「論文試験とその評価」について

石 井 巖*

ESSAY EXAMINATION AND ITS EVALUATION

Iwao ISHII*

国公立大学向けの共通 1 次試験の実施に伴って、国公立大学の入学試験に変化と改善の努力がみられるようになった。その主要なもの 1 つに論文試験の採用がある。1981 年度の国公立大学入学試験において、3 割に近い 76 大学 118 学部で小論文が課せられるとのことである（1980 年 8 月 3 日読売新聞朝刊）。私立大学においても独自の入試改革の努力がみられる。筆者の属する立教大学文学部でも、1980 年度の入学試験から論文試験が行われた。従来の外国語・国語・社会の 3 教科試験による私立大学文系の一般的な入学試験は、文学部 A（文 A）入学試験として存続させながら、それとは別に 1980 年度から小論文を重んじる文学部 B（文 B）入学試験が実施された。文 A と文 B とは、それぞれ別個の入試であり、試験日と試験方式が異なっている。しかし、入学後は、どちらの入試で入っても同じ文学部学生となり何ら区別はない。同一の文学部に対する入口としての入試が 2 本立てになったのであり、どちらからでも自分の得意とする方から入れる、併願も可能である。文 A の入試は、私立文系の多くで従来から採られてきた方式である。文 B の入試は、中に 30 分間の休憩をはさんでの正味 3 時間の小論文試験と、1 時間の外国語試験とから成る。可否の決定に当っては小論文の結果が重視される方式になっている。もちろん国公立大学のような 1 次試験はない。したがって論文入試と言っても過言ではないと思われる。このような論文入試の実施に関与した者として、論文入試について実情に即して書くようになった。しかし、入試の結果の資料については、まだ公表することが許されていないので、公表されている限りの実施状況をふまえて論文入試の実施の経験に基づいて感じてい

ることを記す程度にとどめることになる。

文 A と文 B との 2 つの入試を並立させ窓口を 2 つにしたことについては、従来の択一式または短答式による 3 教科入試方式一辺倒に対する批判があったためである。国公立の 2 次試験および他の私立大学の入試において、小論文を採用するようになったのも、従来の客観テスト方式の入試問題だけに対する批判や反省からであると思われる。従来の一般の入試問題の多くは、客観性を重視するために、客観性の得られ易い択一式問題や単語や短いことばで答える問題が主であった。択一式問題は客観性も得られ易く、また数量的扱いもし易いので、多くの研究もなされているようである。しかし、この種の問題だけで、大学の求めている能力・人材がとらえられているかという疑問である。ユニークな発想、多面的な思考、思考の推移や構想、複雑でまとまったすぐれた思考、それらの表現力、長文の読解・要約力などが充分に把握されていたとは言えないであろう。それらを把握しようとする試みとして、小論文試験が採用されるに至ったのであろう。もちろん、択一式テストにも長所があり、たとえば、多人数について基礎学力・基礎能力のある面を客観的に限られた時間内に測定するのには向いている。そのために、広く採用されるようになったのであろう。しかし、現状では、補完が必要であり、択一式テストと同等またはそれ以上にまで強調されるべきであると思われる。

小論文試験が、現在では国公立大学入試の 2 次試験や私立大学の入試で多く採用されるようになったが、その中味となると小論文試験といっても種々のようである。

(a) 「愛」「旅」などの題を与えて自由に書かせる作文に類するもの

(b) 「早期教育の是非」「現代社会の技術革新と芸術、自然保護のかかわりについて」「現代社会で言葉が

* 立教大学文学部心理学科 (Department of Psychology, Faculty of Literature, Rikkyo University)

- 果たす役割について」などのテーマだけ与えて論じさせるもの
- (c) 日本の就業構造と所得構造の対比のグラフを示し論評させるもの、勤労者世帯の家計グラフと生活時間表を示し「近年の勤労者生活の特徴、生活時間の特徴、共働きについて」論じさせるもの、石炭や原油などの統計表を示して「わが国の一次エネルギー供給源の基本的特徴を輸入エネルギー依存の観点から論じよ」とするものなど、資料となる図表等を示して論じさせるもの
- (d) 「生物の適応」について具体例をあげて論述させるもの、インフレが企業の経営と労働者の生活に及ぼす影響について論述させるもの、老化現象の説明を求めるもの、一直線上の2球の衝突を例にとり運動量保存と力学的エネルギー保存の両法則について論じさせるものなど、教科に直結した内容のことを論じたり説明させたりするもの
- (e) 寺田寅彦の「科学上における権威の価値と弊害」を読ませ感想を書かせるもの、ベルナルの「実験医学序説」の抜粋を読ませ感想を書かせるものなど、ある文章を読ませ感想を書かせるもの
- (f) 中村真一郎の「芥川・堀・立原の文学」を読ませてその要約を書かせるもの、坂田昌一の「科学と現代」を読ませその要約を書かせるものなど、ある文章を読ませその要約を書かせるもの
- (g) 森有正の「古いものと新しいもの」を読ませ、その主張内容を自分の思考経験と結びつけて論じさせるもの、辻清明の「政治の精神」の一節を読ませ、秩序と対立との関係について論じさせるもの、柴谷篤弘の「反科学論」の一節を読ませ、現代社会における科学のあり方について論じさせるものなど、ある文章を読ませ、それをふまえて、ある論題について論じさせるもの
- (h) 養護教員志望の抱負と期待を書かせるもの、教育学部における家庭科をどのように認識し、そこで何を学びたいかを書かせるもの、心理学とはどんな学問で、何を期待するか、なぜ選んだかを書かせるものなど、志望学科に対する理解、抱負、動機などを書かせるもの
- (i) ベートーベン第9の最終楽章を聞かせて、知っていることや感じたことを書かせるもの、間宮芳生作曲の「子供の領分」を聞かせて感想を書かせるものなど、音楽を聞かせたり絵を見せたりしてその感想や所見を書かせるもの

などがある。(a)では、発想もさることながら、言語的な流暢さがかなりの比重を占める。(b)は、テーマだけしか与えられていないため、自由度が大きい。したがって、結果として得られる論述答案に対する評価の採点者間のずれも生じやすくなる。(c)は図表等の資料が示され、それに基づいての記述となるので、枠が与えられ、しかも、発散的思考というよりはむしろ「正答」を求める収束的思考の試験に近いものである。その点では、一般入試の教科試験の思考に近いものを対象にする。(d)は教科試験そのものであり、記述式教科試験といえよう。したがって、教科の知識に基づく推論過程、「正答」に至る過程、および文章表現力をみるものが主になる。(e)は与えられた文章を正確に読みとりまとめる力と、与えられたものに対する「感想」が主である。感想は、与えられたものに対する感想であるから、与えられたものが中心である。自分の考えを中心にした論の展開は求められていない。(f)は、与えられた文章を正確に読み要約する力をみるものである。(g)は与えられた文章を読み、その論点をふまえながら、別に与えられたテーマに従って論じさせるものである。自分の考えを中心にした論の展開が要求される。与えられた文章の要約やそれに対する感想程度に留まっているものは望ましくない。(g)のような論題に対する論述には、唯一の「正答」があるというものではない。着想や論点は多様であり、したがって、発散的思考のよさと、その展開と、文章表現力をみるものである。しかし、(b)とは異なり、与えられた文章の論点をふまえ、それと絡ませて論を展開しなければならないので、同一答案に対する採点者間の評価のずれが大きくなる可能性は少くなる。(h)は、志願学科に対する理解や抱負をみることに、文章表現力も若干みることになる。(i)は、芸術作品に対する感じ方をみるものが主である。

このように、小論文または論述形式の試験といっても、出題形式や内容に種々あり、それらによってみられる能力も異なる。ただ漠然と文章で書かせる問題であればといったようなことではなく、出題者はみたい能力に適した形式と内容にすべきであることは当然である。と同時に、受験生に一律に明確に出題の意図が伝えられることが必要である。出題者の意図は1つであるのに、その出題の意図がどちらにでもとれるような出題の表現になっていたり、出題の意図のとり方が受験生に任されていたりすることは避けなければならない。概して、小論文形式の試験では、問題の指示文は短かくて、多量の答を求めている。それだけに、出題の意図をはっきりと受験

図1 (Guilford, 1967)

生に伝えるべきである。たとえば、作文でも論述でも、独創的なものを求めているのかオーソドックスな標準的な思考が着実にできることをみようとしているのかを、受験生に明確に伝えるべきである。

上記のことから、小論文式試験が対象にしている能力は次のようなものであると言える。着想・発想、独創的であるかどうか、思考の構想と過程、深さ複雑さ、多様性、知的活動性、具体的鮮明さと抽象性、感性、および文章表現力などである。文章表現力には、語彙、字句、流暢さ、構成、文体などが考えられる。出題によっては、教科の知識を問うこともあるが、小論文形式の出題の魅力の1つは、独創的な着想に基づく一貫性と構想のある論の展開であり、知的活動性を示す多様な展開であろう。知的能力の構造については、Guilford (1967) が図1のような立体的構造を提出している。これによると、知的能力のうち、操作に関して収束的思考とは別に発散的思考の能力があることになる。収束的思考とは、通常の教科試験のように、唯一の「正答」がある問題でそれに向って考え答を導くことである。発散的思考は、あらかじめ予定された「正解」がないか、解決方法が1つとは限らないような問題における思考である。新しい思考を開拓するものであり、創造的思考につながるものである。大学における思考には、収束的論理的思考も必要であるが、新しい思考様式を創造する思考、解決策のわかっていない問題に取り組んですぐれたエレガントな方策を創出する営みも重要である。小論文試験のみが発散的

思考をみる試験であるとか、小論文試験が創造的思考をみるための最良の方式であるとか言っているのではないが、最近の入学試験で小論文試験が多く採り入れられている背景には、文章表現力もさることながら、従来の客観テスト方式の教科試験に対する補完の必要性の認識があるものと思われる。

上記の (h) や (g) のような、論述を求める小論文試験の場合には、その出題意図に合うように、論述答案の長さ・字数もある程度の量を指定する必要がある。独創的発想に基づき構想と一貫性のある展開、しかも、きめ細かな知的活動性や具体的鮮明さや複雑多岐にわたる多様な展開のような知的能力、および、それらと密接に関連しそれらを適切に伝える文章表現力などをみるためには、100字や200字程度の記述では不足である。論述をする受験生の方も、その能力を発揮できないであろう。あらかじめ準備してきたキーワードや殺し文句を並べただけでも100字や200字はすぐうまってしまい、本来の能力と見分けがつきにくい。少なくとも1000字程度の長さは必要であろう。小論文である程度の長さを求めるのであれば、それを書くための試験時間もそれ相応の長さが必要となる。1000字程度のものであれば、1時間半か2時間程度は必要であろう。下書き用紙や草稿用紙も必要となる。言うまでもないことであろうが、要約や感想の場合には、それほど長くなくてもよいであろうし、特に要約の場合には、短かくて要を得たまとめ方をみるのも当を得ていると思われる。与えられた図表等の資料

から「正答」を出させる問題や、教科に関する記述式の問題では、それぞれの「正答」に見合った字数の指定と試験時間とが必要となる。

筆者の属する立教大学の文B入試の小論文試験は、上記の(g)のタイプに属する論述を求めるものである。中に30分間の休憩を挿入での正味3時間の試験である。休憩前の1時間は、与えられた長文の読解と読解力テストである。休憩後の2時間は、論述文の作成である。前部の1時間の読解と読解力テストでは、長文の冊子と読解力テスト問題とが与えられ、その長文を読み、その内容に関する読解力テスト問題に答える。読解力テストは、与えられた長文の内容を正確に把えたかどうかをみるものである。1980年度の入試では、与えられた長文は高階秀爾氏の文章で、A5判でおよそ12ページのものであった。高階氏の長文は、日本人の自然観について論じた論理的な文章である。論理的な文章ではあっても、自然科学の実験報告のような平易な文章ではなく骨のある文章である。独自の見解の展開と論点の多元性のある文章である。受験生の読解力の得点分布からみても、それほど難解でも平易でもなく、長さも無理なものでもなく、妥当な程度のものであったと思われる。文B入試の小論文試験には、長文読解力をみることも含まれていることが、入試要項等にも明記されているが、この種の読解力が求められているのである。

長文の読解および読解力テストの後、30分間の自由な休憩に入る。読解力テストは既に回収されているが、長文の冊子の方は、後の論述文作成のときにも使用されるので、30分間の休憩中は、各自が所持保管することになっている。この間にその冊子を見ることも自由である。

休憩後、正味2時間の小論文作成の試験が行われる。ここでは、短文の参考論文と小論文のテーマとがあらたに与えられる。この参考論文と、すでに前に与えられて持っている冊子の高階氏の長文とを併読して、それぞれの内容をふまえて、与えられたテーマについて1000字前後の小論文を作成するのである。1980年度に用いられた短文は、谷崎潤一郎氏の「陰翳礼讃」からの抜粋で、およそ1700字程度の長さのものであった。内容は、日本古来の風俗・文化に見出される陰翳について論じた文章である。具体的で精細で、文化についての広い視野を背景に感覚的機能を十分に発揮して論じた文章であり、感性豊かな文章であった。小論文のテーマは、「文化の特質を自然と人間生活との関係において論ぜよ」であった。

論理的な文章と感性的な文章との2つを読み、それらの内容をふまえて、与えられたテーマについて論じるこ

とが求められる。論述するに際して考慮すべき出題の意図は、入試要項にも、試験の際に配布される冊子にも明記されている。入試要項には、「長文の読解力、および小論文作成をとおして独創的な発想、論理的構成力、文章表現力などを判定する」ためと記されている。小論文テーマ等を記した冊子にも、小論文作成に当たっての指示文が記されている。「与えられた長文と短文とを併読して、それぞれの内容をふまえて、与えられたテーマについて論ぜよ」となっている。これらの記載事項によって、小論文作成試験の出題意図は尽されているように思われる。

与えられた条件や他者の論点を正確に把握し、それらとしっかりと取り組みながら、しかも独自の見解・ユニークな発想で一定の方向性を打ち出し、全体をまとめているものが求められている。テーマについて論じていて、そのかぎりでは優れていても、与えられた長文と短文とに無関係に論を進めているのはだめである。与えられた2つの文章の内容をまとめた程度のもので、抽象論・一般論だけのもの、常識的なもの、月並みの発想のものなども、出題の意図からはずれている。具体性豊かで、緻密で知的活動性が高く、複雑で多様性があり、広く諸事例に精細に言及しながら、全体的に一定の構想のもとにしっかりとまとまっているものがよい。精細な論理的な文章でも、感性的な文章でも可とされるが、論述においては構成力もポイントの1つになっているので、刹那的に感じるままに書いた表面的な流暢な文章とか、思い浮かぶままに書き流した文章などの類の感性的な文章ではだめである。感性を十分に発揮した文章であっても、谷崎氏の文章のように、全体をまとめているしっかりとした論理的構成があることが必要である。誤字、脱字は言うに及ばずであろう。豊富な語彙、微妙な差異までの確に伝えられる表現力、文章のリズム、文体等も出題の意図に含まれていることが記されている。文章表現の形式面のみ整っていても、内容の伴わないものはだめなのは当然である。

小論文式試験の長所、すなわち、客観テスト式教科試験では把えにくく小論文式試験で把えられる能力には、まず文章表現力があげられる。これは、程度の差はあっても、上記の(a)から(i)のすべてのタイプに当てはまる。小論文式試験の長所の中には、論述を通してみられる姿がある。着想とか発想、構想、それらに基づく展開の姿である。これには、上記の(b)と(g)が適している。自分勝手の論を唱えるだけでなく、与えられた条件や他者の論点を正確に把握し、それらとしっかりと取り

組みながら、しかも独自のユニークな見解で一定の方向性を打ち出し全体をまとめる能力の重要性、あらかじめ準備して暗記してきたいくつかの小論文のうちの1つを書き写すに類する弊をさけること、採点の基準と公平性などからみて、(b)のタイプよりは(g)のタイプの方を推したい。文Bの小論文試験では、(g)のタイプに準拠している。文Bで作成される小論文の長さは1000字前後と指示され、論述のための試験時間は2時間である。数1000名という多数の受験生を対象にした場合、試験当日の実施上の条件、試験終了後の採点上の条件からも、2時間で1000字前後が1つの目安と思われる。

すべての方式の試験に当てはまることではあるが、小論文式試験においても、出題における検討と配慮と並んで採点に関する検討や配慮も重要である。出題における努力も採点の仕方いかんでは活かされないことも生じる。特に小論文式試験においては、客観テスト方式の試験に比べて、問題文や指示文など条件規定に当るものが少く、求められる答の部分が多い。したがって、答の自由度が大きい。客観テスト方式の試験の場合には、テストそのものの中に自づと採点の基準が含まれている。つまり、採点基準は、採点者の中よりもテストそのものの中にあると言える。小論文式試験においては、むしろ、採点基準は、採点者の中にあるとも言える。したがって、小論文試験の採点に当っては、試験の信頼性・妥当性・公平性のために、特に配慮が必要である。採点者の頭の中にある基準が、出題の意図と合致するようにしなければならない。出題の段階で十分な検討がなされ、問題文や指示文がよいものになり、受験生がそれをよく理解して小論文を作成したとしても、採点の基準の段階でずれが生じていたのでは救われない。その小論文試験の意図は達せられないばかりか、別な作用を及ぼすことになる。採点者の採点基準が出題の意図に合致していても、恒常的で安定したものでなければならない。多くの受験生の比較的長い記述の答案を、長時間にわたり、場合によっては数日をかけて、採点しなければならない。その間に、ときどき基準が変動したり、ある方向へ変っていくようなことがあってはならない。前半の採点の結果によって後半の採点のための基準を変えるとか設定しなおすことがあってはならない。受験生が数1000名と多いのが通常であるので、採点者各人が、すべての受験生の答案を採点することはできないことが多い。その場合には、採点者のちがいがよる採点の公平性の問題も存在する。採点者間で、採点基準の比重のおき方のちがいが、甘辛のちがいが、採点幅のちがいがあがある場合に

は、同一受験生がどちらの採点者によって採点されたかによって、結果として得られる得点に差が生じる可能性がある。これらの問題に対しては、十分な配慮と処置をして、試験の妥当性・信頼性・公平性が得られるようにしなければならない。

採点に先立って、まず、採点者を一堂に集めて、試験問題、出題の意図、採点基準などについて、出題責任者が十分に説明する。その後、全員で基準について討論をしてつめておく。これを十分に行うかどうかによって、後の採点の良否がかなり左右される。発想の独創性とか構想等の内容面、(g)タイプであれば、与えられた文章との関連の有無についてはもちろんのこと、誤字・脱字や論文の長さ、漢字の多少などの文章表現力の細部についても基準を明確にしておくべきである。さらに、評定における各種の恒常誤差(田中、1965)についても説明をし、それを避けるように配慮すべきである。寛大性の誤差、中心化傾向の誤差、光背効果、論理誤差、対比誤差などがあげられる。

寛大性の誤差は、採点者がよく知っている人などを評価する場合に、実際より高く評価する傾向があることであり、入試においては厳に避けるべきことであり、そのためには、説明だけでなく、受験番号だけにして受験者の氏名を答案に書かせないとか、受験番号を暗号化したりとじ込んで目に触れなくすることが必要になる。中心化傾向の誤差とは、採点者は一般に極端に高い点や低い点をつけることを避ける傾向にあるため、採点結果が中央に集中する傾向が生ずる。これを除くために、両端に非常に極端なカテゴリーを加えておくことも一策である。光背効果は、答案の一般的な印象のために、ある特性の評定がひっぱられてしまうことである。たとえば、上手な字でびっしり書いてあるために、内容も実際よりもよい方向に偏らせて判定したり、内容がよいために誤字・脱字があっても甘めに評定したりすることであり、排除しなければならない。論理誤差は、採点者が論理的に関連していると考えている特性については、類似した評定を下しやすという誤差である。各特性を事実に従って評定していかなければならない。対比誤差とは、採点者自身も持っている能力や特性とは反対の方向に評定する傾向のことである。たとえば、緻密な論の進め方を得意としている採点者は、概して、他の人の論述を緻密でないと判定する傾向のあることである。

出題の意図や採点基準などについて説明と十分な討議と理解の後に、採点者は、各々採点にとりかかる。採点に入っても、いきなり本番の点とはしないことである。

最初の 50 枚程度を本番のつもりで採点してみても、採点基準が安定し変動がなくなり、現実の答案に接して基準が確定したところで、その基準で改めて最初から採点しないとして、それを本番の採点とするのがよい。最初の 50 枚の試みの段階で、頭の中で考えていた採点基準の現実への適用の仕方、現実の答案の幅、甘辛さの基準などを十分に検討して、採点基準を確定した後に、改めて本番の採点に入るべきである。最初からいきなり本番の採点をし、採点の進行に伴って基準を変えていくようなことは絶対にしてはならない。

採点の段階数をどの程度にすべきかの問題がある。論述文の採点段階は、順序づけられたカテゴリーであると考えられる。したがって、論述文の採点は、各答案を採点カテゴリーのいずれかに入れることであると考えられる。1 点きざみの 100 段階のカテゴリーをたてることもできるが、限られた採点時間と労力との条件のもとで信頼性の高い採点となると、100 段階設定しておいても、実際に使用されるのは、せいぜい 10 点きざみの 10 段階であったり、5 点きざみの 20 段階でも実際に使われているのはそのうちのある部分（たとえば、45 点から 90 点の 10 段階）であったりする。結局、無理のない信頼性の高い段階数は 5 段階から 7 段階のようである。（田中、1965）。筆者の文 B の経験からも、そのように思われる。時間と労力が制限なくかけられるのであれば別である。すべての答案を一対比較できたり、品等法的にすべての答案を並べたり、第 1 段のカテゴリーで分け次に各カテゴリー内に第 2 段のカテゴリーを設けて細分化していき、隣接するカテゴリー間で調整する方法がとれたりすれば、実働する信頼性のあるカテゴリー数は多くなれるが、時間と労力の制約のある状況においては不可能に近い。現実的には、5 段階から 7 段階が妥当であろうと思われる。

受験者数が多い場合、各採点者が全受験生の答案をみることはできなくなるので、手分けして採点することになる。すべての答案が同一採点者によって採点されずに、答案によって採点する採点者が異なることが生ずる。採点者が異なることによる採点のちがいが生ずる可能性がある。この可能性を救うために、同一答案を複数の採点者によって採点することがなされる。複数は、2 人か 3 人、多くて 5 人である。2 人の場合は、たとえば採点者 A と B で採点第 1 チーム、採点者 C と D で採点第 2 チーム、以下同様にして採点チームをつくり、第 1 チームは受験番号 1 番から 200 番までの 200 人分の答案を採点し、第 2 チームは受験番号 201 番から 400 番までの 200

人分の答案を採点するという具合に採点すべき答案を割り当てて、各採点チーム内の採点者 2 人は、そのチームに割り当てられた同一の答案を採点する。たとえば、第 1 チームの A も B も、受験番号 1 番から 200 番までの答案を採点する。採点チームの編成に当っては、チーム内ではなるべく採点傾向の互に異なる採点者を、チーム間では各チームごとに全体としてみた場合に採点傾向が同じになるように配置すべきである。チーム内の採点者間の相関が高くなるように配置すべきではない。そのようにすれば複数採点の意味がなくなる。複数採点の場合、各採点者の採点は互に独立に行うべきである。たとえば、第 1 チームの採点者 A と B とは別の採点室で、しかも A は B の採点結果が見えないようにして、B は A の採点結果が見えないようにして採点すべきである。そのようにして、複数採点によって得られた結果についての処理については、たとえば、採点はすべて終了した後に、2 人の採点者の採点を突合わせて、一致・開きをみる。開きが大きい場合には、採点者にフィードバックして、不注意等によるものでないかどうかなどの確認を求める。その後の処理としては、その後の処理の必要に応じて、たとえば、各採点者ごとに採点値の尺度変換や甘辛の調整等を行うなどが考えられる。採点者 2 人ずつによる複数採点方式で、2 採点者間に大きな点の開きが多く出るようであれば、採点者 3 人ずつによる複数採点方式、または体操の採点のように採点者を 5 人ずつにし、結果として得られた採点のうち、両端の採点 2 つをはずして中 3 つの採点を採用する方式などが考えられる。複数採点方式をとることによって、1 採点者内における採点基準の不測の変動をもチェックすることができる。

複数採点方式をとり、しかも、採点チームごとに異なった答案が割り当てられることになると、採点者間の甘辛の程度のちがいが問題になる。経験からみると、これは、採点における観点のちがいがよりもしばしば生じることであり、より現実的な問題である。この問題の処理法として、採点者ごとに粗点を尺度変換した後、一般的には、採点者ごとに平均値を等しくしたり標準化したりする方法がとられるようであるが、それは仮定を含んでおり望ましくはない。つまり、チームごとに、割り当てられる受験生の等質性（少なくとも平均において等しいなど）を仮定しているのである。この仮定が成り立たないことがあり得る。別の能力ではあるが、たとえば、純客観テストの択一式テストの結果だけについて、受験番号順に同人数ずつチーム分けして、各チームごとの平均点を求めてみると、チーム間に平均値のかなりのち

表 1

	A	B	C	D	E		N
A		$\bar{x}_b^a - \bar{x}_a^b$	$\bar{x}_c^a - \bar{x}_a^c$	$\bar{x}_d^a - \bar{x}_a^d$	$\bar{x}_e^a - \bar{x}_a^e$		$\bar{x}_n^a - \bar{x}_a^n$
B	$\bar{x}_a^b - \bar{x}_b^a$		$\bar{x}_c^b - \bar{x}_b^c$	$\bar{x}_d^b - \bar{x}_b^d$	$\bar{x}_e^b - \bar{x}_b^e$		$\bar{x}_n^b - \bar{x}_b^n$
C	$\bar{x}_a^c - \bar{x}_c^a$	$\bar{x}_b^c - \bar{x}_c^b$		$\bar{x}_d^c - \bar{x}_c^d$	$\bar{x}_e^c - \bar{x}_c^e$		$\bar{x}_n^c - \bar{x}_c^n$
D	$\bar{x}_a^d - \bar{x}_d^a$	$\bar{x}_b^d - \bar{x}_d^b$	$\bar{x}_c^d - \bar{x}_d^c$		$\bar{x}_e^d - \bar{x}_d^e$		$\bar{x}_n^d - \bar{x}_d^n$
E	$\bar{x}_a^e - \bar{x}_e^a$	$\bar{x}_b^e - \bar{x}_e^b$	$\bar{x}_c^e - \bar{x}_e^c$	$\bar{x}_d^e - \bar{x}_e^d$			$\bar{x}_n^e - \bar{x}_e^n$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	$\bar{x}_a^n - \bar{x}_n^a$	$\bar{x}_b^n - \bar{x}_n^b$	$\bar{x}_c^n - \bar{x}_n^c$	$\bar{x}_d^n - \bar{x}_n^d$	$\bar{x}_e^n - \bar{x}_n^e$		
Σ	Σ_a	Σ_b	Σ_c	Σ_d	Σ_e		Σ_n
$\frac{\Sigma}{n-1}$	$\frac{\Sigma_a}{n-1}$	$\frac{\Sigma_b}{n-1}$	$\frac{\Sigma_c}{n-1}$	$\frac{\Sigma_d}{n-1}$	$\frac{\Sigma_e}{n-1}$		$\frac{\Sigma_n}{n-1}$

がい出たり、その出方にある傾向のみられる可能性があり得る。これらのことを考慮した調整方式の理論的な議論は、たとえば、Kamp et al. (1980) に委ねることにして、現実的には、次のような対比較法的な処理も考えられる。採点者 2 人ずつの複数採点における採点者のチーム分けを、採点の始めから終わりまで同じ顔合わせに固定しないで、あらゆる顔合わせになるように組む。採点者が A から N までの n 人いるとする。採点開始から終了までの間に、全体として、表 1 の対角線の上半分または下半分のますの数だけ採点チームが生成されることになる。つまり総チーム t 数は、

$$t = {}_n C_2 = \frac{n(n-1)}{2}$$

となる。受験者数を s とすれば、 t チームの各 1 チーム当りの採点枚数は、 s/t となる。各採点者別にみると、たとえば、採点者 A は、他のすべての採点者とチームを組んだことになるから、 $n-1$ チームに参加したことになり、1 チーム内での採点枚数は s/t 枚なので、採点者 A の合計採点枚数 w は

$$\begin{aligned} w &= \frac{s}{t} \times (n-1) \\ &= \frac{s}{\frac{n(n-1)}{2}} \times (n-1) \\ &= \frac{s}{\frac{n}{2}} \end{aligned}$$

となり、単純なチーム分けの複数採点方式のときと同じになる。その後の処理については、表 1 に従って進められる。表 1 で、たとえば、 $\bar{x}_a^b$ とあるのは、採点者 A と B とのチームに割り当てられた s/t 枚の答案に対する A の採点の平均点であり、 $\bar{x}_b^a$ は、同じ答案に対する B の採点の平均点である。同様に、 $\bar{x}_a^c$ は、採点者 A と C とのチームに割り当てられた別の s/t 枚の答案に対する A の採点の平均点であり、 $\bar{x}_c^a$ は、その同じ答案に対する C の採点の平均点である。説明の簡単化のため、 s/t 枚を含む任意の q 枚の答案に対する採点者 A と B との採点結果は、分布の形は同じで平行移動した関係にあり、平均値がそれぞれ $\bar{x}_a$ と $\bar{x}_b$ とであるとする。A と B 両者の採点の間に一貫性がある（相関が完全である）ならば、

$$\bar{x}_a^b - \bar{x}_b^a = \bar{x}_a - \bar{x}_b$$

となる。このような仮定がどの採点者間についても成り立つ状況においては、 $\bar{x}_a^b - \bar{x}_b^a$ のような部分平均点の差が $n-1$ 個得られれば、各採点者の採点の平均間のちがいはすべて求まることになる。しかし、現実には、諸変動が伴うので、現実的な処理としては、表 1 の各列において、それぞれ対角成分を除く $n-1$ 個の部分平均点の差の合計をし、それを Σ 欄に記入し、それを $n-1$ で除した値を $\Sigma/(n-1)$ の欄に記入する。この最終欄の値のちがいが、平均的にみて、それぞれの採点者の採点の平均値のちがいとして扱うのである。たとえば、採点者 A と B との採点の甘辛さのちがいは $\Sigma_a/(n-1)$ と $\Sigma_b/(n-$

1) とのちがいが程度であろうとするのである。

文B入試の採点に当たっても、信頼性・妥当性・公平性が確保できるように、できるかぎりの努力が払われた。まず、採点に先立って、出題および採点の責任者から全採点者に対して、出題のねらいや採点の基準・ポイント等について、詳細な解説がなされた。その後、質疑応答・討論等をすませてから、採点に入った。採点は、同一の答案を複数の採点者が採点する複数採点方式で行なわれた。複数採点の採点チームは22チーム編成された。チーム編成に当たっては、採点者の専攻分野、年令その他の諸要因を考慮し、チーム内ではなるべく採点者が異質になるように、チーム間では同質になるように編成された。各採点者は合計で約220枚の答案を採点した。これらのチームとは別に、採点調整チームも設けられ、複数採点間の開きのチェックと原採点者へのフィードバックを行なった。複数採点者の各々は、別室で互に独立に採点し、採点に当たっては他採点者の採点がわからないように配慮された。50名を超える採点関係者が動員され、足かけ3日間にわたって慎重に採点された。複数採点者間の採点の開きのチェックは調整チームで行なわれたが、結果として、一致度は高く、複数採点終了後に採点のふたをあけて突合せたときに関係者は皆一様にその一致度の非常に高いのに驚き感心したものであった。複数採点者間で大きな開きが出たのは、平均して100枚に1件あるかないかであり、全体でわずか数10件であった。全答案数が5000枚に近かったことから考えると、一致度がきわめて高かったことがわかる。このような成功をおさめられたのは、1つには、採点に先立っての解説の適切さとその徹底にあったと思われる。もう1つの成功因として、論文入試に対する考え方と採点の経験について、過去数年の積み重ねがある。規模と運用上の差はあるが、立教大学の文学部の3年編入試験でも、大学院の文学研究科の入学試験においても、専門試験の他に、共通論文という試験が、数年にわたって行われてきた。特定専門領域に偏らない内容について、長文の読解と要約と論述という形成の記述試験である。文B入試の小論文試験は、これらの積み重ねをふまえたものと見ることもできる。このようにして採点された後、ある一定方式の尺度変換と甘辛の調整がなされ小論文論述の得点とされた。詳細は記せないが、文Bで小論文を重視すること、甘辛のちがいが現実のものであること、および、可否の境界などの現実を考えると、採点後の甘辛等の調

整は必要であることが痛感される。

論文入試の採点に実際に関与した者として、論文採点に先立っての適切な解説が重要であることを感じる。出題の意図、採点基準の明確化、それらと採点段階との明確な関連づけ等を、採点者に徹底させることである。次に、採点者の訓練・経験の積み重ねも重要な要因であるように思われる。論文入試の施行の初年度や次年度の困難さをのり越えると、次第に採点者側の経験も積み重ねられ、採点者内の尺度もよりよいものになり、より信頼性・妥当性の高い論文入試が行われるようになると思われる。年次を経るに従って、採点に関する資料も蓄積されてくるので、その面からの分析・検討が必要であるように思われる。たとえば、採点者をテスト項目と見て、採点者の項目分析のようなことも考えられる。採点者のタイプがあるとすれば、採点者のクラスター分析も考えられる。後のチーム編成に役立つ。文B入試について、採点に関する資料の蓄積とその分析、およびその結果を後の入試に活かすことが始動している。

入試改善の努力の現れとして、いくつかの方式の入試が行われ、その1つとして小論文入試も行われるようになった。それぞれの方式の入試を行う努力をした後に、当然その成果についての検討・反省をしなければならない。入試方式の予測的妥当性の検討もその1に加えられるべきものであろう。この検討は、息の長い追跡調査をまわって可能になるものと思われる。文B入試では、外国語はある程度の得点があればよく、小論文の結果が重視される。拙速な評価は慎まねばならないが、文Aと文Bとの併願者の状況を調べてみると、文Aの不合格者で文Bに合格した者が49名もあった。その中には、文Aの高点順の席次では合格圏よりもかなり下で、一般入試では入ることが困難と思われる者もいる。このことは文B入試の目的の1つが達せられたことを示す資料の一部とも思われる。小論文入試の成果についての全般的な評価は、確実な資料をふまえての、今後の息の長い検討が必要とされる。

参 考 文 献

- Guilford, J.P. *The Nature of Human Intelligence*. 1967, McGraw-Hill.
 Kamp, L.J. et al. eds. *Psychometrics for Educational Debates*. 1980. John Wiley & Sons.
 田中良久. 心理学的測定法. 1965, 東大出版会.