

昭和62年度共通第1次学力試験 「理科」各科目のテスト特性の比較

研究開発部助手 鈴木規夫
(情報処理研究部門)

1 はじめに

異なるテストの結果を比較する場面はよくみかける。共通第1次学力試験の「理科」は、昭和62年度から従来の2科目選択制から1科目選択制となった。どの科目を選択するかは、試験当日の受験者の判断に委ねられているが、選択による有利・不利をなくすためには、どの科目を選択しても難易度に差異が生じないことが望まれる。しかし、実際には科目間に難易差がないよう問題を作成することは大変難しく、たとえ問題作成者が科目間で難易差が生じないように努力しても、おのずとそれには限界があり、科目によって差異が生じることは十分予測されるところである。そこで、科目間の差異を除去し、その上で相互に比較可能なよう同一の尺度上に表すことが望まれる。

ここでは、項目反応理論 (Item Response Theory) にもとづいて、「理科」の各科目間の差異を除去し、相互

に比較可能なよう同一の尺度上に表わすことにより「理科」の各科目のテスト特性の比較分析を行うと共にテストやそれを構成する項目の評価に関して1つの手続きを指示する。

2 項目反応理論

一般に、テストを実施したとき、項目に正答したときに1を与え、誤答したときに0を与えるならば、そのテストに対する反応パターンは0と1の組み合わせによって表される。そのとき、ある項目に対し、学力の低い者は正答する確率は低く、学力が高くなるにつれ正答する確率は高くなり、やがて1に近づくと考えるのは自然である。このような考え方に立ったテスト得点のモデルに項目反応理論 (Lord, 1980) がある。項目反応モデルは、ある潜在特性 (latent trait) または学力 (ability) を仮定し、ある項目に対し、どの程度の確率で正答するかを関数で近似したものである。ここでは項目反応モデルとして、2パラメタ・ロジスチッ

ク・モデル式

$$P_i(\theta) = \frac{1}{1 + \exp\{-1.7a_i(\theta - b_i)\}}$$

を考える。モデルによれば、学力 θ をもった者の項目*i*に対する正答確率は θ の関数 $P_i(\theta)$ で与えられ、各項目は2つの項目パラメタ a_i , b_i によって特性化される。項目パラメタ a_i は項目識別力と呼ばれ、項目パラメタ b_i は項目困難度と呼ばれている。

モデルは、次の特徴をもっている。①推定された項目パラメタの値は集団の性質によらず不変であり、②推定された学力は特定の選択された項目によらない。すなわち、どのような集団から項目パラメタを推定しても、推定誤差を除けば同じ結果が得られ、項目プールからどのような項目が選ばれたとしても、推定誤差を除けば同一被験者からは本質的に同じ学力の推定値が得られるということを示している。ただ、項目反応理論においては学力を表す尺度の原点と単位は任意であるため、異なる集団を用いて推定された項目パラメタの値を比較するためには学力を共通尺度上に表すための等化の手続きが必要となる。

このような特徴をもつために、いくつかの仮定が設けられている。その1つに学力 θ の1次元性の仮定がある。もし、テストが1次元の学力を測定しているのであるならば、テストを構成

している項目の内容は一貫しており、どの項目からも項目反応曲線で示したような反応を得ることができる。もちろん、テストの結果に与える要因はこの1つの学力に限らず他の別の学力やテストを受けたときの状況 (動機、不安、テスト慣れ) 等が微妙に関係してきている。しかし、その学力がテストに与える影響が支配的と考え、主として1つの学力によってテストの結果は決定され则认为することにする。

また、もう1つの仮定として、局所独立の仮定がある。これは、同一の学力をもつ者に対して、異なる項目に正答する確率は互いに独立であるとする仮定である。

3 分析の手順

分析は、以下の手順を踏んで行う。

- ① 等化を考慮したテストの設計
- ② 項目パラメタの推定
- ③ 項目特性曲線へのあてはまりの検証
- ④ 学力の1次元性の確認
- ⑤ テストの等化のための学力の1次元性の確認
- ⑥ テストの等化
- ⑦ テスト特性の比較

3.1 等化を考慮したテストの設計

テストの等化を行うためには、

- ① 等化の対象となるテストを同一受験者集団が受験しているか (Single-group Design)
- ② 同一受験者集団ではないがそれと同等とみなせる集団が受験しているか (Equivalent-group Design)
- ③ 受験者集団は異なるが、対象となるテストの中に共通な項目が含まれているか (Anchor-test Design)
- のいずれかの条件が必要である (Hambleton & Swaminathan, 1985)。

共通第1次学力試験の性格と昭和62年度以降の「理科」の各科目が1科目選択制であることを考慮して、「理科」各科目間の等化の問題を考えた場合、上記の条件を照らし合わせてみると直接そのまま適用することはできない。しかし、「社会」、「理科」を除く他の教科(例えば、国語、数学や英語)はほぼ全ての受験者が受験している教科であるので、これらの教科の項目を共通項目として利用することが可能である。そこで、これらの教科のうち「理科」と最も相関の高い「数学」の中の「数学I」を「理科」各科目の共通選択項目として利用し、③の方法を適用することによって、「理科」の各科目間の等化を試みることにする。

テストは、「物理」、「化学」、「生物」、「地学」の各科目および共通項目として利用する「数学I」の単独テストと、

テストの等化を行うため、数学Iを共通項目とした「数学I—物理」、「数学I—化学」、「数学I—生物」、「数学I—地学」の合成テストの計9つのテストから構成されている。単独テストあるいは合成テストは、いくつかの項目から構成されているが、分析ではこれらの項目に対する受験者の反応を、完全正解なら1、それ以外では0というように、重み付き得点を1-0に変換した項目得点を用いる。各単独テストおよび合成テストの項目数を表1の左欄に示す。また、分析に用いた項目の項目番号と実際の解答において利用された解答番号との対応表を表2に示す。分析では、「理科」の各科目を受験した者についてそれぞれ満点を除き5,000人をランダムに抽出し、これらの集団を対象とした。ちなみに、物理の受験者は、105,074人、化学は112,960人、生物は109,102人、地学は20,013人であった。

表1 数学Iおよび理科テストの項目数

テスト名	項目数	選別後の項目数	除外した項目番号
物理	23	12	1, 2, 5, 11, 12, 14, 15, 18, 19, 22, 23
化学	31	27	11, 12, 14, 15
生物	33	27	10, 17, 20, 25, 27, 31
地学	34	26	1, 6, 7, 10, 25, 26, 27, 28
数学I	17	13	11, 13, 15, 16
数学I—物理	40	25	
数学I—化学	48	40	
数学I—生物	50	40	
数学I—地学	51	39	

表2 項目一覧

項目番号	解答番号				
	物理	化学	生物	地学	数学I
1	1	1	1	1	1—ア
2	2	2	2	2	1—イウ
3	3	3	3	3	1—エオカ
4	4	4	4	4	1—キ
5	5	5	5	5	1(ク—ツ)
6	6	6-7	6	6	2—ア
7	7	8	7	7	2—イ
8	8	9	8	8	2—ウエ
9	9	10	9	9	2—オカ
10	10	11	10	10	2—キク
11	11	12	11	11	2—ケコ
12	12	13	12	12	2—サシ
13	13	14	13-14	13	2—スセソ
14	14	15	15	14	3—アイ
15	15	16	16	15	3—ウエオ
16	16	17	17	16	3—カキク
17	17	18	18	17	3—ケコサ
18	18	19	19	18	
19	19	20	20	19	
20	20	21	21	20	
21	21	22	22	21	
22	22	23	23-24	22	
23	23	24	25-26	23	
24		25	27-28	24	
25		26	29	25	
26		27	30	26	
27		28	31	27	
28		29	32	28	
29		30	33	29	
30		31	34-35	30	
31		32	36	31	
32			37	32	
33			38	33	
34				34	

3.3 項目特性曲線へのあてはまりの検証

モデルを適用する際には、テストの対象となる項目データがそのモデルを十分説明しているか、モデルへの適合性について検討を加える必要がある。ここでは、等人数(500人)になるよう学力 θ を10の階級に分けたときのそれぞれの階級の正答率と各階級の中央の値に対応するモデルの期待値との適合度を χ^2 統計量によって検定(df=9, $\alpha=0.05$)する。検定の結果有意となった項目のうち特に適合度の悪い項目を分析の対象から外すことにする。検定は、項目パラメタが各科目から推定された場合と、合成テストから推定された場合のいずれの場合についても行い、いずれかで適合度の悪かった項目を除外する。

3.4 学力の1次元性の確認

項目反応理論の大きな特徴である1次元性の仮定の確認については、この理論を適用する際の必要不可欠の条件である。この仮定を確認する方法として、一般には因子分析を利用する場合が多いが、ここでは、項目パラメタの性質を利用して、次のような手続きによって確認する。

- ① 単独テスト(例えば物理のテスト)と合成テスト(例えば数学I—物理からなる合成テスト)の2つのテ

3.2 項目パラメタの推定

単独テストおよび合成テストの項目パラメタを「GETAB」(Baillie, 1980)によって推定する。

トを考えてみる。

- ② 今、単独テスト（項目数 n ）が 1 次元の学力を測定していると仮定する。この仮定の下で、分析の対象となる受験者によって、単独テストの項目パラメタ $a_i, b_i, (i = 1, 2, \dots, n)$ を推定する。
- ③ 次に、合成テスト（項目数 $m, m > n$ ）も 1 次元の学力を測定していると仮定して、同一受験者を利用して、項目パラメタ $a_i^*, b_i^*, (i = 1, 2, \dots, m)$ を推定する。
- ④ もし、2 つのテストが同一の学力を測定しているのであるならば、2 つのテストに共通な項目の項目パラメタは同じ値をもつはずである。すなわち、 $a_i = a_i^*, b_i = b_i^*$ の関係が成り立つはずである。言い換えれば、2 つのテストの項目パラメタをプロットしたとき、各点 (a_i, a_i^*) および (b_i, b_i^*) は傾斜 1 で切片 0 の直線上 ($Y = X$) に散布する。
- ⑤ しかし、もし項目パラメタの推定値が $Y = X$ の直線上から外れるならば、その項目に正答するか否かは、テストに含まれる項目に依存することになり、すべての項目が一つの学力で説明されるという仮定と矛盾が生ずることになる。すなわち、その項目は 1 次元ではなく、他の学力も測っていることになる。

3.5 テストの等化のための学力の 1 次元性の確認

テストの等化を行う際の条件として、対象となるテスト間で測定している学力が同一であること (Lord, 1980, Hambleton & Swaminathan, 1985) があげられている。前節の 1 次元性の確認は、単独テストおよび合成テストへのモデルの適用可能性を確認するためのものであって、科目間での 1 次元性の確認は行っていない。そこで、この点について、先の各テスト毎に確認した方法を利用し、次のような手続きで確認する。

- ① 今、全ての単独テスト（数学 I、物理、化学、生物、地学）が 1 次元であるという仮定の下で、「数学 I—物理」、「数学 I—化学」、「数学 I—生物」、「数学 I—地学」の合成テストの項目パラメタを利用してテストの等化を行う。そのとき、等化によって得られた項目パラメタをそれぞれ $a_i, b_i (i = 1, 2, \dots, n)$ とする。
- ② 次に、各単独テスト毎に 1 次元性を仮定して、それぞれの単独テストの項目パラメタを推定する。その値を $a_{ij}, b_{ij} (j = \text{数, 物, 化, 生, 地}, i = 1, 2, \dots, n)$ とする。
- ③ もし、①の仮定が正しいのであれば、項目特性曲線の性質により、 a_i と a_{ij} および b_i と b_{ij} は線型変換で一致

するはずである。すなわち、 (a_i, a_{ij}) と (b_i, b_{ij}) の相関係数は全ての i に関して 1 となるはずである。

3.6 テストの等化

合成テスト j が項目 i を含むとき、その合成テストを用いて推定された項目パラメタの値を a_{ij}, b_{ij} 、また、同様に推定される受験者の学力の値を θ_j とすると、これらは、等化後の共通尺度上でのパラメタの値 a_i, b_i, θ とすると、次の関係をもつと想定される。

$$a_i = a_{ij} / r_j$$

$$b_i = r_j b_{ij} + q_j$$

$$\theta = r_j \theta_j + q_j$$

q_j および r_j は等化を行うための係数で、等化前の項目特性曲線 $P_{ij}(\theta_j)$ と等化後の項目特性曲線 $P_i(\theta)$ との 2 乗誤差を最小にするように選ばれたものである（詳しくは、前川・鈴木 (1988) を参照）。一度、 q_j および r_j が推定されると、等化後の値は上式によって簡単に変換することができる。

このとき、学力 θ_j が平均 0、分散 1 の分布を仮定すると、合成テストを受験した者の共通尺度上における学力分布の平均 $E(\theta)$ は、下式で与えられる。

$$E(\theta) = q_j$$

また分散 $V(\theta)$ は、

$$V(\theta) = r_j^2$$

で与えられる。

3.7 テスト特性の比較

等化後の各科目のテストの特性について比較分析を行う。

4 分析の結果と考察

4.1 項目パラメタの推定

図 1 は推定された単独テストの項目パラメタの分布と合成テストの項目パラメタの分布を平行箱形図 (Tukey, 1977) によって示したものである。

分布をみると、項目識別力の中央値は、物理で大きく、分布も拡がりを見せており、他の 3 科目と異質な傾向を示している。しかし、共通項目として数学 I を加えた場合、数学 I の影響で各科目共中央値は大きくなる傾向を示し、分布も類似した傾向を示すようになる。これは、数学 I の項目の項目識別力が全般に大きく、このことが影響したものと考えられる。物理と数学 I の項目識別力は他の科目に比べ項目識別力の大きい項目から構成されていることが分かる。

これに対し、項目困難度の分布をみると、中央値は科目間の差異は小さく、分布も類似した傾向を示している。共通項目を加えた場合も同様な傾向を示している。

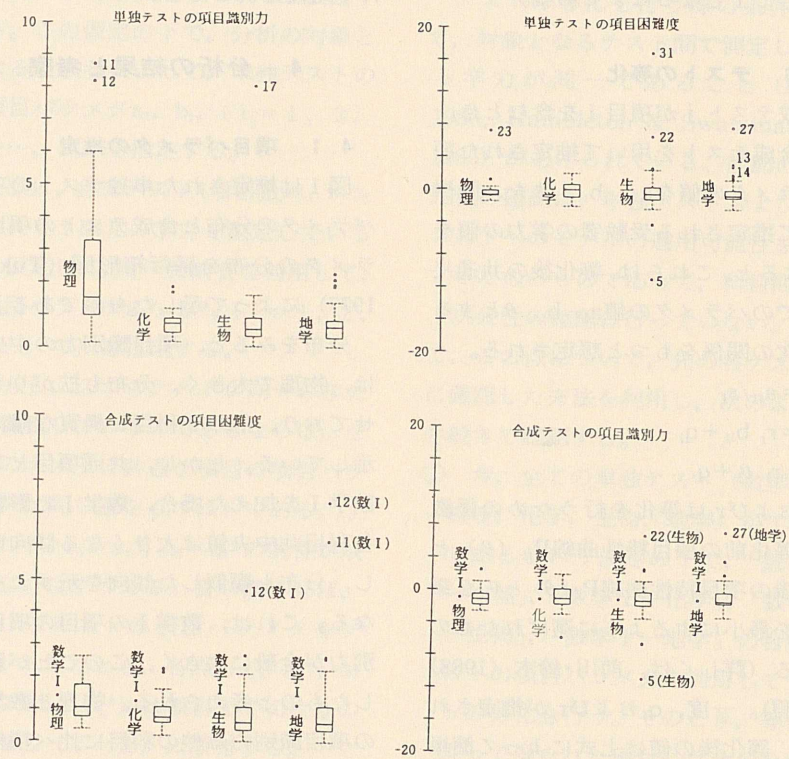


図1 項目パラメタの分布

4.2 項目特性曲線へのあてはまりの悪さによる項目の選別

項目特性曲線へのあてはまりの悪い項目は、物理で2、化学で2、生物で5、地学で4、数学Iで4つあった。これらの項目は以降の分析の対象から

除外した。除外した項目は、表1の右端の欄に下線を付した項目番号がそれである。

4.3 学力の1次元性の確認と項目の選別

学力の1次元性の確認を行うため単

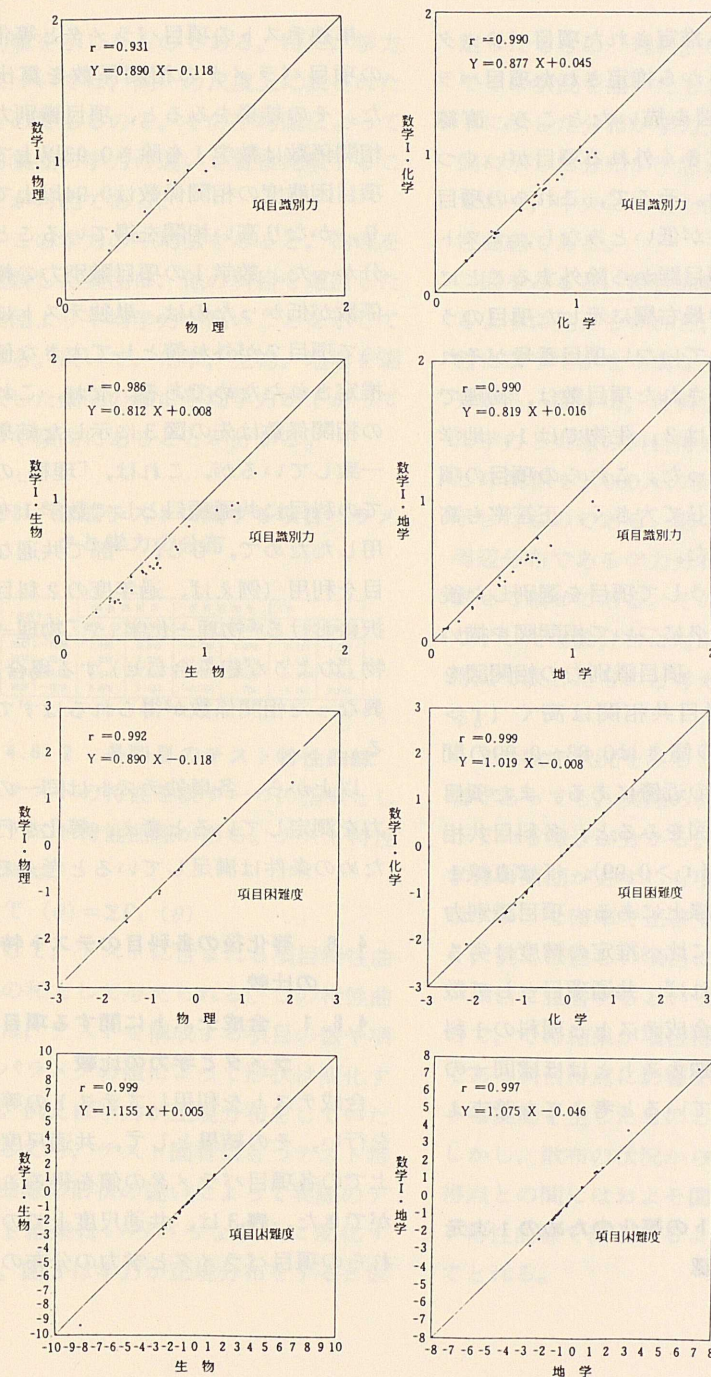


図2 項目識別力等の相関図

単独テストから推定された項目パラメータと合成テストから推定された項目パラメータの散布図を描いたところ、直線 $Y=X$ から大きく外れる項目がいくつか見つかった。そこで、これらの項目は、1 次元性が低いとみなし、テストを構成する項目群から除外することにした。表 1 の最右欄に示した項目のうち下線がついていない項目番号がそれである。除外された項目数は、物理では 9、化学では 2、生物では 1、地学では 5 つであった。これらの項目の項目識別力は総じて大きく、正答率も高い傾向を示した。

図 2 は、こうして項目を選別した後、項目パラメータについて相関図を描いたものである。項目識別力の相関図を見ると、各科目共相関は高く ($r > 0.93$)、直線の傾きは $0.83 \sim 0.89$ の間で、切片も 0 の近傍にある。また項目困難度の相関図を見ると、各科目共相関係数は高く ($r > 0.99$)、ほぼ直線 1 で傾き 0 の直線上にある。項目識別力は項目困難度に比べ推定の精度は劣ることを考慮すれば、共通項目として数学 I を加えた合成テストと理科の 1 科目からなる単独テストとはほぼ同一の学力を測定していると考えても差支えないと考える。

4.5 テストの等化のための 1 次元性の確認

単独テストの項目パラメータと等化後の項目パラメータの相関係数を算出した。その結果をみると、項目識別力の相関係数は数学 I を除き 0.93 以上で、項目困難度の相関係数は 0.94 以上であり、かなり高い相関を得ていることが分かった。数学 I の項目識別力の相関係数が低かったのは、単独テストにおいて項目 7 が外れ値として大きな値で推定されたためである。なお、これらの相関係数は先の図 3 に示した結果と一致しているが、これは、「理科」の全ての科目に共通項目として数学 I を利用したためで、もし、一部で共通な項目を利用（例えば、過年度の 2 科目選択における「物理—化学」や「物理—生物」のような組み合わせ）する場合は、異なった相関係数が得られるはずである。

以上から、各単独テストは同一の学力を測定していると考え、等化を行うための条件は満足していると考えられる。

4.6 等化後の各科目のテスト特性の比較

4.6.1 合成テストに関する項目パラメータと学力の比較

合成テストを利用してテストの等化を行い、その結果として、共通尺度の上での各項目パラメータの値を得ることができた。表 3 は、共通尺度上でのこれらの項目パラメータと学力の分布の要

約値を示したものである。特に、学力については、同一の尺度上に表されたものであるので、その平均値によって各集団の学力の違いを直接比較することが可能である。

この学力の平均値をみると、物理を受験した集団は、他の科目を選択した集団より平均学力が高いことを示している。次いで、化学、生物、地学を選択した集団の順に平均学力が下がっていく傾向にあることが分かる。

表 3 合成テストに関する項目パラメータと学力の分布

合成テスト	項目識別力		項目困難度		学 力	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
数学 I—物理	1.069	0.640	-0.776	1.039	0.339	1.089
数学 I—化学	1.037	0.532	-0.408	0.993	0.101	0.757
数学 I—生物	0.844	0.608	-1.125	2.129	-0.346	1.001
数学 I—地学	0.809	0.635	-0.758	1.704	-0.467	0.944

4.6.2 各科目のテスト特性曲線

テストの特性を表す一つの指標としてテスト特性曲線がある。テスト特性曲線は、

$$T(\theta) = \sum P_i(\theta)$$

として、テストに含まれる項目特性曲線の和として与えられる。この特性曲線は、テストを構成する項目の数や項目パラメータの値によって形状は変化する。例えば学力が正規分布をしていたとしても、テスト固有にもつテスト特性曲線の形状の違いによって実際のテスト結果はいろいろな形状に変化する。図 3 は学力が正規分布をすると仮

定したときの「理科」の各科目についてその状況を描いたものである。図中、下に示した分布が学力分布であり、左側に示した分布が予想される得点分布であり、中央に示した曲線がテスト特性曲線である。

図をみると、特性曲線は類似しているように見えるけれど、得点分布は科目によってかなり異なっていることが分かる。図 4 は、各科目毎に推定された学力と実際の科目得点（100 点満点）との関係を 1,000 人の受験者について示したものである。図中には、さらに周辺分布である学力分布と得点分布も併せて載せてある。その意味で、図 3 で示した理論分布に対応して、実際の得点で表したものと考えることができる。

この散布状況をみると、同一学力の者であっても、実際の科目得点は変動している様子が分かる。これは、その学力の集団が必ずしも項目特性曲線で示すような確率で正答するわけではなく、例えば易しい項目に誤答し、難しい項目に正答するといった反応をするので、その結果が項目得点の積み重ねである科目得点に影響を与え、このような変動を生じたものと推測される。しかし、散布の状況から、学力と科目得点との間にはおよそ図 3 で示すテスト特性曲線で表されるような関係が見てとれる。

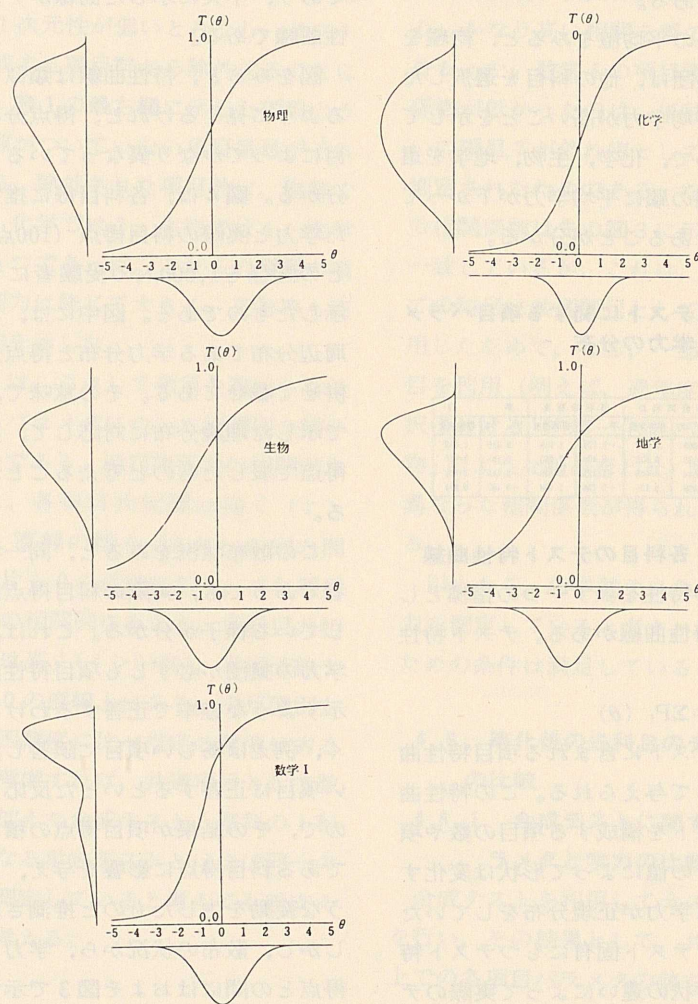


図3 テスト特性曲線

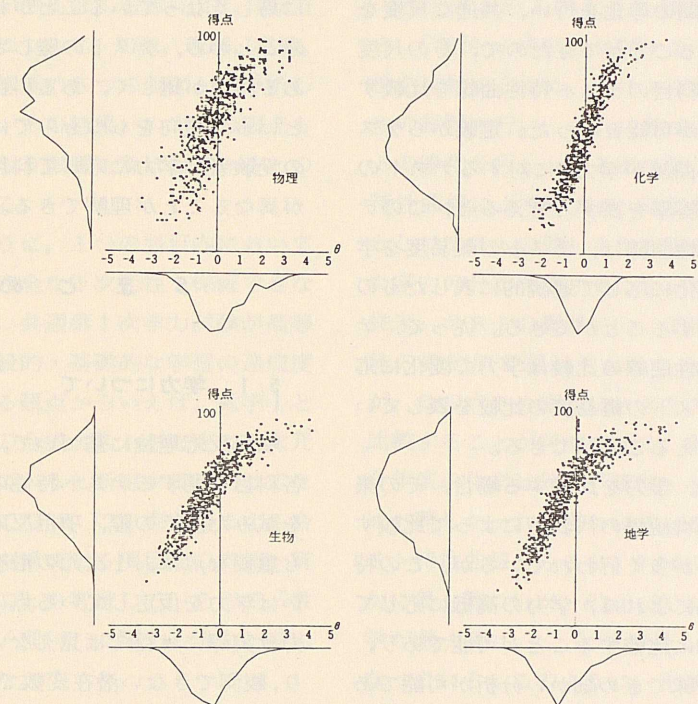


図4 学力 θ と科目得点との関係

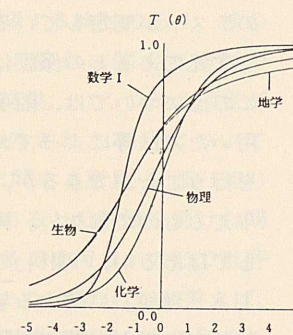


図5 等化後の数学 I・理科各科目のテスト特性曲線

4.6.3 テスト特性曲線の比較

科目間の等化を行い、共通な尺度を構成することができたので、その尺度上で各科目のテスト特性曲線を比較することが可能となった。定義からテスト特性曲線が学力 θ に対するテストの平均正答率を表すと考えると、このテスト特性曲線は、テストの難易度を学力の変化に応じて連続的に表したものと解釈することができる。従って、テスト特性曲線の比較は学力の変化に応じたテストの難易度の比較を表していると考えることができる。

従来、学力を比較する場合、その集団の平均値等の代表値によって比較することが多く行われているが、この特性曲線によれば、学力の高低に応じて連続的に比較することが可能であり、その意味できめ細かい分析が可能である。図5は、各科目のテスト特性曲線を比較するためまとめて示したものである。

図をみると、 $\theta=0$ では、テスト特性曲線の値は数学Ⅰで最も大きく、生物、地学、物理、化学の順となっている。すなわち、学力 θ が0の者にとっては、数学が最も易しい科目であり、次いで、生物、地学、物理、化学の順に難しくなっていることが分かる。同様に $\theta=-3$ では、数学Ⅰが最も難しく、化学、物理の順に易しくなっている。さらに、 $\theta=3$ では、地学が最も難しく、

次いで生物、物理、化学、数学Ⅰの順に易しくなっていることがわかる。このことから、テストの難しさは一様にある科目が難しく、ある科目が易しいといった傾向をもつものではなく、その受験者の学力に応じて科目の難しさが異なることが理解できる。

5 ま と め

5.1 学力について

項目反応理論に基づいて、「理科」の各科目に関するテスト特性の比較分析を試みた。その際、項目反応理論で最も重要な点は、1次元の潜在特性あるいは学力を仮定している点にある。学力は実際には目では見えないものであり、観測できない潜在変数である。我々は、その学力を実際に観測するテストの得点によって推定する。しかし、そのテストが測定している学力が完全に1次元であるとの保証は得られない。この点については、因子分析やここで用いた方法等によっておおよその検証を行うことができるが、100%の保証を与えるものではない。実際、当然のことではあるが、「理科」の各科目は数学Ⅰと「理科」に関する学力以外に、科目独自の学力に関する因子もあるはずである。また、科目内の大問毎にまとめられた項目間では、それぞれ異なっ

た学力に関する因子も含まれているはずである。例えば、石塚・山田(1987)は、化学について、知識、理解、実験、思考力、態度の5つの側面から分析しているが、その結果、いずれの観点も大なり小なり影響を及ぼしているとの報告もある。

このように、1つの科目内においてさえも、完全な1次元性を保証できないことは、共通第1次学力試験が高等学校の一般的・基礎的な学習の達成度を測定する観点からいえば、数学Ⅰと「理科」の合成テストが完全に1次元を測定していると断定することはできない。ここでは、この点を踏まえ、合成テストを構成する項目間の内部一貫性が高いことを考慮して、あえて「数学Ⅰ」と「理科」に関する学力と呼ぶことにした。

分析では、この学力の1次元性を2つの側面から検討を行った。1つは、理論の適合に関する検討であり、他の1つはテストの等化に関する検討であった。その結果、いくつかの項目が除外された。除外された項目については、試験問題に関する意見・評価(1987)の中で良問であるとの評価を受けた問題もあるが、多くは何らかの形で批評が加えられていた。その点から言えば、この項目反応理論を利用した項目分析手法は、問題作成の段階における有用な情報を提供するものと思われる。

5.2 テスト特性の比較について

各テストの1次元性を確認しながら、「理科」の各科目に関するテスト特性を比較するため、テストの等化のための設計を行い、これに基づいてテストの等化を行った。その結果、テスト特性曲線により学力 θ の尺度上の関数としてテストの難しさを表現でき、また学力の分布を定めることができた。従来、テストの難易については集団全体の要約値で議論されることが多いが、この点きめ細かにテストの難易を比較することが可能である。例えば、今年度の化学は他の「理科」の科目に比べ難しかったと言われているが、テスト特性曲線をみると、 $\theta<0.5$ では確かに最も難しいが、 $\theta>0.5$ では逆に地学が最も難しく、化学が最も易しくなっている。このように、化学が難しい科目であったとしても、学力の違いによってその難しさは異なることに留意する必要がある。

項目反応理論は、大問主義からなる日本のテストには適していないかもしれないが、テストの1次元性を確認しながらテストの等化を行い、「理科」の各科目のテスト特性をきめ細かに比較することができた。テストの比較を行うといった場面はよくみかける。テストを比較する際には、先に述べた条件が揃ってこそ、正しく判断が下せるのであって、比較を行う際の状況を考慮

に入れずに行うべきではないと考える。ここでは、その一つの手続きを提示した。

参考文献

大学入試センター (1987) 昭和62年度共通第1次学力試験の試験問題に関する意見・評価

石塚智一・山田文康 (1987) 共通第1次学力試験の試験問題の分析—昭和57年度「化学I」を中心として—
大学入試センター研究紀要 No16

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タの共通尺度上での等化
大学入試センター研究紀要 No17

Baillie, R. 1980, *GETAB: A com-*

puter program for estimating item and person parameters of the one- and two-parameter logistic model on the PLATO system. Urbana: University of Illinois, Computer-based Education Research Laboratory.

Hambleton R.K. & Swaminathan H. 1985, *Item Response Theory—Principles and Applications*. Kluwer・Nijhoff Publishing.

Lord F.M. 1980, *Applications of Item Response Theory*. Lawrence Erlbaum Associates, Inc.

Tukey J.W. 1977, *Exploratory Data Analysis*. Addison-Wesley.