

# 教科科目複合型総合試験作成の試み

— 東北大学 倉元 直樹\*1 大学入試センター 柳井 晴夫 —

## 1. 問題\*2

大学生の基礎学力への懸念が取り沙汰されている。全国95の国立大学、362学部の学部長を対象とした最近の調査によると、半数以上の学部で新入学生の学力低下が認識されていた(大学入試センター、1999)。問題の背景要因には、教育環境の変化の複合作用が挙げられる。最も顕著なのは、進学率の上昇と18歳人口の減少である。すなわち、全体として、以前ならば学力的に進学が困難だった層が大学に流入していると考えられる。

時代の変化に伴う社会構造的な変化と同時に、政策的な要因も学力不足問題に拍車をかけている。具体的には、高等学校のカリキュラムと大学入試の軽量化である。昭和35(1960)年度告示(昭和38〔1963〕年度実施)の学習指導要領をピークとして、高校のカリキュラムは必修科目と卒業必履修単位数が削減の傾向にある。平成15(2003)年度実施予定の新しい学習指導要領案では、週5日制の完全実施ともあいまって、削減傾向がさらに強まっている。大学入試においても受験科目数が削減の傾向にある。私大入試や国公立大学の個別試験では2～3科目程度が一般的であり、1科目のみの試験も珍しくなくなっている。共通試験においても、昭和54(1979)年度からの共通第1次試験(以下、「共通1次」と記す)が、平成2(1990)年度に大学入試センター試験(以下、「センター試験」と記す)へと移行した時点でア・ラ・カルト方式と言われる4教科以下の受験が本格化した。

大学入試が教育を歪めているという議論がなされるが、裏を返せば大学入試に向けての受験勉強に力点が置かれるのが進学校の教育

の実態ということであろう。「ゆとりの教育」という目的で設定された幅広い科目選択の自由も、いったん現実の荒波に揉まれれば受験科目に即応したカリキュラムへと変質せざるを得ない。結果的に、大学進学志望者は自分が志望するところが設定する少数の受験科目のみに絞り込んだ勉強を行い、受け入れる大学側は新入生の基礎学力の不足に頭を痛めることになる。少子化現象が大学の危機感を煽り、最も手っ取り早い受験生確保の手段として科目削減が成されていると考えるならば、新入生の学力不足に対する大学の責任も考える必要があるだろう。

さて、ア・ラ・カルト方式の拡大と呼応するかのように、大学の個別試験で「総合問題」、「総合科目」といったタイトルの試験科目が目立つようになってきた。これらの試験の問題内容は、高等学校の教科科目の枠とは独立の「教科科目フリー型(Curriculum-free type)」と複数の教科科目の内容を統合した「教科科目複合型(Cross-subject type)」に分類することが可能と考えられる。個々の大学にはそれぞれの事情があると思われるが、教科科目複合型総合試験導入の動機の一部には、受験生の学習の偏りを少しでも矯正したいという願望があるのではないだろうか。ただし、現行制度のセンター試験と個別試験の役割を考えたとき、大学進学者の基礎学力への配慮は、共通試験であるセンター試験が主として担うべきだという考え方もあるだろう。

本研究は、以上のような問題意識に基づき、共通試験で実施可能な総合試験を模索するための基礎研究として実施されたものである。

## 2. 目的

本研究の目的は、共通試験における総合試験の実施を想定した場合に生ずる技術的問題の基礎研究である。したがって、直接的に利用可能な問題の開発を目指したわけではない。

具体的には以下の諸点に焦点を当てて研究を行った。

- ①問題作成の限界と可能性を探索すること
- ②モニター調査<sup>\*3</sup>で得たデータから問題の難易度や構造を統計的に解析すること
- ③学力形成に関する副次的な情報と成績との関係を明らかにすること

## 3. 方法

### 3.1 問題作成

本研究では、2つの科目を組み合わせて問題を作成した。問題作成は、教科科目の専門家のペア（2名1組）に委託して行った。作題者として高校教員25名、大学教員1名（計13組のペア）の協力を得ることができた。作題者の専門領域は、政治・経済を除く全ての主要科目に互っている。文系科目同士の組み合わせを「文系（科目）複合型」、理系科目同士の組み合わせを「理系（科目）複合型」、文系科目と理系科目との組み合わせを「文理（科目）複合型」と呼ぶこととすると、文系複合型のペアが2組（「国語—英語」、「倫理—世界史」）、理系複合型が2組（「数学—地学（自然地理）」、「化学—生物」）、文理複合型が9組（「国語—物理」、「国語—生物」、「英語—物理」、「英語—地学」、「倫理—生物」、「地理—数学」、「日本史—化学」、「日本史—地学」、「世界史—物理」）形成された。

まず、作題者に対して作題方針に関する説明会を行った。出題形式はセンター試験の地歴、公民の形式をモデルとし、マークシートに移植可能な形に限定した。また、問題が難しすぎないように、と特に注意を促した。依頼した作題数は、各作題ペア毎に大問1～2

問、問題作成の期間は平成9（1997）年9月～11月の約3ヶ月間であった。平成9（1997）年12月には、作成した問題の検討会を2回に分けて行った。最終的に各組大問1～4題、計24題の問題が得られた。また、自分が作題した問題に関するアンケートを実施した。

### 3.2 モニター調査

モニター調査は2回に分けて実施した。問題冊子は内容的なバランス等を考慮して2セット作成した。それぞれ、B1セット（大問9問、小問78問）、B2セット（大問10問、小問80問）と呼ぶ。B1セットは、平成10（1998）年3月に首都圏の3つの国立大学1、2年生397名を対象として調査を行った。B2セットは、同年9～10月に同じ大学の1、2年生413名を対象として調査を行った。試験時間はそれぞれ135分間であった。

### 3.3 モニター生へのアンケート

問題に関するアンケートでは、大問ごとに大学入試センター研究開発部（1993）で作成された「27の資質」から解答に必要と思われるものを最大3個選択させた。高校時代の履修経験については、教科・科目ごとに「1. 全く履修していない」～「5. 個別試験や私立大学入試のための受験勉強をした」の5段階評価を求めた。

## 4. 結果

### 4.1 問題構成<sup>\*4</sup>

B1セットの問題構成は、第1問が「倫理—世界史（小問4問）」、第2問が「倫理—生物（13問）」、第3問が「世界史—物理（5問）」、第4問が「国語—物理（11問）」、第5問が「日本史—化学（9問）」、第6問が「英語—物理（7問）」、第7問が「地理—数学（12問）」、第8問が「英語—地学（11問）」、第9問が「国語—生物（6問）」の計78問、B2セットは、第1問が「倫理—世界史（6問）」、

第2問が「日本史—地学（9問）」、第3問が「世界史—物理（6問）」、第4問が「国語—物理（13問）」、第5問が「化学—生物（10問）」、第6問が「地理—数学（7問）」、第7問が「英語—物理（12問）」、第8問が「数学—地学（3問）」、第9問が「国語—英語（7問）」、第10問が「国語—生物（7問）」の計80問である。

#### 4.2 問題作成の過程

問題作成を放棄したペアは1組もなかった。一部に単独で問題作成をしたペアも見られたが、全体としてはほぼ期待通りの方法で問題が作成された。作題に関する感想を簡単にまとめると以下のようなになる。

作題作業に関しては、2つの科目内容の組み合わせ方に苦労したという感想が多かった。科目の難易度を対等に組み合わせることは困難で、どちらかの科目に基礎を置くことになる。また、リード文に既成の文章を引用した場合、素材探しが難しい。

実際に共通試験で実施する場合の問題点としては、知識問題に偏る可能性、出題内容が限定される可能性、教科書の範囲の逸脱や難しい問題になる可能性が挙げられていた。

#### 4.3 モニター調査データによるテスト問題内容分析

##### 4.3.1 問題の難易度

小問の問題内容に関わらず、正答を1点、誤答を0点と得点化して採点した。以下の分析はその採点結果に基づく。

B1セットについては、合計得点78点満点による成績を算出した。平均得点率49.6%（平均得点38.7点）、最高得点率79.5%（62点）、最低得点率20.1%（16点）であった。文系・理系別では、文系の平均得点率が45.1%（平均得点35.1点）、理系の平均得点率は53.7%（平均得点41.9点）と理系の成績が良かった。作成された問題は全体としてかなり難しかった。

た。

B2セットは、合計得点80点満点による成績を算出した。平均得点率55.8%（平均得点44.6点）、最高得点率81.3%（65点）、最低得点率16.3%（13点）であった。B1セットとのバランスは取れていたが、作成された問題は全体として難しかった。文系・理系別では、文系の平均得点率は54.9%（平均得点43.9点）、理系の平均得点率は53.7%（平均得点45.4点）とほぼ同じであった。

##### 4.3.2 問題の識別性能と信頼性係数

次に、古典的項目分析による識別力（識別性能）の算出と信頼性係数の推定値の算出を行った。識別力の指標には、項目得点と合計得点との点双列相関係数を用いた\*5。

テストB1の識別力は全体的に低く、最高で.50、最低で-.05であり、識別力が負の小問が2問あった。また、55問（全体の70.5%）が識別力.25に満たなかった。テストB2の識別力も全体的に低く、最高で.39、最低で-.08であり、識別力が負の小問が5問あった。61問（全体の76.3%）が識別力.25に満たなかった。

信頼性係数\*6の推定値にはアルファ係数（KR-20）を用いた。テストB1は $\alpha = .80$ 、テストB2は $\alpha = .77$ であり、項目数が多い割には十分な値が得られたと言いがたい。

以上のことから、本研究のテストセットに、テスト全体として1次元の能力を識別する通常のテストの手法を用いた分析を適用していくことは無理があることが示唆された。そこで、以後、テストセットの持つ多次元的な評価構造を探っていくこととした。

##### 4.3.3 探索的因子分析による問題構造の分析

多次元的構造の探索には因子分析\*7の手法を用いることとした。テストB1、B2とも、3～6因子構造のいずれかの仮定が適当であ

ると思われたが、バリマックス回転を行って因子構造を求めたところ、ともに4因子構造で解釈することが内容的に最も適切であった。

テストB1の第1因子は21項目 ( $\alpha = .89$ ) である。含まれる小問は、いずれも数学、物理、化学の内容に関わっており、「数学・物理・化学系」から構成された因子である。第2因子は4項目 ( $\alpha = .83$ ) から成る。いずれも時差計算の問題であり、「時差」問題の因子である。問題構成に依存しており、因子の一般性は低い。第3因子は18項目 ( $\alpha = .57$ ) からなる。生命や人間に関わる社会・生物系の問題と国語、社会の問題が集まった因子であり、「生命・人文・社会系」の因子である。第4因子は15項目 ( $\alpha = .57$ ) から成る。理科的な知識を問うものがほとんどで、「生物・地学・化学系」の問題の因子と言える。

テストB2の第1因子は19項目 ( $\alpha = .74$ ) である。小問の問題内容は、数学、物理、化学に関わっている。「数学・物理・化学系」の因子である。第2因子は17項目 ( $\alpha = .64$ ) から成る。全てに共通する属性を見出すのは難しいが、国語的な設問と一般教養的な知識に関わる問題と言える。したがって、「国語・一般教養系」の因子と考えることが可能であろう。第3因子は15項目 ( $\alpha = .61$ ) から成る。歴史に関わる問題と英語的な設問の組み合わせであり、「英語・歴史系」の因子と言える。第4因子は3項目 ( $\alpha = .43$ ) から成り、いずれも第5問の「生態系生化学」に関わる小問である。問題構成に依存して現れた特殊な因子と考えることができよう。

#### 4.3.4 尺度特性に関する分析

尺度間相関係数は、テストB1においては、尺度3と尺度4 ( $r = .30$ ) 以外は値が小さく ( $r = .03 \sim .21$ )、独立な尺度となっていることが伺える。それに対して、テストB2では、テストB1ほどには尺度間の区別が判然としていない。

## 4.4 アンケートデータの分析

### 4.4.1 履修経験と成績

高校での科目履修経験は、「1：全く履修していない」、「2：授業は受けたが、特別の勉強はしなかった」、「3：受験勉強はしなかったが、授業以外にも日常的に勉強した」、「4：センター試験のための受験勉強をした」、「5：個別試験や私立大学入試のための受験勉強をした」という形で得点化した。履修科目経験の成績への影響は、重回帰分析<sup>\*)</sup>による修正重決定係数  $R^2$ 、および、単純相関係数の値で検討することとした。

テストB1では、合計得点に対する説明力は  $R^2 = .27$ と中程度であり、数学関連4科目、物理、化学の相関係数の大きさが目立った。テストB2では、 $R^2 = .16$ と教科・科目の履修経験の影響は全体として小さく、国語、英語、地歴等の文系諸教科でも統計的に有意な程度の相関係数の値が得られた科目があった。

尺度ごと、文系・理系別に履修経験の影響について相関係数を用いて見た場合、概ね、問題内容に対応した関係が見られた。理系モニター生の国語の履修経験が多くの科目の成績に影響していたことが特筆される。

### 4.4.2 テストが測定する能力

モニター調査で用いられたテストが、どのような能力を測定していると考えられたのか、ということ調べるために、大学入試センター(1993)で同定された大学の専門分野への適応に必要とされる27の資質を用いた。作題者に対しては、大問毎にそれぞれの資質をどの程度測定しているか、5段階評価で回答を求めた。

作題者、被験者それぞれ、2つのテスト冊子に対する評価は酷似していた。一方、作題者と被験者の評価に異なった特徴としては、作題者が「知識・教養」、「論理的思考力」、「探求心・好奇心」、「読書力」、「集中力・持続

力」、「判断力」、「発想力」、「社会問題への関心」、「自然と人間との関わり合いへの関心」、「歴史への関心」といった多様な資質を評価しているとしているのに対し、被験者の回答は「知識・教養」に集中し、それ以外には「数理能力」、「歴史への関心」、「語学力」といった項目が挙げられた程度である。質問の形式が異なるので単純な比較はできないが、特徴的な傾向と言える。

## 5. 考察

本研究の結果、易しい問題を作成する困難さが浮き彫りになった。モニター生は入学試験を突破して難関大学に入学した者である。達成動機が少々低かったとしても、約50%の平均正答率は難しすぎる。大学志願者層全体から見れば、超難問と言えらるだろう。本研究では問題作成過程そのものを充分検討した訳ではないが、難易度を高くしないための工夫は今後に残された研究課題と言える。また、上記は複数の科目の組み合わせによる問題作成に付随して発生したものである。万が一、実際的な解決が非常に困難であれば、個々の大問が特定科目にほぼ対応するような形式でもやむを得ないかもしれない。テストセット全体で幅広い領域をカバーすれば、総合問題化の目的は部分的にでも達成可能と言えよう。

本研究で作成されたテスト問題は、B1セット、B2セットともに識別力が充分ではなかったが、本研究の目的と被験者の学力分布の偏りを考慮すると、それは大きな問題ではない。むしろ、因子分析を通して得られた構造が現実の大学入試に示唆的な興味深い傾向を示している。まず、B1セットは独立性が高い4つの尺度に分かれたが、学力テストのデータとしては異例の結果と言える。原因は、雑多な問題を集めた項目内容にもあるが、被験者の学力分布が狭かったことも一因である。大学入試に対する「偏差値輪切り」という批判が表しているように、共通試験と比較した

場合、個別試験では多くの受験者の学力分布が狭い範囲に集中する傾向は否めない。その場合、合計得点がどれだけの受験者の情報を的確に表しているかは疑問である。現在よく見られる教科科目単位の試験であっても、合計得点よりも個々の科目の成績を別々にプロットした学力プロフィールの方が「多様な学力の個性」という豊かな情報を持つと考えるべきである。本研究の結果は、合計得点による選抜だけではなく選抜資料の持つより詳細な情報を生かした評価の可能性を示唆している。例えば、学力がほぼ同等と考えられる層に学力とは違った観点からの評価を行う方法も考えられる(池田・平, 1995)。その場合には、適切なプロフィールを描くための標準化の方法、新しい選抜のルールを検討する必要があることに留意すべきであろう。

B1セット、B2セット共通の特徴は、「数学・物理・化学」という因子が得られたことである。他の因子は冊子に固有の様相が強いが、生物、地学を含まない「数学・物理・化学」的分野の問題、というまともは安定しているように思われる。教科科目複合型総合試験の構成を考える場合、既成の教科の枠にこだわらずに、例えば「教理物質科学分野」と「生命環境科学分野」というような分類が合理的であることを示唆している。

高校時代の履修経験は被験者の学習歴を要約した情報と考えることができる。教科科目複合型の問題に対して、科目単位の学習成果とは独立に別種の能力特性を測定することを期待するのは無謀である。設問の内容に鑑みて、それらの経験が影響を与える方が自然な結果と言える。その意味では、全体としてほぼ納得できる結果が得られたように思われる。

文系・理系別の分析結果を見たときに、理系における「国語」の履修経験が様々な尺度の成績に影響していたのは興味深い。大学間の基礎学力の差が反映されたものとも考えられるが、理系にとっては幅広い学習をした被

験者の方が有利になる問題構成であったことを示唆している。

次に、教科科目複合型総合試験が測る能力は何か、という問題について検討する。測定可能な能力が明らかに「学力」であることは疑いない。それは、履修経験の成績に対する影響からも裏付けられた。しかし、それ以上の判断は慎重にすべきである。作題者と被験者の双方の見解が一致したのは、「知識・教養」を測っているということであるが、このことは、逆に、問題が知識依存型であったことを示しており、反省点である。本研究では「27の資質」という形式で、テストが測定しようとしている能力を聞いた。この方法では、結果的に作題者と被験者の見解が一致した場合でも、その結果は単なる一つの主観的な見方に過ぎないことに留意すべきである。例えば、「発想力」を測っていると判断された問題で好成績を挙げた者が、現実の場面で実際に能力を発揮するかどうか、誰にも保証できるものではないし、検証も難しい。評価の方法に関して具体案のないまま、やみくもに目新しい方法を採用することには弊害が伴う。現実の大学入試を考える上で、テストで測定できる能力、測定すべき能力とともに、テストの限界についても問題を真剣に議論して整理すべき時期が来ている。諸外国の本格的な取り組みを見習うべきであろう。

最後に、将来的な展望として現在のセンター試験のような共通試験の性格の大学入試に科目単位の試験以外のメニューを加えることも検討していく必要がある。その中で、教科科目複合型総合試験の導入は、有力な現実的選択肢の一つと言える。たとえ、測定される能力や資質の性質に現在の学力テストを越える飛躍的なものを求めなくても、教科科目複合型総合試験は大学での学習活動に必要な学力の保障と公平性の確保の面での実利が大きい。試験の内容が幅広い分野をカバーするために受験勉強の範囲が自然と広がる効果

が期待できる。また、多くの受験生に共通の問題を用意して選択科目数を減らすことができるならば、得点調整のリスクを軽減することも期待できるのである。逆に言えば、実施条件を具体的に検討することで、総合試験に対する過大な期待を現実的な可能性に修正していくことも必要であろう。

#### 脚注

- \* 1 : 平成11年3月まで大学入試センター研究開発部に所属。旧姓平。
- \* 2 : 本研究は、平成8~10(1996~1998)年度における大学入試センター研究開発部の共同研究「大学の各専門分野への適性の評価を目的とする総合試験のあり方に関する共同研究」の一環として行われたものである。研究の詳細は大学入試センター研究開発部(1999a)を参照のこと。
- \* 3 : 大学生のモニター(アルバイト)を募集し、テスト問題を解かせてデータを得る研究方法。
- \* 4 : 問題内容の詳細は大学入試センター研究開発部(1999b)を参照のこと。
- \* 5 : 識別力とは被験者の学力差が設問の得点に反映される程度。選抜試験では高い方がよい。本稿の指標では.25以上が目安。
- \* 6 : 信頼性係数はテスト結果の再現性の指標。9程度以上が望ましい。
- \* 7 : 本稿では、質問項目を分類、整理するための方法として用いている。
- \* 8 : 履修科目経験のデータから成績が予測できる程度を示す。 $0 < R^2 < 1$ であり、値が大きいくほど予測力が大きい。

#### 文献

- 大学入試センター(1999). 学生の学力低下に関する調査結果
- 大学入試センター研究開発部(1993). 大学の各専門分野の進学適性に関する調査研究報告書—大学入学者選抜資料としての

適性検査のための基礎資料一

大学入試センター研究開発部(1999 a). 平成8～10年度「大学の各専門分野への適性の評価を目的とする総合試験のあり方に関する共同研究」最終報告書

大学入試センター研究開発部(1999b). 総合試験モニター調査問題集 平成8～10年度「大学の各専門分野への適性の評価を目的とする総合試験のあり方に関する共同研究」最終報告書別冊

池田輝政・平直樹(1995). 合否ボーダー層における多面的評価法, '95大学入試フォーラム, 18, 5-16.