

文系学生向き総合論述問題による能力測定の試み

平井 洋子*
椎名 久美子**
柳井 晴夫***

要 約

大学入学後の学習活動を基準とした入試問題の試みとして、文系大学生の学習プロセスを意識した総合論述問題が作成された。そして、モニター調査（大学入試センター、1999a, 1999b）の結果から試作された総合論述問題の測定論的な特徴を明らかにし、このような総合論述問題の可能性と課題について検討した。分析によると、回答文字数が多く主観的な判断にもとづいて採点された設問において、採点者間の評価の一致度がかなり低かった。この不一致は、主として評価の観点や重視する側面が採点者ごとに異なることが原因であった。しかし全体としては、表現力、理解力、視野の広さ、展開力、要約力などの方向が測定可能であることが示され、文系学生の学習プロセスをほぼカバーした測定を行うことができた。また3つの測定方向（因子）のうち2つの因子で文系・理系による差がみられ、文系の学業適性を測定している可能性が示唆された。これにより、教科学力試験でも小論文試験でもない、総合的な試験問題の実用可能性が示されたといえる。

1 はじめに

小論文試験は、その採点の信頼性と妥当性がかねてより問題にされてきた。採点の信頼性は、採点者バイアスの面から論じられることが多い（Blok, 1985, van der Kamp and Mellenbergh, 1976, など）。たとえば渡部・平・井上（1988）は、全体的評価の採点者間一貫性がきわめて低いこと、採点者内の安定性にも個人差があること、分析的評価は全体的評価との重相関係数が低いものの全体的評価と同程度の信頼性を持つこと、などを見いだした。一方妥当性の面では、どのような能力が測定されているのかが問

題とされてきた。Benton et.al.(1984) や Benton and Kiewra (1986)は、文や段落の並べ替え課題が良くできる被験者は小論文の全体的評価も高いことを見いだしたほか、Huot (1990)は、全体的評価にあたって採点者は、小論文の構成のほか内容や着想にも着目することを指摘している。これらのことから小論文試験では、主に論理的構成力や展開力、表現力などが測定されていると推察することができる。とはいえ外的基準との関連を実証的に検討した研究は少なく、実際にどのような能力が測定されているのかはあまり明らかになっていないようである。

全国の国公立大学を対象とした調査（大学入試センター研究開発部, 1999a）によると、小論文

* 東京都立大学人文学部（心理学専攻）

〒192-0397 東京都八王子市南大沢 1-1 e-mail: yhirai@bcomp.metro-u.ac.jp

** , *** 大学入試センター 研究開発部 進学適性研究部門

〒153-8501 東京都目黒区駒場 2-19-23 e-mail: shiina@rd.dnc.ac.jp

yanai@rd.dnc.ac.jp

文形式の入学試験は調査で回答のあった学科コースの 85.7%で実施されており、中でも教育系を中心とした文系学部で多く実施されている。また小論文試験は主として「理解力」「読解力」「論理的思考力」「文章表現力」「思考の柔軟性」の測定を意図して作成され、高校での教科学習にとらわれることなく、「出典資料を用い、日本語によって問題文を提示する」形式で出題されることが多いという。

このように日本の大学入試の中で小論文試験は、測定論上の課題はあるものの、教科学力の試験では測定しにくい能力を測定することが期待され、また一定の成果をあげているといえる。

ところで、文系の大学生が研究活動を行う際には、数多くの文献を調べてその内容を正しく理解し、その意義を評価して既知の情報と組合せ、それをもとに自分の考えを構築していくというプロセスをたどることが多い。教科学力による入学試験を、入学後の学習活動の土台となる知識をみるものと位置づければ、学力検査は「土台を測定することで間接的に大学における学業適性を測定するもの」と捉えることができる。一方、大学入学後の学習活動そのものを基準とした入学試験を考えることも可能である。学習プロセスのそれぞれに対応した設問によって学業適性を直接測定するのである。そうすれば、入学試験における遂行が入学後の学業生活に一般化しやすくなり、生態学的妥当性 (Bracht and Glass, 1968) が高まるほか、学業成績の予測もより正確になって予測的妥当性が高まることが期待できる。ただし大学生の学習プロセスをなぞるには、テーマを与えエッセイを書かせる形式の小論文試験では間に合わない。小論文試験は知識以外の、主として表現に関わる能力を測定することができるが、要点の把握や的確な理解、情報の評価や応用については測定力が限られるからである。

そこで本研究では入学後の学習活動を基準とした試験問題の試みとして、文系大学生の学習プロセスを意識した総合論述問題を作成した。そして作成された問題のモニター調査 (大学入試センター, 1999a, 1999b) の結果からその測定論的な特徴を明らかにし、このような総合論述

問題の可能性と課題について検討を行う。

2 総合論述問題

2.1 設問の概要

問題文と設問の全文を付録に掲げた。問題文の素材は新聞社の立場から少年犯罪と報道のありかたを述べた論説文である。分量は約 3,000 字、問題冊子にして 2.5 ページ分とかなり長い。標準的な回答時間は全体で約 30 分を想定した。

設問は 4 つある。本文中に記述されているルールを当てはめて判断させる設問 (設問 1, 回答 100 字程度)、報道された人物像について行間や文体まで含めた評価・鑑賞的な読解を求める設問 (設問 2, 回答 200 字程度)、新聞記事の見出しを作らせる設問 (設問 3, 10 字を 2 つから 4 つ)、総合的な立場から少年報道のありかたを論じる設問 (設問 4, 回答 400 字程度) である (付録参照)。これらの設問では、大量の長文から必要な情報や要点を即座にバランス良く把握すること、その意味を自分なりに統合して位置づけること、それをもとに自分の考えを構成し自分の言葉でまとめること、などが求められる。

2.2 データ

被験者は首都圏の 3 つの国立大学に在学する 1, 2 年生で、文系 191 名、理系 205 名の合わせて 396 名である。平成 10 年 3 月の第 1 回総合試験モニター調査「テスト A」(大学入試センター, 1999a, 1999b) の一部として実施された。

採点にあたり、はじめに問題作成者 1 名とモニター試験担当者 2 名との間で採点基準を作成した (表 1)。また、採点がどの程度文字のきれいさに影響されるのかを調べるため、「字のきれいさ」についても別に評定ポイントを設けた。採点作業には首都圏にある公立大学で心理学を専攻する大学院生 3 名 (男性 1 名女性 2 名) があたった。3 名には予め合同で採点基準の説明を行った上で、個別に採点作業を行って

表 1 採点基準

採点ポイント		内 容	配 点	満点
文 字 ^a		手書き文字のきれいさ	5段階	
設 問 1	q1_1	本文中のルール適用	0～1	5 点
	q1_2	本文中のルール適用	0～1	
	q1_3	本文中のルール適用	0～3	
設 問 2	q2_1	自分の言葉でまとめている	0～2	5 点
	q2_2	多面性に着目	0～1	
	q2_3	視点移動に着目	0～1	
	q2_4	記事の文体効果に着目	0～1	
	q2_5	個人的な判断が混入	-2～0 (減点)	
設 問 3	q3_1	必要要素にふれている	0～2	8 点
	q3_2	必要要素にふれている	0～2	
	q3_3	必要要素にふれている	0～2	
	q3_4	優れたセンスを感じさせる	0～2	
	q3_5	形式上の不備	-2～0 (減点)	
	q3_6	内容の冗長性	-2～0 (減点)	
	q3_7	内容の誤り	-2～0 (減点)	
	q3_8	強調点のずれ	-2～0 (減点)	
設 問 4	q4_1	自分の意見を述べている	0～3	10 点
	q4_2	多面的な視点から考察	0～3	
	q4_3	主題にのっとった考察	0～2	
	q4_4	優れた表現力	0～2	
	q4_5	首尾一貫性	-2～0 (減点)	
	q4_6	視野が狭い (独断的・感傷的)	-2～0 (減点)	
	q4_7	語の用いかたの誤り	-2～0 (減点)	
読みやすさ		文章の読みやすさ (含、句読点)	5段階	

a:内容的な評定への影響を調べるために設定

もらった。各採点者には答案以外、被験者に関する情報を一切与えなかった。以下の分析の対象となったのは、採点にミスが見つかった 9 名を除いた 387 名分のデータである。そのうち、文系は 190 名 (男性 104 名, 女性 86 名), 理系は 197 名 (男性 152 名, 女性 45 名) であった。

3 分析 1: 採点者間の評定の一致度

3.1 設問にみられる採点者間の一致度

採点者間の評定の一致度をみるため、2 者間の積率相関係数と、3 者の級内相関係数を求めた (表 2)。級内相関係数は採点者を変量効果とする 2 要因変量計画にもとづいたものとし、具体的には以下の式で定義されるものを用いた

(Shrout and Fleiss, 1979)。

$$ICC(2,1) = \frac{BMS - EMS}{BMS + (k-1)EMS + k(JMS - EMS) / n}$$

ただし BMS: 被験者間平均平方 (df=n-1)

JMS: 採点者間平均平方 (df=k-1)

EMS: 残差 (df=(n-1)(k-1))

n: 被験者数, k: 採点者数

採点者間の食い違いには、評価観点の違いなどの“意味のある不一致”と、評点の原点と単位の任意性という“意味のない不一致”の 2 通りがある。ICC (2,1)は、両方の要因の影響を受けて級内相関係数が低められる。一方採点者を固定効果として扱うと、採点の甘さ・辛さという得点の平行移動で説明される要因が採点者効果として除外され、その分高めの級内相関係数

表2 各採点における採点者間の一致度

	2 者間の積率相関係数			3 者間の 級内相関係数
	採点者 1－ 採点者 2	採点者 1－ 採点者 3	採点者 2－ 採点者 3	
字のきれいさ	0.692	0.577	0.435	0.517
設問 1	0.705	0.733	0.724	0.684
設問 2	0.439	0.468	0.430	0.424
設問 3	0.910	0.909	0.898	0.899
設問 4	0.490	0.563	0.516	0.481
読みやすさ	0.186	0.147	0.178	0.091
大問 (問 1 から 4 の合計点)	0.745	0.810	0.758	0.766

注) 欠測値を変数のペアワイズで削除したため、被験者数は組合せごとに異なる。

(325 ≤ N ≤ 387)

が得られる (Fagot, 1991, 1993)。ここでは採点そのものの一致度をみるのが目的なので、採点者を変量効果として扱うことにする。

表2によると、設問3において積率相関係数や級内相関係数の値が高く、採点者間で評価が一致していたことがうかがえる。この設問は解答に要する文字数が少なく、かつ採点基準が比較的客観的に定められているものである。逆に設問2 (200字) や設問4 (400字) のように解答文字数が多く、採点ポイントとしても「多面性への着目」や「自分の意見の有無」を評価するなど、主観的な判断にもとづいて採点する設問では採点者間の一致度が低かった。一致度の低い設問2や設問4では、積率相関係数が0.430から0.563と、どの採点者の組合せにおいてもあまり高くない。つまり、ある採点者だけが特異な採点を行ったわけではなく、三者が三様の評価観点にもとづいて採点を行ったと思われる。

表2では、「文章の読みやすさ」に関する評

定の一貫度がきわめて低いことが目を引く。

「文章の読みやすさ」は、細かな採点基準を設けずに採点者の印象のままに評定したものである。答案の読みやすさには、論理の一貫性や表現力、説得力など能力の本質的なものに由来する部分もあるが、句読点や段落分けなど二次的なものの影響も考えられる。実際、「読みやすさ」の評定と、個々の採点ポイントとの相関係数をみると、「q4_4 表現力」とは0.435、「q4_5 首尾一貫性」とは0.525であったが、一方で「字のきれいさ」という本質的でない要因とも0.381という、無視できない相関がみられた。

表3は、採点者ごとの平均値に違いがあるかどうかを、採点者を被験者内要因とした1要因の分散分析によって検討したものである。どの得点においても1%水準で採点者効果が有意であった。また表から、採点者の“甘い・辛い”のパターンが設問ごとに異なっていることがわかる。ひとりの採点者がある設問で甘く、別の

表3 各採点における評定の平均・標準偏差と一致度

	採点者ごとの平均 (標準偏差)			3 者間の 級内相関係数 (標準化後)
	採点者 1	採点者 2	採点者 3	
*字のきれいさ (N=386)	3.51(1.15)	3.18(0.84)	2.93(1.00)	0.569
*設問 1 (N=387)	2.50(1.61)	2.91(1.29)	2.33(1.62)	0.721
*設問 2 (N=368)	3.03(1.10)	2.61(1.12)	2.68(1.00)	0.443
*設問 3 (N=375)	2.19(1.91)	2.33(1.95)	2.53(1.84)	0.906
*設問 4 (N=334)	4.37(1.89)	4.68(2.30)	5.25(1.83)	0.517
*読みやすさ (N=334)	3.05(0.63)	3.10(0.62)	1.94(0.71)	0.171
*大問 (N=387)	11.30(4.07)	11.74(4.28)	11.93(4.10)	0.772

*1つの被験者内要因の分散分析において有意 (p<.001)

設問で辛く採点したということは、合計点の中で甘い設問の得点が占める割合が相対的に大きくなることを意味する。本論述問題では、採点者ごとの大問得点が結果的に同じであっても、それが表す能力の“内訳”が異なっている可能性があるわけである。

採点者ごとの評定の違いが同じ評価次元における平均とばらつきの任意性に由来するならば、採点者内で標準化することで調整が可能である。各設問得点や大問得点は近似的に間隔尺度とみなすことができるから、標準化は許容される変換である。そこで各採点者内で標準化したデータからあらためて級内相関係数を求めたが、標準化前より若干数値が高くなったものの、大きく上昇することはなかった(表3)。標準化後の ICC(2,1)は、“意味のない不一致”が標準化によって除外され、採点者の観点の違いという“意味のある不一致”だけが考慮されたものである。このときの級内相関係数は、各採点者が共通して測定している部分の大きさを反映していると考えられる。つまり設問2や設問4、読みやすさなど、主観的評定が含まれる得点では、3名の採点者が共通して評価した内容がかなり少なかったということになる。

3.2 分析1のまとめ

採点ポイントが比較的客観的に定められている設問では、採点者間の一致度が高い。逆に、「多面性に着目している」「自分の意見を述べている」など、主観的な判断がもとになるような採点箇所では、採点者間の一致度が低い。また、採点者によって評価の甘い・辛いのパターンが異なるだけでなく、採点にあたって異なる側面に着目していたことが明らかになった。このことは、評定者による大問得点が結果的に同じ値であっても、それが表す能力の“内訳”が異なることを示唆する。また「文章の読みやすさ」という全体的な印象の評定では、採点者間にほとんど一致がみられないだけでなく、「字のきれいさ」の影響も受けている可能性があることがわかった。

4 分析2：本論述問題で測定された能力

4.1 採点ポイントが測定する方向

採点者ごとに、採点ポイント間の単純な順位相関係数(スピアマン)と、そこから字のきれいさの影響を取り除いた偏順位相関係数の2通りを求め、計6つの相関行列を多次元尺度法の個人差モデル INDSCAL (Carroll and Chang, 1970, Arabie, Carroll, and DeSarbo, 1987) に投入した。計算には SAS の MDS プロシジャを用いた。順位相関係数を用いたのは、2段階で評価する採点ポイントがあることを考慮したからである。「答案の読みやすさ」の評定に関しては、それが答案の首尾一貫性や表現力とも関連しているため、統制する変数には含めなかった。多次元空間の次元数は因子分析などの予備分析を参考にして3次元と定めた。

図1(a)(b)は、INDSCAL モデルで得られた採点ポイントの布置である。解釈を容易にするため、いずれの軸も得られた解の符号を逆転させてある。それぞれ縦軸・横軸にそった解釈を試みると、次元1はプラス方向に表現・産出・主観性に関する採点ポイント、マイナス方向に理解・受動・客観性に関する採点ポイントが多く布置されており、「表現 vs. 理解」の次元と解釈される。

次元2は、プラス方向に全体を見渡した総合的な課題処理に関わる採点ポイント、マイナス方向に局所的・断片的な課題処理に関わる採点ポイントが多く布置されている。したがってこの次元2は、「統合 vs. 断片」の次元と解釈される。

次元3はあまり傾向が明らかではないが、プラス方向に情報を応用したり展開したりする思考が要求される採点ポイント、マイナス方向に全体をまとめるような思考が要求される採点ポイントが多く布置されており、「展開 vs. 要約」の次元と解釈される。

次元2

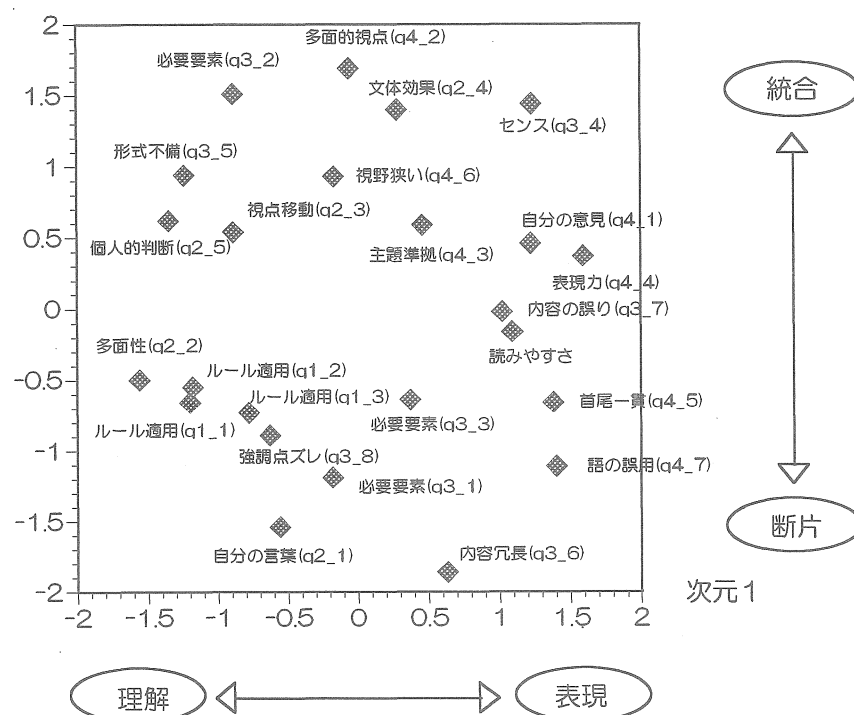


図 1(a) 各採点ポイントの布置 (次元1と次元2)

次元3

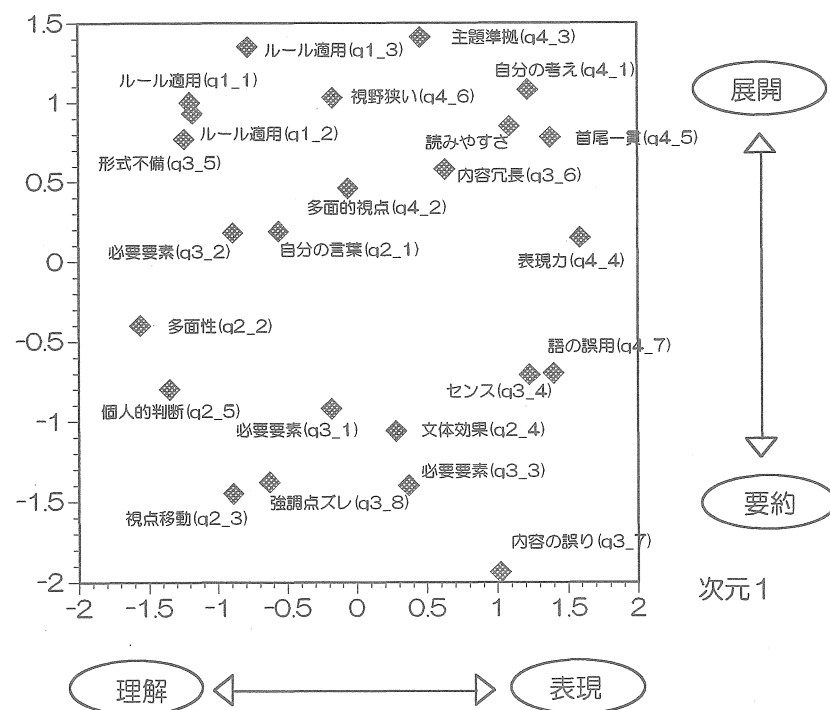


図 1(b) 各採点ポイントの布置 (次元1と次元3)

以上のことから本問題の採点ポイントは、表現力、理解力、視野の広さ、展開力、要約力などの方向性で説明できるといえる。これは当初想定した文系の研究活動に必要な能力にかなり対応した内容となっており、設定した採点ポイントが測定意図にそって機能していたことの1つの表れと考えられる。

4.2 字のきれいさの影響

図2はINDSCALモデルによって得られた各採点者の重みである。これによると、各採点者とも、採点そのままの場合と字のきれいさの影響を取り除いた場合とで重みが非常に近く、採点者3にいたっては両者が一致していることがわかる。このことから、各採点者とも字のきれいさによって採点の本質が影響を受けることは、ほとんどなかったと考えられる。これは、はじめに「字のきれいさ」について評定したのち内容面の採点に移ったため、内容の採点と字のきれいさとは別のものであるということを常に採点者に思い起こさせる結果となったためと考えられる。

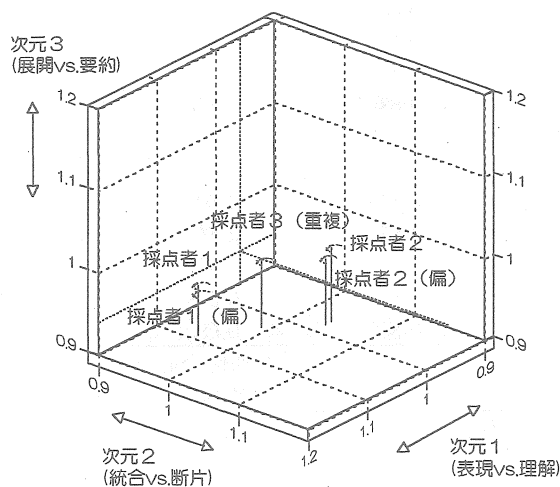


図2 部分点に表れた採点者の個人差

各採点者の個人差をみると、採点者1は他者と比較して相対的に次元1を重視していることがわかる。一方採点者2は相対的に次元2を重視しているようである。採点者3はどの次元に

も同じような比重を置いており、平均的な位置づけである。このことから採点者1は、被験者の理解の程度や文章表現力に対して得点を相対的にばらつかせる傾向があったといえ、一方で採点者2は、被験者の理解が断片的で視野が限られているか、全体を見渡しているかに対して得点をより変化させる傾向があったと推測できる。三者のなかで相対的にこのバランスがとれていたのが採点者3ということになる。

分析1において、一部の設問得点で三者が三様の採点を行っていた可能性が高いことが明らかになったが、以上のような評定上の比重の違いによってそれを説明することができよう。

4.3 分析2のまとめ

多次元尺度法のINDSCALモデルによる分析の結果、本論述問題が測定した能力は表現力、理解力、視野の広さ、展開力、要約力などの方向にまとめることができ、採点ポイントの設定意図はほぼ達成されたといえる。また「字のきれいさ」は全体として本質的な内容の採点にはほとんど影響していなかった。採点者にかかる重みから、採点者1は被験者の理解の程度や文章表現力に対して得点を相対的にばらつかせる傾向があり、採点者2は、被験者の理解が断片的で視野が限られているか、全体を見渡しているかに対して得点をより変化させる傾向があることがわかった。このような個人差が分析1における級内相関係数の低さにつながっていたと考えられる。

5 分析3：高校における履修科目の影響

5.1 斜交プロクラステス回転

図1(a)(b)における各採点ポイントの座標をターゲットとして、因子数3の斜交プロクラステス回転(芝, 1979, 柳井・繁樹・前川・市川, 1990, 前川, 1997)を行った。因子分析を行ったのは、各被験者の因子得点を求め属性別の分析を行うためである。相関行列は採点者3名の平均値による相関行列を用い、主成分を初期解と

した。このとき第3因子（主成分）までによって説明される分散は7.03となり、全分散に占める割合は $7.03/24 = 0.293$ であった。プロクラステス回転にさいしては、ターゲットの座標を強調するため、その絶対値を2乗した。これらの分析にはSASを用いた。

表4(a)は回転の結果得られた因子パターン、表4(b)は因子間相関である。もともと直交する軸をターゲットにしたためか、因子間相関はあまり高くならなかった。因子負荷の大きいもの、

小さいものに注目すると、第1因子は「表現・主観 vs. 理解・客観」、第2因子は「全体・包括 vs. 局所・断片」と解釈され、ほぼ多次元尺度法による次元に対応する因子が得られた。第3因子は「展開 vs. 基礎的理解」と解釈され、多次元尺度法の第3次元とやや方向の異なる因子が得られた。しかし解釈的には似た内容の因子と考えられるため、以下の分析ではプロクラステス回転によって得られた3因子を用いることにする。

表4(a) 斜交プロクラステス回転による因子パターン

採点ポイント	因子1	因子2	因子3	共通性
Q1_1	-0.587	-0.376	0.352	0.728
Q1_2	-0.562	-0.319	0.379	0.679
Q1_3	-0.303	-0.247	0.355	0.331
Q2_1	-0.172	-0.075	0.296	0.148
Q2_2	-0.141	-0.094	0.268	0.118
Q2_3	-0.265	0.180	0.157	0.149
Q2_4	-0.048	0.413	0.051	0.179
Q2_5	-0.417	0.306	0.106	0.293
Q3_1	-0.047	-0.068	0.238	0.067
Q3_2	-0.148	0.440	0.096	0.232
Q3_3	0.018	-0.165	0.221	0.067
Q3_4	0.053	0.401	-0.031	0.166
Q3_5	-0.214	0.248	0.135	0.140
Q3_6	0.053	-0.272	0.064	0.074
Q3_7	-0.014	-0.169	-0.124	0.047
Q3_8	-0.141	-0.141	0.257	0.122
Q4_1	0.545	0.039	0.689	0.574
Q4_2	-0.139	0.623	0.222	0.489
Q4_3	0.115	0.169	0.529	0.308
Q4_4	0.537	-0.055	0.532	0.407
Q4_5	0.604	-0.225	0.640	0.565
Q4_6	-0.233	0.378	0.403	0.430
Q4_7	0.159	-0.119	0.131	0.039
読みやすさ	0.527	-0.079	0.805	0.680
因子寄与 ^a	2.364	1.817	2.961	共通性合計 7.031

因子負荷の大きいものから4つ小さいものから4つのものにアミをかけて示した

a 他の因子の効果を無視したときに説明される分散

表4(b) 斜交プロクラステス回転による因子間相関

	因子1	因子2	因子3
因子1	1.000		
因子2	0.069	1.000	
因子3	-0.276	0.099	1.000

5.2 文系・理系と性別

各因子得点について、文理別と性別を入れた 2 要因の分散分析を行った。表 5 は、性別×文理別の組み合わせによる因子得点の平均と標準偏差である。各セルの度数が偏り、disproportionality が疑われるため、タイプⅢの平方和を用いた分析を行った。タイプⅢの平方和とは、当該効果以外の効果すべてについて調整された平方和で、合計が総平方和に一致しないが、要因の効果が非直交の場合にも不偏推定値が得られる点で好ましいとされている (Lomax, 1992, Overall, Lee, and Hornick, 1981)。

表 5 性別×文理の組み合わせによる因子得点の平均と標準偏差

		文系	理系
男性	人数	N=87	N=118
	因子 1	-0.06 (1.06)	0.01 (1.03)
	2	-0.08 (1.02)	0.26 (0.96)
	3	0.02 (1.05)	0.46 (0.90)
女性	人数	N=76	N=38
	因子 1	-0.00 (1.00)	0.11 (0.73)
	2	-0.36 (0.97)	0.11 (0.91)
	3	-0.46 (0.90)	-0.52 (0.67)

表中の数字は平均 (標準偏差)

分散分析の結果、第 1 因子「表現・主観 vs. 理解・客観」はいずれの効果も有意にならなかった (図 3(a))。第 2 因子「全体・包括 vs. 局所・断片」は文理別の主効果のみが有意となった ($F_{1,315}=11.573$, $p<0.01$) (図 3(b))。第 3 因子「展開 vs. 基礎的理解」は、性別の主効果 ($F_{1,315}=42.390$, $p<0.01$) と文理×性の交互作用 ($F_{1,315}=4.971$, $p<0.05$) とが有意になった (図 3(c))。そこで下位検定として男女別に文理の単純主効果、文理別に性別の単純主効果を検定した。その結果、男性では文理差が有意であった ($p<0.01$) が、女性では有意でなかった。また文系でも理系でも、性差が有意となった (いずれも $p<0.01$)。

これらの結果から一般的傾向を述べると、男女ともに文系のほうが全体・包括的な課題処理

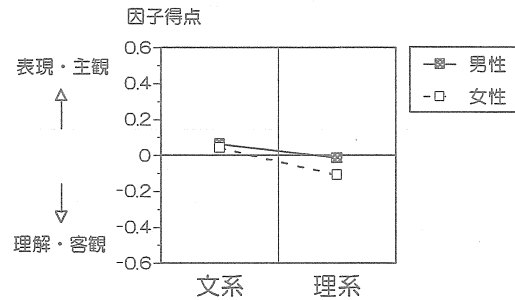


図 3(a) 第 1 因子の平均値プロット

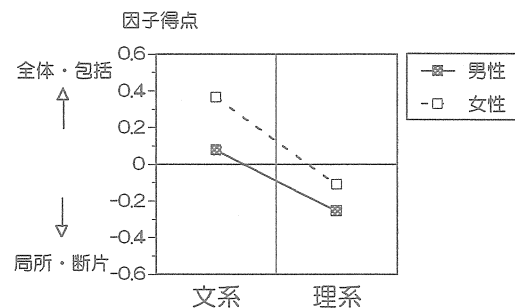


図 3(b) 第 2 因子の平均値プロット

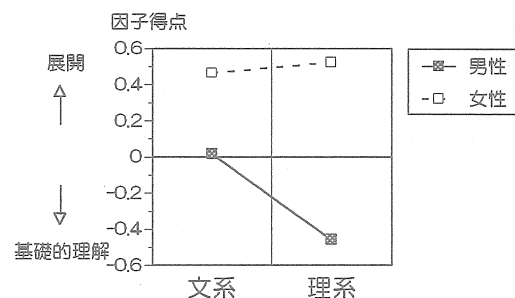


図 3(c) 第 3 因子の平均値プロット

を行い、理系のほうは局所・断片的な課題処理を行う傾向があるといえる (第 2 因子)。また男性では、文系のほうが自分の意見を展開することが得意で、理系は基礎的理解が得意なようである (第 3 因子)。ただし女性ではこのような文理差はみられず、文理ともに男性より女性のほうが意見の展開が得意 (第 3 因子) という傾向がみられた。

5.3 高校での履修科目の影響

一部の因子で性別×文理別の交互作用がみられたので、以下、男女別に履修経験と因子得点との関連をみることにする。表 6(a)(b)は、男性について、高校での履修経験によるグループ別因子得点と、その差を 1 要因の分散分析 (5%水準) によって検討したものである。また表 7(a)(b)は同様のことを女性について示したものである。履修経験によるグループ分けは、科目により履修のしかたが異なったため一律の分類は行えなかった。

男性の場合 (表 6(b))、第 1 因子は理科の受験科目数で有意になり、下位検定の結果、理科を 1 科目以上受験した群が受験しなかった群にくらべて「表現・主観」の傾向が強いことがわかった。第 2 因子は、地歴と物理で有意になった。地歴では、地歴を 2 科目受験した群が受験しなかった群にくらべて「全体・包括」的に課題処理を行う傾向が強かった。物理は下位検定で有意な集団差が得られなかった。第 3 因子は、国語、数学、地歴、物理で有意になった。国語では 2 次試験で受験した群が、数学では 2 次試験で数学Ⅱまで受験した群が、地歴では 2 科目以上受験した群が、それぞれ国語を 2 次試験で受験せず、数学は 2 次試験で微積まで受験し、地歴は 1 科目以下しか受験しなかった群と比較して「展開」が得意な傾向がみられた。物理では下位検定で有意な集団差が得られなかった。また、本論述問題と内容的な関連があると思われた公民の科目は、いずれの因子においても履修状況による違いがみられなかった。

以上のことから男性では、第 1 因子を除いて基本的に文系と理系の履修科目や受験科目で因子得点の違いが説明できるといえる。このことは、第 2、第 3 因子の因子得点に反映される能力や思考様式の違いが、高校において教科科目の得意・不得意をもたらした、学力の本質的な部分と関連があることを示唆している。

一方女性の場合 (表 7(b))、有意になったのは第 2 因子における理科の受験科目数だけであった。しかし下位検定の結果、有意な集団差は得られなかった。

このように女性の分析結果には、男性でみられた文系・理系の履修・受験科目による差がみられない。もし男性の因子得点の差が文系と理系に適した能力や思考様式の違いを反映しているのなら、女性においても文理差が見られるはずである。男性と女性とで能力のありかたが異なる可能性も考えられるが、この点に関しては今後の検討課題としたい。

5.4 分析 3 のまとめ

男性を文系・理系で分けた場合、文系のほうが全体・包括的な課題処理を行い、理系のほうが局所・断片的な課題処理を行う傾向があった。また文系のほうが自分の意見を展開することが得意で、理系のほうは基礎的な理解が得意であった。この傾向は高校における履修状況や受験科目別にみても得られたことから、第 2 因子と第 3 因子が文系と理系に適した能力や思考様式の違いを反映している可能性が考えられる。

一方女性では、文系のほうが全体・包括的な課題処理を行う傾向がみられたが、意見の展開に関しては文系も理系も一貫して得意という傾向がみられ、高校での履修状況の影響もほとんど見られないなど、男性と女性とで異なる結果となった。

表 6(a) 履修経験グループごとの因子得点の平均と標準偏差 (男性)

グループ 番号		1	2	3	4	5
国語	履修歴 人数	履修のみ N=33	センター試験 N=71	2 次試験 N=100		
	因子 1	-0.34(1.12)	-0.08(1.21)	0.12(0.86)		
	2	0.29(0.79)	0.20(1.07)	0.00(1.01)		
	3	0.64(1.05)	0.39(1.00)	0.07(0.92)		
数学	履修歴 人数	数学 I まで N=32	数学 II 2 次 N=57	数 II & 微積 N=115		
	因子 1	-0.16(1.17)	0.11(0.93)	-0.05(1.06)		
	2	0.09(1.24)	-0.10(0.87)	0.23(0.98)		
	3	0.30(1.28)	-0.08(0.98)	0.45(0.85)		
現代社会	履修歴 人数	履修せず N=97	履修のみ N=95	入試 N=12		
	因子 1	0.08(0.88)	-0.13(1.21)	-0.04(0.78)		
	2	0.20(0.96)	0.08(1.04)	-0.26(1.01)		
	3	0.20(0.99)	0.36(0.94)	0.19(1.32)		
倫理政経	履修歴 人数	両方履修せず N=62	残り N=99	どちらか入試 N=43		
	因子 1	-0.21(1.16)	0.05(0.96)	0.09(1.03)		
	2	-0.03(1.09)	0.13(0.89)	0.31(1.10)		
	3	0.34(1.08)	0.28(0.94)	0.17(0.98)		
地歴受験	履修歴 人数	なし N=39	世界史 N=29	日本史 N=39	地理 N=54	2 科目 N=43
	因子 1	0.00(1.17)	-0.32(1.28)	0.04(0.99)	-0.01(0.95)	0.09(0.93)
	2	0.45(0.97)	0.08(0.92)	-0.02(1.05)	0.29(1.02)	-0.25(0.92)
	3	0.40(1.02)	0.46(1.04)	0.37(0.76)	0.50(1.01)	-0.34(0.86)
物理	履修歴 人数	履修せず N=54	履修のみ N=35	センター試験 N=11	2 次試験 N=104	
	因子 1	0.16(1.03)	-0.22(1.15)	0.06(0.70)	-0.06(1.04)	
	2	-0.03(0.94)	-0.13(1.20)	-0.20(0.95)	0.31(0.94)	
	3	0.09(0.97)	0.11(1.15)	-0.26(1.08)	0.48(0.89)	
化学	履修歴 人数	履修せず N=26	履修のみ N=49	センター試験 N=30	2 次試験 N=99	
	因子 1	0.31(1.14)	-0.06(0.93)	-0.22(1.08)	-0.03(1.06)	
	2	-0.10(1.04)	0.00(0.96)	0.11(1.20)	0.23(0.95)	
	3	0.28(0.99)	0.03(1.03)	0.27(1.18)	0.40(0.89)	
生物	履修歴 人数	履修せず N=100	履修のみ N=74	センター試験 N=23	2 次試験 N=7	
	因子 1	-0.08(1.11)	0.14(0.95)	-0.36(1.12)	0.25(0.46)	
	2	0.15(1.03)	0.20(0.89)	-0.14(1.22)	-0.47(0.72)	
	3	0.37(1.02)	0.22(0.93)	0.03(1.09)	0.34(0.54)	
地学	履修歴 人数	履修せず N=139	履修のみ N=42	受験あり N=23		
	因子 1	-0.04(1.06)	-0.10(1.08)	0.23(0.87)		
	2	0.15(1.10)	0.22(0.77)	-0.26(0.68)		
	3	0.35(1.00)	0.24(0.96)	-0.12(0.88)		
理科受験	履修歴 人数	なし N=16	1 科目 N=83	2 科目以上 N=105		
	因子 1	0.59(0.85)	-0.13(1.09)	-0.03(1.01)		
	2	0.09(1.10)	-0.03(1.02)	0.24(0.97)		
	3	0.21(0.80)	0.15(1.14)	0.38(0.87)		

表中の数字は平均 (標準偏差)

表 6(b) 履修経験による因子得点の分散分析結果 (男性)

要因	因子 1 表現主観 vs. 理解客観	因子 2 全体包括 vs. 局所断片	因子 3 展開 vs. 基礎的理解
国語	$F_{2,201}=2.618$ ($p=.075$)	$F_{2,201}=1.364$ ($p=.258$)	$F_{2,201}=5.079$ ($p=.007$) (1-3)
数学	$F_{2,201}=0.730$ ($p=.483$)	$F_{2,201}=2.184$ ($p=.115$)	$F_{2,201}=5.770$ ($p=.004$) (2-3)
現代社会	$F_{2,201}=1.002$ ($p=.369$)	$F_{2,201}=1.280$ ($p=.280$)	$F_{2,201}=0.637$ ($p=.530$)
倫理政経	$F_{2,201}=1.473$ ($p=.232$)	$F_{2,201}=1.520$ ($p=.221$)	$F_{2,201}=0.349$ ($p=.706$)
地歴受験	$F_{4,199}=0.746$ ($p=.562$)	$F_{4,199}=3.292$ ($p=.012$) (1-5)	$F_{4,199}=6.003$ ($p=.000$) (1234-5)
物理	$F_{3,200}=1.056$ ($p=.369$)	$F_{3,200}=2.807$ ($p=.041$) ^a	$F_{3,200}=3.617$ ($p=.014$) ^a
化学	$F_{3,200}=1.257$ ($p=.290$)	$F_{3,200}=1.039$ ($p=.376$)	$F_{3,200}=1.579$ ($p=.196$)
生物	$F_{3,200}=1.660$ ($p=.177$)	$F_{3,200}=1.555$ ($p=.202$)	$F_{3,200}=0.849$ ($p=.468$)
地学	$F_{2,201}=0.833$ ($p=.436$)	$F_{2,201}=1.910$ ($p=.151$)	$F_{2,201}=2.319$ ($p=.101$)
理科受験	$F_{2,201}=3.257$ ($p=.041$) (1-2)	$F_{2,201}=1.639$ ($p=.197$)	$F_{2,201}=1.262$ ($p=.285$)

N=205

アミかけ部分は、5%水準で有意になったもの

カッコ内は下位検定 (Scheffé) で有意になった集団。太字は「表現主観」「全体包括」「展開」方向に平均値の高かった集団。

a:分散分析では有意になったが、下位検定では有意な集団差が見いだせなかった

表 7(a) 履修経験グループごとの因子得点の平均と標準偏差 (女性)

グループ 番号		1	2	3	4	5
国語	履修歴 人数	履修のみ N=18	センター試験 N=38	2 次試験 N=58		
	因子 1	-0.01(0.81)	0.01(1.00)	0.06(0.92)		
	2	-0.39(1.10)	0.01(0.79)	-0.29(1.02)		
	3	-0.44(0.88)	-0.41(0.71)	-0.55(0.89)		
数学	履修歴 人数	数学 I まで N=31	数学 II 2 次 N=50	数 II & 微積 N=33		
	因子 1	0.03(0.84)	0.02(1.09)	0.06(0.72)		
	2	-0.27(1.16)	-0.25(0.95)	-0.09(0.81)		
	3	-0.57(1.02)	-0.42(0.78)	-0.51(0.71)		
現代社会	履修歴 人数	履修せず N=58	履修のみ N=50	入試 N=6		
	因子 1	0.01(0.84)	0.06(1.04)	0.02(0.70)		
	2	-0.19(1.08)	-0.19(0.83)	-0.53(0.97)		
	3	-0.59(0.79)	-0.39(0.81)	-0.20(1.25)		
倫理政経	履修歴 人数	両方履修せず N=29	残り N=49	どちらか入試 N=36		
	因子 1	0.04(0.75)	0.08(0.94)	-0.04(1.04)		
	2	-0.12(0.76)	-0.31(1.15)	-0.13(0.86)		
	3	-0.31(0.87)	-0.57(0.83)	-0.51(0.79)		
地歴受験	履修歴 人数	なし N=28	世界史 N=22	日本史 N=21	地理 N=18	2 科目 N=25
	因子 1	-0.11(1.08)	0.09(0.96)	0.00(0.82)	0.15(0.82)	0.08(0.90)
	2	-0.11(0.98)	-0.00(0.99)	-0.41(1.02)	-0.26(0.81)	-0.28(1.03)
	3	-0.50(0.76)	-0.76(0.82)	-0.35(0.81)	-0.32(0.87)	-0.45(0.89)
物理	履修歴 人数	履修せず N=51	履修のみ N=38	センター試験 N=3	2 次試験 N=22	
	因子 1	0.17(0.97)	-0.13(0.96)	-0.38(1.16)	0.05(0.68)	
	2	-0.20(1.09)	-0.25(0.95)	-0.67(0.35)	-0.08(0.76)	
	3	-0.37(0.84)	-0.60(0.90)	-0.39(0.51)	-0.56(0.72)	
化学	履修歴 人数	履修せず N=12	履修のみ N=54	センター試験 N=11	2 次試験 N=37	
	因子 1	0.24(0.83)	0.06(0.92)	-0.18(1.43)	-0.02(0.78)	
	2	-0.41(0.84)	-0.40(1.05)	0.13(0.72)	0.05(0.89)	
	3	-0.39(0.99)	-0.54(0.87)	-0.49(0.91)	-0.43(0.71)	
生物	履修歴 人数	履修せず N=17	履修のみ N=40	センター試験 N=40	2 次試験 N=17	
	因子 1	0.13(0.88)	-0.08(1.04)	0.14(0.88)	-0.06(0.79)	
	2	-0.06(0.78)	-0.34(0.77)	-0.27(1.11)	0.12(1.18)	
	3	-0.69(0.66)	-0.58(0.90)	-0.38(0.77)	-0.28(0.92)	
地学	履修歴 人数	履修せず N=76	履修のみ N=24	受験あり N=14		
	因子 1	0.09(0.93)	-0.18(0.87)	0.08(0.97)		
	2	-0.18(1.00)	-0.28(0.96)	-0.21(0.84)		
	3	-0.38(0.78)	-0.60(0.86)	-0.84(0.96)		
理科受験	履修歴 人数	なし N=8	1 科目 N=69	2 科目以上 N=37		
	因子 1	0.20(0.88)	0.02(1.03)	0.02(0.71)		
	2	-0.75(0.77)	-0.32(0.99)	0.13(0.89)		
	3	-0.72(0.96)	-0.47(0.87)	-0.46(0.73)		

表中の数字は平均 (標準偏差)

表 7(b) 履修経験による因子得点の分散分析結果 (女性)

要因	因子 1 表現主観 vs. 理解客観	因子 2 全体包括 vs. 局所断片	因子 3 展開 vs. 基礎的理解
国語	$F_{2,111}=0.050$ ($p=.951$)	$F_{2,111}=1.457$ ($p=.237$)	$F_{2,111}=0.356$ ($p=.701$)
数学	$F_{2,111}=0.020$ ($p=.980$)	$F_{2,111}=0.352$ ($p=.704$)	$F_{2,111}=0.317$ ($p=.729$)
現代社会	$F_{2,111}=0.048$ ($p=.953$)	$F_{2,111}=0.359$ ($p=.699$)	$F_{2,111}=1.132$ ($p=.326$)
倫理政経	$F_{2,111}=0.172$ ($p=.842$)	$F_{2,111}=0.502$ ($p=.607$)	$F_{2,111}=0.889$ ($p=.414$)
地歴受験	$F_{4,109}=0.287$ ($p=.886$)	$F_{4,109}=0.568$ ($p=.686$)	$F_{4,109}=0.960$ ($p=.433$)
物理	$F_{3,110}=0.980$ ($p=.405$)	$F_{3,110}=0.367$ ($p=.777$)	$F_{3,110}=0.641$ ($p=.590$)
化学	$F_{3,110}=0.454$ ($p=.715$)	$F_{3,110}=2.269$ ($p=.084$)	$F_{3,110}=0.179$ ($p=.910$)
生物	$F_{3,110}=0.494$ ($p=.687$)	$F_{3,110}=1.082$ ($p=.360$)	$F_{3,110}=1.096$ ($p=.354$)
地学	$F_{2,111}=0.805$ ($p=.450$)	$F_{2,111}=0.093$ ($p=.912$)	$F_{2,111}=2.164$ ($p=.120$)
理科受験	$F_{2,111}=0.142$ ($p=.867$)	$F_{2,111}=4.279$ ($p=.016$)	$F_{2,111}=0.350$ ($p=.706$)

N=114 アミかけ部分は、5%水準で有意になったもの

a:分散分析では有意になったが、下位検定 (Scheffé) では有意な集団差が見いだせなかった

6 全体的考察：可能性と問題点

論述式問題は、測定対象となる能力・資質の豊かさという点で非常に可能性に富んでいる。設問形式にあまり制約がないため、さまざまな問いかけが可能となり、被験者の思考を多面的にみることができる。大学入学後の学習プロセスをなぞった総合試験を考えると、この論述式問題の柔軟性と可能性は大いに魅力である。本研究においても、文系大学生の学習プロセスをほぼカバーする、表現力、理解力、視野の広さ、展開力、要約力などの方向が測定可能であることが示された。教科学力試験や小論文試験でもこれらの能力を測定できないわけではないが、本研究のような総合論述問題では、それらをより重点的、直接的に測定できるといえる。

一方小論文試験に関して繰り返し指摘されていることであるが、論述式問題は採点の客観性に問題を抱えている。本研究においても、主観的な判断を要する採点箇所、先行研究と同様に採点者間の一致度がかなり低くなった。相関係数や INDSAL モデルによる分析の結果、採点者間の不一致は、主として評価の観点や重視する側面が採点者ごとに異なることに由来していると判断された。

採点者間の一致度を高め、採点の実質的な内容を同一に保つためには、採点ポイントを具体

的かつ細分化して定めることが有効である。しかし採点ポイントを多くすれば、それだけ何度も答案に目を通すことになり、採点者の時間的、精神的負担が大きくなる。論述式問題の良さをいかしながら、一方で採点の信頼性を高め、採点者バイアスを低減するためには、ひとつの長い論述答案をいくつもの角度から評定することを止め、設問を分割設定して、測定したい能力を絞った問いかけと採点を行うことが有効ではないだろうか。5 段階程度の評定でも 100 点満点の評定と同程度の信頼性を得ることが可能である (平井・渡部, 1994)。各設問や採点ポイントにおける得点段階はあまり細かくせず、しかも各得点段階の基準となる回答側を示すなどして採点作業をできるだけ具体化・簡素化していけば、論述問題の測定論的な欠点はある程度克服できるものと考えられる。

本研究で用いた総合論述問題では、「全体・包括 vs. 局所・断片」と「展開 vs. 基礎的理解」の 2 つの因子得点に文理差がみられた。これらの因子が文系の学生に適した能力や思考様式を反映しているのなら、文理差の存在は文系学部への学業適性を測定していることの表れと考えられ、好ましいものといえる。しかし「展開 vs. 基礎的理解」の因子得点には性差も存在する。もしこの因子において本質的な性差があるとすれば、このような論述問題は女性に有利

なバイアスを持つことになり、公平さの点で問題が生じる。ただし現段階では、データにみられた性差が本質的に性別に由来するのか、被験者の他の要因に由来するのかは判断できない。今後の研究課題としたい。

本研究では、題材によるバイアスは検討されなかった。どのような題材にしる、「好き」と感じる被験者と「嫌い」と感じる被験者が生じることは避けられないが、題材によって一部の被験者が不利になってはならない。題材の持つバイアスとそれへの対処法についても今後の検討課題としたい。

謝 辞

東北大学アドミッション・センターの倉元直樹助教授には、採点ポイントの検討をはじめ、本研究の遂行にあたり多大なご協力と助言をいただきました。ここに感謝の意を表します。また査読者からは、数々の有用で教育的なご指摘をいただきました。合わせて感謝いたします。

参考文献

- Arabie, P., Carroll, J. D., and DeSarbo, W. S. (1987) *Three-way scaling and clustering*. Sage.
- Benton, S. L., Kraft, R. G., Glover, J.A., and Plake, B. S. (1984) Cognitive capacity differences among writers. *Journal of Educational Psychology*, 76, 820-834.
- Benton, S. L. and Kiewra, K. A. (1986) Measuring the organizational aspects of writing ability. *Journal of Educational Measurement*, 23, 377-386.
- Blok, H. (1985) Estimating the reliability, validity, and invalidity of essay ratings. *Journal of Educational Measurement*, 22, 41-52.
- Bracht, G. H. and Glass, G. V. (1968) The external validity of experiments. *American Educational Research Journal*, 5, 437-474.
- Carroll, J. D. and Chang, J. J. (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.
- 大学入試センター研究開発部 (1999a) 「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) 「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
- Fagot, R. F. (1991) Reliability of ratings for multiple judges: Intraclass correlation and metric scales. *Applied Psychological Measurement*, 15, 1-11.
- Fagot, R. F. (1993) A generalized family of coefficients of relational agreement for numerical scales. *Psychometrika*, 58, 357-370.
- Huot, B. (1990) The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60, 237-263.
- 平井洋子・渡部洋 (1994) 小論文評点のカテゴリ化に関する測定論的考察 行動計量学 21 巻 21-31.
- van der Kamp, L. J. T. and Mellenbergh, G. J. (1976) Agreement between raters. *Educational and Psychological Measurement*, 36, 311-317.
- Lomax, R. G. (1992) *Statistical concepts: A second course for education and the behavioral sciences*. Lawrence Erlbaum Associates.
- 前川眞一 (1997) SAS による多変量データの解析 東京大学出版会
- Overall, J. E., Lee, D. M., and Hornick, C. W. (1981) Comparison of two strategies for analysis of variance in nonorthogonal designs. *Psychological Bulletin*, 90, 367-375.
- 芝祐順 (1979) 因子分析法 (第2版) 東京大学出版会
- Shrout, P. E. and Fleiss, J. L. (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- 渡部洋・平由実子・井上俊哉 (1988) 小論文評価データの解析 東京大学教育学部紀要

28 巻 143-164.

柳井晴夫・繁樹算男・前川眞一・市川雅教
(1990) 因子分析－その理論と方法－ 朝倉
書店

付録 本研究で用いた総合論述問題

以下の文章は、ある新聞社の少年犯罪の報道のしかたを述べたものである。この文章を読んで、以下の三問に答えなさい。なお、下書き用の原稿用紙が十一ページ（十二ページにあるので、適宜利用して下さい）。

少年（20歳未満）は、法律が匿名にすることを報道機関に義務付けている唯一のもので、その趣旨は「未成年少年を保護し、その将来の更生を可能にする」ことにあるからです。

問題は、次々と人を殺したような場合は例外を認めるべきではないのか、被害者が少年であった場合も同じなのか、家族の氏名、学校名、住所などについてどの程度報道を求ますべきか、などにあります。

「本紙の少年報道の記述原則」

① 未成年者による犯罪は、容疑者はもちろん、保護者の氏名も伏せる。住所、学校名、職場名は、本人が特定できない程度にとどめる。容疑者が未成年者一人の場合はアルファベットを使わず「少年（16）」のように書く。

② 知人家庭内の未成年者による重大犯罪で、社会的衝撃の度合いが強い事件は、未成年者は匿名だが、保護者は実名で書くことができる。未成年者の学校、職場名は伏せる。

③ テロなど社会的衝撃が大きく、かつ歴史的意味のある事件を引き起こした未成年者は、実名で書くことができる。保護者も必要がある場合は同じ。未成年者の学校、職場名も同じ。

④ 凶悪犯罪を引き起こした未成年者が、逃走中に凶悪犯罪を重ねることが予想され、他人を害する危険性が高いときなど、事件が社会に大きな衝撃を与え、広く恐怖心と呼び起こした場合は、実名で書くことができる。

⑤ 未成年者が否かは犯行時を基準とする。

⑥ 交通事故を起こした未成年者は、生存の場合は匿名で書く。死亡の場合は実名で書く。交通事故の被害者は原則として実名で書く。

事例A（恐喝をそそのかす母）

奈良県警は六日まで、下級生から貯金など約百五十万円を脅し取っていた奈良市内の中学三年の女子生徒（14）とその母親（31）を恐喝の疑いで逮捕、被害者の友人で中学一年の女子生徒（12）を同容疑で逮捕した。

調べでは、二人の女子生徒は今年三月二十日から四月十六日までの間、同じ中学校一年の女子生徒（12）から五回にわたり約八十二万円を脅し取った。恐喝を知った母親が「お金いるねん。四十万円準備して」と中学三年の長女に催促。長女が被害者の女生徒に電話で脅し、母親の運転する車で市内の郵便局まで連れて行き、被害者名義の貯金五十八万八千円を引き出させ、（94年6月7日付朝刊）

この母親は容疑者の少女の共犯ですが、少女の身元を推測させないために、とんでもない母親も匿名にし、住所もボカさざるを得ません。

事例B（有名な棋士宅の悲劇）

二十三日午前八時五十五分ごろ、兵庫県××市××の将棋棋士△△九段（実名（44）が、自宅二階の書斎兼寝室で、左胸を刺されて死んでいるのを、妻の〇〇さん（実名（40）が見つけた。一一〇番しようとしたところ、ついできた中学一年の長男（12）がいきなり包丁で〇〇さんに切りつけて逃げた。〇〇さんは右首などに二週間のけが。西宮署は、長男が犯行をはめかすようなことを〇〇さんに言い残していることなどから、重要参考人として行方を追っている。長男は学校を休みがちで、同署ではこのことで△△九段が、長男と口論になり殺されたのではないかとみている。（93年11月24日付朝刊）

見出しは「△△九段刺され死亡 ××市の自宅 中一の長男、娯楽す 母親にも切りつけ休みがちかられて?」。△△九段の顔写真と自宅の光景を撮った写真が大きく扱われている。他紙もほぼ同じ表記方法と扱いでした。少年本人の特定は容易です。

この報道をどのように考えるべきでしょうか。結論から言うと、やむをえない例外的なケースとして両親の実名表記が許されると考えます。

もし少年法の規定の趣旨を厳格に考えるとすれば、「将棋棋士△△九段（44）が兵庫県下の自宅で胸を刺され、妻（40）も包丁で切りつけられ……」といった書き方にならざるをえません。しかし、これでは何が起きたのか、あまりに漠然として、もはや事実の報道と言えないものになります。

しかし、例外的範囲は厳格に考えるべきです。具体的には、犯罪の内容が殺人など、とくに重い場合に限って例外とすべきです。「殺人など」としたのは、殺人事件以外にも社会的な反響という面から「例外」とすることが認められてよい犯罪があるからです。大きな被害を出した放火、身代金誘拐など何の罪もない人を入質にとつて死傷させたり、ハイジャックのように多数の人を死と隣り合わせの目にあわせた場合も含まれるでしょう。ただし、これらは罪名だけから区別するのではなく、犯罪内容と引き起こした社会的衝撃から実質的に判断する必要があります。

ところで、△△九段の事件の真相は明らかになりました。長男がやったことを裏付けるものは皆無だったことから、93年暮れ、聴取は打ち切れ、長男は児童相談所の一時保護に引き継がれました。

そうすると、匿名といっても本人を推測させる書き方であったことに加え、警察の見方としてであっても本人が父親を刺殺した疑いが強いとする紙面づくりをした点も問われかねません。長男を父殺しの犯人として強く印象付ける紙面をつくりながら、そうであることが証明できなかったことになりす。

それでは、各紙はこぞって長男に対して名刺裏拍を犯したことになるのでしょうか。現場の状況、長男の言動、事件の背景、捜査当局の見方などを総合すると、長男がやったと

(下書用原稿用紙) (答は解答用紙の所定の位置に記入)

A full page of blank graph paper with a uniform grid of small squares. The grid consists of 20 columns and 20 rows, creating a total of 400 small square units. The lines are thin and black, set against a white background. There are no margins or additional markings on the page.

(下書用原稿用紙)(答は解答用紙の所定の位置に記入)

A full-page sheet of white graph paper with a light gray grid. The grid consists of small squares, approximately 1 cm by 1 cm each. There are 20 columns and 20 rows of squares, creating a total of 400 small square units. The grid lines are thin and evenly spaced.

中見出し

雪 頁

要約記事

小見出し

大見出し

本文

(次頁に問題は続きます)

Developing a general type of extensive open-ended questions to assess the abilities of college students majoring in the humanities or social sciences

Yoko HIRAI*
Kumiko SHIINA**
Haruo YANAI***

Abstract

Here we present an attempt to develop a new type of college entrance test based on the learning processes of students majoring in the humanities or social sciences. Four open-ended questions were devised around a passage about a boy's crime to assess the candidates' ability to a) grasp the points of the passage; b) apply the knowledge written in the passage; c) use his/her own words to summarize the contents; d) develop his/her own thoughts based on the passage; and e) express his/her own thoughts effectively and logically. We found that three raters did not agree when it came to the rating of categories such as "whether stating his/her thoughts," "whether considering from two or more points of view," or "easy or not easy to read." In rating these points, each rater seemed to evaluate different aspects of the subjects' responses from the others. However, although the rater bias existed, we found that the four questions successfully assessed the abilities required of college students of the humanities and social sciences; i.e., comprehension, expression, versatile point of view, summarization, and developing his/her own thoughts. Factor analysis showed differences between students of the humanities and social sciences and students of science and technology. This suggests that this type of test may successfully measure the academic aptitude for the humanities or social sciences at college.

Key Words : Extensive Essay Test, Agreement of Ratings, Humanities/Social Sciences Aptitude

* Faculty of Humanities and Social Sciences, Tokyo Metropolitan University.

** , *** Research Division, The National Center for University Entrance Examinations.