

小論文課題の複数人による採点の基礎的な分析 — 採点者による得点の違いについて —

荒井清佳, 石岡恒憲 (大学入試センター研究開発部)

課題及び制限字数の異なる小論文試験を行い、3名の採点者がどのように採点をしたのかについて、基礎的な分析を行った。採点は、11の観点による分析的評価と総合的評価を行った。採点に当たっては事前に得点の一致を図るために得点基準の調整を行った。その結果、分析的評価については、3名の得点分布は互いに異なるものもあったが、総合的評価については似たような形の分布になった。また、得点分布は、課題の違いよりも採点者の違いによる方が違いが大きかった。

1 はじめに

日本の大学入学者選抜をみると、平成27年度には国公立大学の約74%が小論文試験を実施している(文部科学省, 2014)。しかし小論文試験には、客観式テストとは異なり、評定者の主観に基づいて採点せざるをえない部分があることから、妥当性や評価の信頼性など測定論的問題があることが古くから指摘されている(渡部ら, 1988; 村上, 2005; 宇佐美, 2008, 2011, 2013)。こうした問題点のうち、課題内容や課題数、制限字数、回答時間などの、小論文試験の作成・実施段階に焦点を当てた研究は、試験を実施する上で必要な観点であるにも関わらず、数が少ない。課題数については、単一の課題よりも複数の課題を課す方が評価の信頼性が高くなるという結果が示されている(大久保, 2013; 宇佐美, 2013)。制限字数や回答時間の違いは、産出される小論文の内容や構成に影響を与える可能性がある。

そこで、制限字数と回答時間及び課題数を変えて小論文試験を実施し、それらの違いが小論文試験の測定に与える影響を検討することを目的に実験を行った。小論文の採点については、3名の経験者に依頼した。本研究は、3名の採点者の採点結果に焦点を当て、採点者一人ひとりの採点がどうであったかの基礎

的な分析を行う。

2 方法

本稿で扱う小論文実験は、平成26年度大学入試センター試験モニター調査内で行われた。実験参加者は都内五つの大学の大学1年生330名、実験日は平成26年1月26日であった。

2.1 小論文実験について

小論文実験は制限字数・回答時間・課題数を変えて実施した。制限字数(少/多)×回答時間(短/長)の4群に実験参加者を分け、回答時間の短い2群にのみ課題を二つ課した。各群の人数及び割り当てを表1に示す。第1群から第4群は、参加者における所属大学及び文理の別の構成比をほぼ同じにしている。

群	人数	課題(字数)・時間
第1群	90	小論文 A (300字)・30分
		小論文 B (400字)・30分
第2群	94	小論文 A (600字)・30分
		小論文 B (800字)・30分
第3群	73	小論文 A (300字)・60分
第4群	73	小論文 A (600字)・60分

回答時間は短い場合は30分とし、第1群と

第2群に割り当てた。長い場合は60分とし、第3群と第4群に割り当てた。

制限字数については、第1群と第3群は少ない方の課題を、第2群と第4群は多い方の課題を割り当てた。課題の一つ目(小論文A)は、短い文章を読ませ、その文章の要約(問1)とその文章に対する自分の意見を書かせる(問2)課題である。制限字数は、問1は80~120字で全群で共通、問2は少ない場合は300字、多い場合は600字とした。課題の二つ目(小論文B)は、課題文の無い、比較的短い問題文を読んで自分の意見を書かせる課題である。制限字数は、少ない場合は400字、多い場合は800字とした。

実際に用いた課題は、荒井ら(2015)が用いた小論文課題と同じ物であり、小論文Bの課題は、宇佐美(2011)で実施された課題Aと同じ趣旨の課題である。

2.2 採点について

小論文の答案(小論文Aの問2及び小論文B)は、課題、回答時間、制限字数を区別すると6種類に分かれる。この6種類からそれぞれ70枚の答案を抜き出し、一つの答案を3名の採点者が独立に採点することとした。すなわち、答案は全部で $70 \times 6 = 420$ 枚であり、採点者が採点する答案枚数も1名当たり420枚であった。

採点基準

宇佐美(2011)の評価観点にならい、11個の分析的評価観点を¹⁾用い、それぞれ1点~5点で採点した。また、総合的評価も行い、0点~6点で採点した。分析的評価、総合的評価ともに、個々の評価点の基準を記した採点基準を準備した。

採点者

3名の採点者はいずれも「小論文の採点や指導を行ったことがある」ことを要件として、

人材派遣会社から紹介された者である。3名とも、予備校等で小論文の指導や添削の経験があった。

採点手順

まず、実験の概要と評価観点と個々の評価点の基準について説明を行った。続いて、小論文の課題と字数ごとに次の1)~5)の手順に従って採点を行うこととした。採点の順番は、「小論文A(問2)600字」→「小論文B・800字」→「小論文A(問2)300字」→「小論文B・400字」とした。採点者には答案をコピーしたものを渡したが、その際、答案の順番が採点者ごとに異なるように並べ替えた。

手順1) 全ての答案に軽く目を通す。

手順2) 別に用意した数枚の調整用答案について、分析的評価と総合的評価を行う。

手順3) 調整用答案の評価点が採点者間で一致するように採点者間で意見を出し合い、調整する。

手順4) 調整後は、採点者は独立して採点を行う。まず、分析的評価を行う。

手順5) 続いて総合的評価を行う。その際、分析的評価の得点は参考にしない。

2.3 本稿での分析対象

本研究では、いずれも回答時間が30分である第1群の小論文A、B、第2群の小論文A、Bの4種類を分析対象とする。以下、それぞれ「A・300字」「B・400字」「A・600字」「B・800字」と呼ぶ。

3 結果

評価観点ごとの得点分布、平均点等を採点者別に示す。

3.1 評価観点ごとの得点分布

分析的評価及び総合評価の得点分布を採点者別に示したものが図1である。折れ線の線種で課題の違いを(実線が小論文A、破線が

小論文 B) , 記号で制限字数の違いを (○が 300 字あるいは 400 字, △が 600 字あるいは 800 字) 表している。

図 1(1)~(11) は, 分析的評価の各評価観点ごとの得点分布である。この図より, 「1. 語句の表記」「2. 文法」「3. 語彙力」では, ほとんどが 4 点か 5 点であり, 得点の差がほとんどないことが分かる。その他の観点では 2 点~5 点にばらついているが, 1 点の答えはほとんどなかった。また得点分布の形は, 課題や制限字数の違いよりも, 採点者に違いによる違いの方が大きいことも分かる。

図 1(12) は, 総合的評価の得点分布である。総合的評価は, 3 点あるいは 4 点の答えが多く, 得点分布は山型であった。

3.2 各評価の採点者間の一致率

表 2 は, 分析的評価と総合的評価の採点者間の一致率をまとめたものである。二名の採点者がつけた点数が 1 点差以内であった答案の割合を示し, また同じ点数であった答案の割合を () 内に示した。「1. 語句の表記」「2. 文法」「3. 語彙力」については, どの採点者間の一致率も 0.97 以上と高かった。また「11. 原稿用紙」については, どの採点者間の一致率も 0.91 以上であった。それ以外の評価については 0.8~0.9 程度が多かった。

3.3 分析的評価の各観点の平均点

図 2 は, 課題・制限字数ごとに, 各採点者の分析的評価の各観点の平均点を折れ線で示したものである。図 2 より, いずれにおいても, 「1. 語句の表記」「2. 文法」「3. 語彙力」「11. 原稿用紙」の平均点が他の観点の平均点よりも高いことが分かる。また, 採点者 1 と 2 は折れ線が似通っているが, 採点者 3 は少し異なり, ほとんどの観点で高めの平均点であることが分かる。

3.4 分析的評価と総合的評価の相関係数

表 3 は, 分析的評価と総合的評価の相関係数を採点者別にまとめたものである。「5. 主張の明確性」「6. 一貫性」「7. 説得力」「8. 構成」の観点では, 総合的評価との相関係数が採点者の 3 名ともに 0.5 以上であった。

4 考察

本研究では, 3 名の採点者の採点がどのようなであったのかについての基礎的な分析を行った。

分析的評価のうち, 「1. 語句の表記」「2. 文法」「3. 語彙力」は, 多くの答案が 4 点か 5 点であった。これらの三つの観点については, 答案の多くが実際によく書けていたのだと考えられ, 採点者間の得点の一致率 (表 2) がいずれも高いのは, このためだと考えられる。しかし, 得点分布を見ると, 「3. 語彙力」については 3 名ともほとんどが 4 点であり, 3 名とも同じような形状になったが, 「1. 語句の表記」「2. 文法」については, 採点者 1 と採点者 3 が 5 点を多く付けたのに対し, 採点者 2 は 4 点を多く付けており, 異なる形状となった。

その他の分析的評価については「11. 原稿用紙」を除いて, 得点分布は 2~4 点が最も高い山型が多く, 採点者間で得点分布が似ている観点もあれば, 似ていない観点もあった。例えば「4. 課題の解釈」では, 採点者 1 と採点者 2 の間の一致率は 0.88 と高いが, 採点者 1 と採点者 3 の間の一致率は 0.74 と表 2 の中で最も低い。「B・400 字」(破線○) が他の種類に比べて得点が高いという点では 3 名の採点者間で同じなのだが, 採点者 1・採点者 2 が 3 点と 4 点を多く付けているのに対し, 採点者 3 は 4 点と 5 点を多く付けており, その違いが得点分布の形の違いや一致率の低さを招いている。このような傾向は「5. 主張の明確性」「6. 一貫性」でも見られる。

また, 分析的評価の全体を眺めてみると, 採点者 2 は 5 点をほとんど付けないのに対し,

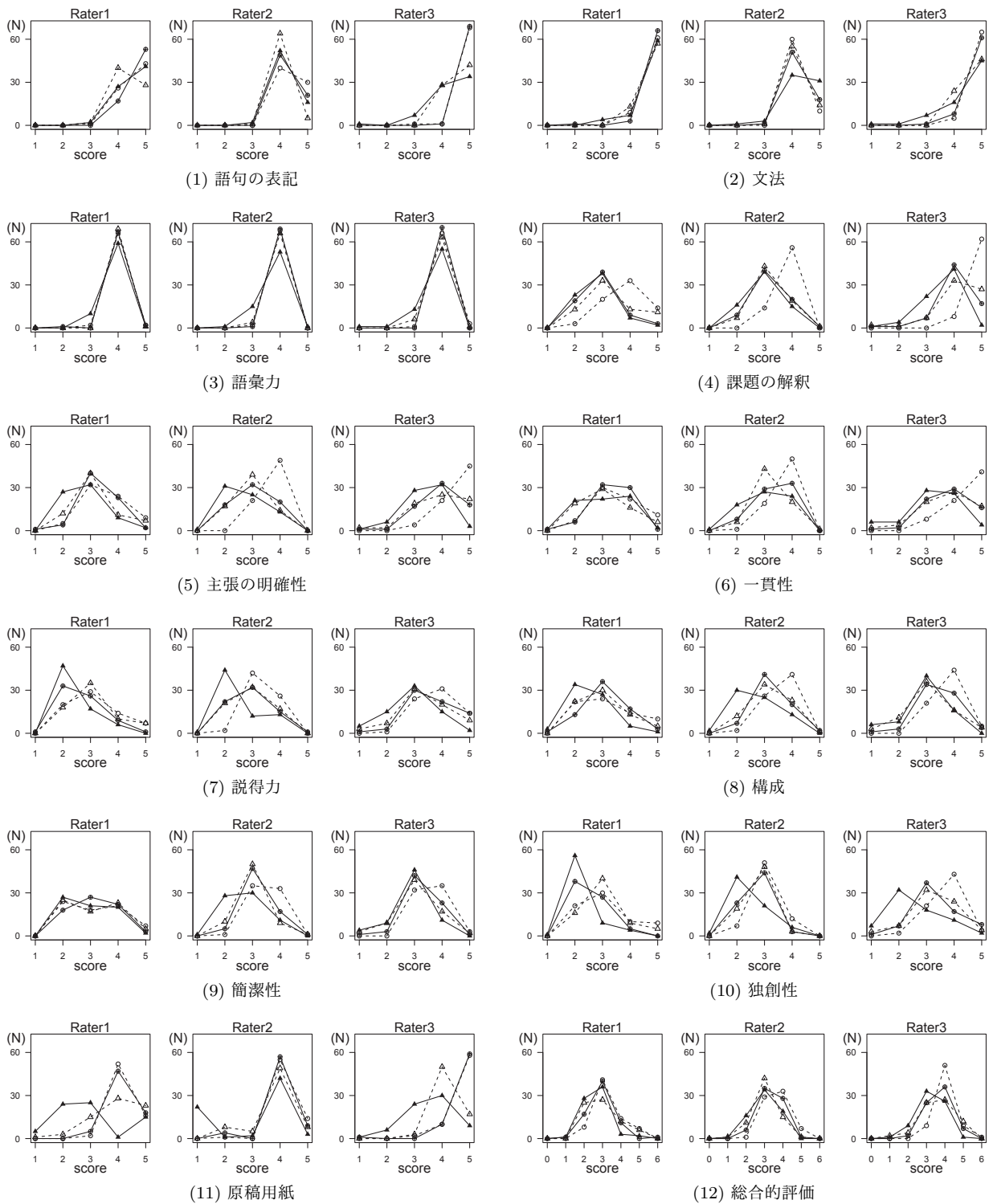


図 1: 評価観点ごとの得点分布 (1)~(11): 分析的評価, (12): 総合的評価
(実線: 小論文 A, 破線: 小論文 B, ○: 300 字/400 字, △: 600 字/800 字)

表 2: 分析的評価と総合的評価の採点者間の得点の一致率

観点	採点者 1 と 2	採点者 1 と 3	採点者 2 と 3
1. 語句の表記	0.99 (0.51)	0.98 (0.61)	1.00 (0.39)
2. 文法	0.98 (0.36)	0.97 (0.72)	0.98 (0.36)
3. 語彙力	1.00 (0.88)	0.99 (0.84)	1.00 (0.88)
4. 課題の解釈	0.88 (0.45)	0.74 (0.23)	0.80 (0.24)
5. 主張の明確性	0.92 (0.47)	0.80 (0.33)	0.77 (0.27)
6. 一貫性	0.93 (0.44)	0.81 (0.36)	0.85 (0.34)
7. 説得力	0.87 (0.47)	0.81 (0.30)	0.83 (0.35)
8. 構成	0.89 (0.41)	0.85 (0.37)	0.93 (0.43)
9. 簡潔性	0.83 (0.34)	0.87 (0.30)	0.96 (0.53)
10. 独創性	0.90 (0.48)	0.81 (0.36)	0.86 (0.35)
11. 原稿用紙	0.97 (0.56)	0.97 (0.38)	0.91 (0.39)
12. 総合的評価	0.89 (0.42)	0.85 (0.34)	0.93 (0.49)

2名による得点の差が1点以内であった答案の割合。()内は2名による得点が同じであった答案の割合。

表 3: 分析的評価と総合的評価の相関係数

	1. 語句の表記	2. 文法	3. 語彙力	4. 課題の解釈	5. 主張の明確性	6. 一貫性
採点者 1	0.09	0.13	0.27	<u>0.65</u>	<u>0.62</u>	<u>0.51</u>
採点者 2	0.25	0.16	0.30	0.47	<u>0.55</u>	<u>0.50</u>
採点者 3	0.32	0.31	0.26	<u>0.58</u>	<u>0.64</u>	<u>0.62</u>
	7. 説得力	8. 構成	9. 簡潔性	10. 独創性	11. 原稿用紙	
採点者 1	<u>0.61</u>	<u>0.55</u>	0.44	<u>0.53</u>	0.30	
採点者 2	<u>0.56</u>	<u>0.55</u>	<u>0.59</u>	0.40	<u>0.52</u>	
採点者 3	<u>0.67</u>	<u>0.61</u>	<u>0.59</u>	<u>0.64</u>	0.46	

下線は相関係数の値が0.5以上であることを示す。

採点者 3 は比較的 5 点を付けやすいという違いがある。この違いは各観点の平均点 (図 2) にも表れており、採点者 1 と 2 の折れ線はほとんど重なっているのに対し、採点者 3 は、重なっている観点もあるが、平均点が高くなっている観点がある。

3 名の採点者は、初対面同士であったが、いずれも小論文の指導・添削の経験者である。また採点の前に、調整用答案を用いて得点の調整を図った。にもかかわらず、3 名の採点者の分析的評価観点ごとの得点分布は互いに異なるものが多く、小論文の種類 (課題や制限字数の違い) よりも、採点者の癖のようなものが現れていた。

一方で、総合的評価では、3 名とも 3 点あるいは 4 点が最も高い山型となり、分析的評価

の得点分布に比べると、3 名とも似たような形となった。総合的評価では、中程度 (3 点, 4 点) の人数を最も多くし、0 点や満点は少なくするという意識が働いたのかもしれない。総合的評価と分析的評価の関係をみると (表 3)、いずれも小論文の本質に近いと考えられる観点 (5. 主張の明確性~8. 構成) での相関係数が高く、採点が適切に行われたことを示唆すると考えられる。

本研究の結果、分析的評価観点ごとの得点分布では、課題や制限字数の違いよりも採点者の違いによる方が分布の違いが大きかった。事前の簡単な調整だけでは、採点者の採点の癖のようなものは取り除くのは難しいことが示された。

大規模試験などで、採点者が全員の答案を

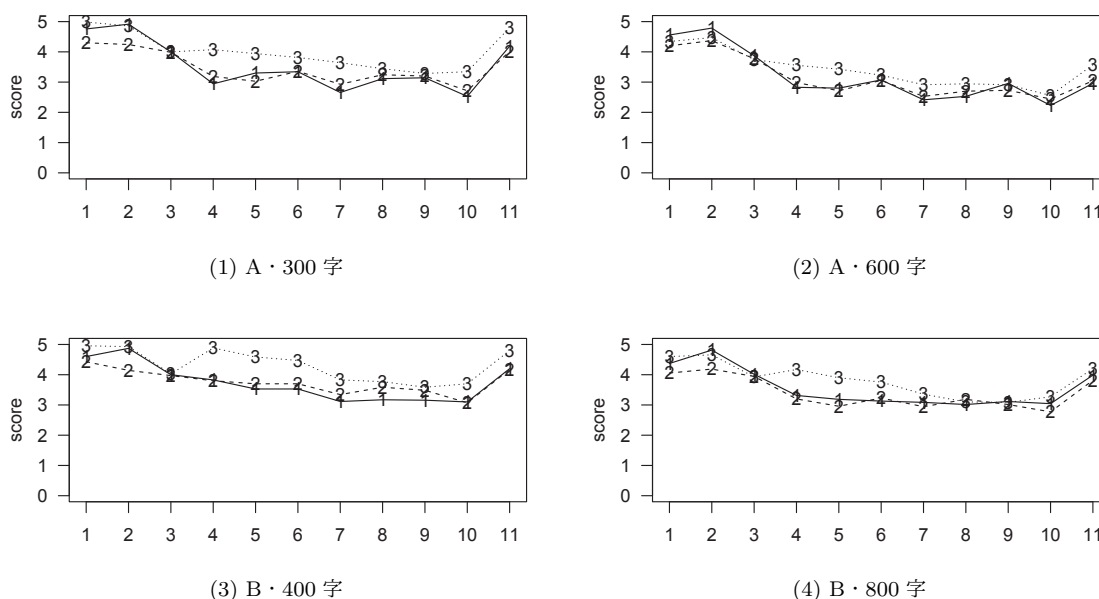


図 2: 分析的評価の各観点の平均点

(折れ線の数字1~3は、採点者1~3を示す。横軸の数字は分析的評価観点の番号である。)

見ることができないような場合には、採点者によって評価基準に違いが生じないように、トレーニングの時間を長く取るなどして採点の一致度を高める工夫が必要になるであろう。今後は、小論文の採点で重視される評価観点は何か、また、短い時間や短い字数による小論文課題で測りたいものを適切に測定できるのか等について検討していく予定である。

- 1) 1. 語句の表記, 2. 文法, 3. 語彙力, 4. 課題の解釈, 5. 主張の明確性, 6. 一貫性, 7. 説得力, 8. 構成, 9. 簡潔性, 10. 独創性, 11. 原稿用紙の11個である。なお、「11. 原稿用紙」とは、回答字数、段落の字下げ等、原稿用紙の使い方が適切かを評価する観点である。

参考文献

荒井清佳・石岡恒憲(2015). 大学入試における小論文の形式について—アンケートを通じて—. 大学入試研究ジャーナル, 169-175.

文部科学省 (2014). 平成 27 年度国公立大学入学者選抜の概要. http://www.mext.go.jp/a_menu/koutou/

senbatsu/_icsFiles/afieldfile/
2014/09/11/1351606_01_1.pdf

村上京子(2005). 日本留学試験「記述問題」が測っているもの. 日本留学試験とアカデミック・ジャパニーズ(2), 306-315.

大久保智哉(2013). 400字論述課題における能力測定の信頼性. 大学入試センター研究紀要, 42, 1-12.

宇佐美慧(2008). 小論文試験の採点における文字の美醜効果の規定因—メタ分析及び実験による検討—. 日本テスト学会誌, 4, 48-59.

宇佐美慧(2011). 小論文評価データの統計解析—制限字数を考慮した測定論的課題の検討—. 行動計量学, 38, 33-50.

宇佐美慧(2013). 論述式テストの運用における測定論的問題とその対処. 日本テスト学会誌, 9, 145-164.

渡部洋・平由美子・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, 33, 7-16.