

【原著】

## 項目困難度による総合試験の問題内容分析

伊藤圭，大久保智哉（大学入試センター），柳井晴夫（聖路加看護大学）

大学入試の多様化に伴い，基礎学力を評価する試験や，問題解決力または志望学部・学科への適性などを総合的に評価する総合試験問題の有用性についての検討が進められている。本研究では教科科目フリー型総合試験のモニター調査データを用いて項目困難度を受験者の学部系統別に比較し，特に差異の大きな項目について問題内容分析を行った。その結果，理系と医学系において項目応答に類似性があること，項目の内容的妥当性が学部系統間の項目困難度の差に関連することなどが示唆された。

### 1 はじめに

近年，大学入試の多様化が進み，教科科目の学習到達度だけではなく，志望学部・学科への適性或基本的な学習能力などを総合的に評価するための総合試験問題を個別試験で利用する大学が増えてきた（伊藤，2006）。また，専門職系の大学院等においても，入学後の高度な学習に耐えうる基礎的能力が備わっているかどうかの評価を幅広い出身分野の受験生に対して行う必要性から，入学試験に特定の教科科目の知識に依らない教科科目フリー型の適性試験や総合試験を利用することがある。

総合試験問題の内容とその性質についての研究は，大学入試センターや総合試験に比較的高関心が高い学部の研究者などを中心に調査研究が行われてきた（柳井ほか，2006；石井ほか，2005；椎名・柳井，2001；倉元・柳井，2001）。著者らは，これまでに得られた知見に基づき，教科科目フリー型総合試験の実際の応用例として，医学部学士編入学者選抜用の問題を作成してモニター調査を行い，得点データから，難易度，識別力，信頼性，妥当性等に関する基本的な特徴を調べた（伊藤ほか，2006）。これらの試験の作題上の課題の一つは，特定の教科科目固有の知識に依存したり，特定の学部・学科（高校の場合は

普通科，専門学科，総合学科などの学科）の出身者だけに有利または不利に働いたりするような内容を問題項目からできるだけ排除することである。

学部系統別の問題項目の特徴については，赤根ほか（2006）によって識別指数を用いた分析が行われているが，さらに困難度の観点から分析を行うことも可能である。今回，著者らは，典型的な教科科目フリー型総合試験の項目困難度を学部系統間で比較し，集団間で項目の持つ意味，すなわち項目の測定内容に差異があるかどうかを確認するとともに，差異が認められた項目については，問題内容分析によってその要因の特定を試みた。

### 2 総合試験とアンケートの構成

今回分析対象とした教科科目フリー型総合試験は医学部学士編入学者選抜用の試験を想定して作成されたものであり，①第1部「情報把握・論理的思考」（35項目，100分），②第2部「コミュニケーション・読解・表現」（35項目，80分）の2部構成である。どちらの問題も医学部の教員の多くが入試で測定する必要があると考えているもので，特定の教科科目や医学の知識を前提とせず，医学生として，あるいは将来医師として様々な課題を遂行していくために必要な基

本的能力を問うものである。問題内容（全設問と選択枝）については先行研究（伊藤ほか, 2006）の付録を参照されたい。

また、様々な課題を遂行するために必要な能力・資質の習得度に関する自己評価アンケートを作成した。能力・資質がどの程度身についているかについて「身につけていない」「あまり身につけていない」「少し身につけている」「身につけている」の4件法で回答を求めるものである。質問は47項目で構成されており、各項目は「情報処理・数理的素養」「創意・多元的判断・論理的思考」「読解力・表現力」「对人的親和性・献身性」「自然・社会・人間への関心」「芸術への関心」という6種類のカテゴリーのいずれかに分類される。

### 3 モニター調査の概要

今回の分析は全国の国公立大学43校から募集した学部3～4年生753名に実施したモニター調査のデータに基づいている。この調査では、先に述べた第1部および第2部の総合試験問題に解答してもらおうと同時に、学力や適性に関するアンケートにも回答してもらった。

受験者の学部系統別の内訳は、文系225人、理系293人、医学系214人、その他21人であり、性別の内訳は男性469人、女性284人であった。また、総合試験の得点率は第1部、第2部ともに60%程度であった。

### 4 項目機能の差異の検証について

項目の特性を表す基本的な量として、困難度と識別力が一般に利用されている。例えば、項目正答率は困難度を、成績上位者25%と下位者25%の正答率の差で表される識別指数は識別力を表す方法の一つである。

項目機能の差異の検証方法として、まずこれらの量を学部系統別に比較することが考え

られる。しかしながら、そもそも学部系統別に分けられた集団の間に正答率や識別指数の差が存在したとしても、その原因が試験の測定対象である能力の差によるものなのか、それとも測定対象以外の要素による特異的な項目機能によるものなのかが明確には判別できない場合も考えられる。そのような場合には、例えば試験の総点、得点に関連する共変量や属性、外的基準等に基づいて各集団を条件づけた上で、集団の等質性や能力の同等性を仮定して比較するなどの工夫が必要となるであろう。

米国のような多民族国家では、例えばテストが測定している能力とは別の要因によって、特定の民族にとって著しく不利になるような項目が生じていないかどうかを検証する場合に、差異項目機能分析が行われることがある。項目機能の差異の分析は測定の等価性の検証として見ることもでき、様々な分析方法が考案されている（田崎, 2007）。その分析方法の一つとして項目応答理論（IRT）に基づく方法も多く用いられている（孫, 2005; 熊谷・脇田, 2003; 前田・野口, 2006）。IRTによる方法では、差異項目機能を検証しようとしている項目以外の項目については、受験者集団間での等質性を仮定する。すなわち、両集団間で項目に対する反応が構造的には等しいとする。その上で、関心下の項目に関して集団ごとに母数を推定し、その値を検討することによって、その項目に集団に対して特異的な作用が働くかを検証する。比較する受験者集団間の等質性を関心下の項目以外で制約として設けるかどうかは正答率の差を検討する方法とは異なる。

本研究では、総合試験問題の各項目について文系、理系、医学系の3つの集団間の差異を検証するため、IRTに基づき、項目困難度の差の評価を行うこととした。具体的には、比較対象となる二つの集団  $f$  と  $r$  に対して、IRTの2パラメータ・ロジスティック・モデル

表1 学部系統間の困難度差（上位5項目）

第1部

理一文			医一文			医一理					
	$\Delta b$	$\beta$		$\Delta b$	$\beta$		$\Delta b$	$\beta$			
第1問-問1	0.959	4.098	理	第6問-問1	-1.308	3.564	文	第8問-問1	0.849	1.722	医
第6問-問1	-0.802	3.457	文	第6問-問3	-1.090	2.995	文	第8問-問3	0.744	1.661	医
第6問-問4	-0.889	3.152	文	第1問-問1	0.719	2.923	医	第5問-問1	-0.289	1.579	理
第6問-問3	-0.949	2.956	文	第11問-問1	1.569	2.890	医	第4問-問4	-0.337	1.560	理
第9問-問1	0.520	1.926	理	第6問-問4	-0.867	2.797	文	第9問-問3	-0.452	1.532	理

第2部

理一文			医一文			医一理					
	$\Delta b$	$\beta$		$\Delta b$	$\beta$		$\Delta b$	$\beta$			
第15問-問3	-0.718	2.294	文	第9問-問2	0.776	4.732	医	第9問-問2	0.763	4.891	医
第15問-問2	-0.572	2.279	文	第10問-問1	-0.385	1.964	文	第15問-問2	0.477	2.178	医
第10問-問1	-0.276	1.568	文	第12問-問3	-0.788	1.611	文	第3問-問1	-1.020	1.380	理
第8問-問2	0.802	1.488	理	第15問-問3	-0.511	1.396	文	第14問-問1	-0.569	1.245	理
第12問-問3	-0.455	1.217	文	第14問-問3	0.543	1.368	医	第14問-問2	0.602	1.171	医

ルを用いて項目識別力母数の等値

$$a_{if} = a_{jr} \quad (j \text{ は項目番号})$$

を仮定した上で項目困難度の差

$$\Delta b_{j,fr} = b_{if} - b_{jr}$$

を推定し、困難度の差の大きさ（絶対値）の指標

$$\beta_{j,fr} = |\Delta b_{j,fr}| / SE_{j,fr}$$

を算出した。ここで、 $SE$ は $\Delta b$ の推定値の標準誤差である。次に、 $\beta$ が大きい項目に対して問題内容分析を行った。また、各項目の内容や計算結果への影響等を検討した結果、総点との相関が低い項目に計算結果の妥当性を低下させるものが散見されたため、分析対象とする項目を総点との相関が0.15以上のものに限った。その結果、第1部で31項目、第2部で16項目となった。

5 困難度差の推定結果と問題内容分析

表1は、理系一文系間、医学系一文系間、医学系一理系間の項目困難度差の推定結果のうち、値の大きい上位5項目を抽出したもの

を第1部および第2部のそれぞれについて示したものである。項目困難度差の数値だけでなく、当該の問題項目の困難度がどの学部系統で高いかを、それぞれ「文」「理」「医」の文字で示してある。また、図1および図2はそれぞれ第1部および第2部の問題項目の正答率を学部系統別に示したものである。学部系統間での正答率の差が大きい順に左から右へ項目を並べてある。これらの正答率に関する結果は先行研究（伊藤ほか,2006）で得られたもので、今回の項目困難度差の推定による結果と比較することができる。なお、学部系統間の差の大きさは各学部系統の得点率の差の有意性を検定する一元配置分散分析によるF統計量の値を用いている。

5.1 抽出された項目について

表1の結果では、第1部、第2部のどちらにおいても、理系一文系間の結果と医学系一文系間の結果に共通項目が多く見られる（第1部では第1問-問1、第6問-問1・問3・問4の4項目、第2部では第10問-問1、第12問-問3、第15問-問3の3項目）。したがって、理系と医学系が比較的

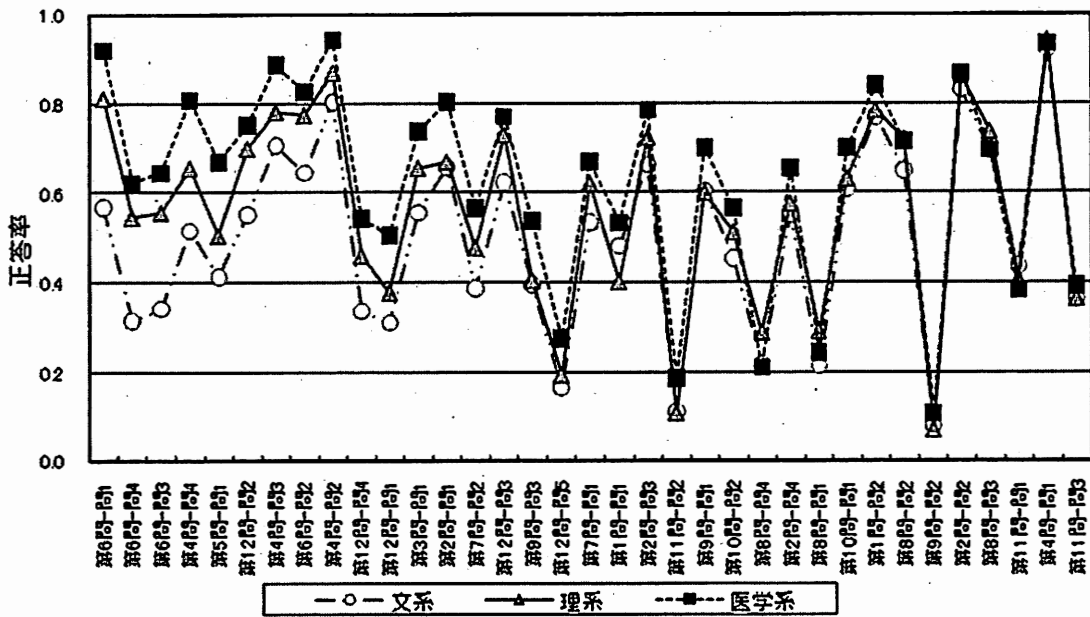


図1  $F$  値の順に並べた学部系統別の正答率 [ 第1部 ] (伊藤ほか, 2006 より)

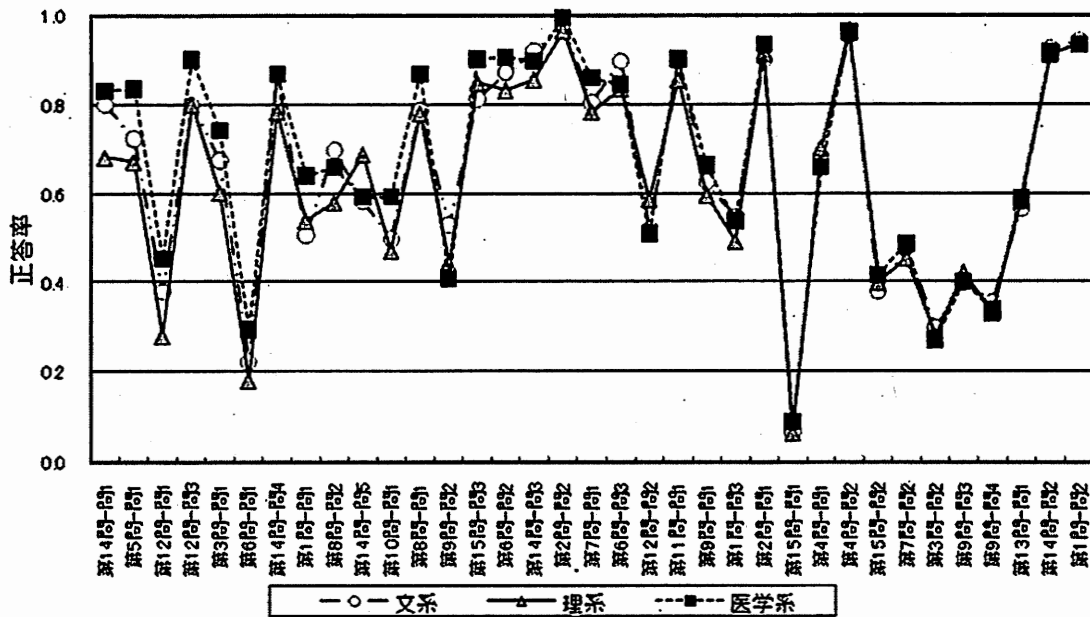


図2  $F$  値の順に並べた学部系統別の正答率 [ 第2部 ] (伊藤ほか, 2006 より)

同じような項目応答をするのに対して、文系だけがやや異なる項目応答をする集団であることが推測される。

次に、今回採用した項目困難度差 ( $\beta$  値) を用いた手法と単純に正答率の差 ( $F$  値) を用いた手法とを比較するために、学部系統間の差が大きい項目として、表 1 の結果と図 1 および図 2 の上位 10 項目とを比較する。第 1 部では、第 4 問-問 4、第 5 問-問 1、第 6 問-問 1・問 3・問 4 がどちらの手法でも上位に挙がっているのに対して、第 1 問-問 1、第 8 問-問 1・問 3、第 9 問-問 1・問 3、第 11 問-問 1 は項目困難度差を用いた場合に上位に挙がっており、今回の手法によって初めて項目機能の差が顕在化したものと考えられる。同様に、第 2 部では、第 3 問-問 1、第 8 問-問 2、第 12 問-問 3、第 14 問-問 1 が共通に上位に挙がっているのに対して、第 9 問-問 2、第 10 問-問 1、第 14 問-問 2・問 3、第 15 問-問 2・問 3 は項目困難度差を用いた場合に上位に挙がっている。

以下、主要な項目の分析結果について述べる。

## 5.2 第 1 部問題

本節では第 1 部問題で抽出された問題項目について分析を行う。次の (1) ~ (3) は、困難度差  $\beta$  を用いた場合の上位 5 項目以内と正答率差  $F$  を用いた場合の上位 10 項目以内のどちらにも含まれた問題、(4) ~ (7) は  $\beta$  を用いた場合の上位 5 項目以内のみ含まれた問題である。また、表 2 はこれらの問題項目と能力・資質の習得度との相関係数を当該項目が属する大問単位で示したものである。

### (1) 第 4 問-問 4

正方形とその内接円の面積比と、正方形内に無作為に落とした針が内接円内に落ちる確

率との関係を利用して、円周率  $\pi$  を繰り返し針を落とす確率実験の結果から求める問題。先に出題されている設問 (問 1 ~ 問 3) から順に解答していく導出的な形式となっている。問 1 ~ 問 3 までは常識的な知識と簡単な計算によって、ほぼ機械的に解答できるが、問 4 では面積比と確率実験を関連づける際に、より論理的思考力が問われる。表 1 では医学系に対して理系の困難度が高いことが示されているが、問題内容は比較的単純であり、能力・資質の習得度との相関 (表 2) の結果からも特に理系で困難度が高くなると思われる目立った要因は見受けられない。また、困難度差の推定値  $\beta$  は 1.560 で、それほど大きな値ではない。

### (2) 第 5 問-問 1

4 種類の果物の「好き」「嫌い」について成立する 3 つの条件を提示し、これらの条件に合わない命題を答えさせる典型的な論理問題である。与えられた条件下で起こり得る全ての場合を書き出した表を作成し、命題の成否を一つずつ確認することによって解答を導くことができる。識別指数の観点から行われた分析 (赤根ほか, 2006) では、論理的思考力だけでなく、複雑な条件文や選択枝の文章表現を正確に把握する必要性が述べられている。但し、読解力・表現力との有意な相関は見られず、情報処理・数理的素養との間に有意な相関が見られる (表 2) ことから、本問で必要な文章表現を正確に把握する能力は、いわゆる長文の読解力や表現力のことではなく、問題文に与えられた情報を適切に処理する情報処理能力に類するものと考えられる。表 1 では医学系に対して理系の困難度が高いことが示されているが、特に理系で困難度が高くなると思われる目立った要因は見受けられず、 $\beta$  も 1.579 で、それほど大きな値ではない。

表2 抽出された項目と能力・資質との相関係数 [第 1 部] (\* :  $p < 0.05$ , \*\* :  $p < 0.01$ )

		情報処理・ 数理的素養	創意・多元的 判断・論理的思考	読解力・表現力	対人的親和性・ 献身性	自然・社会・ 人間への関心	芸術への関心
第1問	問1	0.061	-0.007	0.084 (*)	-0.185 (**)	-0.065	-0.070
	問2	0.175 (**)	0.067	0.087 (*)	-0.058	0.038	-0.047
第4問	問1	0.095 (*)	0.060	0.006	0.006	0.031	0.057
	問2	0.157 (**)	0.030	0.065	-0.097 (**)	-0.002	-0.046
	問3	0.161 (**)	0.072	0.027	-0.129 (**)	0.006	-0.066
	問4	0.211 (**)	0.058	0.013	-0.125 (**)	-0.044	-0.054
第5問	問1	0.169 (**)	0.042	0.060	-0.107 (**)	0.030	-0.054
第6問	問1	0.267 (**)	0.042	0.040	-0.130 (**)	-0.030	-0.053
	問2	0.154 (**)	0.081 (*)	0.060	-0.048	0.020	-0.058
	問3	0.186 (**)	0.077 (*)	0.122 (**)	-0.090 (*)	0.037	0.008
	問4	0.236 (**)	0.064	0.085 (*)	-0.044	0.014	0.006
第8問	問1	0.096 (**)	-0.037	0.023	-0.110 (**)	0.004	0.005
	問2	0.092 (*)	0.013	0.005	-0.071	-0.042	0.002
	問3	0.037	0.008	0.058	-0.069	0.004	0.035
	問4	-0.014	-0.015	-0.019	-0.059	0.009	-0.021
第9問	問1	0.132 (**)	-0.025	-0.012	-0.097 (**)	-0.042	0.011
	問2	0.014	0.020	0.003	-0.030	-0.018	-0.003
	問3	0.116 (**)	0.006	0.056	-0.098 (**)	-0.016	0.002
第11問	問1	0.087 (*)	0.071	0.048	-0.060	0.049	0.048
	問2	0.025	-0.004	-0.009	-0.070	0.047	0.001
	問3	0.081 (*)	0.054	0.121 (**)	-0.065	0.001	0.037

(3) 第6問—問1・問3・問4

第6問は、問題文を読みながら文中に現れるいくつかの物理量の間関係を比較的簡単な数式で表すことができるかを問う問題で、4つの小問「問1」～「問4」が設けられている。このうち3つが困難度差が大きい問題項目として抽出された。教科科目の固有の知識に依らずに解答できるように、解答に必要な情報は全て問題文中に与えられている。しかしながら文系の困難度が高く、理系や医学系に優位な結果となった。問題文中で物理量や物理概念などが説明されているという意味では知識の有無は影響しないと思われるが、このような題材に触れる機会が少ない文系の受験者には、ある程度の心理的抵抗感があった可能性がある。また、単純ではあるが数式による表現が求められているので、ある程度

数理的な能力が必要であることも文系における困難度が高くなった要因と考えられる。

能力・資質の習得度との関係を見ると、「問3」と「問4」で、読解力・表現力との相関が見られるが、情報処理・数理的素養との相関が支配的であった。「問2」は問題文中の空欄に適切な語句を挿入するもので、あまり数理的な能力が必要なかったものと思われる。

(4) 第1問—問1

第1問は、複数の条件下で、ある命題が成立するために必要な条件を考えさせる論理問題として出題されたものであり、二つの小問「問1」「問2」が設けられている。このうち、「問1」のみ困難度差が大きい問題項目として抽出された。論理問題であるため、直

感的には理系受験者が優位であると予想されるが、「問1」は文系よりも理系や医学系で困難度が高い。この問題は問題冊子の先頭で出題される比較的困難度が低い問題で、内容も複雑ではない。与えられた条件のうち、明らかに無関係なものを消去するだけで正答を導くことができる。複雑な論理関係を把握する必要は無く、むしろ問題文を正確に読んで消去すべき不必要な情報を選ぶことができるかどうかのポイントとなる。事実、能力・資質の習得度との関係を見ると、情報処理・数理的素養との相関に有意性が見られなかったが、読解力・表現力との相関には有意性が見られた。 $\beta$ は理系—文系間で4.098、医学系—文系間で2.923と大きな値を示した。試験の導入部分に配置する上で困難度を低めに設定したために、論理能力を測るという作題意図が反映しにくかった可能性がある。一方、「問2」は素直な論理問題で、条件の包含関係を把握することがポイントである。この問題の困難度については理系—文系間、医学系—文系間に目立った差は見られなかった。

#### (5) 第8問—問1・問3

第8問は、鏡に向き合った時に、実像と鏡像では実際には体の前後方向だけが逆になっているのに、何故、左右方向だけが逆になったように感じるのかということに対する説明文の空所を正しく補充する問題で、4つの小問「問1」～「問4」が設けられている。このうち、2つが困難度差が大きい問題項目として抽出された。空欄には実像と鏡像の関係を説明するための方向に関する言葉などが入る。説明文の論理展開を把握するとともに、想像内で体を回転させることや、実像から正しい鏡像を判断するための思考力が問われる。理系に対して医学系で困難度が高い。鏡による光の反射という物理現象に伴う鏡像や実像の方向関係を扱う話題が理系に比べて医学系ではそれほど馴染みがなかった可能性が

あるが、 $\beta$ の値は問1で1.722、問3で1.661と小さく、理系と医学系を分ける大きな要因は見受けられない。

#### (6) 第9問—問1・問3

第9問は、親族構造に関する基本的な定義を与えた上で、婚姻によって生じた新たな親族関係を正しく理解できるかを問う問題で、3つの小問が設けられている。このうち、2つが困難度差が大きい項目として抽出された。論理的思考を測ることが狙いであるが、複雑な問題文を読みながら、多数の定義と婚姻操作に関する情報を混乱せずに処理する必要があり、問題文に与えられた情報を適切に処理するための短期的記憶量と情報処理能力が問われる内容となっていると思われる。能力・資質の習得度との関係を見ると、いずれも情報処理・数理的素養との相関が見られる。表1の結果では理系で困難度が高いことが示されているが、 $\beta$ の値は問1で1.926、問3で1.532と小さく、理系で不利であると結論づける積極的な理由は見当たらない。

#### (7) 第11問—問1

第11問は、貯食する習性のある鳥が、餌の種類と貯蔵場所、貯蔵時期などを記憶できるかを調べた研究の実験結果を提示し、そこから得られる結論を正しく選択できるかどうかを問う問題である。3つの小問が設けられており、このうち「問1」が困難度差が大きい項目として抽出された。解答には、主張したい内容と必要な根拠を正しく対応づける能力が必要である。文系に対して医学系で困難度が高く、 $\beta$ も2.890と比較的大きな値となった。文系受験者のうち行動科学の知識がある者にとっては取り組み易い問題であった可能性も考えられるが推測の域を出ない。

### 5.3 第2部問題

本節では第2部問題で抽出された問題項目に

ついて分析を行う。次の (1) ~ (3) および (4) の第 14 問-問 1 は、困難度差  $\beta$  を表 3 抽出された項目と能力・資質との相関係数 [第 2 部] (\* :  $p < 0.05$ , \*\* :  $p < 0.01$ )

		情報処理・ 数理的素養	創意・多元的 判断・論理的思考	読解力・表現力	対人的親和性・ 献身性	自然・社会・ 人間への関心	芸術への関心
第3問	問1	-0.036	-0.081 (*)	0.006	-0.025	-0.045	-0.047
	問2	0.000	-0.084 (*)	-0.062	-0.039	-0.037	-0.075 (*)
第8問	問1	-0.017	0.026	0.041	0.063	0.045	0.086 (*)
	問2	-0.045	0.045	0.064	0.038	0.072	0.056
第9問	問1	0.015	0.023	0.122 (**)	0.030	0.057	0.095 (*)
	問2	0.021	-0.002	0.104 (**)	-0.138 (**)	0.055	-0.058
	問3	0.043	0.017	0.042	-0.003	-0.045	-0.066
	問4	-0.054	-0.023	-0.022	-0.026	-0.038	-0.026
第10問	問1	0.127 (**)	0.056	0.131 (**)	-0.013	0.101 (**)	0.030
第12問	問1	0.045	0.022	0.170 (**)	-0.038	0.099 (**)	0.032
	問2	0.029	0.052	0.014	0.014	-0.008	-0.013
	問3	0.080 (*)	0.015	0.051	-0.056	0.085 (*)	0.025
第14問	問1	-0.003	0.097 (**)	0.125 (**)	-0.040	0.130 (**)	0.057
	問2	0.075 (*)	0.033	0.070	-0.024	0.055	-0.022
	問3	0.002	0.073 (*)	0.085 (*)	-0.049	0.046	0.028
	問4	0.056	0.032	0.078 (*)	-0.051	0.051	0.044
	問5	0.105 (**)	0.019	-0.011	-0.047	-0.053	-0.021
第15問	問1	0.062	0.042	0.067	-0.048	0.034	0.015
	問2	0.064	0.009	0.026	-0.089 (*)	0.007	-0.028
	問3	0.108 (**)	0.071	0.121 (**)	-0.083 (*)	0.085 (*)	0.023

用いた場合の上位 5 項目以内と正答率差  $F$  を用いた場合の上位 10 項目以内のどちらにも含まれた問題、(4) の第 14 問-問 2・問 3 および (5) ~ (7) は  $\beta$  を用いた場合の上位 5 項目以内のみ含まれた問題である。また、表 3 はこれらの問題項目と能力・資質の習得度との相関係数を当該項目が属する大問単位で示したものである。

### (1) 第 3 問-問 1

第 3 問は、医療場面における患者の家族と主治医の短い会話文を提示し、主治医の行動および患者の家族の心理を読み取る問題である。2 つの小問中、問 1 が困難度差がある問題項目として抽出された。正答を得るには文脈から場面を正確に想起する必要がある。医学系に対して理系で困難度が高いが、 $\beta$  は 1.380 と小さい値であった。能力・資質との

関係では有意な正の相関を示すものはなかった。題材が医療場面であったことから、医学系にとって馴染みがあった一方で理系にとってはやや戸惑いがあったという単純な心理的要因によって困難度に差が生じたのかもしれない。

### (2) 第 8 問-問 2

第 8 問は、セリフ付きの挿絵によって診察場面における医師と患者のやり取りを提示し、その後続く医師の発言を問う問題である。文章によって場面の描写をするだけでなく、挿絵から医師が直面している状況を把握させるところが特徴的である。2 つの小問中、問 2 が困難度差がある問題項目として抽出された。しかしながら、 $\beta$  は 1.488 と小さく、能力・資質との関係でも高い相関を示すものは見られない。第 3 問と同様に、理系に



としては不慣れな問題で解答に戸惑った可能性はある。

### (3) 第 12 問—問 3

第 12 問は、フェールセーフとフォールトトレランスに関する長文問題。読解力とともに、文脈を正しく捉える談話的能力が必要であり、典型的な「読解力・表現力」の問題である。3つの小問のうち、問3が困難度差がある問題項目として抽出された。文系で困難度が高いが、 $\beta$ の値は理系—文系間で 1.217, 医学系—文系間で 1.611 といずれも小さく、能力・資質との関係からみても文系に対して公平さを欠くような測定対象以外の要素が多く含まれていたとは考え難い。強いて言えば、安全性・信頼性工学に関する話題であったことが文系受験者に不利に働いたのかもしれない。

### (4) 第 14 問—問 1・問 2・問 3

第 14 問は、先進国首脳会議における日本語の通訳の仕方や、大学の講義に用いる言語についての著者の意見を述べた長文を提示し、その内容および著者の意図を読み取る読解問題。5つの小問のうち、3つが困難度差がある問題項目として抽出された。理系および医学系で困難度が高いが、 $\beta$ の値は医学系—文系間で 1.368, 医学系—理系間で 1.245 (問1) および 1.171 (問2) と小さい値であった。第 12 問と同様に、典型的な「読解力・表現力」の問題であるが、特に理系や医学系に不利な問題であったとは考え難い。

### (5) 第 9 問—問 2

第 9 問は、社会科学における実証主義と解釈主義の説明文を与え、その定義を正確に理解することや、馴染みの無い専門用語の意味を文脈から判断する問題である。4つの小問のうち、問2が困難度差がある問題項目として抽出された。非常に抽象度の高い文章の読

解力および抽象概念を把握する能力が問われる。文系および理系との比較において、医学系で特に困難度が高く、 $\beta$ が 4.7 を超える大きな値となっている。能力・資質の習得度との関係を見ると、読解力・表現力との相関が強い。抽象度が高いために、医学系の受験者において解答に対する動機づけが弱かった可能性が考えられるが、識別指数による問題分析(赤根ほか, 2006)では解答率に顕著な傾向は見られていない。いずれにしても抽象度の高い文章を利用する場合には注意が必要であると思われる。

### (6) 第 10 問—問 1

第 10 問は、疾病構造の変化を経済社会の転換の中で捉える健康転換の概念についての説明文を読み、転換の各段階において関連の深いキーワードの組み合わせを問う問題である。1つの小問だけからなる。文系で困難度が高いが、 $\beta$ は理系—文系間で 1.568, 医学系—文系間で 1.964 とそれほど大きな値ではない。問題文中に含まれない用語については、文脈から推測する必要がある。能力・資質との関係を見ると、読解力・表現力だけではなく、情報処理・数理的素養との間にも有意な相関が見られる。このことから文系での困難度が高まった可能性がある。

### (7) 第 15 問—問 2・問 3

第 15 問はサブリミナル効果とプラシーボ効果に関する長文読解問題である。3つの小問のうち、問2と問3が困難度差がある問題項目として抽出された。問2は両方の概念の関係を正しく把握できるかを問う問題で、理系に対して医学系で困難度が高いが、 $\beta$ は 2.178 でそれほど大きな値ではない。問3はサブリミナル効果とプラシーボ効果に関する事象の因果関係を正しく判断できるかを問う問題であり、医学系に対して文系での困難度が高いが、 $\beta$ は 1.396 と小さい値であった。

但し、能力・資質の習得度との関係を見ると、読解力・表現力と同程度に情報処理・数理的素養との相関が見られ、このことが文系の困難度を高めた可能性が考えられる。

## 6 まとめと考察

今回、教科科目フリー型総合試験について、学部系統間で困難度に差が見られる項目を抽出し、その要因の特定を試みた。困難度の評価は項目応答理論を用いて行ったが、本研究で分析した総合試験のような大問形式の問題に項目応答理論がどの程度有効であるかは厳密に言えば不明である。従って、今回用いた手法は、あくまで学部系統間で測定内容の同等性が保たれていない危険性がある項目を警告するためのものと考え、得られた結果が偶然では起こり得ないと判断される系統的な特徴を有し、かつ合理的な解釈が可能かどうかによって手法の妥当性を認めることとした。

以下、分析結果から得られた主な知見をまとめておく。まず、項目困難度差  $\beta$  の値が 2.50 を下回る問題項目については、困難度差が生じた有力な要因を見出すことがほとんどできなかった。一方、3.00 付近から上の値を示した項目については、比較的顕著な特徴を示すケースが少なからずあった。このことから、 $\beta$  の値として 2.50~3.00 を境にして、これより小さい値を示す項目については、学部系統間での差異をほぼ無視し得るものとみなす方が適当であると思われる。次に、第 1 部「情報把握・論理的思考」の問題であっても、問題文や選択枝の文章が複雑であったり、解答中に比較的多くの情報を考慮したりしなければならぬ問題では、理系での困難度が高くなる傾向が見られた。また、第 2 部「コミュニケーション・読解・表現」の問題であっても、情報処理・数理的素養との相関が見られる項目では、文系での困難度が高くなると考えられる。つまり、表面的な妥当性

が見られたとしても、具体的な内容についての妥当性に注意した作題を行わなければ、相当程度に意図したものとは異なる能力を測る問題となってしまう危険性があると考えられる。作題作業に割ける労力や時間が限られていることから、出題範囲や題材だけが議論の対象となり、解答過程のチェックがあまり検討されないことも考えられるので、問題の形式も含めた出題基準を作成することが望まれる。さらに、原因の特定までには至らなかったが、抽象度の高い文章の読解問題では、医学系で困難度が高くなるという特徴が見られた。問題の題材については、学部系統間でその馴染みの程度の差が大きいものはできるだけ避けることが望ましいと考えられる。たとえ解答に必要な情報が全て問題文中に与えられていたとしても、それらの情報の運用に慣れているかどうかや、問題に対峙した際の心理的抵抗感などが項目困難度に影響を及ぼす可能性がある。

総合試験問題の利用が増えてきたとはいえ、教科科目フリー型の作題経験者は決して多くはないと思われる。今後も、作題時の参考として、今回のような問題内容分析の事例を提供していく予定である。

## 参考文献

- 赤根敦・伊藤圭・林篤裕・椎名久美子・大澤公一・柳井晴夫・田栗正章 (2006). 「識別指数による総合試験問題の項目分析」『大学入試センター研究紀要』 35, 19-47.
- 石井秀宗・椎名久美子・柳井晴夫・岩坪秀一・荒井克弘 (2005). 「基礎学力評価のための国語, 数学, 英語試験問題の開発研究」『大学入試センター研究紀要』 34, 1-17.
- 伊藤圭 (2006). 「大学入試における総合試験および適性試験の動向」『大学入試研究ジャーナル』 16, 149-155.

- 伊藤圭・林篤裕・椎名久美子・大澤公一・石井秀宗・柳井晴夫・田栗正章・岩坪秀一・赤根敦・麻生武志・岩堀淳一郎・内田千代子・川崎勝・齋藤宣彦・武田龍司 (2006). 「医学部学士編入学者選抜のための総合試験の開発とその評価」『大学入試センター研究紀要』35, 49-108.
- 熊谷龍一・脇田貴文 (2003). 「特異項目機能検出方法の比較－BILOG-MG と SIBTEST を用いた検討－」『名古屋大学大学院教育発達科学研究科紀要, 心理発達科学』50, 83-90.
- 倉元直樹・柳井晴夫 (2001). 「教科科目複合型総合試験の問題内容分析」『大学入試センター研究紀要』30, 83-108.
- 前田忠彦・野口裕之 (2006). 「第6章「法科大学院統一適性試験」のDIF分析」『法科大学院統一適性試験テクニカル・レポート 2005』商事法務, 64-85.
- 椎名久美子・柳井晴夫 (2001). 「教科科目フリー型総合問題を用いたモニター調査の分析」『大学入試センター研究紀要』30, 39-66.
- 孫媛 (2005). 「学力テストデータの差異項目機能 (DIF) 分析」『教育心理学年報』44, 35-36.
- 田崎勝也 (2007). 「文化的自己観は本当に「文化」を測っているのか : 平均構造・多母集団同時分析を用いた特異項目機能の検証」『行動計量学』34, 79-89.
- 柳井晴夫 [研究代表者] 他 (2006). 『平成15 - 17 年度共同研究報告書「総合試験問題の分析的研究」』大学入試センター研究開発部.