

小論文試験の採点における文字の美醜効果の規定因

—メタ分析および実験による検討—

What factor moderates the effect of handwriting quality on essay test scoring:
Investigation by meta-analysis and experiment

宇佐美 慧

Satoshi Usami

東京大学大学院教育学研究科

Graduate School of Education, University of Tokyo

小論文試験の採点における文字の美醜効果の規定因

—メタ分析および実験による検討—

小論文試験の採点において問題視される文字の美醜効果が、実際に生じるか否かに関しては一貫した知見が得られていない。そこで本研究では、文字の美醜効果を規定しうる要因として、(1)項目の解答内容の自由度、(2)受験者の年齢、(3)採点者のスキルの3つの要因をとりあげた。

本研究ではまずメタ分析を実施して3つの要因が文字の美醜効果に及ぼす影響を調べた。その結果、受験者の年齢が低いほど、文字の美醜効果の影響が大きいことが示唆された。

そこで次に、受験者の年齢は、受験者の作成する文章の質の差を介して文字の美醜効果の影響度を規定するという仮説を立て、その仮説を検討するために実験的に20名の採点者に小論文試験の採点をさせた。採点結果を従属変数とし文字の美醜と文章の質を要因とした2要因分散分析の結果、2つの要因間に交互作用は見られず、文章の質が文字の美醜効果を規定していない可能性が示唆された。

No consistent results have been shown about whether the handwriting quality affects essay test scoring. We hypothesized that the following factors may moderate the effect of handwriting quality: (1)degree of freedom for answer, (2)ages of examinees, (3)skills of scorers.

Then, we performed a meta-analysis to evaluate the effect of these factors. The result suggested that the younger the examinee, the larger the effect of handwriting quality. Based on this result, we hypothesized that the factor of ages virtually means the factor of quality of essay, and that the quality of essay mediates the moderator effect of age. To test the hypothesis, we let 20 participants score essay tests with different levels of essay quality and handwriting quality. The result of a two-way ANOVA showed no interaction effect, which indicated that essay quality may not mediate the effect of handwriting quality.

キーワード： 小論文試験 テスト 文字の美醜 メタ分析 誤差

Keywords : essay test, bias, handwriting quality, meta-analysis, error

1 問題と目的

1.1 小論文試験の普及と問題点

近年の小論文試験の普及は目覚ましく、現在日本の国公立大学の約8割が大学入試において小論文試験を利用している（文部科学省、2005）。また、それだけにとどまらず小論文試験は就職試験などにも頻繁に用いられており、小論文試験の利用範囲はさらに拡大している。

しかし、このように小論文試験の利用が拡大する一

方で、小論文試験による評価には多くの測定論的問題があることが古くから指摘されている。それには、例えば同一採点者の採点結果が時間的に極めて不安定であるといった信頼性の問題（平・江上、1992）や、採点が受験者の性別や文字などの様々な外的誤差要因に影響されるといった公平性の問題がある（例えば渡部・平・井上、1988）。

これらの問題に対処するためには、項目数・採点者数・採点方法などの様々な観点から小論文試験の測定論的な妥当性を詳細に検討していく、小論文試験の利

用を改善する手立てを講じることが重要である。ところが、小論文試験を測定論的観点から捉えた実証的研究は日本ではまだ少なく、項目の作成から実施・評価に至るまで小論文試験が適切に運営されるために必要な理論は、客観式テストに比してほとんど構築されていないのが実状である（渡部, 1994）。

1.2 小論文試験の採点に影響する外的誤差要因

前述のように、小論文試験の採点においては採点の公平性を脅かす様々な外的誤差要因が存在する。海外ではこの外的誤差要因について比較的古くから議論されており、多くの知見が既に得られている。外的誤差要因の種類には、例えば文字の美醜効果 (Briggs, 1970 ; Bull & Stevens, 1979 ; Chase, 1968, 1979 ; Eames & Loewenthal, 1990 ; Hughes, Keeling, & Tuck, 1983a ; James, 1927 ; Markham, 1976 ; Marshall & Powers, 1969 ; Soloff, 1973 など) や、評定する順序によって同じ文章の採点結果が異なる現象としてよく知られる系列効果(Hughes, Keeling, & Tuck, 1980, 1983b など)、さらには時間制限、採点日数、受験者の人種・性別・性格・魅力、項目の選択方式（すべての項目に必答か、一部選択式か）、教師の期待、文法成熟度(これらの要因については、Chase, 1983, 1986 ; Cooper, P. L. 1984 ; Hughes *et al.*, 1983b など)など様々な外的誤差要因が知られている。

1.3 文字の美醜効果を扱った研究

特に海外で古くから議論の俎上に載せられていた外的誤差要因に文字の美醜効果がある。ここでいう文字の美醜とは、採点者それぞれが主観的に判断する手書き文字の美しさのことを指す。文字の美醜が評価に与える影響の大きさは様々であるが、例えば Hughes *et al.*(1983a)の実験においては、単一の作文の評価において文字の美しい文章はそうでない文章に比べ、採点の結果に25点満点中平均1.32点(100点満点であると約5点分)の差が生じたことが示されており、その影響度は決して小さくはないことがわかる。また CBT(Computer Based Test)のようなテスト形式でない限り、入学試験や就職試験などにおいて、常に文字の美醜効果が生じる可能性があり、さらにその影響度も未知数である(Chase, 1986 など)。文字の美醜の外的誤差要因は一般に認識されているよりも非常に現実的な問題である可能性がある。

1.4 先行研究の結果の不一致とメタ分析の適用

しかし、採点時に文字の美醜効果が実際に生じるかどうかについては未だ一貫した結論が得られていない。例えば、Markham(1976)の研究では45名の熟練教師と36名の学生が、小学5年生の張子の動物の描写に関する作文を採点したところ、文字の美しい作文に対して高い得点を与える傾向が示された。それに対し Eames & Loewenthal (1990)では、16名の大学教師が大学生の書いた(記憶・言語などの専門分野に関する)小論文を採点しているが、文字の美醜に影響はされていないことを示している。

このような違いについて、Chase(1979)は教師の生徒への期待が小さいときには文字の質が良いほど評価が高く、逆に生徒への期待が大きいときには文字の質が悪いほど評価が高いといった、文字の美醜要因と教師の生徒に対する期待要因の交互作用効果によって説明した。しかしその後の研究において交互作用が示されないという知見も得られており(Hughes, *et al.*, 1983a)，さらに教師の期待効果のない実験的な採点場面においても、文字の美醜効果が示された研究とそうでない研究が混在していることから、教師の期待要因だけでは文字の美醜効果の有無を説明することは困難であるように思われる。

このように同一の要因の効果（本研究の場合では文字の美醜効果の有無）を検証した一連の研究の分析結果に一貫した傾向が見られないときに、各研究の条件の違いや用いられた被験者群の属性などの特徴を独立変数とし、効果量を従属変数として独立変数の従属変数に及ぼす効果をメタ分析によって統計的に検討することが可能である。ただし、複数の研究を統合すれば統計的な信頼性は高まるものの、独立変数とした要因以外では異質な研究の結果を統合することになり、それ自体に意味がないといった apples and oranges 問題が生じうる。他にも、統計的に有意でない結果は発表されずにしまい込まれる傾向があるといった file drawer 問題などがメタ分析の問題点としてしばしば指摘されるが、利用する先行研究の選択やコード化を十分に吟味し、データの性格に合わせた適切な分析をすれば結果を統合的に解釈するための参考となる。

このような利点を踏まえて、本研究ではまず、文字の美醜効果について実験的に検討した先行研究を収集してメタ分析を実施する。文字の美醜効果と関連の深い要因を検討することにより、文字の美醜効果に影響を受けにくいテストの作成・実施・評価のあり方を考

えていくための一助となるだろう。また、メタ分析に関する研究では Lipsay & Wilson(2000), Mullen(小野寺訳, 2000), 芝・南風原(1990), 渡部(2002)などを参考されたい。

1.5 文字の美醜効果を規定すると考えられる要因

本研究におけるメタ分析では、文字の美醜要因の効果量を従属変数として、文字の美醜要因に関連が深いと思われる①項目の解答内容の自由度、②受験者（答案作成者）の年齢、③採点者のスキル、の3つの要因を独立変数としてとりあげる。以下では、これら3つの独立変数について具体的に説明する。

(1)項目の解答内容の自由度

小論文試験の中でも、想定される解答がある程度客観的な事実内容を多く含むため、比較的客観的な採点が期待できる項目もあれば、想定される解答があらかじめあまり定まっていない項目もある。前者のような項目では、採点者側は得られる答案の解答をあらかじめ見積もっており、さらに評価基準があらかじめ決まっていることが多いため、評価に無関係な文字の美醜の影響を意識的に避けることができると考えられる。

先行研究においては項目の具体的な内容が記述されていない研究もいくつかあるが、大別して歴史科目的論述や、具体性の高い専門分野の論述の項目は、想定される解答が比較的客観的な事実内容を多く含むため前者のタイプである。例えば文字の美醜効果が示されなかった Eames *et al*(1990)における項目は客観的テストに近い性格を持った記憶・言語などの専門的な内容であり、想定される解答の内容にある程度の客観的な事実内容が含まれていると考えられるテストであった。逆に文字の美醜効果が見られた Markham(1976)における項目は動物の張子を見てその描写をすることが課題であるため、受験者のアイデアの創造性や表現力に評価の力点が置かれており、想定される解答には比較的広い幅があったと思われる。

(2)受験者(答案作成者)の年齢

先行研究における受験者（答案作成者）は小学生である場合もあれば、大学生の場合もある。たとえば、文字の美醜効果が示された Briggs(1970)では小学生が受験者であり、逆に文字の美醜効果が示されなかつた Eames *et al*(1990)では大学生が受験者であった。これは、受験者の年齢の違いが文字の美醜効果の有無に影響を及ぼしていた可能性を示唆している。しかし、受験者の年齢が直接的に文字の美醜効果に影響してい

ると考えるよりも、受験者の年齢が他の何らかの要因を介して文字の美醜効果に影響を与えると考えるほうが自然であり、この点について検討することで実際のテスト作成において有用な知見が得られると思われる。

たとえば、一般に年齢の異なる集団から得られる文章には何らかの質的な違いがあると考えられることから、文章の質の差を介して文字の美醜効果に影響を与える可能性がある。すなわち、全体として質の高い文章を採点する場合には、文字の美醜ではなく文章の質そのものが評価されやすく、全体として質の高くない文章の場合には文字の美醜効果が表れやすくなるという可能性が考えられる。ただし、文章の質については具体的に先行研究の結果からコード化することは困難であるため、まずは受験者の年齢に注目して、メタ分析で検討する。

(3)採点者のスキル

先行研究において答案の採点者は、学校や塾の採点の経験が比較的長い教師の場合と、経験が比較的短い大学院生・大学生が採点をしている場合がある。

熟練の教師であるほど、自分の経験に基づいて本来の評価とは無関係な要因をより意識的に排除することができることが考えられる。したがって、採点者のスキルの違いが文字の美醜効果に影響を及ぼしている可能性も検討する価値があるだろう。

2 メタ分析による検討

2.1 論文の収集

論文の選択基準

メタ分析のソースとなる論文の収集においては以下の6つの観点を基準に選んだ。

- ① 小論文試験の採点における、文字の美醜の高低群差の検定のための t 統計量、 F 統計量、ないしは評価得点と美醜得点の相関係数のいずれかの統計量が報告されており(若しくはこれらを計算するための平均値や標準偏差などの基礎統計量が報告されており), 効果量(Fisher の Z 変換値)が算出できること。
- ② 採点が文章の内容の質を基準としてなされていること。

表 1. メタ分析データベースにおける関連項目の値

研究	項目内容	N	Z _{fisher}	解答内容の 自由度	受験者の 年齢	採点者の スキル
Chase(1968)	詳細不明	128	0.24	不明	5	1
Marshall <i>et al</i> (1969)	歴史論述	420	0.11	1	3	0
Briggs(1970)	作文「深い霧の日」	500	0.14	0	1	0
Soloff(1973)	歴史論述	64	0.54	1	3	0
Markham(1976)	動物の張子の描写	648	0.31	0	1	0
Bull <i>et al</i> (1979)	詳細不明	48	0.11	不明	5	0
Chase(1979)	テスト理論に関する文章	62	0.04	1	5	1
Hughes <i>et al</i> (1983a)	作文「10 年後の希望について」	224	0.30	0	2	1
Eames <i>et al</i> (1990)	(記憶・言語の) 専門分野に関する文章	64	0.03	1	4	0
Z				-1.09	-1.86*	-0.24

*両側 10% 水準で有意傾向

- ③ 文字の美醜得点の水準（高群と低群）が明確に区別されていること、または得点化されていること。
- ④ 各群で 10 人以上の標本サイズが確保されていること。
- ⑤ 図書館やインターネット上から、直接或いは注文をして入手することが可能であること。
- ⑥ 3 つの要因のうち少なくとも 1 つの要因の水準が判断できる記述が論文中にあること。

検索方法

1920 年 1 月～2007 年 10 月の国内外の雑誌に掲載された論文を、DATA BASE, EBSCOHOSTS, ERIC, FELIX, GeNii, Info Trac Custom , JSTOR, PCI, PILOT, PsycARTICLES, PsycINFO, Web of Science のデータベースと Google などの検索 エンジンを用いて調べた。キーワードは handwriting, handwriting quality, writing, penmanship, essay, test, evaluation, neat, neatness, character, letter, scoring, rating, bias, factor, error (国内の論文に関してはテスト, 試験, 小論文, 文字, 字, 評価, 採点, バイアス, 誤差などの用語)をもとに検索した。

その結果、文字の美醜効果について検討した 17 件の論文が検索された。これらの論文の中にはワープロ文字と手書き文字の比較をした研究は含まれていない。そして、これらの論文の中の引用文献からさらに新しく見つかった 1 件の論文を含めた合計 18 件の論文から、上述の①～⑥の基準により 9 件の論文を選択した。具

体的には①で 2 件、②で 3 件、⑤で 4 件の論文が条件に合わなかったため、それぞれメタ分析から除外した。

日本で行われた研究については吉村(1991,1992)や渡部・曹(1992)の論文があるが、いずれもワープロ文字と手書き文字の採点の比較を目的とした研究である。吉村(1991,1992)では手書き文字の美醜得点の高群と低群における採点の比較もしているが、文章の丁寧さや温かさの印象など、採点が文章の内容の質とは無関係な項目に基づいており、さらに実験における材料の収集の手順に不明な点が多くいたためメタ分析には含めていない。

2.2 独立変数のコード化

解答内容の自由度

先行研究においては歴史科目や専門分野に関する記述といった、解答内容の自由度が比較的小さい項目と、張子の描写や将来の目標に関する作文など自由度の大きい項目があった。多くの先行研究において項目の具体的な内容については記されていなかったため、本研究では歴史や専門分野に関する項目であれば 1 に、そうでなければ 0 とカテゴリ化した。また、項目の内容に関して全く記述されていなかった論文に関しては「不明」とした。

受験者の年齢

受験者の年齢については、先行研究によってはその平均値が報告されている場合と、「大学生の受験者である」などといった受験者の属性のみ報告されている場合があった。そのため、わが国における学校の進学の仕組みにあわせてコード化した。すなわち、小学生(年齢の平均値だけ報告されている場合はその値が6歳以上12歳未満の場合)=1、中学生(12歳以上15歳未満の場合)=2、高校生(15歳以上18歳未満の場合)=3、大学生(18歳以上22歳未満の場合)=4、大学院生(22歳以上)=5のようにした。また、著者自身が文章を作成している場合(Chase, 1968, 1979)は5とした。

採点者のスキル

前述のように、採点者は学校や塾の比較的採点の経験が長い教師か、経験が比較的短い大学院生・大学生の2通りがある。本研究におけるメタ分析では、採点者を採点の経験が短い大学院生・大学生か否かを基準にして2値でコード化（大学院生・大学生の採点者は1でそれ以外は0）した。

データベースの作成

ここで、各先行研究における研究者名と、項目内容・標本サイズ(N)・効果量(Z_{fisher})・解答内容の自由度・受験者の年齢・採点者のスキル、のコード化の結果についてまとめたものを表1に示す。 Z_{fisher} とは標本相關係数 r に対するFisherのZ変換値を指す。すなわち、

$$Z_{fisher} = \frac{1}{2} \log \frac{1+r}{1-r} \quad (1)$$

である。

2.3 分析

①変数間の相関関係

表1の結果から諸変数間の相関係数を算出した表が表2である。これによると受験者の年齢が効果量と関連が比較的深いといえる。ただしこの相関係数自体を要因の強さの直接的な指標として利用するには様々な問題があることが知られている（例えばCooper, H.M. 1984）ので、参考程度にとどめるべき指標である。

また、本来、重回帰分析などによってそれぞれの変数の独自の効果を検討すべきところであるが、1件あたりの被験者数が異なるために効果量の等分散性の仮定が満たされない場合が多い。そのため、メタ分析で

表2 相関係数

	Z_{fisher}	自由度	年齢	スキル
Z_{fisher}	1			
自由度	-0.20	1		
年齢	-0.37	0.86*	1	
スキル	-0.04	-0.09	0.36	1

*は5%水準で有意。

（「スキル」は採点者のスキル、「自由度」は解答内容の自由度、「年齢」は受験者の年齢をそれぞれ指す。）

効果量を分析する際に分散分析や回帰分析を利用することは研究結果に重大な歪みが生じる恐れのあることが指摘されている（Hedges & Olkin, 1985）。さらに本研究ではソースとなる論文の件数が少ないとこと、独立変数間に高い相関があることによる多重共線性の影響も考慮して、本研究では以下に述べる、メタ分析を実行する上で奨励されている指標（Mullen, 2000）によって各独立変数の効果を検討する。

②Zによる分析

Mullen(2000)に示されているように、メタ分析において各独立変数が従属変数に及ぼす影響を1変数ずつ検討していく場合に、帰無仮説のもとで標準正規分布に従う以下の統計量Zが利用できる。

$$Z = \frac{\sum \lambda_j Z_{fisher}}{\sqrt{\sum \frac{\lambda_j^2}{N_j - 3}}} \quad (2)$$

ここで、 N_j は研究jにおける標本サイズであり、 λ_j は研究jの対比重みづけである。 λ_j は効果の検定をする独立変数の値が、全ての研究間における平均値が0になるように平均からの偏差を算出すること求められる。

表1の最下行には、各独立変数に対するZの値の計算結果を記してある。解答内容の自由度と採点者のスキルについては統計的に有意ではなかったが、受験者の年齢に関しては $Z = -1.86(p < .10)$ であり、有意傾向であった。また受験者の年齢のコード化において、著者自身が文章を作成している場合(Chase, 1968, 1979)を不明として分析から除外した場合でも同様の傾向がみられた。

2.4 考察

本研究のメタ分析では、解答内容の自由度と採点者のスキルの要因についてはその効果は有意でなかった。しかし、これらの要因は2値カテゴリであり、また要因の評定自体の信頼性の問題も考慮すると、元来有意になりにくい測度であった可能性がある。また、ソースとなった論文の件数が少なかったことからも、メタ分析の結果について過度な一般化は避けるべきであろう。特に表2において自由度は年齢と高い相関を持っていたことからも、自由度については今後さらに検討する必要があるだろう。

一方、受験者の年齢の要因についてはその効果が有意であり、受験者の年齢が何らかの形で文字の美醜効果に影響をしていることが示唆された。これは、これまでの先行研究において指摘されてこなかった点であり、今後文字の美醜が採点に及ぼす影響を考慮する上で1つの有用な観点になると思われる。

しかし、上述したように、受験者の年齢の要因についてはそれが直接的に文字の美醜効果に影響を及ぼしているというよりは、受験者の年齢の要因が他の何らかの要因を介して影響を与えていていると考えられる。

この、他の要因としては、まず前述した文章の質の要因が挙げられる。具体的には、採点する答案の質が高いほど、答案と評価基準との一致度を意識的に検討して採点する傾向が強まり、評価に無関係な文字の美醜が影響しにくくなるというプロセスが考えられる。もしこのようなプロセスが存在するのであれば、質の良い文章は正当に評価されやすい一方で、質の悪い文章は本来の評価に無関連な文字の美醜に影響されやすいといった、文字の美醜と文章の質の交互作用効果が生じることが予想される。

また、採点者が大人の書く文章に対しては文章の質に基づいて採点する一方で、子どもの書く文章に対しては単に文章の質だけでなく文字の美醜にも注意を払うといった採点者の構え、より広く言えばそのような採点者を取り巻く教育現場の環境要因も、受験者の年齢の要因の実質的な内容として考えられる。実際、文字を綺麗に書くための子ども向けの通信添削指導や習字の習い事など、文字を美しく書くこと自体が肯定的に評価されるケースが子どもには多い。

この他、受験者の年齢が低い場合には、その同年齢内での文字の美醜の程度の差が大きいために、受験者の年齢が高い場合に比べて文字の美醜効果が現れやすいという可能性もある。さらには、いま述べたことと

も関係するが、受験者の年齢が低い場合には、文字の質の悪さが判読の困難を生じるようなレベルのものもあって、そのために評価が下がるという可能性もある。

本研究では次に、これらの可能性のうち、文章の質の要因に注目して、文章の質の要因と文字の美醜効果の関連を実験的に検討することとする。具体的には、文章の質に応じて採点結果に及ぼす文字の美醜の効果が異なるか否か、すなわち文字の美醜と文章の質との間の交互作用効果を検討する。この点については、過去にMarkham(1976)が文字の美醜と文章の質の要因の交互作用効果を検討しているが、そこでは交互作用効果は見出されていない。ただしMarkham(1976)と後述する本研究における実験では受験者の属性など多くの異なる点があり、Markham (1976)とは異なる結果が得られる可能性がある。本研究の結果と、Markham(1976)の結果を合わせて文章の質と文字の美醜効果の関連について再検討する。

3 実験による検討

3.1 概要

本節では、メタ分析において文字の美醜効果との有意な関連が示唆された受験者の年齢の要因について、より直接的には文章の質の要因が文字の美醜効果を規定しているという仮説を検証するために実験を行う。そのためには、文章の質と文字の美醜の2要因の組み合わせを反映した材料を用意する目的で予備実験を行う。そして本実験では、文章の採点結果を従属変数とし、文字の美醜と文章の質を独立変数とした2要因分散分析を行い、交互作用効果を検討する。

また、一般に学年が異なると文章の評価の観点や項目の困難度が大きく異なるために、異なる年齢集団に実験を行うことが困難であることを考慮し、本研究では被験者を大学生に限定して実験を行う。

3.2 予備実験

3.2.1 基礎的な材料の収集（予備実験1）

受験者

国立N大学の学生20名。(男性10名、女性10名、平均年齢20.8歳)

材料

国立T大学の入学試験として数年前に実施された項目である、“「遊び」とは何かについて、自分の考えるところを400字以内で分かりやすく定義せよ。”を用

いた。この項目を選択した理由は、①実際の生活場面に根ざしたテーマであり、誰にでも比較的書きやすく、採点者にも身近な話題であること、②また「定義せよ」と具体的な指示を与えてるので受験者は答案の作成において何を求められているのかが判断しやすく、採点者にとって予想外の答案がくる可能性が低くなる、の2点を考慮したためである(Christian, Timothy, Richard,& Bud 2002 を参考)。

手続き

大学内の実験室において、一度に4～5名単位で一斉に試験を行った。採点は内容の質に基づいて評価されることをあらかじめ各受験者に伝えておいた。制限時間は40分であった。また下書き用の原稿用紙も1枚ずつ配布した。テストの回答が終わった者から、統計学について書かれた文章(200字程度で漢字・仮名のみが用いられている)を普段書いているような書き方で別途配布した原稿用紙に模写するよう教示した。これは、後に文字の美醜要因の操作に用いるためである。実施には教示(試験の目的など)を含めて1時間程度かかった。得られた小論文はこの後の内容の質に関する採点で文字の美醜の影響を防ぐためにすべてパソコンでタイピングした。

3.2.2 文章の質の要因の操作の準備（予備実験2）

採点者

A県で勤務する塾講師5名(国語科担当3名、数学科1名、英語科1名)、大学教員3名(言語系専攻2名、理工系専攻1名)の計8名(このうち男性5名、女性3名であり平均年齢42.3歳)である。塾講師の勤続年数は平均13年である。

手続き

予備実験1で収集・タイピングされ、順番はランダムに配置された各20枚の小論文とフェイスシートを綴じた冊子を8名の採点者に渡し、それぞれの小論文を包括的評価法(7点満点)で採点してもらった。フェイスシートには「本研究の目的は大学生の文章作成能力の変化を検討すること、採点を3時間以内に終了すること、内容の質にもとづいて採点すること、最初にすべての文章を読んでから採点すること」と記しておいた。7点満点を用いたのは5カテゴリ程度でほぼ100点満点と同様の信頼性が得られるといった知見による(平井・渡部1994)。そして採点後に感想などを自由記述により書いてもらった。

文章の質の高低を表す小論文の選択

まず、20枚の小論文ごとに8名の採点者の平均点を求めた。その平均点の平均値は4.30点で標準偏差は0.52点であった。この平均点の最上位と最下位それぞれ2名の文章を本実験で用いることにした(下位に関しては得点が同一の者が2名いたが、採点者間の標準偏差の小さいほうを選択した)。抽出された4名分の文章には内容に関する特別な問題点や、誤字・脱字は見受けられなかった。

なお、20枚の小論文について、採点者間の得点の相関係数は-0.32～0.39(中央値-0.06)であった。内容の質に基づくよう教示した答案にもかかわらず、採点者間でかなり評価基準が異なることがわかる。これは渡部ら(1988)でもほぼ同様の結果が示されている。

3.2.3 文字の美醜の要因の操作と準備（予備実験3）

採点者

予備実験2とは異なる17名(大学生から社会人まで、年齢は19～52歳で、平均30歳)。

手続き

予備実験1で20名の受験者に模写してもらった文章のコピーをランダムな順番で配置し、それらを綴じた冊子を採点者にそれぞれ配布した。A4シートに3名分の文章が並ぶようにし、あわせて7枚、そしてフェイスシート1枚の計8枚からなる冊子であった。それぞれの模写文章のコピーの大きさや濃さは統一して使用した。そして、文字の美醜をそれぞれ5件法で評定してもらった。

文字の美醜の高低を表す書き手の選択

まず20枚の文章ごとに17名の採点者による美醜得点の平均点を求めた。その平均点の平均値は3.20点で標準偏差は0.64点であった。この平均点の最上位、最下位それぞれ2名が本実験で用いる文章の書き手として選択された。選択された4名分の文字に関して、書いてある文字が全く判別できない、極端に文字の太さが違う、などの違いは見られなかった。なお、17名の採点者間の得点の相関係数は0.00～0.83(中央値0.44)であった。

3.3 本実験

3.3.1 方法

材料

予備実験3において選択された文字の書き手の4名に依頼し、予備実験2で選ばれた4枚の小論文を

0.5mmの同じ濃さのシャープペンで400字(20×20)の原稿用紙に書いてもらった。予備実験1で書いてもらった模写文章と一緒に配布し、「この文字のように、普段書いているような字体で模写する」旨を依頼した。

この手続きによって4種類の文章×4種類の文字、の合計16種類の小論文が作成された。そして、A：文章の質・高&文字・高、B：文章の質・高&文字・低、C：文章の質・低&文字・高、D：文章の質・低&文字・低、からなる4種類の小論文を1名の採点者が採点するようにした。したがっていずれの要因も被験者内要因となる。ここでは1名の採点者が、同一の文章や文字を重複して採点するがないように工夫している。そして、4種類の小論文(計4枚)と、1枚のフェイスシートからなる計5枚の冊子が20部作成された。

採点者

予備実験2、3とは異なる塾講師4名(国語科3名英語科1名)、高校国語教師4名、大学教師4名(全員、言語系専攻)と、経験2年以上の塾のアルバイト講師の大学生8名(全員国語科担当経験者)の計20名。

アルバイト講師の大学生は男性3名、女性5名、平均年齢21.8歳。それを除く12名は男性4名、女性8名、平均年齢40.8歳である。この中で高校国語教師・塾講師8名の平均勤続年数は20年である。

手続き

採点者は5枚からなる冊子をそれぞれ受け取り、7件法の包括的評価法で採点した。フェイスシートには「本研究の目的是大学生の文章作成能力の変化を検討すること、採点を40分以内に終了すること、内容の質にもとづいて採点すること、最初にすべての文章を読んでから採点すること」を記しておいた。そして採点後に感想などを自由記述により書いてもらった。

3.3.2 結果

本実験で得られたデータに欠測値は含まれていなかつたので20名全員のデータを使用した。A～Dの4つの群間で評定得点の分布の歪み方に大きな違いは見られなかった。そして、文章の質の得点(高×低)と文字の美醜の得点(高×低)を被験者内要因とした2要因の分散分析を実施したところ、文字の美醜要因の主効果は有意でなかった($F(1,19)=0.01$, n.s.)。また、内容の質の主効果に関しては有意であった($F(1,19)=39.59$, $p<.01$)。そして、交互作用効果については有

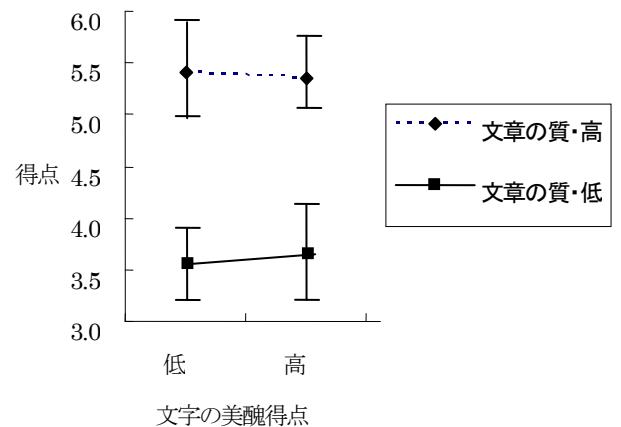


図.1 各水準における平均と標準誤差

意でなかった($F(1,19)=0.11$, n.s.)。各水準における平均と標準偏差を図.1に示す。

さらに、採点者をスキル別(大学生か否かで分け、大学教師・塾講師・高校教師は同等に扱う)でさらに群分けし、スキルを被験者間要因、文字の美醜と文章の質を被験者内要因とする3要因分散分析をした。その結果、スキルと文字の美醜の間に交互作用はみられなかった($F(1,18)=0.99$, n.s.)。

3.4 考察

実験による検討の結果、Markham(1976)の場合と同様に、文字の美醜と文章の質の要因間の交互作用効果が示されなかった。このことから、受験者の年齢が文章の質を介して文字の美醜効果を規定しているのではない可能性が示唆された。したがって被験者の年齢の実質的な要因として他の可能性を検討する余地があるだろう。これには2節で述べたような採点者の構えや、異なる年齢集団における文字の美醜の質的な差異などが考えられる。

ただし、メタ分析で用いられた年齢の幅は小学生から大学院生と幅広い一方で、実験においては大学生に書かせたものであり、質が低いという条件であっても、低年齢者の質の低さと同等のものとは言えない。そのため、そのことが実験の結果に影響を与えた可能性は否定できない。

しかし、採点者の感想では、採点した文章に対して「中学生の書いた文章だと思った.」、「大学生の文章とは思えない.」といった感想をもった採点者が4名(大学生・国語教師それぞれ2名)いたことから、文章の

質が低群の文章は少なくとも中・高校生程度の文章の代表性はあったと思われる。そのため、受験者が大学生に限定されない場合においても、受験者の年齢が文章の質を介して文字の美醜効果を規定している可能性は低いと考えられる。この点については、より経験的な結果を蓄積していく必要があるだろう。

なお、採点者を職業別に分けた分散分析の結果は、採点者のスキルは文字の美醜効果を説明する要因とはならないことを示しており、メタ分析で得られた結果を再確認するものであると言えよう。

4 総合考察

本研究は、小論文試験の採点の公平性に影響を与えるとして問題視されている、文字の美醜効果を規定する要因を検討することを目的に行われた。以下、文字の美醜効果を規定する要因として考えられた、「解答内容の自由度」、「受験者の年齢」、「採点者のスキル」の3点について個別に考察する。

まず、解答内容の自由度についてはメタ分析の結果では有意でなかったが、有意傾向が見られた受験者の年齢と高い相関が見られた。そのため、今後さらに実験的な結果を蓄積していくことにより、解答内容の自由度が文字の美醜効果に及ぼす影響を検討していく必要があるだろう。また、今後メタ分析や実験により解答内容の自由度の効果を検討していくためには、解答内容の自由度の測度をより具体的に検討し、さらに複数人による評定を行うなど測定の信頼性に対する配慮が必要であろう。

受験者の年齢については、メタ分析の結果では有意な傾向が見られた。そして受験者の年齢は、受験者の作成する文章の質を介して文字の美醜効果に影響すると仮定して、実験的な小論文試験の採点を行った。その結果、文字の美醜と文章の質の交互作用効果は有意でなく、受験者の年齢は文章の質の要因を介して文字の美醜効果に影響を与えていた可能性は低いことが示唆された。ただし上で述べたように、この結果については、本実験で用いた質・低群の文章が低年齢者の文章と同様の役割を果たしていたかについては考慮する必要がある。それでもなお、本実験ではある程度の年齢差を反映するぐらいの文章の質の差は実現できていたと考えられるため、今後は被験者の年齢が、文章の質の要因だけでなく、採点者の構えや、異なる年齢集団における文字の美醜の質的な差異を介して文字の美

醜効果に影響を及ぼす可能性についても積極的に検討していくべきであろう。

採点者のスキルについては、メタ分析と実験のいずれにおいても、採点者のスキルと文字の美醜効果の直接的な関連性は見られなかった。しかし、いずれにおいても学生か否かという2値カテゴリを採点者のスキルの測度としていたために、採点者のスキルの効果を検討するうえでは限界があったと思われる。小論文の採点作業の慣れ、自信、訓練歴の長さ、等を反映する測度の利用も検討していく必要がある。

そして、今後いずれの要因の効果を実験的に検討していく場合にも言えることであるが、包括的評価法だけでなく分析的評価法もを利用して、個別の観点から文字の美醜効果との関連を検討していく方法も検討していくことが有効であろう。

本研究で行われたメタ分析と実験だけでは文字の美醜効果の規定因を完全に検討する上で決して十分ではないが、本研究の結果により少なくとも①文字の美醜効果に受験者の年齢が何らかの形で影響を与えており、年齢が高いほど文字の美醜効果の影響は弱いこと、②受験者の年齢は文章の質の要因を介して文字の美醜効果に影響を与えていた可能性は低いこと、の2点が指摘できると思われる。小論文試験の採点は、実験的な統制を加えても様々な外的誤差要因に敏感があるので、より安定した結果を得るために、今後より様々な観点から経験的な結果を蓄積する必要がある。その際、この2点を考慮していくことが、文字の美醜効果の規定因を検討する上で有用であろう。

本研究で検討した文字の美醜効果は、小論文試験の作成・評価法における問題の1つに過ぎないが、小論文試験が広く普及してきた昨今、小論文試験の作成・評価法への地道な理論的構築が必要になってくると思われる。文字の美醜効果に関しては、その影響度は未知数であり、本研究の結果を踏まえてより詳細な検討が望まれる。

5 引用文献

- Briggs, D. (1970). Influence of handwriting on assessment. *Educational Research*, 13, 50-55.
Bull, R. & Stevens, J. (1979). The effects of attractiveness of writer and penmanship on

- essay grades. *Journal of Occupational Psychology*, 52, 53-59.
- Chase, C. I. (1968). The impact of some obvious variables on essay test scores. *Journal of Educational Measurement*, 5, 315-318.
- Chase, C. I. (1979). The impact of achievement expectations and handwriting quality on scoring essay test. *Journal of Educational Measurement*, 16, 39-42.
- Chase, C. I. (1983). Essay test scores and reading difficulty. *Journal of Educational Measurement*, 20, 293-297.
- Chase, C. I. (1986). Essay test scoring : Interaction of relevant variables. *Journal of Educational Measurement*, 23, 33-41.
- Christian, M. R., Timothy, W. B., Richard, R. S., & Bud, W. (2002). How to prepare effective essay questions, BYU Faculty Center, *Brigham Young University Testing Services*.
- Cooper, H.M. (1984). The integrative research review: *A social science approach*. Beverly Hills, CA: Sage.
- Cooper, P. L. (1984). The assessment of writing ability: A review of research No.82-15R ETS Research Report 84-12
- Eames, K., & Loewenthal, K. (1990). Effects of handwriting and examinee expertise on assessment of essays. *The Journal of Social Psychology*, 130, 831-833.
- 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, 21-31.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*, Orlando, FL: Academic Press.
- Hughes, D. C., Keeling, B. F., & Tuck, B. F. (1980). The influence of context position and scoring method on essay scoring. *Journal of Educational Measurement*, 17, 131-135.
- Hughes, D. C., Keeling, B. F., & Tuck, B. F. (1983a). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*, 20, 65-70.
- Hughes, D. C., Keeling, B. F., & Tuck, B. F. (1983b). The effects of instructions to scorers intended to reduce context effects in essay scoring. *Educational and Psychological Measurement*, 43, 1047-1050.
- James, H. (1927). The effect of handwriting on grading. *English Journal*, 16, 180-205.
- Lipsey, M. W. & Wilson, D. B. (2000). Practical meta-analysis. Sage.
- Markham, L. F. (1976). Influence of handwriting on teacher evaluation of written work. *American Educational Research Journal*, 3, 277-283.
- Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors and essay grades. *Journal of Educational Measurement*, 6, 97-101.
- 文部科学省(2005). 平成18年度国公立大学入学者選抜の概要,
http://www.mext.go.jp/b_menu/houdou/17/08/05083001.htm
- Mullen, B. (2000). 小野寺孝義 訳 基礎から学ぶメタ分析 ナカニシヤ出版
- 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
- Soloff, S. (1973). Effect of non-content factor on the grading of essays. *Graduate Research in Education and Related Disciplines*, 6, 44-54.
- 平直樹・江上由実子 (1992). ESSAY TEST の方法論的諸問題に関する研究の動向について 教育心理学研究 40(1) 108-117.
- 渡部洋 (1994). 小論文試験の特徴とその利用法について 学校教育研究所年報 38, 48-59.
- 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, 143-164
- 渡部洋・曹亦薇 (1992). 小論文評価における字の美しさの影響について 東京大学教育学部紀要 32, 253-256.
- 渡部洋(編) (2002). 心理統計の技法 福村出版.
- 吉村英 (1991). ワープロ文字と手書き文字の違いが文章の内容の評価に与える影響(I)日本心理学会第55回大会発表論文集 753
- 吉村英 (1992). ワープロ文字と手書き文字の違いが文章の内容の評価に与える影響(II)日本心理学会第56回大会発表論文集 539

謝 辞

本論文は 2006 年度に名古屋大学教育学部に提出された卒業論文の一部を加筆修正したものです。本論文を作成するにあたり、御指導くださった名古屋大学教

育学研究科教授の野口裕之先生と東京大学教育学研究科教授の南風原朝和先生に心から御礼申し上げます。

そして、実験に協力していただいた名古屋大学教育学部の皆様にも、この場を借りて御礼申し上げます。