

大学入試センター試験の得点調整

—基本的な考え方と方法—

大阪大学副学長 真 弓 忠 範
名古屋大学教育学部教授 村 上 隆
大阪大学大学院基礎工学研究科教授 白 旗 慎 吾

東京理科大学工学部教授 吉 村 功
大学入試センター研究開発部助教授 前 川 眞 一

平成9年度における「旧数学Ⅱ」と「数学Ⅱ・数学B」の間に発生した得点較差が社会問題化した結果、大学入試センター（以下「入試センター」という。）は、一旦は行わないことに決定していた得点調整を再開することになった。後述のように、入試センターでは平成9年度初めから、委員会等において調整の方法について検討を行ってきたが、11月下旬に最終案が決定され公表された。

平成10年度試験においては、地理歴史において「日本史B」と「地理B」の間に、調整を発動させる基準とされていた20点以上の平均点差が発生し、この方法による得点調整がいきなり実施されることとなった。共通1次試験以来2度目、大学入試センター試験（以下「センター試験」という。）としては初めての得点調整が実施されたのである。幸いにして、今回の得点調整に関しては、社会的にさほど大きな議論を巻き起こすことはなく、新たな方法はおおよそ社会的に受け入れられた

と考えられる。

ここでは、今回決定され、実施された得点調整の方法について、その理念と手続きをやや詳細に解説するとともに、なお残された問題点について述べる。その際、考えられる他の選択肢がなぜ選ばれなかったのかということについても論ずる。

1. 得点調整の方法とその特徴

1.1 調整の方法の決定までの経緯

共通1次試験に代わってセンター試験が始められた平成2年の時点において、入試センターは、理科と社会（当時）の科目間に30点以上の平均点差が生じた場合に対して、これを調整するための方法を用意していた。幸いにして、以後、平成8年度までの7年間、調整を必要とするような大きな得点差は発生せず、この方法が実際に発動されることはなかった。

平成9年度から、高校教育の多様化を受けて、センター試験の科目数は34科目に激増した。そこには、ちょうど

新旧の課程の変わり目にも当たっていた数学における「旧数学Ⅰ」と「旧数学Ⅱ」も含まれていた。この多様化に際し、入試センターは、試験開始以来得点調整を行った例がないこと、従来の方法がそのままでは適用できないこと等を理由に、得点調整の廃止を決定し公表した。

不幸にして、平成9年度において、「旧数学Ⅱ」と「数学Ⅱ・数学B」の間に、約22点の平均点差が発生し、理科においても19点の平均点差が生じた。それに対して、入試センターは既定方針通り、得点調整を行わなかったが、その処置は社会的に大きな批判を呼び起こした。2年間の経過措置である新旧数学の共存の時期に、その間で較差が生じたことは皮肉であったが、その差が浪人生の多くが選択した「旧数学Ⅱ」の平均点の方が低くなる方向で生じたことも不運であった。

いずれにせよ、このことは多様化の

下でも、放置することのできないほど大きな較差が生じ得ること、以前の方法で明示されていた調整実施の基準となる平均点差30点を下回る差であっても、社会的に許容されない場合があること等をはからずも示すこととなった。入試センターは、一旦廃止した得点調整を復活することを決定したが、そのことは、多様化の下で新たな得点調整の方法を開発することを強いられることを意味していた。

平成9年度当初、得点調整検討委員会が組織された。委員会は、5月以降集中的に審議を行い、その結論は入試センター内外の各種委員会との調整を経て、11月下旬には公表される運びとなった。

1.2 方法の概要

ここではまず、入試センターが11月20日付けで公表した文書、「大学入試センター試験における得点調整について」を引用しておく。

大学入試センター試験における得点調整について

平成9年11月20日

大学入試センター

1 基本的な考え方

- (1) 著しい平均点差が生じないよう作題の段階でできる限りの努力を払うが、それでも科目間で大幅な平均点差が生じ、これが試験問題の難易差によるものと認められる場合に得点調整を行う。
- (2) 得点調整は、行えば行っただけ新たな不公平感が生じる措置なので、調整後できるだけ受験者全体の不公平感が少なくなるような方式とする。

また、得点調整を実施する場合は、受験者の心理に配慮し、素点は下げないことを原則とする。

2 得点調整の対象とする教科・科目

教科間を超えての調整は行わない。具体的には、次の各教科内の科目間に限る。

- (1) 地理歴史の「世界史B」、「日本史B」、「地理B」の間
- (2) 公民の「現代社会」、「倫理」、「政治・経済」の間
- (3) 数学①の「数学Ⅰ・数学A」と「旧数学Ⅰ」の間
- (4) 数学②の「数学Ⅱ・数学B」と「旧数学Ⅱ」の間
- (5) 理科の「物理ⅠB」、「化学ⅠB」、「生物ⅠB」、「地学ⅠB」の間

3 得点調整の方法

- (1) 上記の各科目の間で、原則として、20点以上の平均点差が生じ、これが試験問題の難易差によるものと認められる場合に得点調整を行う。なお、公平性の観点から20点以上の平均点差が生じた科目だけでなく、平均点がある間にある科目についても調整を行う。
- (2) 得点調整に当たっては、最高平均点科目と最低平均点科目との平均点差が、15点（通常起こり得る平均点差の範囲）となるよう調整する。その際、平均点がある間にある科目についても素点の平均点差の比率に応じて調整する。
- (3) 得点調整に当たっては、最高平均点科目の累積分布を目標分布とし、調整すべき科目の累積分布を目標分布の方向へ移動させる方法（分位点差縮小法）を用いて行う。
- (4) 上記(1)から(3)については、図1・2を参照。

4 得点調整の是非の判断

上記の方針を踏まえ、具体的には関係者による委員会を設け、試験終了後に、その委員会において試験問題の難易差等をも分析して、得点調整の是非を個別に判断する。

5 受験者への周知方法

得点調整を行う場合は、国公立大学の個別学力検査の出願前にできるだけ早く、具体的な方法を新聞等により周知する。（引用終わり）

この後に、図1と図2が付されている。

得点の累積分布図

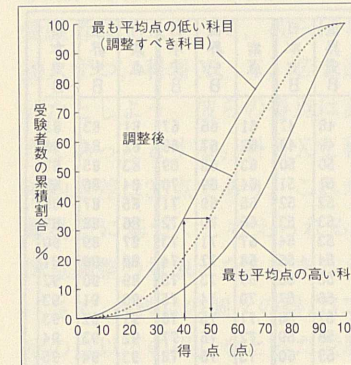


図1 分位点差縮小法の説明

以下において、調整方法の中心である「分位点差縮小法」について少し詳細に説明したい。しかし、その前に、実際に行われた調整について見ておく方が、具体的様相はつかみやすいであろう。

1.3 実際に行われた調整

平成10年度センター試験の地理歴史の「B科目」の平均値と標準偏差は表1の「調整前」の欄に示したようになっていた。もっとも、この数値は最終集計ではなく、その直前の約50万人段階の中間集計の結果である。得点調整は時間的に極めて切迫した状況で行われざるを得ず、従来の経験から最終集計と（決定を覆すほどには）大きな差

15点＝通常起こり得る平均点差の範囲

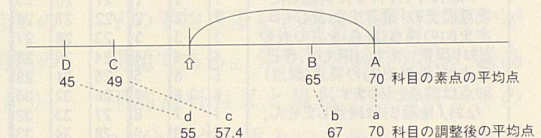


図2 複数科目の調整方法の説明図

違が生じないことが確認されているこの時点が、判断の時期と定められていた。「日本史B」と「地理B」の平均点の間には、約21点の差がある。

1月23日（試験終了後5日目）、入試センターは、この点差が「問題の難易差で生じた」と認め、地理歴史の得点調整の実施を発表した。そして、翌日の朝刊には、図3のような分位点差縮小法に基づく「得点換算表」が掲載された。

分位点差縮小法では、最高平均点科目に他の科目を近づける方向で調整が行われるから、この場合、最高平均点科目である地理Bの得点は修正されない。

図3 1998年1月24日付け朝刊に掲載された「お知らせ」

大学入試センター試験 得点調整のお知らせ	素点	世界史B	日本史B	素点	世界史B	日本史B	素点	世界史B	日本史B	素点	世界史B	日本史B	素点	世界史B	日本史B
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
平成10年1月17日に実施した 地理歴史の「世界史B」及び「日 本史B」の得点(素点)を右の表の とおり調整します。(例えば、自己 採点で「世界史B」の得点(素点) 50点は56点となります。) なお、「地理B」は調整しません。	0	0	0	21	26	25	41	48	47	61	66	67	81	83	85
	1	1	1	21	26	25	41	48	47	61	66	67	81	83	85
	2	2	2	22	27	26	42	49	48	62	67	68	82	84	86
	3	3	3	23	28	27	43	50	50	63	68	69	83	85	87
	4	4	4	24	30	28	44	51	51	64	69	70	84	86	88
	5	5	5	25	31	29	45	52	52	65	69	71	85	87	88
	6	6	7	26	32	30	46	53	53	66	70	72	86	88	89
	7	7	8	27	33	32	47	54	54	67	71	73	87	89	90
	8	9	9	28	34	33	48	54	55	68	72	74	88	90	91
	9	11	10	29	35	34	49	55	56	69	73	75	89	90	92
	10	12	12	30	36	35	50	56	57	70	74	75	90	91	93
平成10年1月24日 大学入試センター	11	13	13	31	38	36	51	57	58	71	75	76	91	92	93
	12	14	14	32	39	37	52	58	59	72	76	77	92	93	94
	13	16	15	33	40	38	53	59	60	73	76	78	93	94	95
	14	17	16	34	41	40	54	60	61	74	77	79	94	95	96
	15	18	17	35	42	41	55	61	62	75	78	80	95	96	96
	16	20	19	36	43	42	56	62	63	76	79	81	96	97	97
	17	21	20	37	44	43	57	62	64	77	80	82	97	97	98
	18	22	21	38	45	44	58	63	64	78	81	82	98	98	99
	19	24	22	39	46	45	59	64	65	79	82	83	99	99	99
	20	25	23	40	47	46	60	65	66	80	83	84	100	100	100

表1 得点調整前後の平均値と標準偏差

科目名	人数	調整前		調整後	
		平均点	標準偏差	平均点	標準偏差
世界史B	98,050	61.12	20.63	65.57	18.81
日本史B	145,880	56.32	15.88	62.17	15.39
地理B	102,873	77.13	14.56	77.13	14.56

修正後の各科目の平均点と標準偏差は表1の「調整後」の欄に記されている。「日本史B」と「地理B」の平均点差は、ほぼ15点まで縮まり、中間にあった「世界史B」もかさ上げが行われている。調整の対象となった科目の標準偏差が(「世界史B」において特に)縮小していることにも注目していただきたい。

図3の換算表の方を見ると、全員に対して一律に同じ得点が加算されているわけではないことが分る。

0点は0点、100点は100点のままであり、0点と満点の近傍の得点はほとんど変化していないのに対し、中間のところでは7点程度のかさ上げがなされるところもある。分位点差縮小法は、このように非線形な得点調整の方法である。

具体的なことを見ていただいたところで、分位点差縮小法の手順について、少し詳細に説明しておきたい。

各科目の受験者を得点の低いものから高いものに順に並べ直したとしよ

う。話を具体的かつ簡単にするために、受験者は30人だけで、彼らはちょうど10人ずつ、科目X、Y、Zを選択していたとしよう。彼らの得点は表2のようであったとする。科目Xと科目Yの平均点差は30点であり、これが問題の難易差に基づくと考えられるかぎり、得点調整の対象となるであろう。

さて、各科目の個々の得点に対して、その点以下の得点をとった受験者の割合 p を求めることができる。これは、相対累積度数とも呼ばれ、図1の縦軸の値に対応する。この p の値は、順位とは逆に、得点が低いほど小さくなる。各科目受験者に対する p の値は表2の右端のようになるであろう。

表2 3つの科目におけるそれぞれ10人の受験者の得点

科目内順位	X	Y	Z	p
10	50	10	25	0.10
9	60	20	30	0.20
8	65	30	35	0.30
7	70	35	40	0.40
6	75	45	50	0.50
5	80	50	60	0.60
4	85	55	85	0.70
3	90	70	85	0.80
2	95	75	95	0.90
1	100	80	100	1.00
平均値	77.0	47.0	61.0	
標準偏差	15.2	22.4	27.6	

値 p に対応する得点を、 p -分位点と呼ぶ。例えば、科目Xの0.3-分位点は65点、科目Yの0.8-分位点は70点である。この表では、各科目とも同点者

はいないので、得点と p とは1対1に対応がつく、すなわち、この例では p -分位点は一意に定まっている。

ここで、いかにして、科目Xと科目Yの平均点差を15点まで調整するかが問題である。分位点差縮小法は、最高平均点科目Xと最低平均点科目Yの p -分位点の重みつき平均を、最低平均点科目の調整後の p -分位点とするものである。

すなわち、 X_p を科目Xの p -分位点、 Y_p を科目Yの p -分位点とすると、調整後の科目Yの p -分位点を、

$$Y_p' = wX_p + (1-w)Y_p \quad (1)$$

のようにして求める。ここで、 w は、 $0 \leq w \leq 1$ の範囲にある重みであって、次の式で求められる。

$$w = 1 - \frac{\alpha}{\bar{X} - \bar{Y}} \quad (2)$$

ただし、 α は達成しようとする平均点差で、ここでは $\bar{X}$ 、 $\bar{Y}$ は、それぞれ、科目XとYの平均点である。この w を用いることによって、(1)で得られた Y_p' と X_p の平均値の差が α まで縮められることは簡単に確認できるであろう。すなわち、

$$\begin{aligned} \bar{Y}' &= w\bar{X} + (1-w)\bar{Y} \\ &= \left(1 - \frac{\alpha}{\bar{X} - \bar{Y}}\right)\bar{X} + \frac{\alpha}{\bar{X} - \bar{Y}}\bar{Y} \\ &= \bar{X} - \frac{\alpha}{\bar{X} - \bar{Y}}(\bar{X} - \bar{Y}) \\ &= \bar{X} - \alpha \end{aligned}$$

というわけである。そしてこれが、図1で示されていることの本質的内容である。

表2の例では、 $\alpha=15$ 、 $\bar{X}=77.0$ 、 $\bar{Y}=47.0$ として、(2)に代入すると、 $w=1-(15/30)=0.5$ となる。この場合は、まさに X_p と Y_p の単純平均をとればよいことになったわけである。例えば、 $Y_{0.1}=10$ は、 $Y'_{0.1}=0.5 \times 50 + 0.5 \times 10 = 30$ と調整される。全体の結果は表3のようになる。実際平均点は15点かさ上げされ、「通常起こり得る平均点差」まで詰められた。 α のとり方次第で、この得点差は任意に調節することができる。

表3 科目Yの得点のみを調整

科目内順位	X	Y'	Z	p
10	50	30.0	25	0.10
9	60	40.0	30	0.20
8	65	47.5	35	0.30
7	70	52.5	40	0.40
6	75	60.0	50	0.50
5	80	65.0	60	0.60
4	85	70.0	85	0.70
3	90	80.0	85	0.80
2	95	85.0	95	0.90
1	100	90.0	100	1.00
平均値	77.0	62.0	61.0	
標準偏差	15.2	18.8	27.6	

ここで、ただちに気づくのは、本来、平均値が最高平均点科目Xと最低平均点科目Yの中間にあった科目Zの平均値が、科目Yに追い越されてしまったことである。常識的に考えてこれは不

合理であり、新たな不公平の発生と見なされてもやむを得ないであろう。考えられるやり方は、Zの得点も同じ w を用いて調整することである。すなわち、

$$Z'_p = wX_p + (1-w)Z_p \quad (3)$$

結果は表4のようになり、16点あった科目Xとの平均点差は8点に縮められ、再び科目Yの平均点を追い越した。

表4 科目Zの得点も調整

科目内順位	X	Y'	Z'	p
10	50	30.0	37.5	0.10
9	60	40.0	45.0	0.20
8	65	47.5	50.0	0.30
7	70	52.5	55.0	0.40
6	75	60.0	62.5	0.50
5	80	65.0	70.0	0.60
4	85	70.0	85.0	0.70
3	90	80.0	90.0	0.80
2	95	85.0	95.0	0.90
1	100	90.0	100.0	1.00
平均値	77.0	62.0	69.0	
標準偏差	15.2	18.8	21.2	

このように同じ重みを用いることによって、最高平均点科目との平均値の差の割合が同じに保たれる点は重要である。すなわち、

$$\frac{\bar{X} - \bar{Y}}{\bar{X} - \bar{Z}} = \frac{30}{16} = \frac{15}{8}$$

であり、

$$\frac{\bar{X} - \bar{Y}'}{\bar{X} - \bar{Z}'} = \frac{15}{8}$$

というわけである。先の引用の図2に

説明されているのはこのことである。

このように、低平均点科目の p -分位点を、対応する最高平均点科目の、 p -分位点の方向に近づける、つまり、同じ p に対応する分位点の差を縮小するのが、分位点差縮小法なのである。

もちろん、現実のデータではこのように簡単にはいかない。受験者の数は科目ごとに異なるし、得点は離散的だから、低得点科目側の p -分位点に対応する最高平均点科目の p -分位点が一意的に定まらない、といったことも、当然起こり得る。すなわち、図1のような滑らかな相対累積度数カーブが常に得られるとは限らないのである。実際の調整点を求める際には、曲線の平滑化の技法を用いているが、この方法は公表されておらず、かつあまりに技術的になるので、その詳細はここでは省略する。

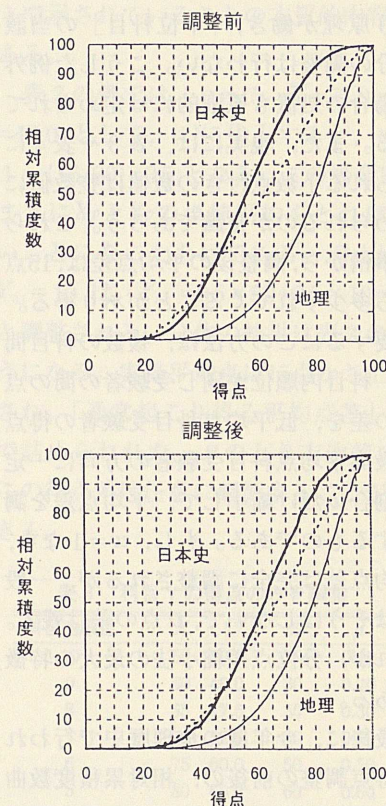
もう1つ、得点分布の形状によっては、平均点の低い方の科目の分位点が、最高平均点科目のそれを上回る、といったケースがあり得る。この場合には、

得点を下げる方向の調整は行わないという原理が働き、「下位科目」の当該部分の調整は行わない。こうした例外的場合の処置も原則的には定められている。また、現実には、表3や表4に見られる小数点付きの数値は整数値に丸められるのが常識であろう。これらの事情から、調整後の平均点差は、15点から多少ずれてくることがあり得る。

要するにこの方法は、複数の科目間で、科目内順位が同じ受験者の間の点数の差を、低平均点科目受験者の得点を最高平均点科目受験者の方向に一定の割合 w だけ縮小して、平均点差を調整するものである。もし、 $w=1$ なら、平均点差は完全に調整されるが、一般にはそうはしないで、 α 点の差は残す。これが、分位点差縮小法の最大の特徴である。

最後に、今年度の地理歴史で行われた得点調整の前後の、相対累積度数曲線を示しておこう。これは、最も問題の少ない得点分布が得られた幸運なケースであったと言うべきであろう。

図4 平成10年度の地理歴史における得点調整



2. あり得る反論とそれへの回答

2.1 基本となる考え方

分位点差縮小法を含む今回の得点調整の方法について、種々の疑問や反論があり得ることは否定できないところである。しかしながら、まずは今回の方法の基本になっている考え方を説明しておくべきであろう。

以下の議論は、やはり平成9年11月20日付けで公表された得点調整検討委

員会の「まとめ」に基づいている。以下、「まとめ」の文章の一部を特に断りなしに引用する。

(1) 素点主義

テストの得点表示の仕方は、必ずしも素点には限られない。例えば、アメリカのEducational Testing Serviceによって行われているSATやTOEFLの得点は素点ではなく、我が国という偏差値に近いものである。しかしながら、センター試験は、この試験を利用する各大学に、受験者の得点を素点で提供することを前提に、問題の作成と答案の採点が行われている。さらに、受験者は志望校の決定に当たって、新聞発表される正解と配点に基づいて自己採点を行っており、その結果もまた素点である。得点調整の方法を考えるに当たってはまずこのことが前提となる。

(2) 得点調整を行うことは基本的に望ましくない

素点が、受験者にとっても大学にとっても判断の基礎となる以上、得点調整はできるだけ行わないようにすべきである。すなわち、入試センターは、各教科・科目間に著しい平均点差が生じないように試験問題の作成段階でできるかぎりの努力を払うことが基本である。

(3) やむを得ない措置としての得点調整

しかしながら、現実の問題として、

入試センターでは、6教科31科目（平成9年度、10年度の経過措置期間中は34科目）と多くの科目が出題されており、それを約60万人に達する多様な受験者が受験するものである。このため、平均点を一定に保つことは極めて困難である。

センター試験の成績は、これを利用する大学の入学者選抜資料の一部であり、できるだけ公平性が求められることから、万一、科目間に著しい平均点差が生じ、これが試験問題の難易差に基づくものと認められる場合には、得点調整を行うことによってできるかぎり不公平感を少なくすることが適当である。

(4) 方法の望ましさを判断する基準

得点調整において、万人が納得する方法を見出すことは困難であるが、

- (1) できるだけ多くの受験者の公平感が保たれるものであること
 - (2) 受験者にとって分かりやすいものであること
 - (3) 調整作業が短期間に処理可能なものであること
- といった条件が重視される。

(5) 平均点差のすべてを調整しない理由

分位点差縮小法の最も目立つ特徴が、これであろう。この点について、「まとめ」は次のように述べている。

得点調整を実施すると、高得点であ

る受験者群からみれば、一度生じた差が縮小されることになるため、その受験者群から不満が出るのは必至である。また、生じた平均点差のすべてを調整の対象とすると、加算点が大きくなり、特に調整すべき科目の得点分布のある部分で平均点差以上の点が加算されることになる。（中略）したがって、平均点差のすべてを調整の対象とすることは、調整の対象とならなかった受験者群に逆に不公平感を与えることになり、公平性の観点からみて適当とは言えない。（引用終わり）

先の引用中にあった、「得点調整は、行えば行ったで新たな不公平感が生じる処置」である、とはこのことを指す。

2.2 採用されなかった考え方

以上のような得点調整検討委員会の基本方針は、幾つかの対立する考え方の否定の上に成り立っている。それら、今回は採用されなかった理念を検討することを通じて、今回の調整方法はより明確なものになるであろう。

(1) アンカー得点の使用

これは、受験者の学力と選択科目の間に関係があることを前提とした議論である。例えば、ある教科（例えば理科）において「科目X」の選択者の学力は「科目Y」選択者のそれより高いという認識がある。それは、もし、国語、数学、外国語を全員が共通に受験していれば、それらの得点を両群間で

比較することによって実証できると考えられる。そうだとすれば、科目間の平均点差を比較するに際して、この点を考慮しないと不公平であるとも考えられる。そこから、両群の学力差を共通テストを用いてコントロールしようという考えが生じてくる。

実際、平成8年度までに準備されていた得点調整の方法は、理科と社会の諸科目に対して、国語、数学、外国語を別々の仕方で重みを付けて合計した得点を共通テストとして定義し、「共通テストを用いた等百分位点法」(例えば、前川、1997)という方法を用いて調整を行うものであった。センター試験では、ここで用いられる共通テストに相当する得点を、アンカー得点と呼んでいた。理系学力と文系学力の2種のアンカー得点を用意されたわけである。前述のように、この方法は、受験科目の多様化と、受験者の科目選択の多様化により、アンカー得点の定義が著しく困難になったことを主な理由として廃止された。

現実の入試では、選択科目間で何らかの比較可能性を考慮しなければならないのは事実である。しかしながら、使用が予定されていた「共通テストを用いた等百分位点法」をはじめとするテスト得点の等化(equating)法(例えば、Kolen & Brennan, 1995)は、本質的に同次元の学力を測定してい

るテストの間の得点調整のための方法であり、選択科目間の得点調整に適用されるには無理がある。実際、次のような問題は無視できないであろう。

- (1) 繰り返しになるが、アンカー得点の定義は、以前に比べてはるかに困難になっている。特に、前年度問題になった「数学Ⅱ・数学B」と「旧数学Ⅱ」のケースなど、適切なアンカー得点は見出しにくいであろう。他の科目についても、アラカルト方式の導入等により、全員が受験している共通科目が存在しない。また、データを見ても、理科のアンカー得点に社会を加えるかどうかで、かなり大きな変動が生ずるなど、問題が多い。
- (2) いかなるアンカー得点を用いるにせよ、アンカー得点と当該科目の得点の間の相関を考慮する必要がある。できるだけ相関の高いアンカー得点を用いる必要がある。それを、試験実施以前に決めることは困難であろうし、試験後では、十分に検討するための時間が不足している。
- (3) いずれの定義によるにせよ、アンカー得点として用いられる科目上での、科目選択による得点差は、それほど大きいものではないと予想される。少なくとも、これが科目間較差の主要な要因とは考えにくい。今回のように、平均点差を完全には調整

しない方式をとるとすれば、その範囲に吸収される程度の差と考えられる。

- (4) 将来、A科目にも調整の範囲を拡張した場合、従来の方法をとれば、それらの科目に対して、著しい修正減(得点を低下させる修正)を行わざるを得ない可能性が高い。このことは、受験生の心理云々とは無関係に、A科目をセンター試験に加えた理念に関して、論争の余地のある困難な問題を表面化させることになる。

(2) 偏差値化

センター試験の得点は、科目・教科を問わず100点(または200点)満点で表示されるために、原則的には相互に比較可能な、一種絶対的な意味をもつように受け取られやすい。多くの大学では、「優」、「良」、「可」、「不可」の評価と試験の得点との間の関係が決められているのが普通である。また、日常用語としても、「今回の交渉の成果は、採点すれば80点だ」(成功)、「この法案の出来は60点」(合格すれすれ)、「この間の経済政策は、せいぜい50点しかつけられない」(不合格)といった表現は定着しており、得点の絶対的な意味付けは社会的に浸透している。

しかしながら、センター試験のように多面的な学力を測定しようとする試験の得点の場合、それらは相対的なも

の、すなわち、せいぜい同一のテストの受験者相互間での上下が比較できるだけであると考えておくのが無難である。また、少なくとも、現在のセンター試験のような競争試験においては、得点は他者との優劣の比較においてのみ意味をもつ。偏差値化は、このテスト得点の相対性に依拠して、素点主義とアンカー得点の使用に反対する論理である。

偏差値化を採用するとすれば、選択科目間で分布の形状が一致しているかぎり、全科目の得点を次の偏差値の公式、

$$X' = 10 \frac{X - \bar{X}}{s_x} + 50 \quad (4)$$

によって変換(調整)することは正当化される。ただし s_x は、テスト X の標準偏差とする。この結果、すべての科目の平均点は50点、標準偏差は10点となる。この際、(4)の変換は、得点較差の大小にかかわらず常に行われることになる。つまり、科目の得点は素点でなく偏差値で与えられることになる。

しかしながら、もし、科目間で分布の形状が著しく異なる場合には、(4)による変換は新たな不公平を生み出す恐れがある。例えば、最高平均点科目は、著しく高得点者の多い、高得点の方向に偏った分布(統計用語では、負に歪んだ分布)になる可能性が高い。この

場合、満点の偏差値が、例えば、60を下回る、といったことが起り得る。受験したテストが易しい場合に、かえって点数が低下することがあるわけである。(すべての科目が同じように偏っていればよいが、値が0点と100点の間に制限されている試験の得点は、平均値が高いほど負の、低いほど正の歪みをもちやすい)。

その点を考慮すると、得点は分布の形状も相互に一致させるように((4)のような線形の変換ではなく)、非線型の変換を行う必要があるであろう。そのため、(共通テストを用いない)等百分位点法、あるいは、「上位 $P\%$ にあたる得点を正規分布の上側 $P\%$ にあたる値に変換する正規化変換」を行うという提案もある(南風原, 1997)。後者によれば、すべての科目の得点は、同じ平均値と標準偏差をもつ正規分布に従うことになり、科目内順位と得点がほぼ1対1に対応する、という意味における公平性が達成される。

この方法が採用されれば、当初は大学側、受験者側の双方にとって一時的に混乱をもたらす可能性もあるが、それは解決不可能なほど困難な問題とは思われない。得点表示の方法が素点とは全く異なるシステムに変えられる結果、調整による得点低下の心理的影響を考えなくてよくなるという利点もある。

偏差値化は、以上のような論拠もあり、処理の上で特別な困難もなく、理解もそれほど難しくないから、得点調整の方法としては有力な方法と言える。しかしながら、現時点で採用するには、次の2つの理由で難点がある。

第1に、現在のセンター試験では、当日会場で問題が配布された後に、科目選択を変更することが許されていることである。もし、受験者が十分有能であれば、問題を見た上で、自分がより高得点を獲得できる科目を選択することが(少なくとも原理的に)可能になっているのである。しかしながら、偏差値化が行われると、この判断は、自分自身が獲得できる得点だけでなく、他者の得点分布まで考慮に入れた、極めて困難なものにならざるを得ない。前述の「まとめ」においても、「大学入試センター試験は、当日、受験者が試験問題を見てから解答科目を選択することができる制度であることにも配慮する必要がある」という形でこのことに触れている。問題を見てから選択を変える受験者の数がどの程度であるのかは、はっきりしないが、これは数の問題ではなく、受験者の権利として保証している以上、考慮に入れないわけにはいかないであろう。

第2に、これはもう少し微妙な問題になるが、正規化のための非線形変換の過程で、得点の精度に得点(偏差値)

段階によって大きな差ができてしまう恐れがあることである。例えば、著しい高平均点科目で、多数の受験者が100点満点をとったとしよう。このとき、まず、正規化変換によっても、100点はそれほど高い偏差値に変換できないという、線形変換で生じた問題は、完全には解決できないことになる。しかし、それとともに問題になるのは、例えば1問だけ誤答であった99点の受験者と、100点の受験者の偏差値に大きな開きができてしまうことである。実質的に誤差範囲と考えられるような点差が拡大されることは、またしても新たな不公平を引き起こすことになる。

3. 問題の根本的解決はあり得るか

結局、問題は、テスト得点の分布を事前にコントロールできないという現在の状況にあると考えられる。少なくとも、50~100点の範囲で、ほぼ斉一な精度をもつ得点を生み出すような出題が常に可能にならないかぎり、すなわち、難易度の水準に多少の高低はあっても、分布の形状が科目間でほぼ一致するようなテストが作成できないかぎり、事後的な得点調整に頼っていたのでは、何らかの問題点が残らざるを得ない。

そうした出題がなされないと、得点調整では救いようのない次のような不

公平も残ることになる。すなわち、ある教科が全く選択の余地のない1科目からなっていたとしても、それが他の教科の得点と加算されるかぎり、例えば、当人にとってその科目が得意科目であった場合、易しすぎる(平均点が高く標準偏差が小さい)テストでは、そこで他の受験者に十分差をつけられないことになる、といったことである。このように、不公平は、選択科目以外のところでも生じ得る。

厳密に言えば、難易度を事前にコントロールするには、問題そのものを事前に小規模な集団に予備的に実施して分析するという、いわゆるプリテストを行うしかないであろう。先に例示したような、Educational Testing Serviceの実施するテストについては、これがなされており、プリテストを行うこと(あるいは、何の先行情報もないテストを行わないこと)は常識であると見なされているようだ(名古屋大学、野口裕之教授による)。しかしながら、それは、本試験の問題を事前に(ある程度不特定多数の)人々の目にさらさなければならないことを意味する。どう考えても、現在のセンター試験でプリテストを実施することは現実的ではない。

米国で可能なものが、なぜ日本で可能でないのかは、文化や制度の比較等を含め、多面的に検討される意味のあ

る問題であると思うが、そういう議論が、今ただちに役立つとは思われない。受験者ごとに問題が異なるコンピュータ化された試験が実現すれば、という考えもあるが、そこでは、現在の大問に基づく出題方式が、大きな隘路となるであろう。かなりまとまった数の小問を、入れ替えなければならないからである。何より現状程度の人員で、コンピュータ化された試験のメンテナンスや実施が可能になるとは到底考えられない。

一方、最初に引用した文書の「基本的な考え方」のところで述べられているように、「著しい平均点差が生じないよう作題の段階でできる限りの努力を払う」ことがまずは大前提であるとすれば、それを可能にするような方向が模索されなければならないであろう。このことは、テストの性質、あるいはテストの得点と能力の関係といったところまで公然と議論する必要が生ずる可能性があることを意味している。例えば、前述の大問形式による出題の是非について議論する必要が生じる可能性もある。いわゆる「共倒れ」、すなわち、連鎖反応的な誤答を誘発し

やすいこの方式は、測定誤差の観点から見て適切なのだろうか。実は、「通常起こりうる得点差」が15点にも達する理由はここにある。このことは、センター試験が学力のどの部分(記憶力、思考力)を測っているのかという問題、マーク・シートという形式の限界についての議論にもつながり得る。

しかしそうした議論について何らかの結論に達するには、なお時間を要するであろうから、当面は、現状のように必ずしも完璧でない方法を、多少の手直しをしながら使っていくしかないであろう。

文献

南風原朝和(1997) 入試の理念と得点調整, 日本行動計量学会代25回大会 発表論文抄録集 154-155.

Kolen, M. J. and Brennan, R.L. (1995). [*Test Equating: Methods and Practice.*] New York, Springer-Verlag

前川真一(1997) 得点調整について セミナー資料(入試研究の基礎知識) 国立大学入学者選抜研究連絡協議会 20-34.