

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

石井 秀宗*
 椎名久美子*
 柳井 晴夫*
 岩坪 秀一*
 荒井 克弘**

要 約

大学教育に必要な基礎学力を診断・評価する試験問題の開発研究として、国語と数学の試作問題を高校生を対象に実施する調査研究を以前に行った（石井他，2004a）。そこで指摘された問題点などをもとに、今回は、国語、数学に加え英語の試験問題を改めて開発し、大学新入生を対象に調査研究を行った。その結果、国語については目標に合致した複数の試験問題を作成し得たが、数学と英語については、内容的に等価な複数の試験問題を作成し得ているものの、難易度を調整する必要性が指摘された。

1 はじめに

平成 12 年 11 月の大学審議会答申「大学入試の改善について」において、「総合的な問題（総合問題）における調査研究」の推進、および「総合問題の大学入試センター試験への導入を考えていくことが適当である」との提言がなされた。この答申を受け、平成 13 年 4 月に大学入試センターに設置された総合問題調査研究委員会（平成 13 年 4 月～平成 15 年 3 月）、および、平成 15 年 4 月に設置された第 2 期総合問題調査研究委員会（平成 15 年 4 月～平成 16 年 3 月）では、年齢人口の約 50% が大学・短期大学に進学するという今日の高等教育の状況を踏まえ、従来の大学入試センター試験よりも幅広い受験者層を対象とした、基礎的な学力を診断・評価する試験を作成する必要があると考え、そのための試験問題の開発研究を進めてきた（荒井他，2003）。そこで考えられた総合問題は、教科

別の基礎的な試験である科目混合型総合試験問題（藤井他，2002 など参照）であった。また、この総合試験問題の用途には、AO 入試や推薦入試など、一般入試を経験しないで大学に入学してくる学生の学力到達度測定などが想定された。

このような総合試験問題の開発研究として、前回の調査研究においては、国語と数学の基礎的な問題（以下、「総合基礎」問題と呼ぶことにする）を作成し、東京近郊の 2 つの県の公立高校の生徒約四千名を対象にモニター調査を実施した。そして、国語については目標としていた 7～8 割の正答率を実現していたが、数学については 4 割程度の正答率であり、問題内容の改善が求められること、また、国語と数学を同一時間内に実施する場合は、受験者は数学の解答により力を注ぎ、国語の成績が低下する可能性があることをなどが明らかにされた（石井他，2004a）。

この結果を受けて、第 2 期総合問題調査研究委員会のもとに、改めて国語、数学の問題が作成さ

* 大学入試センター

** 東北大学

2004 年 10 月 27 日 受理

れた。また、国語、数学とともに大学教育において必要とされる割合が高い英語（柳井他，1993；2002）の問題が新たに開発され、これらの問題を用いて、想定される受験者に近い大学新入生を対象とする調査研究が実施された。本稿では、この調査研究で収集されたデータに基づいて、試験問題の難易度や内容の等価性などについて検討する。

2 問題作成

2.1 作題方針

平成 15 年秋に、各教科ごとに高等学校と大学の教員からなる作題ワーキング・グループ（以下、作題 WG）が組織され、「総合基礎」問題の作成が開始された。

出題範囲は、義務教育の範囲、高等学校学習指導要領における該当教科のミニマム・エッセンス、および、大学教育に必要な基礎的な内容とされ、以下のように設定された。

国語：慣用語句／故事成語／諺・格言／手紙文・挨拶文／語句の意味・用法／漢字の書き取りと読み・意味／漢語・漢熟語の構造／説明的文章の読み取り／文学的文章の読み取り／文法・ことばのきまり

数学：小数／分数／文字式の計算／覆面算／方程式／数学的文章の読み取り／（一次）グラフの読み取り（組合わせ問題，推定問題）／相対的な考え（パーセント，等比数列）／空間の把握と三角比

英語：代名詞／形容詞／副詞／動詞の活用／時制／過去分詞／現在分詞／否定文／疑問文／命令文／前置詞／疑問詞／関係詞／接続詞の用法／英作文／発音／基本語彙と綴り／会話／会話の流れ／文章理解

問題形式は、基礎的な事項の理解を評価することを目的としているため、基本的には単問形式とされたが、とくに数学については、1つの素材に関連して段階的にいくつかの設問を構成することが考えられるため、大学入試センター試験と同様に大問形式で問題が作成されることとされた。

難易度については、前回の問題作成と同様に、平均得点率 7 割程度（100 点満点に換算した場合、平均点 70 点程度）、最頻値 8 割程度、および、大学

進学のために十分準備した受験生であればほぼ全問正答となることが目標とされた。ただし、前回、目標とする難易度の試験問題を作成し得なかった数学と、今回初めて問題を作成する英語については、目標とする難易度と、それよりもやや難易度が高いと想定される試験問題を作成することが求められた。国語については、前回の問題作成でほぼ目標とする試験問題を作成し得ているので、それを参考にして、目標とする難易度の複数の試験問題を作成することが求められた。

解答方式については、ある程度多人数の実施を視野に入れ、マークシートによる多肢選択式とされた。ただし、数学に関しては、大学入試センター試験の数学と同様に、解答欄の各マスに対応する数字を塗りつぶす形式が原則とされた。なお、位取りの計算問題等に対応するために、塗りつぶしの候補として「小数点」が加えられた。

解答所要時間に関しては、当初は前回同様各教科 30 分ずつとされていたが、問題作成の過程で、数学については基礎的な学力を評価するのに 60 分程度は必要であるという意見が作題者から出され、また、国語に関しても前回の作題で所要時間 30 分の試験を作成し得ているということが確認されたため、最終的には、解答所要時間は、国語 50 分、数学 60 分、英語 30 分と設定された。ただし、調査協力校における実施時間上の制約から、国語に関しては、所要時間を 30 分とする短縮版も作成することとされた。

以上の作題方針のもとに、平成 15 年秋から作題 WG による「総合基礎」試験問題の作題が開始され、数ヶ月をかけて、国語 4 冊子（うち 2 冊子は短縮版）、数学と英語はそれぞれ 2 冊子ずつが作成された。数学と英語については、問題作成の過程で、大学付属の高校において予備調査が実施され、問題の調整が行われた。また、大学入試センター研究開発部の 2 名の教官が、問題文中に解答が書かれていないかなどの確認作業にあたった。

各教科の問題例を表 1 に示す。なお、表 1 に示した各問題の正答率は 0.8 程度であった。

2.2 問題冊子の構成

国語 4 冊子（F, G, H, I 冊子と名付ける）のうち、F 冊子と G 冊子はそれぞれ 50 問の設問からなる問題冊子である。F 冊子と G 冊子に共通問題

表 1 問題例

五語玉

次の文の空欄に共通に入れることばとして最も適当なものを、あとの①～⑤のうちから一つ選びなさい。

「修学旅行にしては、その旅館は至れ 尽くせ のもてなしであった。」

① つ ② ぬ ③ よ ④ り ⑤ ば (正解④ 正答率 0.805)

数学

これは持ち家を欲しいと思っている人の初夢である…

ある中古マンションの現在価格は 1000 万円である。このマンションは、一年毎に値段が下がり、新しい値段は古い値段の 90%になるという。いま 500 万円の預金を持っている人が、この預金だけでこのマンションを購入したい。この預金には年利 8%の利息（複利）がつくものとして次の問いに答えよ。

現在の預金額は現在のマンション価格の %である. (正解 50 正答率 0.732)

英語

英文が日本語と同じ内容になるように、下に与えられた語の中から最も適当なものを選んで に補いなさい。

あなたはあの少女を知っていますか？ はい、以前彼女に会ったことがあります。

「Do you know that girl? Yes, I have met before.」

① she ② that ③ her ④ him (正解③ 正答率 0.795)

はなく、試験時間は 50 分と設定されている。H 冊子と I 冊子は、それぞれ F 冊子、G 冊子から、出題範囲が元冊子とほぼ同等になるように冊子全体から抽出した 30 問の設問からなる短縮版であり、試験時間は 30 分である。各冊子の難易度については、F 冊子と G 冊子の難易度は同等であることが想定されている。H 冊子と I 冊子についても同等の難易度が想定されているが、I 冊子には長文読解問題が含まれており、H 冊子よりも難易度が高い可能性もある。

数学2冊子(J, K冊子)は, それぞれ大問4問からなる. 設問数はJ冊子27問, K冊子29問であり, 大問3(設問数9問. 数学的文章の読み取り/相対的な考え(パーセント, 等比数列)に関する問題)が共通問題である. 各冊子の難易度については, J冊子よりもK冊子の方が難易度が高いことが想定されている.

英語 2 冊子 (L, M 冊子) は、それぞれ 50 問、48 問の設問からなり、35 問が共通問題である。共通問題は、代名詞／形容詞／副詞／動詞の活用／時制／過去分詞／現在分詞／否定文／疑問文／命令文／前置詞／疑問詞／関係詞／接続詞の用法／英作文に関する問題であり、冊子 L においては設問 16～設問 50、冊子 M においては設問 6～設問 40

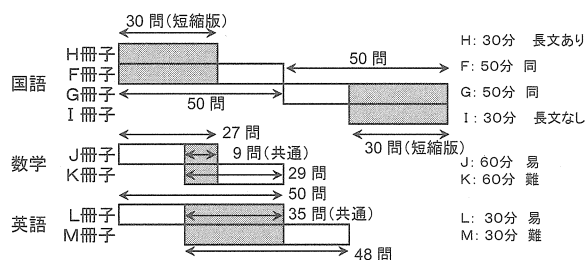


図 1 試験問題冊子の設問構成の概念図

各教科において、網掛け部分は他の冊子との共通項目であることを示す。ただし、共通項目の配置は概念図とは異なる（本文参照）。

に配置されている。各冊子の難易度については、L冊子よりもM冊子の方が難易度が高いことが想定されている。

各問題冊子の設問構成の概念図を図 1 に示す。図 1 において、各冊子ごとに横に伸びる帯がその冊子に含まれる設問の集合を表しており、帯が重なる部分（網掛け部分）は、他の問題冊子との共通項目であることを表している。例えば、国語において、H 冊子のすべての設問は F 冊子との共通項目になっているが、F 冊子と G 冊子の帯には重なる部分がなく、これらの冊子間には共通項目がないことを表している。なお、図 1 の概念図におい

ては、共通項目は冊子の先頭または最後に集中しているかのような表示をしているが、実際の問題冊子内における共通項目の配置は図とは異なる。

3 モニター調査の実施

3.1 調査大学と受験科目

平成 16 年 4 月上旬に、大学新入生を対象としたモニター調査が実施された。モニター調査の実施校は、一般入試の他に推薦入試や AO 入試を実施している、関東、中部、関西地区にある 2 つの私立大学と 1 つの私立短期大学である。

調査に用いた試験問題は、今回試作された問題、および、前回作成された国語 30 問、数学 16 問からなる国数問題（C 冊子。60 分を 30 分ずつに区切って国語から実施）である。今回作成された国語の冊子間（F-G 間、または、H-I 間）には共通項目が含まれないことから、単純には冊子間の難易度の比較はできない。そこで、前回の調査研究で作成された C 冊子を一部の受験者に実施し国語の共通項目とすることで、今回作成された国語試験問題の冊子間の難易度の比較を行うこととした。なお、C 冊子は国語の問題と数学の問題から構成されているが、国語問題の分析に利用することを考え、共通項目としては国語の部分だけを用いることとした。国語 C 冊子の出題範囲は、慣用語句／故事成語／謬・格言／手紙文・挨拶文／語句の意味・用法／漢字の書き取りと読み・意味／漢語・漢熟語の構造／説明的文章の読み取り／文学的素材の理解／文法・ことばのきまり、である。

今回作成された国語、数学、英語の 8 つの試験問題、および、前回作成された国数試験問題のすべてを同一受験者に解答させるのは、時間的制約や疲労効果の問題などがあり現実的ではない。そこで、各受験者は、国語 1 冊子、数学または国数（J, K, C のいずれか）1 冊子、英語（L か M のいずれか）1 冊子の合計 3 冊子に解答することとした。同一校で試験時間割にずれが生じないようにするため、国語については、2 つの私立大学においては F 冊子か G 冊子のいずれか 1 冊子（50 分）、短期大学においては H 冊子か I 冊子のいずれか 1 冊子（30 分）に解答させることとした。各協力校から、事前に学部・学科等に関する情報の提

供を受け、上記の問題冊子を受験する集団の所属学部・学科がなるべく同等になるように、各受験者への問題冊子の割当てを計画した。

受験者 1 人当たりの解答時間は国語 50 分または 30 分、数学または国数 60 分、英語 30 分である。各教科の間には 15～20 分の間隔を空けた。また、協力校側の都合などもあり、教科の実施順は協力校によって異なった。

3.2 受験者数

欠席者や留学生を除いた有効受験者（以下、単に受験者と言う）の人数は 1,051 名であった。冊子別の受験者数は、国語については、F 冊子 227 名、G 冊子 249 名、H 冊子 275 名、I 冊子 279 名であった。数学については、J 冊子 361 名、K 冊子 310 名であった。英語については、L 冊子 506 名、M 冊子 529 名であった。また、前回作成した国数 C 冊子の受験者は 283 名であった。

調査実施大学によって試験時間割が異なり、また、モニター調査への協力が必ずしも受験者に義務づけられていなかったため、遅刻や早退などにより、個人に割り当てられた 3 冊子すべてには解答しなかった受験者もいたため、各教科の受験者の合計（数学については、数学と国数の合計）は全体の受験者数に一致しない。割り当てられた 3 冊子すべてに解答した受験者は 939 名であった。約 1 割の受験者は割り当てられた 3 冊子のすべてには解答していないことになる。ただし、3 冊子全てに解答しなかった受験者が特定の学部・学科に偏るということはなく、系統的な欠測は生じていないものと考えられる。

4 分析方法

今回作成された「総合基礎」試験問題は、各教科ごとに作題 WG が組織され作成されたこと、また、多くの場合、教科ごとに時間を区切った形で試験が実施されることから、試作問題についての分析は教科ごとに行うことにする。つまり、個人に割り当てられた 3 冊子すべてに解答した受験者だけのデータではなく、当該科目に解答した受験者のデータはすべて用いて、分析を行うことにする。

新しい試験問題の開発研究であるという点は各教科共通であるので、分析の基本的なプロセスは

以下のように統一する。

1) まず、大学新入生においてどの程度の得点を獲得することができるかを確認するため、得点分布および得点率（満点を1とした場合の平均点）についての検討を行う。項目に正答した場合を1点、誤答した場合を0点として各被験者の得点を冊子ごとに求め、得点分布を作成する。各冊子の満点は項目数であり、国語F、Gは50点、H、Iは30点、数学Jは27点、Kは29点、英語Lは50点、Mは48点がそれぞれ満点となる。得点を満点で割った値が得点率である。

目標とされた得点率は、先に述べたように、平均7割、最頻値8割程度、および、それよりやや低い得点率（難易度を高くした試験問題）であるが、実際の難易度がどの程度になっているかを確認する。なお、冊子ごとに受験者群が異なるため、ここで得られる得点率の分布から直接各冊子の難易度を比較することはできない。その分析は5)で行う。

2) 各試験問題は複数の設問で構成されていることから、各設問の項目正答率を計算し、その分布について検討する。1)で求める全体の得点率がどんなに目標に合致するものであっても、項目正答率が極端に低い項目と極端に高い項目に分かれるなど、項目正答率の分布に著しい歪みがある場合は、得点率が中程度の受験者の能力を識別することが困難になるため、よい試験問題であるとは言えない。平均7割、最頻値8割程度の得点率の試験を作成することが当座の目的であり、この得点率付近の受験者の能力をよく推定するためには、項目正答率が0.7前後の項目を多数用意する必要がある（芝、1991など）。作成された各設問の項目正答率がこのように分布するかを確認する。

3) 次に、内的整合法による信頼性の確認（ α 係数の算出）を行う。信頼性の低い試験問題から得られる得点は、受験者の真の能力を十分反映することができず、受験者の能力について誤った診断・評価を導いてしまう可能性が高い。能力テストの信頼性の下限は0.8程度と言われており（渡部、1993など）、今回作成された試作問題がこの基準を満たし、受験者の能力を診断・評価する用途に用い得るかを確認する。

4) 異なる問題冊子の内容的等価性を確認するため、まず、共通項目得点と冊子全体の得点との相

関を確認する。それぞれの冊子において、合計得点と共通部分の得点との相関が高ければ、異なる2つの問題冊子間の相関は高いことが予想され、内容的等価性が示唆される。なお、本稿で言う内容の等価性とは、同一の能力を測定していることを指し、テスト理論においては「同族性」という用語が用いられている（池田、1994）。

5) 共通項目の得点に基づいて受験者を群分けし、各群ごとに2つの問題冊子の得点率を比較する。これは、内容的には等価である2つの冊子の難易度が同等か、または、異なるかを確認するものである。各群ごとに、2つの冊子の得点率が同程度であれば、2つの冊子の難易度は同等であると考えることができる。また、各群ごとに、一貫して一方の冊子の得点率が高い（低い）という傾向が確認できれば、内容的には等価であるが難易度が異なる2つの冊子が作成されたと考えることが可能である。

受験者の群分けは、各群の推定誤差が同等になるように、共通項目の得点率分布において、なるべく受験者数が20%ずつになるように5群に分割するものとする。

6) 異なる2つの問題冊子の内容的等価性がどの程度のものであるかを把握するため、2つの問題冊子の得点間の相関係数の値を推定する。難易度が異なる試験問題であっても、各試験問題の得点間の相関が高ければ、内容の等価性が示唆され、得点の等化を行うことにより、異なる問題冊子に解答した受験者間の得点の比較が可能となる。得点間の相関が低ければ、内容の等価性は示唆されず、得点の等化を行い異なる問題冊子に解答した受験者の得点の比較を行うことの妥当性は失われる。

今回の調査では、同一教科の2つの冊子に解答した受験者はおらず、データから直接相関係数を計算することはできない。そこで、共通項目得点と同程度である受験者群の中で、異なる問題冊子に解答した者同士の対をいくつか作り、各対をひとりの受験者とみなして同一受験者が2つの冊子に解答したかのようなデータを作成し、相関係数の推定を行う。受験者の群分けは、5)の分析で利用した方法を用いることにする。

すなわち、国語については、共通項目とするC冊子に解答した受験者を、受験者数なるべく20%ずつになるようにC冊子の得点分布に基づいて5群

に分け、各群の中で、F (H) 冊子に解答した受験者と G (I) 冊子に解答した受験者の対をいくつか作り、各対における F (H) 冊子の得点と G (I) の得点を相関データとみなして、F (H) 冊子と G (I) 冊子の得点間の相関係数を計算する。数学と英語については、各冊子に含まれる共通項目の得点分布に基づいて受験者を 5 群に分け、国語の場合と同様にして、J 冊子と K 冊子の得点間の相関係数、L 冊子と M 冊子の得点間の相関係数を計算する。

各群内において異なる冊子に解答した受験者の対を作るに当たっては、復元抽出法で一様乱数を利用して無作為に行うものとし、各群 10 名、合計 50 名分の相関データを作成し、相関係数の値を計算する。同じ操作を 100 回繰り返し、推定値の標本分布を作成して、2 つの冊子間の相関係数の推定を行うこととする。

7) 試験問題の信頼性を低めるような、特異的な解答パターンが得られた設問について検討する。合計得点を用いて受験者の能力を推定する場合、試験の信頼性を低める項目が混入することは、テスト理論 (Lord & Novick, 1968 など) の観点から言えば避けなければならないことである。信頼性を低くする設問がどのような問題形式であるか、どのような点が解答に混乱を来しているかなどについて、設問内容と分析結果に基づいて考察する。

本研究においては、信頼性の 1 つの指標として、内的整合性を確認する α 係数を用いるので、具体的には α 係数を小さくする効果をもつ項目についての検討を行う。なお、 α 係数は内的整合性の指標であることから、当該の設問に欠陥がない場合でも、他の項目との相関が小さければ α 係数を小さくする方向に機能し、また、 α 係数への貢献度は相対的なものであるということには注意しておく必要がある。

以上のようなプロセスで、以下、国語、数学、英語の試作問題の分析を行う。

5 国語問題の分析

国語 H 冊子および I 冊子は、それぞれ F 冊子、G 冊子の短縮版であり、F (G) 冊子に解答した受験者は必然的に H (I) 冊子に含まれるすべての設問に解答していること、また、調査において国語の解答時間はどの被験者においても十分であった

表 2 国語問題の得点率の基本統計量

	F 冊子	G 冊子	H 冊子	I 冊子
人数	227	249	502	528
平均値	0.707	0.605	0.719	0.618
中央値	0.740	0.620	0.767	0.633
最頻値	0.780	0.680	0.833	0.667
標準偏差	0.140	0.134	0.146	0.145
四分位範囲	0.160	0.180	0.200	0.200

様子が観察されたことから、H 冊子および I 冊子の分析に当たっては、F 冊子、G 冊子に解答した受験者のデータのうち、H 冊子、I 冊子に相当する部分のデータも利用して、分析を行うことにする。

5.1 得点分布および得点率

国語試験問題の得点率の基本統計量を表 2、得点分布を図 2 に示す。図 2 をみると、得点分布は最頻値の周りに分布していることがわかる。得点率については、F および H 冊子で平均約 7 割、最頻値約 8 割となっており、目標としていた難易度を達成していることがわかる。しかし、G および I 冊子の難易度は、平均約 6 割、最頻値 7 割程度となっており、F および H 冊子よりも難易度が高いことが予想される。得点率の散らばりを見ると、どの冊子の標準偏差も 0.14 前後となっており、平均得点率を中心に ± 0.14 の範囲におよそ 70% の受験者が分布していることが推察される。以上のことから、難易度に関しては、目標としていた難易度をよく実現している試験問題と、やや難易度が高いと予想される試験問題が作成されたと言えるであろう。

5.2 項目正答率

国語問題の各冊子に含まれる設問の項目正答率を計算し、その分布の基本統計量を表したものを表 3 に示す。平均項目正答率は、F および H 冊子では 0.7 程度、G および I 冊子では 0.6 程度であり、また、標準偏差の値は F および H 冊子で 0.22 程度、G および I 冊子で 0.27 程度であることから、F および H 冊子においては項目正答率 0.7 ± 0.22 の範囲、G および I 冊子では項目正答率 0.6 ± 0.27 の範囲に、多くの項目が分布していると考えられる。

項目正答率が 0.1 以下と極端に低い項目は、F および H 冊子には見られないが、G 冊子で 4 項目、I 冊子で 2 項目観察される。

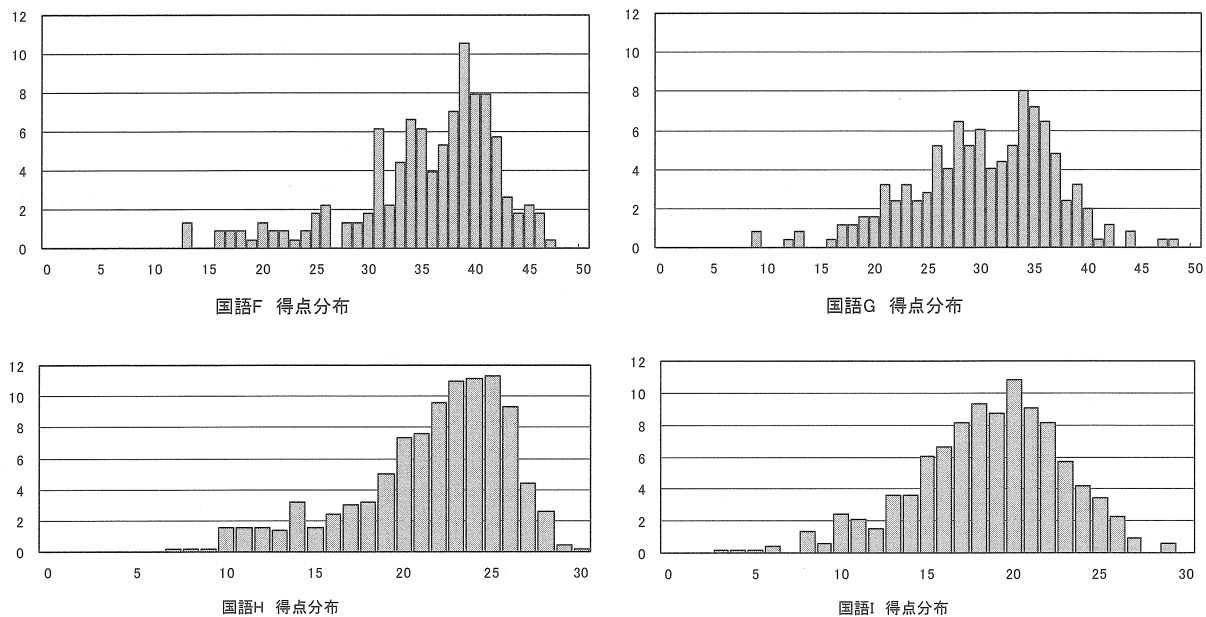


図 2 国語問題の得点分布

表 3 国語問題の項目正答率の基本統計量

	F 冊子	G 冊子	H 冊子	I 冊子
項目数	50	50	30	30
平均	0.707	0.605	0.719	0.618
標準偏差	0.221	0.269	0.221	0.274
最小値	0.145	0.060	0.131	0.064
最大値	0.974	0.984	0.968	0.985
0.10 以下の項目数	0	4	0	2
0.95 以上の項目数	4	2	4	2

G 冊子および I 冊子に含まれる長文読解問題の項目正答率は 0.6~0.8 程度であり、また、それらの設問以降の設問の項目正答率が、時間が足りなくなったなどの理由で系統的に低くなるということは観察されなかったことから、とくに I 冊子において、長文読解問題の存在が試験問題の難易度を高くしているわけではないことが推察される。G 冊子および I 冊子には正答率が 0.1 以下の項目が含まれており、むしろこれらの設問の存在が、これらの冊子の得点率を低くしているものと考えられる。

5.3 信頼性

各問題冊子の α 係数を計算すると、F 冊子 0.854、G 冊子 0.829、H 冊子 0.783、I 冊子 0.765 という結果である。一般に、設問数が多くなるほど α 係数の値は大きくなるので、50 設問からなる F、G

冊子の方が、30 設問からなる H、I 冊子よりも α 係数の値は大きくなる。とはいえ、H、I 冊子の α 係数も 0.8 程度となっている。2.2 節で述べたように、H 冊子、I 冊子の出題範囲も多岐に渡っており、そのような場合でも 30 設問で α 係数が 0.8 程度となる試験問題を作成しうることが確認される。

5.4 共通項目得点との相関

共通項目となる前回作成した国語問題 (C 冊子) の得点との相関を見ると、各冊子の相関係数の値 (および受験者数) は、F 冊子 0.693 (89 名)、G 冊子 0.744 (81 名)、H 冊子 0.628 (147 名)、I 冊子 0.712 (136 名) となっている。C 冊子の α 係数は 0.695 であり、信頼性の大きさによる相関係数の希薄化 (芝・南風原, 1990 など) を考慮すると、共通項目得点との相関が一番低い H 冊子でも 0.6 以上、F、G、I 冊子では 0.7 前後という値になっていることから、今回作成した試験問題の冊子間の内容的等価性が示唆される。

F (G) 冊子の短縮版として H (I) 冊子が作成されていることから、F (G) 冊子に解答した受験者において、F (G) 冊子 50 項目の合計得点と、H (I) 冊子に含まれる 30 項目の合計得点の間の相関係数を求めると、F-H 間 0.947、G-I 間 0.941 という高度に高い相関係数の値が観察される。また、F (G) 冊子において、H (I) 冊子に含まれる 30 項

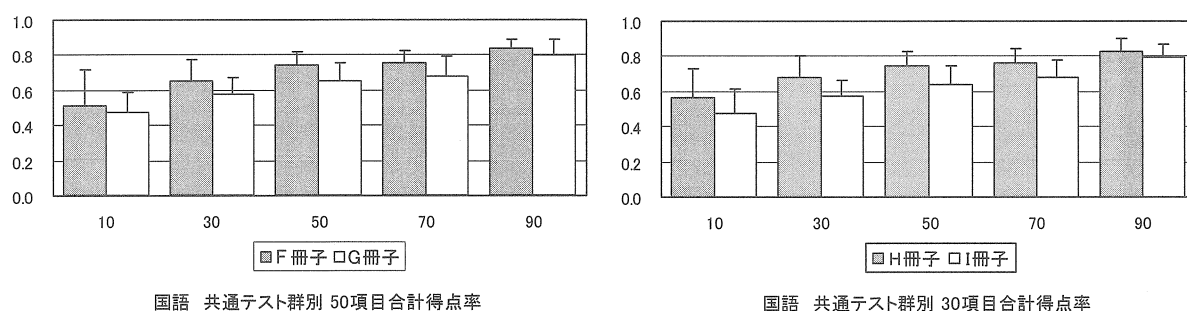


図3 国語問題の群別得点率の比較

目の合計得点と、H (I) 冊子に含まれない部分の合計得点間の相関係数は、それぞれ、0.712, 0.707 という高い値である。このことと、難易度および信頼性についての結果をあわせて考えると、H (I) 冊子は内容的に等価な F (G) 冊子の短縮版であると言えることができると考えられる。

5.5 群ごとの得点率の比較

共通項目として使用した前回作成した30設問からなる国語問題 (C 冊子) を受験した283名の受験者の得点分布 (1問1点、満点30点) は、平均値 20.940, 標準偏差 3.609, 最小値 10, 最大値 29 であった。この得点分布において、なるべく受験者数が20%ずつになるように受験者を5群に分け (階級値を10, 30, 50, 70, 90とする), F, G, H, I 冊子の平均得点率を群ごとに比較したものを図3に示す。図3は、項目数別に、F, G 冊子の図と、H, I 冊子の図に分けてある。図3を見ると、共通項目得点が高い群ほど、それぞれの冊子の平均得点率は高くなっており、また、どの群においても F (H) 冊子よりも G (I) 冊子の方の平均得点率が低くなっていることが観察される。以上のことから、F (H) 冊子と G (I) 冊子は、内容的に等価であるが難易度が異なる問題となっていることが確認される。

5.6 冊子間相関係数の推定

分析方法6) で述べた方法に従って、F 冊子と G 冊子間、H 冊子と I 冊子間の相関係数の値を推定すると、F-G 間の相関係数の推定値の平均値は0.509, 中央値 0.509, 標準偏差 0.088, 四分位範囲 0.106 であった。また H-I 間の相関係数の推定値の平均値は 0.466, 中央値 0.466, 標準偏差 0.090,

四分位範囲 0.129 という結果であった。

共通項目得点が同程度である受験者群として、共通項目の得点に基づいて受験者数を約20%ずつの幅に区切るという、ある程度能力分布に幅のあると考えられる群を用いた中で相関データの対応付けを行っていること、また、F (H) 冊子と G (I) 冊子には共通項目が含まれないことを考えれば、0.5程度という相関係数の推定値は比較的高い値であると解釈でき、冊子間の内容的等価性が支持されたと言えることができるであろう。このことから、異なる冊子 (F か G のいずれか、または、H か I のいずれか) に解答した受験者の得点を、得点の等化 (前川, 1999 など) を行うことによって互いに比較可能にすることができることが示唆される。

5.7 特異的な解答パターンを示す設問の分析

国語の問題において試験問題の信頼性を低めるような特異的な解答パターンを示す設問には、項目正答率が必ずしも低くはないものが含まれていた。これらの項目では、受験者を共通項目の得点に基づいて5群に分けたとき、群別の項目正答率がそれほど変わらないという現象が見られた。

例えば、F および H 冊子に含まれる「あの時のきみの一打は、『』逆転かと思わせるような惜しい当たりだったね。」という文章の空欄 (『』内) に入る最もふさわしいことばを選ぶ問題の項目正答率は0.6程度であるが、正解である『あわや』以外に、『よもや』を選択する受験者がどの群においても一定程度以上観察された。確かに最もふさわしいのは『あわや』であるが、『よもや』でも理解できてしまうところが難点であったと考えられる。

また、G 冊子に含まれる「彼女から心のこもっ

た手紙をいただいた。水茎の跡も『』筆遣いだった。」という文章の空欄に入れる最もふさわしいことばを選ぶ問題では、正解である『うるわしい』を選択した受験者はどの群でも5割程度であり、『ものうげな』『はかなげな』『このましい』『かぐわしい』など他の選択肢を選ぶ受験者もそれぞれ一定程度観察された。選択率が最も低かったのは『このましい』で選択率8.8%であるが、それ以外の用語は、日常あまり用いられないため馴染みがなく、どの群においてもいずれが正答か判断がつかない受験者が多くいたのではないかと考えられる。

項目正答率が著しく低く、かつ、試験問題の信頼性を低める設問は、GおよびI冊子に含まれる「怒り心頭に『』して、席を蹴りそのまま帰国の途に就いた。」の空欄に入る漢字を選ぶ問題である。正解の『発』を選んだ受験者は全体の9%以下であり、7割以上の受験者は誤答である『達』を選んでいた。とくに、階級値が高い群ほど『達』を選ぶ割合が高くなっていたことは注目すべきことである。「怒り心頭に～」という用語を知っているものは、ほぼ確実に「怒り心頭に達して」と記憶していることが推察される。

以上の例から、正答でなくても理解できてしまうもの、受験者にとって馴染みの薄いもの、間違っただけで記憶されているものなどが含まれる設問は、試験問題の信頼性を低くする可能性が高く、テスト理論の立場からは検討の余地があると考えられる。ただし、国語的な意義と、大学教育において必要な基礎能力とのすり合わせは別に検討されるべき課題である。

6 数学問題の分析

6.1 得点分布および得点率

数学試験問題の得点率の基本統計量を表4、得点分布を図4に示す。平均得点率は、J冊子0.460、K冊子0.383となっており、前回の調査研究（荒井他，2003；石井他，2004a）と同様に、目標としていた難易度を大きく下回っていることが確認される。ただし、難易度の調整という点では、J冊子よりもK冊子の方が平均得点率が低くなっており、異なる難易度の冊子が作成されていることが予想される。

標準偏差の値は、J冊子0.234、K冊子0.229となっており、国語の問題に比べ得点率の散らばりが大きいことがわかる。実際、図4をみると、得点分布は低得点から高得点まで広く分布していることが確認される。

6.2 項目正答率

数学問題の各冊子に含まれる設問の項目正答率を計算し、その分布の基本統計量を表したものを表5に示す。平均項目正答率は、J冊子では0.460、K冊子0.383、標準偏差の値は、J冊子0.216、K

表4 数学問題の得点率の基本統計量

	J冊子	K冊子
人数	361	310
平均値	0.460	0.383
中央値	0.444	0.383
最頻値	0.333	0.345
標準偏差	0.234	0.229
四分位範囲	0.370	0.310

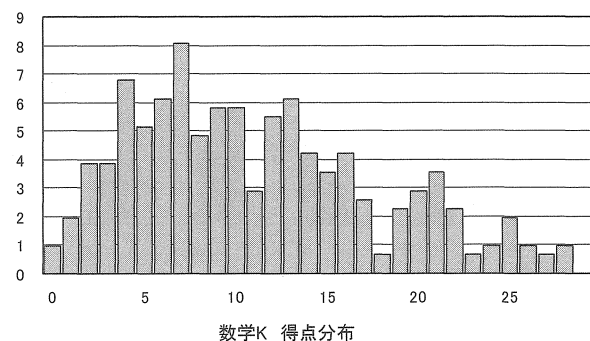
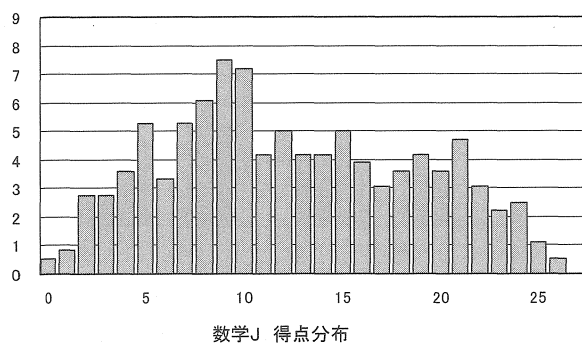


図4 数学問題の得点分布

表 5 数学問題の項目正答率の基本統計量

	J 冊子	K 冊子
標本数	27	29
平均	0.460	0.383
標準偏差	0.216	0.176
最小値	0.025	0.048
最大値	0.823	0.745
0.10 以下の項目数	2	2
0.95 以上の項目数	0	0

冊子 0.176 であり、さらに、項目正答率の最大値が、J 冊子 0.823、K 冊子 0.745 であることから、J 冊子、K 冊子ともに、項目正答率は低い範囲に多く分布していると言える。項目正答率が 0.1 以下と極端に低い項目は、J 冊子、K 冊子ともに 2 項目観察される。

6.3 信 頼 性

各問題冊子の α 係数を計算すると、J 冊子、K 冊子ともに 0.895 という値になっている。数学問題の出題範囲は 2.1 節に述べたようになっているが、そのような場合にも 30 問弱の設問で α 係数の値が 0.8 程度となる問題を作成し得ることが確認される。

6.4 共通項目得点との相関

J 冊子、K 冊子に共通である大問 3 (設問数 9 問) の共通項目得点と、J 冊子、K 冊子の合計得点との相関を見ると、相関係数の値は、J 冊子 0.893、K 冊子 0.870 という高い値になった。また、共通項目得点と、大問 3 の得点を除いた各冊子の得点との相関係数の値も、J 冊子 0.668、K 冊子 0.643 という高い値であった。このことから、冊子間の内容的等価性が示唆される。

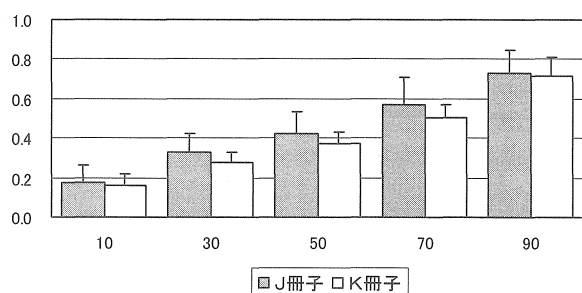
6.5 群ごとの得点率の比較

J、K 冊子のいずれかに解答した 671 名の受験者における共通項目 (全 9 問) の得点分布は、平均値 4.429、標準偏差 3.079、最小値 0、最大値 9 であった。この得点分布において、なるべく受験者数が 20% ずつになるように受験者を 5 群に分け (階級値を 10, 30, 50, 70, 90 とする)、J、K 冊子の平均得点率を群ごとに比較したものを図 5 に示す。図 5 の a) の縦軸は、冊子全体の合計得点の平均得点率、b) の縦軸は、非共通部分の合計得点の平均得点率である。

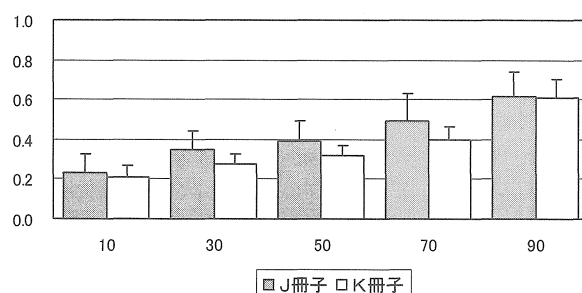
図 5 の双方の図を見ると、共通項目得点が高い群ほど、それぞれの冊子の平均得点率が高くなっていることが観察される。各群において J 冊子と K 冊子の平均得点率を比較すると、最も共通項目得点の低い群 (階級値 10) と、最も共通項目得点の高い群 (階級値 90) においては、冊子間の平均得点率の違いがほとんど見られないものの (J 冊子よりも K 冊子の方が平均得点率が低くはなっている)、共通項目得点为中位の 3 群 (階級値 30, 50, 70) においては、J 冊子よりも K 冊子の方が平均得点率が低く難易度が高くなっていることが確認される。以上から、J 冊子と K 冊子は、内容的に等価であるが難易度が異なる問題となっていることが示唆される。ただし、低階級値群と高階級値群においては、得点に床効果、天井効果が働いている可能性が考えられる。

6.6 冊子間相関係数の推定

分析方法 6) で述べた方法に従って、J 冊子と K 冊子の合計得点間の相関係数の値を推定すると、推定値の平均値は 0.750、中央値 0.750、標準偏差



a) 数学 共通項目得点群別 冊子全体合計得点率



b) 数学 共通項目得点群別 非共通部分合計得点率

図 5 数学問題の群別得点率の比較

0.052, 四分位範囲 0.068 であった。また共通項目を除いた J 冊子と K 冊子の得点間の相関係数の推定値の平均値は 0.418, 中央値 0.422, 標準偏差 0.113, 四分位範囲 0.146 という結果であった。設問数の少ない J 冊子でも合計 27 設問あり, その 1/3 である 9 設問だけが共通項目であることを考えれば, 0.75 という J 冊子と K 冊子の合計得点間の相関係数の推定値はかなり高い値であると解釈できる。また, 非共通部分の相関係数の推定値は 0.42 程度となっているが, 共通項目得点が同程度である受験者群として, 共通項目の得点に基づいて受験者数を約 20% ずつの幅に区切るという, ある程度能力分布に幅のあると考えられる群を用いた中で関連データの対応付けを行っていることを考えれば, 決して低い値ではないと考えられ, 冊子間の内容的等価性が支持されたと言うことができるであろう。このことから, 異なる冊子 (J か K のいずれか) に解答した受験者の得点を, 得点の等化を行うことによって互いに比較可能にすることができることが示唆される。

6.7 特異的な解答パターンを示す設問の分析

数学の問題において試験問題の信頼性を低めるような特異的な解答パターンを示す設問は, 多肢選択式問題であった。

例えば, J 冊子と K 冊子で, ある平方根の大きさと, 2 次方程式の 2 つの解の大小関係を選択肢の中から選ぶ問題は, 項目得点と合計得点との相関が低く信頼性を低める項目となっていた。

やはり J 冊子と K 冊子で内容は異なるものの, 図形の形状を選択肢の中から選ぶ問題についても, 試験問題の信頼性を低くする可能性が伺われた。

また, K 冊子に含まれる「グラフの読み取り (組合わせ問題, 推定問題)」に関する問題に挿入された, 標本から母集団の特性を完全に言い当てることのできるかを問う設問については, 多くの受験者が解答に混乱を来した可能性が伺われた。この問題は二者択一問題であったが, どの群においても, 正答と誤答の選択率が同等か誤答の方が選択率が高くなるという結果であった。このような解答パターンが得られた背景には, 通常の数学の問題におけるグラフに関する問題が, 与えられたグラフ (だけ) に注目して解答を求めるのに対し, この設問は, 与えられたグラフ以外のグラフが存在

する可能性を問う設問であったため, 受験者が混乱し, 当て推量により解答を選択したということが, 1 つの理由として考えられる。

数学問題の解答形式は, 解答欄の各マスに対応する数字を塗りつぶす形式が原則とされており, 当て推量により正解となる可能性は低い。以上の例を見ると, そのような解答形式の問題に, 当て推量により正解となる可能性が一定程度以上ある多肢選択式の設問を挿入することは, 試験問題の信頼性を低くする可能性が高く極力回避すべきことであると考えられる。

7 英語問題の分析

7.1 得点分布および得点率

英語試験問題の得点率の基本統計量を表 6, 得点分布を図 6 に示す。平均得点率は, L 冊子 0.600, M 冊子 0.507 となっており, 目標としていた難易度をやや下回っていることが観察される。難易度の調整という点では, L 冊子よりも M 冊子の方が平均得点率が低くなっており, M 冊子の方が難易度が高いことが予想される。

標準偏差の値は, L 冊子 0.196, M 冊子 0.222 となっており, 国語の問題に比べ得点率の散らばりが大きい, 数学の問題に比べると散らばりは小さいと言える。図 6 をみると, 得点が 10 点未満 (得点率約 0.2 未満) という極端に低得点の受験者は少ないものの, 10 点以上 (得点率約 0.2 以上) から高得点まで広く得点が分布していることが確認される。

7.2 項目正答率

英語問題の各冊子に含まれる設問の項目正答率を計算し, その分布の基本統計量を表したものを

表 6 英語問題の得点率の基本統計量

	L 冊子	M 冊子
人数	506	529
平均値	0.600	0.507
中央値	0.600	0.507
最頻値	0.600	0.479
標準偏差	0.196	0.222
四分位範囲	0.300	0.354

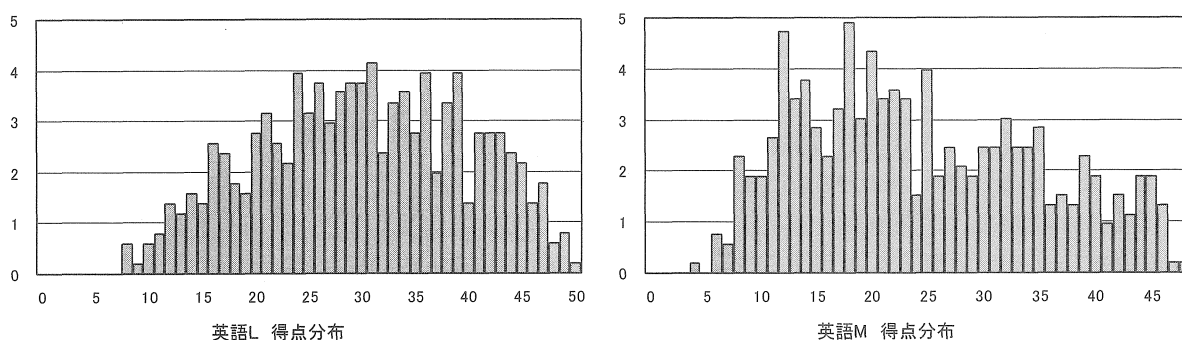


図 6 英語問題の得点分布

表 7 英語問題の項目正答率の基本統計量

	L 冊子	M 冊子
項目数	50	48
平均	0.600	0.507
標準偏差	0.196	0.155
最小値	0.176	0.189
最大値	0.941	0.800
0.10 以下の項目数	0	0
0.95 以上の項目数	0	0

表 7 に示す。平均項目正答率は、L 冊子 0.600、M 冊子 0.507、標準偏差の値は、L 冊子 0.196、M 冊子 0.155 である。また、いずれの冊子においても、項目正答率が 0.10 以下、または、0.95 以上の項目は見られないことから、L 冊子、M 冊子ともに項目正答率は中程度の範囲に多く分布していると言える。

7.3 信 頼 性

各問題冊子の α 係数を計算すると、L 冊子 0.912、M 冊子 0.923 という値になっている。英語問題の出題範囲は 2.1 節で述べたようになっており、それらに関する 50 問程度の設問数で、 α 係数の値が 0.9 を超える問題冊子が作成されたことが確認される。

英語問題と国語問題の設問数はともに 50 問（または 50 問程度）であり、同程度の項目数を用いて国語問題よりも α 係数が大きい試験問題が作成されていると言える。しかし、国語問題は目標としていた平均得点率 7 割程度を実現しており、その分、易しい項目を多く含み、それらの項目間の相関が低くなるので、国語問題は英語問題ほど信頼性が高くなかったということも考えられる。

7.4 共通項目得点との相関

L 冊子、M 冊子に共通である 35 設問の共通項目得点と、L 冊子、M 冊子の合計得点との相関を見ると、相関係数の値は、L 冊子 0.978、M 冊子 0.979 という高度に高い値であった。また、共通項目得点と、共通項目を除いた各冊子の得点との相関係数の値は、L 冊子 0.693、M 冊子 0.737 という高い値であった。50 設問中半数以上の 35 設問が共通であることから冊子間の内容的等価性は示唆されるが、共通部分と非共通部分の相関も 0.7 程度と高い値であることから、冊子間の内容的等価性は支持されと考えられるであろう。

7.5 群ごとの得点率の比較

L、M 冊子のいずれかに解答した 1,035 名の受験者における共通項目（全 35 問）の得点分布は、平均値 18.422、標準偏差 7.830、最小値 2、最大値 35 であった。この得点分布において、なるべく受験者数が 20% ずつになるように受験者を 5 群に分け（階級値を 10、30、50、70、90 とする）、L、M 冊子の平均得点率を群ごとに比較したものを図 7 に示す。図 7 の a) の縦軸は、冊子全体の合計得点の平均得点率、b) の縦軸は、非共通部分の合計得点の平均得点率である。

図 7 の双方の図を見ると、共通項目得点が高い群ほど、それぞれの冊子の平均得点率は高くなっていること、また、どの群においても、L 冊子よりも M 冊子の方が平均得点率が低く難易度が高くなっていることが確認される。以上から、L 冊子と M 冊子は、内容的に等価であるが難易度が異なる問題となっていることが示唆される。

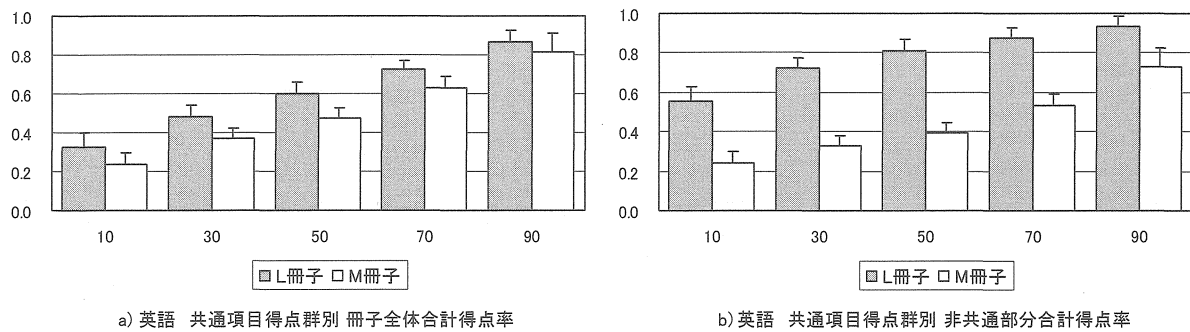


図 7 英語問題の群別得点率の比較

7.6 冊子間相関係数の推定

分析方法 6) で述べた方法に従って、L 冊子と M 冊子の合計得点間の相関係数の値を推定すると、推定値の平均値は 0.904、中央値 0.904、標準偏差 0.018、四分位範囲 0.026 であった。また共通項目を除いた L 冊子と M 冊子の得点間の相関係数の推定値の平均値は 0.463、中央値 0.475、標準偏差 0.090、四分位範囲 0.129 という結果であった。L 冊子 50 設問、M 冊子 48 設問中、半数以上の 35 設問が共通項目であることを考えれば、両冊子間の相関係数の推定値が 0.9 程度になることは理解できる。非共通部分の相関係数の推定値は 0.46 程度となっているが、共通項目得点が同程度である受験者群として、共通項目の得点に基づいて受験者数を約 20% ずつの幅に区切るという、ある程度能力分布に幅のあると考えられる群を用いた中で相関データの対応付けを行っていることを考えれば、決して低い値ではないと考えられる。冊子間の内容的等価性が支持されたと言ったことができるであろう。このことから、異なる冊子 (L か M のいずれか) に解答した受験者の得点を、得点の等化を行うことによって互いに比較可能にすることが示唆される。

7.7 特異的な解答パターンを示す設問の分析

英語の問題において試験問題の信頼性を低めるような特異的な解答パターンを示す設問は、群別の平均項目正答率が、中階級値群で変化なしに低くなり、高階級値群においても正答率がそれほど高くないということが観察される項目であった。受験者の群分けは共通項目の合計得点に基づいて行われており、合計得点が高い群ほど各項目

の正答率も高くなることと一般的に予想される。平均項目正答率が中階級値群で低くなるということは、何らかの影響で、低階級値群の正答率が高くなっているか、中階級値群の正答率が低くなっていることが考えられる。

例えば、L 冊子に含まれる「Do we have any (feed, food, foods, foot) in the house? (家に食べ物はある?)」という文で正解 (food) を選ぶ問題では、各群の正答率が、共通項目得点が低い群から順に、0.539, 0.500, 0.462, 0.337, 0.495 となっていた。不正解答の大多数は foods であり、その選択率は順に、0.382, 0.479, 0.539, 0.652, 0.495 という値であった。この選択率から、低階級値群では「食べ物 = food」と単純に考えて解答し正解となり、階級値が高い群では「any の後の名詞は複数形だから foods」と考え不正解となる受験者が多く、「食べ物という意味では food は単数形」ということまで理解して解答した受験者は高階級値群においても半数以下であったという可能性が推察される。food が馴染みある単語でありながら、単数形と複数形の使い方が十分に理解されていないことにより、正答率が中階級値群で低下し、高階級値群でも正答率がそれほど高くないという現象が観察されたと考えられる。

また、「He looked at us very (anger, angry, angrily, much anger). (彼は大変腹立たしげに私たちを見た)」という文で正解 (angrily) を選ぶ問題では、各群の正答率が、共通項目得点が低い群から順に、0.232, 0.196, 0.287, 0.369, 0.636 となっていた。誤答として低階級値群では much anger を選ぶ割合が高い傾向が見られ、中階級値群では angry、高階級値群では anger を選ぶ割合が若干高く観察された。「大変 = very much」と考えられた

ことと, anger, angry, angrily の品詞の区別ができていないことが, 解答に混乱を来した原因であると考えられる。

「そのケーキはあまりおいしくなかった。」という文を英訳する問題で, 「The cake (nice, taste, very, didn't).」の () 内の単語を並べ替えて最初の 2 つの並びを解答する問題の正答率は, 共通項目の合計点が低い群から順に, 0.317, 0.353, 0.413, 0.413, 0.667 となっていた。正答は「The cake didn't taste very nice.」であるが, 「The cake didn't very nice taste.」または「The cake taste didn't very nice.」と作文していると思われる受験者がそれぞれ約 3 割, 1 割ずつと多数観察された。この問題では, 文には動詞が必要であり taste を動詞として使うことに気がつくかどうか問われているが, 「taste = 味」という観念強く taste を名詞として扱いたいために, 助動詞である didn't を動詞であるかのように扱ったということが考えられる。

以上の設問は, 試験問題の信頼性を低くする項目ではあったが, 英語を学習するという点ではどれも興味深い問題であると考えられる。しかし, 正答率が高階級値群でもそれほど高くないという点からすれば, 「総合基礎」問題としては検討の余地があると考えられよう。

8 全体的考察と今後への提言

「総合基礎」問題は, AO 入試や推薦入試など, 一般入試を経験しないで大学に入学してくる学生の学力到達度測定などへの用途を考えて開発が進められているものであり, 得点率の目標値として平均約 7 割, 最頻値約 8 割が掲げられている。今回の調査研究では, AO 入試や推薦入試などを実施している私立大学, 私立短期大学の新生生に対し 4 月上旬に調査を実施しており, 想定される受験者にほど近い被験者を用いて調査が実施されたと考えられる。

国語の試験問題は, 目標としていた得点率をよく実現した試験問題が作成され, また, 難易度は多少高くなっているものの, 内容的に等価な別の試験問題も作成された。それぞれの試験問題と関連の高い短縮版の試験問題も作成された。

長文読解問題の存在が試験問題全体の得点率を低下させるということは観察されず, 得点率を低

下させるのは, むしろ正答率が著しく低い項目であることが推察された。つまり, 単問形式が原則である「総合基礎」試験問題であっても, 長文読解問題を挿入することは可能であることが確認された。

以上のことから, 国語については「総合基礎」問題の試作が概ね完了したと言えるであろう。

ただし, 今回作成された試験問題冊子間には共通項目が含まれないことから, 複数の受験者が異なる冊子に解答した際の得点の比較は容易ではないことが予想される。今回の調査では, 前回作成した国語問題を共通項目として実施したため, これを利用して異なる冊子間の得点を比較し, 難易度の検討, および, 異なる冊子の得点の比較可能性の検討を行うことができたが, 共通項目が存在しない場合にはこれらの検討は困難である (異なる冊子の双方に解答する共通受験者がいれば可能ではある)。

複数の試験問題を作成して提供する場合には, それぞれの試験問題の難易度の比較ができ, また, 異なる試験問題における得点の比較を行うことができるが必要となると考えられる。今後の作題では, 共通項目を一定程度挿入した複数の試験問題の作成が望まれるであろう。

数学の試験問題は, 平均得点率が 4 割前後となり, 内容的には等価であっても難度が高い問題が作成された。これは前回の試作問題と同様の結果であり (石井他, 2004a), 難易度については改善がみられなかったと言える。

数学問題の得点分布は低得点から高得点まで広く分布しており (図 4), すべての受験者にとって難しい試験問題であったというわけではないことがわかる。つまり, 全体を平均すれば約 4 割の得点率であるが, 難しいと感じた受験者もいれば, 易しいと感じた受験者も多数いたということである。高等学校においてクラスは理系/文系と分けられ, 数学の履修度には学生により差があることから, 数学の得点が広く分布することは容易に予想されることである。

このような状況において, 平均得点率 7 割程度の試験問題を作成することは, 易しい項目で試験問題を構成することによりとりあえずは可能となるが, 大学教育においては, いわゆる文系学部と理系学部 (経済学部も含む) では, 数学的能力の

必要性に差があることが示唆されており（石井他、2004b）、全体的な平均値として得点率7割を目指すのは適当なことではないことと考えられる。数学問題は、例えば理系用と文系用など、数学的能力の必要性に応じて、難易度がある程度異なる問題を作成するという方向性が検討される必要があると考えられる。

英語の試験問題は、内容的には等価であるが、平均得点率が約6割、または、約5割となり、難易度が目標よりもやや高い2つの問題が作成された。しかし、信頼性を多少抑えることにはなるが、易しい項目を入れるなどして、難易度を調整することは可能であると考えられる。

2つの冊子間に共通項目が半数以上あり、難易度および得点の比較は十分に行えたが、複数の異なる試験問題を作成するという観点からすれば、共通項目の数が多少過剰気味であると言えよう。数学の問題では全設問のうち約1/3が共通項目であったが、これだけでも難易度および得点の比較は行い得ている。また、前回の作題でも、国語、数学において、それぞれ1/3程度の共通問題を挿入して問題が作成されている。英語についてもこれを1つの目安として問題の作成を試み、難易度および得点の比較が可能であるかどうかを検討することが必要であると考えられる。

なお、今回の調査では、同一教科の2つの冊子に解答した受験者がおらず、異なる冊子の得点間の相関係数の値をデータから直接求めることができなかったため、相関係数を推定する方法として、共通項目得点が同程度である受験者群の中で、異なる問題冊子に解答した者同士の対をいくつか作り、各対をひとりの受験者とみなして同一受験者が2つの冊子に解答したかのようなデータを作成し、相関係数の推定を行うという方法を考えた。その際、受験者の群分けは、共通項目の得点率分布において、なるべく受験者数が20%ずつになるように5群に分割した（4.分析方法5）、6）参照）。この分析方法について、群の数を増減させる、また、受験者数ではなく得点区間に基づいた群分けを行うなど、群の分け方を変えた場合に、相関係数の推定値がどのように変化するかについて、今後、検討する必要もあるであろう。

9 ま と め

大学教育に必要な基礎学力を診断・評価するための、平均得点率7割、最頻値8割を目標とする、国語、数学、英語の試験問題の作成が試みられ、大学新入生を対象に調査研究が実施された。

国語に関しては、目標としていた得点率を実現する50項目からなる試験問題を作成することができた。また、それよりもやや難易度の高い平均得点率6割、最頻値7割程度の50設問からなる試験問題、および、30項目からなるそれぞれの試験問題の短縮版も作成され、それらの信頼性、内容的等価性が確認された。国語問題において、長文読解問題を挿入し得ることも確認された。ただし、設問数が同じ2つの冊子間に共通の設問が含まれていないため、それぞれの冊子に解答する受験者の得点を比較することには困難が予想された。

数学については、難易度は異なるが内容的に等価である2つの試験問題が作成されたと言えるものの、その難易度は目標を大きく下回り平均得点率は4割前後という低い値であった。しかし、得点分布は低得点から高得点まで広く分布していた。大学教育に必要な数学の基礎的能力は学部によっても異なると考えられることから、難易度がある程度異なる複数の試験問題を開発し、数学的能力の必要性に応じた適用を検討する必要性が示唆された。また、解答形式として、解答欄の各マスに対応する数字を塗りつぶす形式と多肢選択形式の混在は、試験問題の信頼性を低下させる可能性が考えられた。

英語については、平均得点率が目標よりもやや低い平均6割、または、約5割という2つの試験問題が作成され、それらが難易度は異なるが内容的には等価であることが確認された。しかし、信頼性を多少抑えることにはなるが易しい項目を挿入するなどし、目標とする難易度を達成する試験問題を作成することが検討された。また、両冊子間には共通項目が半数以上含まれており、複数の試験問題を提供するという観点からは共通項目の削減を検討する必要性が示唆された。

謝 辞

今回の調査研究は大学入試センター研究開発部共同研究Ⅱ「総合試験問題の分析的研究」の一環として実施されました。「総合基礎」試験問題の作成に携わって下さった作題ワーキング・グループの先生方、及び、本調査における「総合基礎」試験問題の使用を許可して下さった第2期総合問題調査研究委員会に厚く御礼申し上げます。また、調査協力校の選定に御助言頂いた関西国際大学の濱名篤教授に感謝申し上げます。調査および予備調査にご協力頂いた大学の先生方、事務の方々、受験して下さった学生の皆さんにも感謝致します。

参考文献

- 荒井克弘他 (2003). 「総合基礎問題調査研究委員会研究報告書」, 大学入試センター.
- 藤井光昭・柳井晴夫・荒井克弘 (2002). 大学入試における総合試験の国際比較, 多賀出版.
- 池田 央 (1994). 現代テスト理論, 朝倉書店.
- 石井秀宗・椎名久美子・柳井晴夫・荒井克弘・中山長年・山本善彦 (2004a). 国語と数学の基礎学力評価試作問題についての検討, 大学入試研究ジャーナル, No.14, pp.127-134.
- 石井秀宗・椎名久美子・柳井晴夫・前田忠彦・鈴木規夫・荒井克弘・大竹洋平 (2004b). 大学生の学力低下に関する教員の意識について, 大学入試センター研究開発部リサーチノート, RN-04-13.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*, Reading MA, Addison-Wesley Publishing Company.
- 前川眞一 (1999). 得点調整の方法について, 柳井晴夫・前川眞一編, 大学入試データの解析—理論と応用, 現代数学社, pp.88-109.
- 芝 祐順・南風原朝和 (1990). 行動科学における統計解析法, 東京大学出版会.
- 芝 祐順 (1991). 項目反応理論, 東京大学出版会.
- 渡部 洋編著 (1993). 心理検査法入門, 福村出版.
- 柳井晴夫・前川眞一・鈴木規夫・石塚智一・豊田秀樹 (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研究, 大学入試センター.
- 柳井晴夫・椎名久美子・石井秀宗 (2002). 大学生の学習に対する意欲等に関する調査研究 (平成 12, 13 年度文部科学省委託研究報告書), 高等学力調査研究会.

Developing new tests of Japanese, mathematics and English to evaluate basic ability

ISHII Hidetoki*
SHIINA Kumiko*
YANAI Haruo*
IWATSUBO Shuichi*
ARAI Katsuhiko**

Abstract

Japanese and mathematics tests were developed in order to evaluate basic ability necessary for higher education. Ishii et al. (2004a) investigated several characteristics of those tests using data from high school students. In the present study, new versions of Japanese and mathematics tests were developed based on the findings obtained by Ishii et al. (2004a), and also tests on English were newly developed. These tests were administered to college freshmen. The results show that the Japanese tests meet the criteria. However, difficulties of the mathematics and English tests need to be adjusted, although the tests are considered equivalent within each of these subject areas.

* National Center for University Entrance Examinations

** Tohoku University

