

大学入試センター試験得点の標準化の試み

ー項目反応理論による方法ー

前川 眞¹、菊地 賢²、内田 照久³、中畝 菜穂子³、石塚 智一³
1 東京工業大学 2 東邦大学 3 大学入試センター

1. はじめに

大学入試センターが実施してきた共通テスト(共通第1次学力試験, 大学入試センター試験)は標準化されたテストではない。すなわち, 毎年新しい問題がプリテストなしに作成・実施されており, その難易度のコントロールは, ひとえに作題者グループの経験と努力に依存している。

他方, 受験生も重複受験者(いわゆる浪人)を除けば毎年新しくなる。したがって, 年度間での各科目のテストの平均点は, 年度ごとの問題の難易度と受験生の学力との, 二つの要因からの影響を受ける。それゆえ, 平均点の差異は, 問題の難易度や受験生の学力の差異をそれぞれ直接的に示すものではない。この平均点の差は, もしも受験生の学力が不変であることを仮定すれば, 問題の難易度の差として, また, 問題の難易度が不変であることを仮定すれば, 学力の差として解釈できるが, これらの仮定が正しいという保証はない。また, 重複受験者(浪人)のデータから推論を行うことも可能であるが, 重複受験者の大部分は学力が伸びている可能性が高いため, この方法もあまり有効ではない。

このようなセンター試験成績の年度間での比較に関連して, 新たな課題が課せられている。具体的には平成12年11月の大学審議会の答申において, センター試験成績の複数年度利用に関する検討の必要性が提言に盛り込まれた(文部省大学審議会, 2000)。この提言を受ける形で, 平成14年度からは, センター試験の前年度成績を入試に用いることができることになった旨が, 文部省高等教育局長から各大学に通知された。

上記の通知に対して, 国立大学協会は, 現在のセンター試験の成績は年度間で比較するには不十分な情報であるため, 成績の取り扱いについては慎重を期すよう促している(国立大学協会第2常置委員会, 2001)。現在は, 前年度成績を利用する大学は少数に留まり, また実施する大学

も利用方法を工夫することで, 現状に即した対応がなされている。

このような状況を鑑みるに, 今後, 大学入試センター試験においても, 年度間の試験成績の比較を可能にするための手続きを検討していく必要がある。そのためには, いわゆる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なる。

ここでは, そのような目標を視野にいたした研究の第一歩として, 実験的に過去の問題を被験者に解答させ, そのデータを基に過去の問題の難易度を推定し, テストの事後標準化を試みた。理想的には, 被験者が過去20年間の問題の全てに解答することが望ましいが, 現実的にそれは不可能である。そこで, 過去の試験問題からいくつかの大問をサンプルして新たな冊子を複数作った。そして, それらを実験者に無作為に被験者に割当てて解答させた。

その際, 新たな冊子が全く別々の問題を含むのではなく, いくつかの大問が複数の冊子に含まれるようにした。そして, 異なるテストの中に配置された共通の問題への解答結果から得られる統計的情報を基に, 複数のテスト間の等化を行い, 相互の成績比較が可能になるように実験をデザインした。

このような大学入試センター試験の標準化については, 前川・石塚・菊地・内田・中畝(2001)が, 古典的テスト理論の文脈に沿った同族モデルを用いて検討を試みたものがある。本研究ではそれをさらに発展させ, より試験の標準化に適しており, 既に米国の各種試験では実務ベースで利用されている新しいテスト理論である項目反応理論(Item Response Theory, IRT)を用いた分析を行う。

2. 実験のデザイン

平成11, 12年度に実施したモニター実験(第I期)では, 昭和56, 57, 59, 61年度, 平成元年度の共通第1次学力試験, ならびに平成3, 4, 5, 7, 8年度の大学入試センター試験の10年度分にわたる数学と英語の問題から25題程度の大問を選び出し, 実験で用いた。

また, 平成13年度のモニター実験(第II期)では, 昭和56, 57, 59, 61年度と平成元年度にかえて, 昭和63年度の共通第1次学力試験, ならびに平成2, 6, 9, 10年度の大学入試センター試験を加えて, 実験を行った。

大問の選定にあたっては, 各年度の試験問題から大問単位で, 英語は2-3問, 数学は2問が選ばれた。その際, 被験者が国立大学に合格している1年生であることを考慮し, ある程度難易度が高く, また大問としての識別力の高かったものを採用した。そして, これらの大問を組み合わせ, 各問題冊子には必ず3つの年度の問題が全て含まれるように, また各年度の問題が必ず3つの冊子に含まれるように, 計10種類の問題冊子を編集した。

その後, 冊子の中での大問の出現位置の影響を除去するため, 各冊子から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, 各大問名は, 和暦(56-63, 1-9)と大問番号(1-9)をハイフンでつないだ形で示した。数学の場合は, 数学IとIIの区別がある年度ではIとIIで区別した。また, AとBの区別がある年度は全てIAとIIBの問題を用いた。なお, 9-0Iは平成9年度の旧数学Iである。その内訳を, 英語について表1aに, 数学について表1bに示す。

被験者は, 平成11, 12年度については, 3つの国立大学から集めた約650人の大学1年生であった。また, 平成13年度は, 大学を1つ入れ換えた3つの国立大学から集めた約600人の学生を用いた。被験者は, 各年度とも大学毎に各問題冊子で均等となるようにランダムに割り振った。平成11, 12年度の各問題冊子は約65名(各版はその1/3)の被験者に割り振られ, 平成13年度の冊子は, 約60名に割り振られた。各大問単位では, それぞれに約170~370人の被験者が受験した。

表 1a 英語の冊子の構成

冊子名	大問数	含まれる大問名								
1-1	8	56-2	56-4	56-5	61-6	61-6	61-7	4-3	4-5	56-5
1-2	8	61-5	61-6	61-7	4-3	4-5	56-2	56-4	56-5	61-6
1-3	8	4-3	4-5	56-2	56-4	56-5	61-5	61-6	61-7	61-7
2-1	7	61-5	61-6	61-7	4-3	4-5	8-3	8-5	8-5	8-5
2-2	7	4-3	4-5	8-3	8-5	61-5	61-6	61-7	4-3	4-5
2-3	7	8-3	8-5	61-5	61-6	61-7	4-3	4-5	59-9	59-9
3-1	7	4-3	4-5	8-3	8-5	59-5	59-7	59-9	4-3	4-5
3-2	7	8-3	8-5	59-5	59-7	59-9	4-3	4-5	8-3	8-5
3-3	7	59-5	59-7	59-9	4-3	4-5	8-3	3-3	3-5	8-5
4-1	7	8-3	8-5	59-5	59-7	59-9	3-3	3-5	8-3	8-5
4-2	7	59-5	59-7	59-9	3-3	3-5	59-5	59-7	59-9	59-9
4-3	7	3-3	3-5	8-3	8-5	59-5	59-7	59-9	3-3	3-5
5-1	7	59-5	59-7	59-9	3-3	3-5	7-3	7-5	59-5	59-9
5-2	7	3-3	3-5	7-3	7-5	59-5	59-7	59-9	3-3	3-5
5-3	7	7-3	7-5	59-5	59-7	59-9	3-3	3-5	57-9	57-9
6-1	7	3-3	3-5	7-3	7-5	57-5	57-8	57-9	3-3	3-5
6-2	7	7-3	7-5	57-5	57-8	57-9	3-3	3-5	7-3	7-5
6-3	7	57-5	57-8	57-9	3-3	3-5	7-3	7-5	1-8	1-8
7-1	8	7-3	7-5	57-5	57-8	57-9	1-4	1-5	1-8	7-3
7-2	8	57-5	57-8	57-9	1-4	1-5	1-8	57-5	57-8	57-9
7-3	8	1-4	1-5	1-8	7-3	7-5	57-5	57-8	57-9	57-9
8-1	8	57-5	57-8	57-9	1-4	1-5	1-8	5-3	5-5	57-8
8-2	8	1-4	1-5	1-8	5-3	5-5	57-5	57-8	57-9	57-9
8-3	8	5-3	5-5	57-5	57-8	57-9	1-4	1-5	1-8	57-9
9-1	8	1-4	1-5	1-8	5-3	5-5	56-2	56-4	56-5	56-5
9-2	8	5-3	5-5	56-2	56-4	56-5	1-4	1-5	1-8	1-8
9-3	8	56-2	56-4	56-5	1-4	1-5	1-8	5-3	5-5	5-5
10-1	8	5-3	5-5	56-2	56-4	56-5	61-5	61-6	61-7	61-7
10-2	8	56-2	56-4	56-5	61-5	61-6	61-7	5-3	5-5	5-5
10-3	8	61-5	61-6	61-7	5-3	5-5	56-2	56-4	56-5	56-5
11-1	7	63-4	63-6	63-7	8-3	8-5	5-3	5-5	63-7	63-7
11-2	7	8-3	8-5	5-3	5-5	63-4	63-6	63-7	8-3	8-5
11-3	7	5-3	5-5	63-4	63-6	63-7	8-3	8-5	8-5	8-5
12-1	6	8-3	8-5	5-3	5-5	2-2	2-3	2-3	2-3	2-3
12-2	6	5-3	5-5	2-2	2-3	8-3	8-5	5-5	5-5	5-5
12-3	6	2-2	2-3	8-3	8-5	5-3	5-5	9-2	9-3	9-3
13-1	6	5-3	5-5	2-2	2-3	9-2	9-3	5-3	5-5	5-5
13-2	6	2-2	2-3	9-2	9-3	5-3	5-5	2-2	2-3	2-3
13-3	6	9-2	9-3	5-3	5-5	2-2	2-3	6-5	6-5	6-5
14-1	6	2-2	2-3	9-2	9-3	6-3	6-5	2-2	2-3	2-3
14-2	6	9-2	9-3	6-3	6-5	2-2	2-3	9-2	9-3	9-3
14-3	6	6-3	6-5	2-2	2-3	9-2	9-3	3-3	3-5	3-5
15-1	6	9-2	9-3	6-3	6-5	3-3	3-5	9-2	9-3	9-3
15-2	6	6-3	6-5	3-3	3-5	9-2	9-3	6-3	6-5	6-5
15-3	6	3-3	3-5	9-2	9-3	3-3	3-5	10-2	10-3	10-3
16-1	6	6-3	6-5	3-3	3-5	10-2	10-3	6-3	6-5	6-5
16-2	6	3-3	3-5	10-2	10-3	6-3	6-5	3-3	3-5	3-5
16-3	6	10-2	10-3	6-3	6-5	3-3	3-5	7-3	7-5	7-5
17-1	6	3-3	3-5	10-2	10-3	7-3	7-5	3-3	3-5	3-5
17-2	6	10-2	10-3	7-3	7-5	3-3	3-5	10-2	10-3	10-3
17-3	6	7-3	7-5	3-3	3-5	10-2	10-3	4-3	4-5	4-5
18-1	6	10-2	10-3	7-3	7-5	4-3	4-5	10-2	10-3	10-3
18-2	6	7-3	7-5	4-3	4-5	10-2	10-3	7-3	7-5	7-5
18-3	6	4-3	4-5	10-2	10-3	7-3	7-5	63-4	63-6	63-7
19-1	7	7-3	7-5	4-3	4-5	63-4	63-6	63-7	7-3	7-5
19-2	7	4-3	4-5	63-4	63-6	63-7	7-3	7-5	4-3	4-5
19-3	7	63-4	63-6	63-7	7-3	7-5	4-3	4-5	8-3	8-5
20-1	7	4-3	4-5	63-4	63-6	63-7	8-3	8-5	4-3	4-5
20-2	7	63-4	63-6	63-7	8-3	8-5	4-3	4-5	63-4	63-6
20-3	7	8-3	8-5	4-3	4-5	63-4	63-6	63-7	8-3	8-5

3. 集団ごとに独立に推定された項目パラメタの等化

異なる年度に実施されたテストの問題項目の困難度や識別力は, 各年度に固有の受験者の試験結果が反映されるため, 集団に依存した独自の尺度上に示される。ここでもしも同一の1次元の尺度上に, これまでに実施された共通テストの全ての問題について, その性質を整理して示すことができたとする。すると, 各問題は共通尺度上で扱うことができるため, 年度間での成績も比較可能になる。

この実験に即して言えば, 各冊子はそれぞれに独自の被験者が割当てられているので, 冊子ごとに集団独自の尺度が構成される。その上で, 全ての冊子に含まれる全問題項目を一元的に整理するための尺度を設定できれば, 共通尺度と

表 1b 数学の冊子の構成

冊子名	大問数	含まれる大問名					
1-1	6	56-1	56-3	61-1	61-4	4-1-1	4-11-1
1-2	6	61-1	61-4	4-1-1	4-11-1	56-1	56-3
1-3	6	4-1-1	4-11-1	56-1	56-3	61-1	61-4
2-1	5	61-1	61-4	4-1-1	4-11-1	8-11-1	
2-2	5	4-1-1	4-11-1	8-11-1	61-1	61-4	
2-3	5	8-11-1	61-1	61-4	4-1-1	4-11-1	
3-1	5	4-1-1	4-11-1	8-11-1	59-2	59-3	
3-2	5	8-11-1	59-2	59-3	4-1-1	4-11-1	
3-3	5	59-2	59-3	4-1-1	4-11-1	8-11-1	
4-1	5	8-11-1	59-2	59-3	3-1-3	3-11-2	
4-2	5	59-2	59-3	3-1-3	3-11-2	8-11-1	
4-3	5	3-1-3	3-11-2	8-11-1	59-2	59-3	
5-1	5	59-2	59-3	3-1-3	3-11-2	7-11-2	
5-2	5	3-1-3	3-11-2	7-11-2	59-2	59-3	
5-3	5	7-11-2	59-2	59-3	3-1-3	3-11-2	
6-1	5	3-1-3	3-11-2	7-11-2	57-2	57-4	
6-2	5	7-11-2	57-2	57-4	3-1-3	3-11-2	
6-3	5	57-2	57-4	3-1-3	3-11-2	7-11-2	
7-1	5	7-11-2	57-2	57-4	1-3	1-4	
7-2	5	57-2	57-4	1-3	1-4	7-11-2	
7-3	5	1-3	1-4	7-11-2	57-2	57-4	
8-1	6	57-2	57-4	1-3	1-4	5-1-3	5-11-1
8-2	6	1-3	1-4	5-1-3	5-1-1	57-2	57-4
8-3	6	5-1-3	5-11-1	57-2	57-4	1-3	1-4
9-1	6	1-3	1-4	5-1-3	5-1-1	56-1	56-3
9-2	6	5-1-3	5-11-1	56-1	56-3	1-3	1-4
9-3	6	56-1	56-3	1-3	1-4	5-1-3	5-11-1
10-1	6	5-1-3	5-11-1	56-1	56-3	61-1	61-4
10-2	6	56-1	56-3	61-1	61-4	5-1-3	5-11-1
10-3	6	61-1	61-4	5-1-3	5-1-1	56-1	56-3
11-1	6	63-2	63-4	8-1-2	8-1-1	5-1-3	5-11-1
11-2	6	8-1-2	8-11-1	5-1-3	5-1-1	63-2	63-4
11-3	6	5-1-3	5-11-1	63-2	63-4	8-1-2	8-11-1
12-1	6	8-1-2	8-11-1	5-1-3	5-1-1	2-1-3	2-11-1
12-2	6	5-1-3	5-11-1	2-1-3	2-1-1	8-1-2	8-11-1
12-3	6	2-1-3	2-11-1	8-1-2	8-1-1	5-1-3	5-11-1
13-1	6	5-1-3	5-11-1	2-1-3	2-1-1	9-01-2	9-11B-1
13-2	6	2-1-3	2-11-1	9-01-2	9-11B-1	5-1-3	5-11-1
13-3	6	9-01-2	9-11B-1	5-1-3	5-1-1	2-1-3	2-11-1
14-1	6	2-1-3	2-11-1	9-01-2	9-11B-1	6-1-3	6-11-1
14-2	6	9-01-2	9-11B-1	6-1-3	6-1-1	2-1-3	2-11-1
14-3	6	6-1-3	6-11-1	2-1-3	2-1-1	9-01-2	9-11B-1
15-1	6	9-01-2	9-11B-1	6-1-3	6-1-1	3-1-3	3-11-2
15-2	6	6-1-3	6-11-1	3-1-3	3-1-2	9-01-2	9-11B-1
15-3	6	3-1-3	3-11-2	9-01-2	9-11B-1	6-1-3	6-11-1
16-1	6	6-1-3	6-11-1	3-1-3	3-1-2	10-1A-2	10-11B-1
16-2	6	3-1-3	3-11-2	10-1A-2	10-11B-1	6-1-3	6-11-1
16-3	6	10-1A-2	10-11B-1	6-1-3	6-1-1	3-1-3	3-11-2
17-1	6	3-1-3	3-11-2	10-1A-2	10-11B-1	7-1-3	7-11-2
17-2	6	10-1A-2	10-11B-1	7-1-3	7-1-2	3-1-3	3-11-2
17-3	6	7-1-3	7-11-2	3-1-3	3-1-2	10-1A-2	10-11B-1
18-1	6	10-1A-2	10-11B-1	7-1-3	7-1-2	4-1-1	4-11-1
18-2	6	7-1-3	7-11-2	4-1-1	4-1-1	10-1A-2	10-11B-1
18-3	6	4-1-1	4-11-1	10-1A-2	10-11B-1	7-1-3	7-11-2
19-1	6	7-1-3	7-11-2	4-1-1	4-1-1	63-2	63-4
19-2	6	4-1-1	4-11-1	63-2	63-4	7-1-3	7-11-2
19-3	6	63-2	63-4	7-1-3	7-1-2	4-1-1	4-11-1
20-1	6	4-1-1	4-11-1	63-2	63-4	8-1-2	8-11-1
20-2	6	63-2	63-4	8-1-2	8-1-1	4-1-1	4-11-1
20-3	6	8-1-2	8-11-1	4-1-1	4-1-1	63-2	63-4

(2)式を最小とするような、 m 個の項目の共通尺度上での項目パラメタ (a_g, b_g) と、各テストでの尺度を共通尺度へ変換する係数 (q_g, r_g) を推定する方法を示している。ただし、 $P(\theta|a,b)$ は項目パラメタ a, b を与えたときのICCを表し、 W_g は集団ごとの重みである。また、 $h(\theta|\mu)$ は集団ごとの特性値分布の密度関数をあらわし、 $\sum_{j \in g}$ は、項目 j が、集団 g に課せられたテストに含まれる場合に、その和を取ることを意味している。

なお、この方法が利用可能であるためには、本研究の実験のように、各冊子に含まれる項目が、冊子間に配置された共通項目を介することによって相互が比較できるように、リンクが取れていることが必要である。ただ、各受験者集団は、等質である必要はない。集団ごとに項目パラメタを推定する際に、その集団における能力値の平均をゼロ、分散を1とおいた場合には、 q_g と r_g がそれぞれ共通尺度上での各集団の能力の平均と標準偏差を表すパラメタとなる。なお、より詳細な解説については、前川(1991)を参照されたい。

4. データ解析

4.1 データのスクリーニング

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を確認した。その結果、本来の実力通りに試験を受けていないもの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。その際、当日の試験監督者の報告書に記載された受験態度や、数学と英語双方の得点をつき合わせた結果も考慮した。最終的に、英語に関しては33名を、数学に関しては44名を分析対象外とした。このスクリーニングを経たデータの平均得点等を、英語について表2aに、数学について表2bに示す。

表2a, 2bでは、各大問に、和暦(56-63, 1-9)と大問番号(1-9)をハイフンでつないだ名称に加えて、科目コード二桁 + 西暦下二桁 + 大問番号一桁 + _ という形で新たなラベルを付与している。以下の分析では、各項目を表すのにこのラベルの後に項目番号を付加したものをを用いる。例えば、H1812_1は、H1が英語を表すため、1981年(昭和56年)の英語の大問2に含まれる最初の項目となる。また、数学の場合、数学IAはコード

して利用が可能になる。

以後の分析では、大問に含まれる小問単位でも分析を行う。ここからは各小問のことをテスト理論の文脈に即して、項目と呼ぶ。一般に、項目反応理論における原点と単位の不定性を考慮すれば、共通尺度 θ と集団独自の尺度 $\theta(g)$ との間には

$$\theta = q_g + r_g \theta_g, \quad a_i = \frac{1}{r_g} a_i^{(g)}$$

$$\text{and } b_i = q_g + r_g b_i^{(g)} \quad (1)$$

という関係がある。ただし、 $(a_i^{(g)}, b_i^{(g)})$ は集団ごとに推定された項目パラメタであり、 (q_g, r_g) は集団ごとの等化係数である。前川(1991)は、

$$RSS = \sum_{g=1}^G \sum_{j \in g} W_g \int_{-\infty}^{\infty} P(\theta|a_i^{(g)}, b_i^{(g)})$$

$$- P(q_g + r_g \theta | a_i, b_i)^2 h^{(g)}(\theta|\mu_g) d\theta \quad (2)$$

D2を、数学ⅡBはコードE2を用いたが、試験科目名の変遷もあり、1984年までは数学ⅠとしてコードD2を、1986-1989年は数学ⅠおよびⅡとしてコードDEを用いる。また、1997年度の旧数学ⅠはコードD3を用いた。

表 2a 英語大問の平均得点・標準偏差等

大問	Label	大問名	N	Mean	Std Dev	Minimum	Maximum
1	H1812_	56-2	190	11.34	2.80	3.00	15.00
2	H1814_	56-4	190	15.68	5.56	0.00	24.00
3	H1815_	56-5	190	8.24	3.82	0.00	12.00
4	H1825_	57-5	190	18.09	3.69	4.00	24.00
5	H1828_	57-8	190	14.72	4.30	0.00	18.00
6	H1829_	57-9	190	15.47	5.02	0.00	20.00
7	H1845_	59-5	190	11.18	3.89	0.00	16.00
8	H1847_	59-7	190	11.13	3.97	0.00	15.00
9	H1849_	59-9	190	28.82	7.82	5.00	40.00
10	H1865_	61-5	193	15.31	4.61	0.00	20.00
11	H1866_	61-6	193	7.27	4.07	0.00	12.00
12	H1867_	61-7	193	14.02	7.08	0.00	24.00
13	H1884_	63-4	168	13.94	3.10	4.00	16.00
14	H1886_	63-6	168	10.88	3.96	0.00	16.00
15	H1887_	63-7	168	17.68	4.22	6.00	24.00
16	H1894_	1-4	187	14.42	4.40	0.00	20.00
17	H1895_	1-5	187	10.46	4.42	0.00	16.00
18	H1898_	1-8	187	15.83	7.28	0.00	24.00
19	H1902_	2-2	172	38.10	7.14	20.00	52.00
20	H1903_	2-3	172	25.01	5.95	6.00	30.00
21	H1913_	3-3	366	18.44	7.36	0.00	30.00
22	H1915_	3-5	366	17.84	5.25	0.00	24.00
23	H1923_	4-3	362	22.38	6.53	0.00	30.00
24	H1925_	4-5	362	20.19	5.89	0.00	24.00
25	H1933_	5-3	359	22.23	6.22	0.00	30.00
26	H1935_	5-5	359	17.10	6.28	0.00	24.00
27	H1943_	6-3	173	22.68	6.81	0.00	30.00
28	H1945_	6-5	173	17.06	5.69	0.00	24.00
29	H1953_	7-3	361	24.14	7.33	0.00	32.00
30	H1955_	7-5	361	14.12	5.99	0.00	21.00
31	H1963_	8-3	364	25.74	5.93	6.00	32.00
32	H1965_	8-5	364	16.90	5.95	0.00	21.00
33	H1972_	9-2	171	31.38	6.33	14.00	40.00
34	H1973_	9-3	171	17.18	4.47	0.00	20.00
35	H1982_	10-2	175	31.31	5.23	12.00	40.00
36	H1983_	10-3	175	17.31	4.23	0.00	20.00

表 2b 数学大問の平均得点・標準偏差等

大問	Label	大問名	N	Mean	Std Dev	Minimum	Maximum
1	D2811_	56-1	192	29.53	9.46	0.00	40.00
2	D2813_	56-3	192	30.83	12.88	0.00	40.00
3	D2822_	57-2	188	29.64	12.63	0.00	40.00
4	D2824_	57-4	188	24.22	10.02	0.00	35.00
5	D2842_	59-2	188	30.26	12.07	0.00	40.00
6	D2843_	59-3	188	33.70	8.35	0.00	40.00
7	DE861_	61-1	192	32.28	10.72	0.00	40.00
8	DE864_	61-4	192	22.80	15.84	0.00	40.00
9	DE882_	63-2	165	22.65	12.76	0.00	40.00
10	DE884_	63-4	165	32.87	8.61	0.00	40.00
11	DE893_	1-3	190	27.07	9.67	0.00	40.00
12	DE894_	1-4	190	30.28	9.31	0.00	40.00
13	D2803_	2-1-3	168	26.83	10.63	0.00	35.00
14	E2901_	2-11-1	168	37.38	14.99	0.00	50.00
15	D2913_	3-1-3	362	26.37	6.71	0.00	30.00
16	E2912_	3-11-2	362	35.82	13.09	2.00	50.00
17	D2921_	4-1-1	354	26.01	10.13	0.00	35.00
18	E2921_	4-11-1	354	30.54	17.06	0.00	50.00
19	D2933_	5-1-3	359	24.79	6.52	6.00	30.00
20	E2931_	5-11-1	359	37.70	12.49	0.00	50.00
21	D2943_	6-1-3	171	22.68	7.96	0.00	30.00
22	E2941_	6-11-1	171	42.54	10.19	10.00	50.00
23	D2953_	7-1-3	168	21.77	10.97	3.00	35.00
24	E2952_	7-11-2	357	37.73	12.20	5.00	50.00
25	D2962_	8-1-2	169	22.12	8.73	0.00	35.00
26	E2961_	8-11-1	357	33.72	13.43	0.00	50.00
27	D3072_	9-01-2	167	18.53	9.01	0.00	30.00
28	E2971_	9-11B-1	167	25.32	5.85	0.00	30.00
29	D2982_	10-1A-2	173	34.16	6.67	10.00	40.00
30	E2981_	10-11B-1	173	24.18	7.08	0.00	30.00

4. 2 実験データの項目反応理論を用いた分析

実験で得られたデータを基に、英語と数学ご

とに、項目反応理論の文脈で定義される困難度(b)や識別力(a)を、全冊子の全ての項目について推定した。推定にあたっては、BILOG-MGの2パラメタ・ロジスティク・モデルを用いて分析した(Zimowski, Muraki, Mislevy, & Bock, 1996)。この分析により、実験で用いられた全ての項目は、困難度(b)や識別力(a)といった統計的なパラメタで表現され、共通尺度上に集約的に整理された。

なお、数学の項目の中で、D2953_01とD2982_01は、この実験群において全員が正答したために、以後の分析からは省いた。

4. 3 過去の共通第1次学力試験・大学入試センター試験の実受験者データによる分析

実験に用いられた問題項目は、過去に実施された試験の中から、大問単位で抜粋されたものである。したがって、同じ年度に実施された試験問題であっても、実験に用いられなかった問題については、統計的なパラメタは未知である。そこで、実験で使用された問題と使われなかった問題を比較できるようにするため、過去に全国で共通テストを実際に受験した実受験者のデータを用いて、年度ごとに全ての項目の統計的なパラメタを推定した。

推定にあたっては、実験冊子に含まれなかった項目を含んだ形で、年度ごとに、30~60万人の当該実施年度の全実受験者データを用い、それぞれを単一母集団として分析を行った。また実験と同じく、2パラメタ・ロジスティク・モデルを用いて推定を行った。

なお、英語のH1902_10は、この実受験者群におけるパラメタ推定値が異常値だったため、下記の等化の分析からは除外した。

4. 4 実験データと過去の実受験者データとで推定された項目パラメタの等化

ここまでに、実験で用いられ、年度内の試験の一部抜粋ではあるが、年度をまたいだ共通尺度上に整理された問題項目のパラメタと、実受験者のデータから推定され、年度ごとの独自尺度という制約があるが、年度内の全問題を網羅した問題項目のパラメタが、共にもとめられた。この二つの分析結果は、相互に共通の項目を媒介とすることで、対応づけて分析することができる。すなわち、実受験者データから得られた

年度ごとに独自の尺度を、実験データから求められた年度間の共通尺度に揃えることで、全ての問題項目のパラメタを一元的に等化することができる。

年度ごとの実受験者群を用いたパラメタの推定値を全体の結果に反映させるため、上記の手続きで求めた項目パラメタについて、(2)式を最小とする方法で等化した。その際の集団の重み $w_g=1, g=1, 2, \dots, G$ としては、全集団ともに同一の重みを用いた。そして、大学1年生である実験群における学力の平均と標準偏差をそれぞれ0と1とするように ($q_1=0, r_1=1$) 共通尺度を定義した。

この(2)式を用いた等化の結果得られた集団ごとの等化係数、(q_g, r_g), $g=1, 2, \dots, G$, すなわち、各集団の学力の平均と標準偏差を表3に示す。なお、この結果は、大学1年生である実験群における学力の平均と標準偏差をそれぞれ0と1としたものとなっており、その学力の高さを反映して、実受験者群の平均は全て負となっている。

5. 考察

5. 1 年度間の試験成績の比較を可能にする共通尺度上での項目パラメタの推定と受験者の学力の推定

本研究では、過去の共通テストの中から選ばれた問題をモニター実験で使用することにより、各年度の問題の項目パラメタと、年度ごとの過去の実受験生の学力を、共通尺度上で推定することを試みた。

まず、平成11-13年度の大学1年生を対象としたモニター実験の結果から、異なる年度の試験から抜粋された問題項目について、年度をまたいだ項目パラメタを求めた。一方、年度ごとの問題項目については、過去の実受験者のデータから、それぞれ項目パラメタのセットを求めた。そして、年度ごとの項目パラメタのセットを、実験を通じて推定された年度間で比較可能な項目パラメタと等化することにより、全体をカバーする共通尺度を定め、最終的な項目パラメタを得た。さらに、実験群と年度ごとの実受験者群の学力も推定した(表3)。

本分析では、実験冊子に含まれる項目以外の項目に関して、年度ごとに実受験者群のデータを用いて項目パラメタを推定した結果に表3で示

した等化係数を(1)式という形で反映させれば、共通尺度上での項目パラメタの推定値を求めることが可能であった。したがって、本研究では、実験冊子・実験群を用いたモニター実験を行うことにより、これまでに実施された英語と数学の項目のほぼ半数を共通尺度に乗せること、すなわちテストの(事後的)標準化を可能ならしめたと言える。

表3 集団ごとの学力の平均と標準偏差 (英語と数学)

英語			数学		
集団	平均 (q_g)	標準偏差 (r_g)	集団	平均 (q_g)	標準偏差 (r_g)
1	exprmt	.0000	1	exprmt	.0000
2	1981 56	-1.4897	2	1981 56	-.9869
3	1982 57	-1.4253	3	1982 57	-.5613
4	1984 59	-1.0822	4	1984 59	-1.5151
5	1986 61	-1.2602	5	1986 61	-.6967
6	1988 63	-1.0907	6	1988 63	-.8947
7	1989 1	-1.0109	7	1989 1	-1.1576
8	1990 2	-.8895	8	1990 2	-.2304
9	1991 3	-1.0621	9	1991 3	-.7108
10	1992 4	-1.3899	10	1992 4	-.7541
11	1993 5	-1.5031	11	1993 5	-.7777
12	1994 6	-1.3724	12	1994 6	-.7131
13	1995 7	-1.5505	13	1995 7	-.8636
14	1996 8	-1.3080	14	1996 8	-.7316
15	1997 9	-1.0306	15	1997 9	-.6169
16	1998 10	-1.3800	16	1997 9	-1.1083
			17	1998 10	-1.0558
					.7385

表3に示した受験生の学力に関しては、その時系列的変化として、経年的な学力の上昇や下降といったシステムティックな傾向を見出すことはできなかった。また、特に数学では、年度ごとの実受験者の平均や標準偏差で、大きな変動が見られた。なお、この結果は、大問単位で古典的テスト理論による分析を行った前川・石塚・菊地・内田・中畝(2001)の結果と、ほぼ同様のもとなっている。

このような結果の解釈として、実験群と実受験者群では、学習指導要領の改訂やカリキュラムの違いによって、内容側面ごとの学習時間が変化し、学力そのものが質的に変化している可能性が考えられる。

図1に、平成11-13年度に大学1年生だった実験群と、昭和56, 57, 59年度に実際に共通第1次学力試験を受験した過去の実受験者群の成績を示す。大学に合格している実験群の成績が全体的に高いのは当然として、実施年度がより古い試験では、同じ年度内での大問間の難易度が、実験群と実受験者群とで反転しているものが見受けられる。これは学力の量的な程度の違いでなく、群間で、いわば学力の形といったものが変化していることを示唆しているのかも知れない。

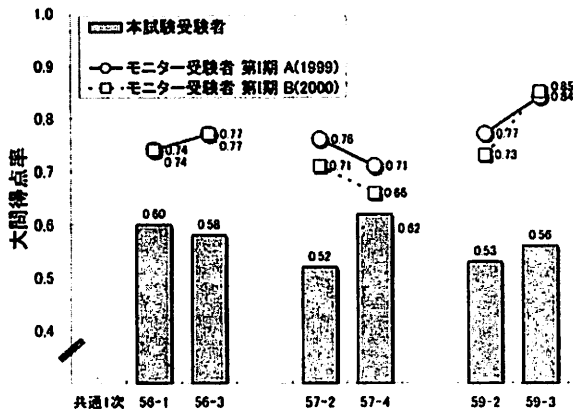


図1 モニター実験の受験者群と過去の実受験者群の成績比較 [数学]

ここで実験群から求められた項目パラメタは、近年の大学生の学力の形を反映し、実受験者群から求められた項目パラメタは過去の受験者の学力のあり方を反映すると考えられる。そこに質的な違いがあるとすると、両者に線形な関係を仮定したパラメタの修正は難しく、表3での数学成績の結果の変動はその微細なぶれによって生じている可能性もある。また実験で用いた数学の問題数は、項目レベルでは若干少なく、それが能力推定の精度を下げている可能性も考えられよう。

5.2 過去の問題の再利用を視野に入れた問題項目についての統合的な統計指標の蓄積

本研究のもう一つの意義としては、過去問題に関する統合的な統計指標の確保とその蓄積がある。先の大学審議会の答申では、良質な試験問題の出題のため、過去の問題の再利用についても検討するよう提言している。本研究は、共通第1次学力試験や大学入試センター試験の過去問題について、年度間の成績比較にも配慮した基礎資料を提供している。

しかし、過去問の再利用については、単に統計的な指標だけを頼りに、安易に行えるものではないことも同時に示された。モニター実験の被験者に対し、試験後にアンケートを行ったところ、過去の問題については、問題形式に慣れておらず、解答しにくいと感じた者が多数いることが見出された(図2)。

さらに、かつて出題された当時は良問だった問題項目でも、時代の趨勢に伴って問題の内容が変質してしまうものが見受けられた。具体的にアンケートの例をあげると、「英文が書かれた

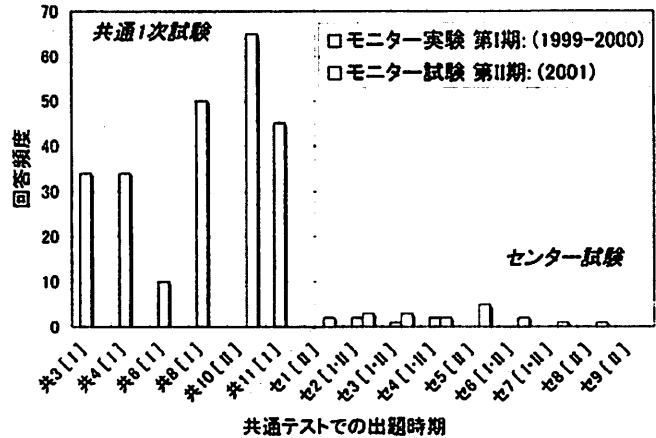


図2 モニター実験のアンケートによる「慣れていない出題形式」の集計結果 [英語]

のが、つい最近のものと思いこんでしまったが、約十数年以上も前の文だったので、困惑してしまった。」「この文章が1980年以前に書かれたものであると注をつけることが、解答のためには必要だと思います。」といった指摘があった。このような指摘を受けたのは、昭和59年(1984年)の英語の長文問題で、近未来の災害予測システムに関わる問題内容であった。そこでは、出題時点では近未来の話題だったのであるが、現時点では既に過去の事項であるというパラドックスを生じてしまっていた。

このように、いわゆる過去の良問であっても、試験問題の鮮度とでもいったような観点から、内容面の精査が必須であることが改めて確認された。

5.3 まとめ

本研究では、過去の問題を用いたモニター実験を実施し、項目反応理論に基づいて項目パラメタを推定した。これは、センター試験標準化の技術的基盤を確立する上での糸口になると考えられる。さらに、これは過去の問題の統合的な指標の確保と蓄積に寄与するものであり、良質な試験問題の出題のための、過去の問題の再利用に供する資料を提供する。

しかし、実験群と実受験群の取り扱いには課題も残る。そこでは、実験群と実受験者群の学力の質的な相違点への配慮、問題形式への順応性の違い、時代の趨勢に伴う問題内容の変質、といった点について、今後の検討が必要である。また、技術的な側面についても、実験冊子・実験群を用いた等化の方法として、全てのデータを

用いてBILOG-MGによる分析を行う方法など、分析手法の一層の改善を図っていく必要があるが、それは次期プロジェクトの課題としていきたい。

参考文献

- 国立大学協会 第2常置委員会 2001 「大学入試センター試験の前年度成績の利用について」の留意事項(平成13年6月1日).
- 前川眞一 1991 パラメタの推定 芝祐順(編) 項目反応理論 東大出版会 Pp. 87-129.
- 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 2002 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－ 大学入試センター研究開発部(編) 大学入学者選抜における評価の標準化に関する研究報告書 大学入試センター Pp. 33-72.
- 文部科学省 大学審議会 2000 大学入試の改善について(答申) (平成12年11月22日).
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. 1996 BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items. Chicago, IL: Scientific Software International.

付記

本研究は、平成11-13年度 大学入試センター研究開発部 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」、及び日本学術振興会科学研究費補助金「基盤研究(B)(展開研究)：大学入試センター試験の標準化の検討(課題番号10551008, 研究代表者：前川眞一)」の援助を受けて行われた。

なお、各大学での実験が実施できたのは、以下の方々のご尽力によるところが大きい。ここに記して感謝の意を表したい。

横浜国立大学 教育人間科学部

総務係長 山中次男(1999年度)

名古屋大学 学務部

入試課長補佐 恒川 亘(1999年度)

同 中村 平(2000-2001年度)

静岡大学 情報学部

教授 山田文康(2001年度)