響を検討した。実験の結果、リスニング・テストにおける問題提示順序は、項目特性には必ずしも著しい影響は与えていなかった。

3. 聴覚障害者に対して、リスニング・テストの代わりの代替的な試験を行うことなどを検討するための基礎資料を得るため、ペーパー・テストでの成績とリスニング・テストでの成績について、それぞれの能力推定値を用いて関係性を分析した。その結果、ペーパー・テスト成績とリスニング・テスト成績の関係の強さは、筆記試験での再テスト時の成績や、入試の1次試験と2次試験での成績の関係の強さと、ほぼ同等であることが見出された。

参考文献

- 大学入試センター (2003). 平成 18 年度からの大学入試センター試験の出題教科・科目等について—最終まとめ— (平成 15 年 6 月 4 日).
- 大学入試センター研究開発部 (2002). 「平成 11 年–13 年度大学入学選抜における評価の標準化に関する研究」共同研究報告書.
- 藤森 進 (2002). Local dependence と部分得点処理 日本教育心理学会第 44 回総会発表論文集, 543.
- ハンブルトン, R.K. 野口裕之 (訳) (1992). 項目応答理論の基礎と応用 リン R.L., (編) 教育測定学第 3 版 (上巻) C.S.L. 学習評価研究所・みくに出版 Pp. 211–282. (Hambleton, R.K. (1989). Principles and selected applications of Item Response Theory. In Linn, R.L., (Ed.) *Educational Measurement*. 3rd ed. The American council on education/Macmillan series on higher education).
- 石塚智一・中畝菜穂子・内田照久・前川眞一 (2001). テストレットモデルによる英語試験問題の分析 大学入試センター研究紀要, No.30, 21–35.
- 石塚智一・中畝菜穂子・内田照久・前川眞一・斎藤栄二・高橋貞雄・Steven Ross・清水裕子・根岸雅史・野澤 健・杉野直樹 (2003). 英語試験問題における出題順序の効果について 大学入試研究ジャーナル, No.13, 73–76.
- 国立大学入学選抜研究連絡協議会 (1983). 国立大学入学選抜研究連絡協議会研究報告書 第 4 号.
- Lord, F.M., & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley.
- 前川眞一・石塚智一・菊地賢一・内田照久・中畝菜穂子 (2001). 大学入試センター試験得点の標準化の試み 大学入試研究ジャーナル, No.11, 15–23.
- 前川眞一・菊地賢一・内田照久・中畝菜穂子・石塚智一 (2003). 大学入試センター試験得点の標準化の試み—項目反応理論による方法— 大学入試研究ジャーナル, No.13, 81–87.
- 前川眞一・鈴木規夫 (1988). 独立に推定されたテスト項目パラメータの共通尺度上での等化法—共通第 1 次学力試験の選択科目の差異の分析への適用— 大学入試センター研究紀要, No.17, 249–272.
- 文部科学省 英語教育改革に関する懇談会 (2002). 「『英語が使える日本人』の育成のための戦略構想」の策定について (平成 14 年 7 月 12 日).
- 村上 隆 (1998). 得点調整における公平性の概念—線形等化法における複数の基準の可能性— 大学入試研究ジャーナル, No.8, 41–46.
- 旺文社エディタ (1997). 英検準 1 級～準 2 級 一次試験対策 CD 旺文社.
- 芝 祐順 (編) (1991). 項目反応理論—基礎と応用— 東京大学出版会.
- 椎名久美子・石塚智一 (2003). 「平成 15 年度大学入試センター試験の本試験と追試験の難易度比較に関する調査研究と標準化の試み」研究報告書, 大学入試センター研究開発部.
- Sugino, N., Saito, E., Takahashi, S., Shimizu, Y., Negishi, M., Nozawa, T., Ishizuka, T., Uchida, T., & Mayekawa, S. (2003). The influence of the test taking strategies on the reliability of test items in English tests. *Annual Review of English Language Education in Japan*, 14, 221–230.
- 鈴木規夫・前川眞一 (1988). 昭和 62 年度共通第 1 次学力試験「理科」の各科目に関する項目反応理論に基づくテスト特性の比較 大学入試センター研究紀要, No.17, 219–248.
- 内田照久・菊地賢一・中畝菜穂子・前川眞一・石塚智一 (2002). 英語リスニング・テストにおける音声の時間構造と提示情報の様式が項目特性に与える影響 教育心理学研究, 50 (1), 1–11.
- 内田照久・林 篤裕 (2004) 韓国教育課程評価院 (KICE) の大学修学能力試験における英語リスニング・テスト 大学入試センター第 240 回研究開発部セミナー資料 (平成 16 年 2 月 25 日).
- 渡辺直登・野口裕之 (1999). 組織心理測定論 白桃書房.
- Zimowski, M.F., Muraki, E., Mislevy, R.J., & Bock, R.D. (1996). *BILOG-MGTM: Multiple-*

group IRT analysis and test maintenance for binary items. Chicago IL: Scientific Software International.

付 記

本研究は、大学入試センター研究開発部の平成11-13年度共同研究「大学入学者選抜における評価の標準化に関する研究」の援助を受けて行われました。

なお本研究の一部は、国立大学入学者選抜研究連絡協議会第23回大会、及び日本教育心理学会第44回総会で発表いたしました。

Appendix

英 語

注 意 事 項

- 1 試験開始の合図があるまで、この問題冊子の中を見てはいけません。
- 2 試験中に問題、解答用紙の印刷不鮮明、落丁・乱丁など不備に気付いた場合には、手をあげて監督者に知らせなさい。
- 3 解答用紙には解答欄以外に、受験番号、所属、氏名の記入欄があるので、監督者の指示に従って、それぞれ正しく記入しなさい。
- 4 解答は、解答用紙に記入しなさい。例えば、第 7 問の【 20 】と表示のある問いに対して③と解答する場合は、第 7 問の解答番号 20 の解答欄に 3 と記入しなさい。
- 5 問題冊子の余白などは適宜利用してよいが、どのページも切り離してはいけません。
- 6 試験終了後、問題冊子は回収するので、監督者の指示に従いなさい。
- 7 試験問題は非公開なので、口外してはいけません。

第1問

次の文章の空欄（【 1 】～【 8 】）に入れるのに最も適当な語を、下に示した①～⑩のうちから一つずつ選べ。ただし、同じ語を繰り返し用いてはならない。

Because of the increases in road traffic and the cost of gasoline in the last few years, people have 【 1 】 the advantages of bicycles as a means of transport. Two of their best points are that they do not need fuel and do not cause pollution; 【 2 】 , bicycles are easy to use and fairly comfortable.

This was not always the case, 【 3 】 . The first bicycles were heavy and hard to ride. They were moved along by the rider 【 4 】 his feet against the ground. Turning left or right 【 5 】 a good deal of muscle power. For stopping, the rider 【 6 】 entirely on his own strength and skill 【 7 】 there were no brakes. He had a rough ride even under the best conditions because the wheels were made of iron. It is not surprising that people mostly rode bicycles for fun. Some still do, of course, but now bicycles are also an efficient means of 【 8 】 around.

- | | | | |
|------------|------------|------------|----------------|
| ① accept | ② although | ③ depended | ④ getting |
| ⑤ however | ⑥ moreover | ⑦ pushing | ⑧ rediscovered |
| ⑨ required | ⑩ since | | |

(63・4)

第2問

次の(1)～(4)の文章は一連のものである。各文章を読み、その内容に最もよく合っているものを、それぞれ下の①～④のうちから一つずつ選べ。

(1) Can computers think? Now that they are used increasingly for functions that in the past only the human brain could perform, this difficult question is often raised.

問1 【 1 】

- ① Thinking has up to now been considered a purely human activity.
- ② With computers becoming common, people do not think as much as they used to.
- ③ It has become necessary for everyone to understand how computers work.
- ④ It is a worrying fact that computers are taking the place of human beings.

- (2) Computers can do some things that human brains can—solving mathematical problems, for instance. Basically, they work, at high speed, by a simple “on-and-off” process. Can this be called “thinking”?

問2 【 2 】

- ① Computers are only good at solving mathematical problems.
 - ② Real thinking is always a careful, rather slow process.
 - ③ The working of computers seems too simple to be called thinking.
 - ④ Computers are merely machines, so they cannot be said to think.
- (3) For certain kinds of problem-solving, the human brain works, less rapidly, in a similar way. But other mental processes are quite different, especially those connected with ideas and creation.

問3 【 3 】

- ① Computers and human brains are totally different from each other.
- ② The human brain is much quicker than a computer at producing ideas.
- ③ Computers would not exist if human brains had not created them.
- ④ The similarities between the human brain and computers are limited.

- (4) To say whether computers can think, it would first be necessary to arrive at a definition of thinking. If only we were able to do that, we might understand better what makes us human.

問4 【 4 】

- ① Better understanding of computers would enable us to understand human beings.
- ② We have not yet discovered exactly what thinking is.
- ③ People talk about computers and the brain without understanding them properly.
- ④ It is difficult to explain clearly how computers work.

(63・6)

第3問

次の文章を読み、問い（問1～4）の空欄（【 1 】～【 4 】）に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

My husband hasn't stopped laughing about a funny thing that happened to me. It's funny *now* but it wasn't at the time.

Last Friday, after doing all the family shopping in town, I wanted a rest before catching the train, so I bought a newspaper and some chocolate and went into the station coffee shop—that cheap, self-service place with long tables to sit at. I put my heavy bag down on the floor, put the newspaper and chocolate on the table to keep a place, and went to get a cup of coffee.

When I came back with the coffee, there was someone in the next seat. It was one of those wild-looking youngsters, with dark glasses and torn clothes, and hair colored bright red at the front. Not so unusual these days. What *did* surprise me was that he'd started to eat my chocolate!

Naturally, I was annoyed. However, to avoid trouble—and really I was rather uneasy about him—I just looked down at the front page of the newspaper, tasted my coffee, and took a bit of chocolate. The boy looked at me closely. Then he took a *second* piece of my chocolate. I could hardly believe it. Still I didn't dare to start an argument. When he took a third piece, I felt more angry than uneasy. I thought, “Well, *I* shall have the *last* piece,” and I got it.

The boy gave me a strange look, then stood up. As he left he shouted out, “This woman's crazy!” Everyone stared. That was embarrassing

enough, but it was worse when I finished my coffee and got ready to leave. My face went red—as red as his hair—when I realized I'd made a mistake. It wasn't *my* chocolate that he'd been taking. There was mine, unopened, just under my newspaper.

問1 The woman telling the story 【 1 】 .

- ① always went shopping with her family on Fridays
- ② had been very busy and needed some time to recover
- ③ wanted a newspaper and some chocolate to take home to her family
- ④ bought a newspaper and some chocolate so that she could keep a place at the table

問2 When the woman came back to the table, she was surprised 【 2 】 .

- ① because she wasn't sure whether the youngster was a boy or a girl
- ② because the youngster was dressed so strangely
- ③ because there was already someone sitting in the next seat
- ④ because the youngster's behavior seemed so rude

問3 When the woman saw the boy go on eating the chocolate, she felt
【 3 】.

- ① too tired to start an argument
- ② too shy to look in the boy's direction
- ③ more and more disappointed at losing the chocolate
- ④ more and more angry with the boy

問4 The woman's face turned red 【 4 】.

- ① because she realized that she had been quite wrong about the boy
- ② because she realized that the boy was poor and hungry
- ③ because she saw everyone staring at her
- ④ because she hated being shouted at

(63・7)

第4問

次の問い（A・B）に答えよ。

A 次の問い（問1～3）において、それぞれ下の①～⑤の語や語句を並べかえて空所を補い、文を完成せよ。ただし、解答は【 1 】～【 6 】に入るものの番号のみを記入せよ。

問1 It was the control of fire and 【 1 】 【 1 】 【 2 】 in the cold northern areas.

- ① allowed ② humans ③ that
④ the use of clothing ⑤ to settle

問2 Speech, as a means of communication, is 【 3 】 【 3 】 【 4 】 and passed on.

- ① because ② culture is shared ③ it is the chief way
④ of major importance ⑤ through which

問3 Biotechnology could offer great advantages to the whole world. If this new technology 【 】, 【 5 】 【 】 【 6 】 【 】 narrowed.

- ① and food production ② population increase ③ the gap between
④ will be ⑤ works

B 次の問い（問1・問2）では、空欄【 1 】・【 2 】に入れるべき四つの文が、順不同で、下のA～Dに示されている。意味の通る文章にするのに最も適した配列を、それぞれ①～④のうちから一つずつ選べ。

問1 Physical gestures may have different meanings in different cultures, and misunderstanding these signals can sometimes be embarrassing. 【 1 】 When I asked him what was the matter, he replied, "In my country, we count people with our eyes. We use our fingers to count pigs."

- A. But one of them became quiet and looked puzzled.
B. I counted them with the index finger, which is common in Japan.
C. I once had an experience which I have never forgotten.
D. Some years ago, I took a small group of foreign students to Kyoto.

- ① C・A・D・B ② C・D・B・A ③ D・B・C・A ④ D・C・A・B

問 2 The ancient Egyptians knew much about the human body.

【 2 】 Through this preparatory treatment of the body, they learned the positions and functions of the main organs.

- A. But if the body was kept in good condition, the soul would come back to live in it again.
- B. Therefore, they prepared the body carefully for the new life.
- C. They acquired this knowledge owing to their religious beliefs.
- D. They believed that after death the soul would leave the body and make a long journey.

① B・A・C・D ② B・C・D・A ③ C・B・D・A ④ C・D・A・B

(8-3)

第5問

次の文章を読み、下の問い（問1～3）の空欄【 1 】～【 3 】に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

Several years ago, certain scientists developed a way of investigating the nature of the atmosphere of the past by studying air caught in the ice around the North or South Pole. According to their theory, when snow falls, air is trapped between the snowflakes. The snow turns to ice with the air still inside. Over the years more snow falls on top, making new layers of ice. But the trapped air, these scientists believed, remains exactly as it was when the snow originally fell.

To find what air was like three hundred years ago, you use a drill in the shape of a hollow tube to cut deep into the layers of ice. When you pull up the drill, an ice core made of many layers comes up inside it. Then, back at the laboratory, you count the layers in the core — each layer represents one year — to find ice formed from the snow that fell during the year to be studied. Using this method, these scientists suggested that the amount of carbon dioxide (CO_2), one of the gases which may cause global warming, had increased greatly over the last two hundred years.

A Norwegian scientist, however, pointed out that there might be a problem with this method. He claimed that air caught in ice does not stay the same. In particular, he said, the quantity of CO_2 does not remain stable, since some of it is absorbed by ice crystals, some enters water, and some locks itself up in other chemicals. If this were true, then there could have

been more CO₂ in the past than we thought. Even so, measurements taken over the past thirty years show that CO₂ has increased by over ten percent during this short period.

問1 Certain scientists claimed that 【 1 】.

- ① atmospheric gases increase the yearly amount of snow
- ② falling snowflakes change the chemical balance of the air
- ③ the action of atmospheric gases causes snow to turn into ice
- ④ the air held between snowflakes keeps its original nature

問2 In order to study atmospheric gases for a particular year, these scientists had to 【 2 】.

- ① count how many cores were necessary for measuring trapped air
- ② examine the different kinds of snowflakes
- ③ identify a particular layer in the ice core
- ④ measure how hollow each layer was within the core

問3 A Norwegian scientist questioned the usefulness of ice-core analysis, claiming that 【 3 】.

- ① ice absorbs more CO₂ when the core is pulled out
- ② ice has effects on global warming by reducing CO₂
- ③ the amount of CO₂ within the ice changes in several ways
- ④ the quality of the ice might be affected during the drilling

(8-5)

第6問

次の問い（A・B）に答えよ。

A 次の問い（問1～3）において、それぞれ下の①～⑤の語や語句を並べかえて空所を補い、文を完成せよ。ただし、解答は【 1 】～【 6 】に入るものの番号のみを記入せよ。

問1 Before you rent an apartment, find out as much about it
【 】 【 1 】 【 】 【 2 】 【 】
misunderstanding.

- | | | |
|---------------|-----------------|------|
| ① as possible | ② in order that | ③ no |
| ④ there | ⑤ will be | |

問2 There is no good reason for 【 】 【 3 】 【 】
【 4 】 【 】 job.

- | | | |
|------------|-----------------|----------------|
| ① if | ② less than men | ③ paying women |
| ④ the same | ⑤ they do | |

問3 In recent years, people have become more health-conscious and consequently 【 】 【 5 】 【 】 【 6 】 【 】 eat.

- ① fat ② have ③ reduced
④ the amount of ⑤ they

B 次の問い（問1・問2）では、空欄【 1 】・【 2 】に入れるべき四つの文が、順不同で、下のA～Dに示されている。意味の通る文章にするのに最も適した配列を、それぞれ①～④のうちから一つずつ選べ。

問1 The population density of France had a rise of four persons per square kilometer (km^2), from 95 to 99, during the period from 1972 to 1981. 【 1 】 Two areas usually affected by such an increase are housing and transportation.

- A. They had rises in density of nine and seven respectively.
B. Its density went up from 281 to 311 in the same period.
C. A country with an even larger figure, though, was Japan.
D. Italy and Spain experienced larger increases.

- ① A・D・B・C ② B・A・D・C ③ C・D・A・B ④ D・A・C・B

問 2 In manufacturing, energy is usually saved when recycled materials are used. 【 2 】 As we develop better recycling systems, we will be able to recover more materials and make ever greater savings in energy.

- A. Take glass, for example. While collection has increased in recent years, we still only manage to recycle about 9% of this product.
- B. One of the major barriers to increased recycling for this and other household products, however, is a lack of good collection systems.
- C. Although more materials are being reused, there is still a lot of progress to be made.
- D. Recycling paper, for instance, cuts the energy required to make this product by 75%.

① A・C・B・D ② B・C・D・A ③ C・D・A・B ④ D・B・A・C

(5-3)

第7問

次の文章を読み、下の問い（問1～4）の空欄【 1 】～【 4 】に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

In many parts of the world, the influence of television is a matter of increasing concern. For years, critics of television have concentrated on the issue of the program content — particularly violence — as it affects viewers. The problem seems especially serious with regard to younger children. There is evidence that TV does in fact lead people to accept more violence in everyday life. How could this not happen when it presents violent acts, often with guns and knives, as normal and common occurrences?

In the last few years criticism of television has moved to a new stage by shifting the focus from the subject matter to the experience of the medium itself. The problem now with TV is not just what is seen but how it is seen. The way children watch it causes them to be passive, and some evidence suggests that such viewing might even affect the development of the brain in small children.

However, the worst aspect of television is the way it can interfere with family life. The “box” has too often become a substitute parent, taking over most of the work of introducing social and moral values to the child and developing them in him. Parents allow this to happen by using television like a drug for the purpose of keeping children quiet. Eventually the child comes to depend on the box and it becomes a necessary, lifelong habit.

While many children go through the “television experience” and

survive, many others are deeply affected by it. Much of the discussion of TV during the next few years will center on how to reduce the dangers which it presents, particularly to younger people. Already there are movements to try to ban TV advertising which is directed at children under a certain age. Perhaps this is just the beginning. In the end, some people may even go to the extreme of demanding the removal of such a powerful medium from the lives of young people. This might not be a practical solution, but we should not ignore the dangers of television.

問1 Television has long been criticized mainly because 【 1 】 .

- ① it causes younger children to waste time in front of it
- ② it encourages people, through advertisements, to want guns and knives
- ③ it makes people less willing to put up with violence
- ④ it shows violence, which influences, above all, younger people

問2 Recent criticism against TV has focused on 【 2 】 .

- ① the active way of living that it promotes
- ② the influence of the content of programs on children's brains
- ③ the medium itself in relation to the development of children
- ④ the subject matter it shows to children

問3 The presence of TV may weaken the family by 【 3 】 .

- ① encouraging children to rely on their parents
- ② giving children a greater sense of moral values
- ③ providing more time for the family to share together
- ④ taking over an essential part of parents' work

問4 The author states that certain people in society might finally attempt 【 4 】 .

- ① to ban TV advertising aimed at young people
- ② to have children survive the "television experience"
- ③ to prohibit children from watching TV
- ④ to recognize the educational merits of the medium

(5-5)

英 語 解 答 用 紙

所属	
----	--

受験番号		氏名	
------	--	----	--

第1問		第2問		第3問		第4問 A		第4問 B	
1		1		1		1		1	
2		2		2		2		2	
3		3		3		3		3	
4		4		4		4		4	
5		5		5		5		5	
6		6		6		6		6	
7		7		7		7		7	
8		8		8		8		8	
9		9		9		9		9	
10		10		10		10		10	
11		11		11		11		11	
12		12		12		12		12	
13		13		13		13		13	
14		14		14		14		14	

第5問		第6問 A		第6問 B		第7問	
1		1		1		1	
2		2		2		2	
3		3		3		3	
4		4		4		4	
5		5		5		5	
6		6		6		6	
7		7		7		7	
8		8		8		8	
9		9		9		9	
10		10		10		10	
11		11		11		11	
12		12		12		12	
13		13		13		13	
14		14		14		14	

解答は楷書でていねいに

Relations between the English ability measured by paper-and-pencil test of the National Center Test and the English communication ability measured by a listening comprehension test, through the common scale of the standardized ability

UCHIDA Teruhisa*
NAKAUNE Naoko**
ISHIZUKA Tomoichi**

Abstract

English test questions of National Center Test (partly of Joint First Stage Achievement Test) were standardized using a design with combination of common examinee method and common item method. Difficulty differed between the most difficult form and the easiest form by 6 percent on the base of correct answer ratio. This was resolved by equating them on a common standardized scale, up to the difference of the average of by 0.05.

Listening comprehension tests were administered on the same groups of examinee at the same time. The product moment correlation between the listening comprehension test and the paper-and-pencil test was about 0.65. The order of listening problems and the degree of continuity of same speaker were also controlled in several manners. The results showed those factors did not affect the test performance significantly.

* Department of Comprehensive Studies on Admission and its Environment, Research Division,

** Department of Applied Statistics and Measurement, Research Division,
The National Center for University Entrance Examinations,
2-19-23 Komaba, Meguro-ku, Tokyo 153-8501, Japan

