

# 南風原・片岡論文に対するコメント

## —規範的な問いへの経験的な答え—

名古屋大学・教育学部 村上 隆

### 1. 得点調整に共通問題の成績を加味すべきか

南風原・片岡(2000)が問題にしている事態は、同論文の3節において設定されている条件を用いて単純化すると、次のようなものであると考えられる。2つの教科の合計点によって合否が決まるような試験において、第1の教科は全員が同一問題の試験を受ける(あるいは、この教科には1つの科目しかない)。その得点を $v$ とする。他方、第2の科目には2種の重なりのない問題(あるいは2つの科目)があり、受験者の半数は $X$ を、他の半数は $Y$ を受験する。それぞれの得点を $x$ 、 $y$ とする。ここで、 $x$ と $y$ の平均値(および、標準偏差)に差があると、不公平が発生していることが懸念され、得点調整がなされる必要が示唆される。すなわち、合否の判定を、 $t = v + x$ と $t^* = v + y$ を一緒にして行うなら、平均点の低いほうの科目の受験者にとって不利になるように見えるからである。

ここで適用される得点調整の方法にはさまざまなものが考えられるが(たとえば、前川, 1999), ここで扱われているのは、線形等化法の中の2種、単純線形等化法とTuckerの線形等化法である。前者は、選択科目の得点を同一の平均値と標準偏差をもつように変換するもの、後者は、共通科目における2つのグループの平均差を最小2乗法的に調整するものである。後者によれば、例えば、科目 $X$ の選択者の共通科目における平均値が、科目 $Y$ の選択者のそれを上回るとすれば、得点調整後において、 $x$ の平均値は $y$ の平均値よりも高くなることが要請される。南風原が、以前、

提起した問題は後者のやり方が、実は新たな不公平をうみだしているのではないか、ということである。すなわち、「同じ情報が二重に用いられることになる……ことによって、共通科目の重みが過度に大きくなる」という可能性である。

南風原と片岡は、共通科目の総点に対する重みを、共通科目の得点と総点との相関係数と、共通科目の総点の中に占める共分散比という2つの指標を用いて、幾つかの条件の下で比較し、この可能性の有無を検討しようとしている。限られた条件下ではあるが、どちらの方法も、素点をそのまま用いる場合よりも、共通科目の重みを大きくする傾向があり、総点との相関においてはTuckerの方法が、共分散比においては単純線形等化法の方が共通重みを大きくすることが認められた。このことは、数理的に証明可能な事実である可能性がある(柳井, 2000)。また、2つの方法の間の効果の差は小さいが、それは、南風原・片岡の条件設定の仕方と、Tuckerの方法が共通得点の平均差を反映させる方法としてはやや保守的であること(村上・前川・菊地, 1999)にもよるのであろう。

### 2. 経験的に問うべき問題なのか?

南風原と片岡の議論は精緻であり、それ自体として興味あるものである。ただし、筆者は、これが最初の問い、Tucker法は共通得点の差を「二重に用いて」却って不公平をもたらしているか、という問いに対する答えを与えているかどうかは疑問であると考えられる。この問いには、このような経験的な方法によってではなく、もっと原理的な方法によって答

えられるべきであると考える。

次のような事態を考えよう。ある学校の2つのクラスで、ある科目の成績が、2日続けて行われる2つのテストの合計点に基づいて(相対)評価されることになったとしよう。1日目の試験は無事終了したが、2日目の試験は一方のクラスで何らかの事故が起こり、そのクラスでは本来の試験の代わりにあらかじめ用意されていた代替問題が用いられたとしよう。はからずも、1日目は共通問題、2日目は選択問題という事態になってしまったわけである。2つのクラスの1日目の得点の間に平均値に差があったとすると、2日目の成績を単純線形等化法で得点調整して2つのクラスの平均差を無くしてしまうことはためられるであろう。この場合、2つのテストの合計点の平均差は、1日目のテスト(共通科目)と同じままであるにもかかわらず、総点の標準偏差は、

$$\sigma(t) = \sqrt{\sigma(v)^2 + \sigma(x)^2 + 2\sigma(x, v)} \quad (1)$$

のように拡大する。ただし、 $x$ は調整後の2日目のテストの得点とする。その結果、南風原・片岡の式(12)、(10)で定義される効果量は、第1日目の得点より合計点の方が縮小することになる。これは、得点の高かったクラスの生徒からみて不公平と感じられるであろう。同一科目の試験であり、1日の間に、大きな能力の変動が起こるとは考えにくい以上、2日目の試験でも、(効果量で見て)同程度の差がつくと期待するのが常識だからである。(実際、2回のテストの平均差が同じで、かつ2回のテストの分散が相互に等しければ、相互相関が1でない限り、(1)から、合計点では効果量は拡大することがわかる。)2つ目のテストの単純線形変換によって、1日目に得点の高かったクラスの生徒の何人かが、得点の低かった方のクラスの生徒に追い越されることになる。この場合、「共通科目」である第1日目の「差を二重に用い」ないことは、明らかに不公平であり、何らかのやり方で2つのクラスの

能力差を反映する調整がなされるべきであると考えられるであろう。単純線形等化法は、やはり、「集団間に学力差がない場合の等化法」なのである。

それでは、南風原・片岡で問われているような本来の形の科目選択事態において、共通科目の得点差を選択科目の得点に反映させることが却って不公平であると感じられるとすれば、それは何故なのだろうか。そこには恐らく、選択科目が、共通科目とは多少とも別の学力を反映しているという事実が考慮されるのであろう。さらには、各受験者が自分の得意科目の方を選択している筈であるという配慮が働いているかもしれない。このことは、Tuckerの方法のような受験者の学力差を考慮した得点「等化」の方法を、得点「調整」のために用いることの無理を物語っているであろう(村上・前川・菊地, 1999)。

ただ、もし「別の学力」の意味を、相互に相関がないこと、と解するとすれば、少なくともTuckerの方法では、 $b=0$ となるから、南風原・片岡の式(2)、(3)からわかるように、単純線形等化法と同じになる。その意味ではTuckerの方法は単純線形等化法を含んでいる。すなわち、「集団間に学力差がない場合の等化法」である単純線形等化法は、「学力差がある場合の等化法」の一種であるTuckerの方法の特殊なケースと見る事が出来る。他方、共通科目と選択科目との間に相関がある場合、共通科目と選択科目が多少とも共通の学力を測っており、それが、求める学力を近似的に表現していると想定される以上(つまり、共通科目だけで選抜を行わない以上)共通科目の得点差を反映させない調整方法の採用を論理的に説明することは困難であるように思われる。南風原・片岡もそうした理論的説明は今回も与えていないように見える。ただし、現実には、ここでなされたような仮定の幾つかは成り立たなくなるであろうから、実際に適用する方法をあらかじめ定式化して

おくことは一層困難な問題となる。

### 3. 得点調整における現実的な対応

ところで、1998年に大学入試センター試験の「地理歴史」において、実際に採用された方法である分位点差縮小法は、共通科目の学力差を反映しないばかりか、科目間の得点差を完全には調整しない、その意味で、不公平を残したままにするものである（真弓、村上、白旗、吉村、前川、1999）。この方法が、南風原・片岡の方法と異なっている点はもう1つ、変換が非線形的に行われることであるが、これは現実のテストの得点分布が、科目間で著しく異なっている可能性を考慮しているためである。

現実問題として、1998年のケースに関する限り、この方法は社会的に受容され、public acceptance という観点からは成功した。このことは、実際の得点調整の方法を考えるにあたっては、従来の等化の理論にもとづく考察が十分なものでないことを物語っているようにも思われるが、こうした問題は南風原・片岡論文や本稿の射程を越えている。

## 文 献

南風原朝和・片岡尚子（2000）得点調整に伴う共通科目の重みの変化 大学入試研究ジャーナル, 10, xx-xx.

前川眞一（1999）得点調整の方法について 柳井晴夫・前川眞一（編） 大学入試データの解析 [理論と応用] 現代数学社 88-109.

真弓忠範・村上 隆・白旗慎吾・吉村功・前川眞一（1999）大学入試センター試験の得点調整について—基本的な考え方と方法—大学入試フォーラム, No. 21, 4-18.

村上 隆・前川眞一・菊地賢一（1999）得点調整における公平性の概念 柳井晴夫・前川眞一（編） 大学入試データの解析 [理論と応用] 現代数学社 110-124.

柳井晴夫（2000）私信.