GPA 算法の比較検証：従前の GPA から functional GPA への移行とその最適互換性をめぐって

半田智久
お茶の水女子大学 教育開発センター

Comparative study on GPA computation methods: Transition from traditional GPA to functional GPA and its optimal compatibility

Motohisa HANDA
Ochanomizu University Center for Research and Development of Education

The GPA (Grade Point Average) is becoming more common among Japanese universities as an index of academic performance. However, in many cases, raw scores are assessed based on a 100-point scale, and if the GPA is calculated by a method using letter grades that correspond to raw score ranges, the resulting GPAs represent unfair evaluation. This issue needs to be resolved for various reasons including for precise performance evaluation. Study 1 of the present research conducted 10 simulations in which grades were generated with random number generation by a computer on the assumption that a group of 1,000 took standard courses for four years. The results showed the average incidence of rank order shift of 96.9%, with the maximum shift of 97 ranks. This disagreement is resolved by taking the functional GPA method in which the GP is calculated directly from raw scores. It is desirable to adopt a computation method with the optimal compatibility, taking such issues as transition from the traditional GPA computation method. In Study 2, four representative computation methods were compared using the same simulations as Study 1: (1) GP = (raw score − X)/10, with X = 50; (2) with X = 60; (3) with X = 55; and (4) GP = (raw score − 60) x .075 + 1. The results demonstrated that Method 3 was found to show the optimal compatibility with the traditional method. In many universities, distributions of grades tend to bias higher scores. Considering this tendency, to recompare Methods 3 and 4, simulations identical to Study 2 were conducted in Study 3 so that the distribution with higher scores bias. The results yielded even a lower compatibility for Method 4, while showing a sustained compatibility for Method 3, reiterating the optimality of Method 3.

keywords : functional GPA, GPA, performance evaluation, computation methods, simulation

はじめに

GPA（Grade Point Average）はいまや日本の大学のおよそ半数が採用するに至り、成績評価制度の一般的な指標として定着しつつある。わが国で GPA を運用している大学が数 % であった状況からここに至るまで約 10 年であった。一方、2009 年時点で、すでに米国の大学ではほぼ全大学が、アジアの代表的な大学[*1]でも 8 ～ 9 割で GPA 制度が運用されている（半田、2011）。こうした経緯や国際的な実情に照らせば、わが国の大学におけるこの制度の全般化はほどなく果たされることが見通せる。

そうした展望下にあって日本における GPA 制度の運用については、Letter Grade（たとえば、S、A、B、C、D; 以下 LG）を Grade Point（たとえば、4、3、2、1、0; 以下 GP）に単に置き換えて算出するのでは、GPA 指標が原成績の構造を的確に反映せず、不公正な評価をもたらすという問題が孕まれている。したがって、全般化を前にこの問題を解決した上での運用が一層、強調される段階にきたといえる。

GPA 制度は文部科学省の審議会答申（大学審議会 ,1998; 中央教育審議会 ,2008）で「厳格な成績評価」という文脈下で例示、推奨され、その影響下で諸大学での導入と運用が図られてきた。それを素直に受けて日頃から厳正厳格な成績評価を履行している教員たちが、なお一層の厳格さをもって学生の学業成果を評価

Copyright 2011. Ochanomizu University. All right reserved.

したとしても、GPA 制度を導入したことで、その評価が不公正な結果を生むとしたら元も子もない話である。

海外諸国では問題にならないことが、なぜわが国の大学では問題が生じるのか。その理由は多くの日本の大学で一般的におこなわれてきた成績評価の仕方にある。その仕方とは 100 点満点、あるいは 100% を最大値とする達成度割合の素点評価で原成績を出す方法である。この素点成績を基に、たとえば 100 ～ 90 点を S、89 ～ 80 点を A、……、60 点未満を D で不合格などと規定し、LG 評価をしてきた。こうして成績証明書に記載される LG は概括的なものだが、その内容の説明責任としては背景にある弁別力のある細かな素点評価とその根拠として試験やレポートの結果があって必要に応じて開示できる具合になっている。これはわが国の大学では一般的な成績評価の慣例ないしは教学文化になっているといってよい。

ところが、このごく当然のように受けとめられている 100 点満点での原成績評価の仕方は、たとえば GPA 制度が全般化している米国では稀な評価の仕方になっている。半田（2011）の調査結果によれば、100 満点評価をして LG との対応関係を規定している米国の大学の割合は約 12% であった（n=153）が、欧州（n=88）ではこれが 26% になり、日本の除くアジア（n=44）では 68% になる。

100 点満点の原成績が稀な米国の大学では、日本の大学で当たり前に受容されている 4 ～ 5 段階の LG というあり方は一般的ではない、100 点満点の原成績をもたない代わりに LG 自体を細分化してそれを原成績にしている。すなわち、各 LG に適宜 + や - の評価をもたせることで 7 ～ 11 段階の評価区分を設定している大学が大勢（先の調査では 65%）になっている。わが国の大学ではそうした評価のあり方も承知していながら、LG を細分化する動きは目立って現れなかった。理由は明らかである。わが国の多くの大学では LG の背後により細かく、しかも連続量評価を許容した 100 点満点の素点評価をしており、LG はその原成績を制度上、丸めた範疇評価として位置づけられてきたからである。したがって、厳格な成績評価の意味を評価行為における姿勢としての厳正さであり、加えて評価尺度としての細かさという点で厳密であるという意に捉えるとすれば、日本の教学文化は質的にそれに見合った性能をもっているといえることになる。

このように原成績としての LG を点数化した米国型の GPA を、そのまま日本に移入して、原成績を丸めた LG にあてがうとすれば、米国では起きない問題が生じてくることは明らかであろう。一定の幅をもつ点数評価の範疇に GP という 1 つの点数をあらたに与えることは、惜しくもその範疇に入ってしまったケースには不本意な再点数化が施されることになるし、幸いその範疇に滑り込めたケースにはかさあげ操作が施されることになる。これは偶然に左右された博打的な評価改めにほかならない。

しかし、この問題の原因は以上のように明白だから、丸める以前の原成績の点数をそのまま GP に変換すれば問題は生じない。しかも、この問題解決をはかれば、米国の多くの大学でおこなわれている 10 段階前後の等級評価よりも、ずっときめ細かな評価ができ、とくに複数の評価手段を組み合わせた多元的評価にも適切に対応できることになる。文字どおり厳格厳正な成績評価としての GPA 制度になるので、GPA 指標を進級、卒業、報奨、履修条件などさまざまな判定機会にも公正かつ安定的に用いてその機能を十全に発揮させていくことができる。その意味で、この問題解決をはかった GPA を、単純に LG と GP を結びつける従前の GPA と区別して functional GPA（以下、fGPA）と呼ぶ。

本研究ではこの fGPA に関連して 3 つの検証をおこなった。上述の原成績順位と従前の方法による GPA 順位との間で生じる齟齬の様子や程度はすでに半田（2006a、2008）が明らかにしている。だが、それらはどちらも 1 回のシミュレーションに基づく結果であった。そこでここでは第 1 に、標本数を増した上で複数回のシミュレーションを実施して、結果の安定性について検討した。

第 2 に、上述のように原成績と GPA の順位間での齟齬を解消するには、原成績を GP に線形変換する fGPA の方法をとればよい。そのためには幾通りかの算法がある。100 点満点の原成績 RS（Raw Score）を GP に変換するには、算式 GP=(RS-X)/10 を用いる。定数 X の値には後述するように複数の考え方がある。当然各々に結果は異なってくるが、いずれの値でも原成績と GPA の順位は一致し、順位攪乱の問題は解消される。ただし、算法の最適性は従前の GPA からの移行やそれらとの通用性を考える必要がある。順位攪乱を起こす方法との互換を考慮することは一見矛盾しているようだが、集団での GPA の平均値といった大域的な代表値については攪乱が相殺されるので、評価指標として妥当な値になる。よって、全学や学部といった集団内での代表値を頼りに互換性を検証できる。

また、原成績合格域の得点範囲（100～60）をそのままGPの4.0～1.0に線形変換する方法（以下、直接法）も原成績とGPAの順位攪乱を防ぐ方法として考えることができる、そのため、この算法についても同様に比較検討した。

第3に、実際の大学における成績分布はしばしば得点の高い方向に偏る傾向がある。そこでそうした高評価に傾いた条件下でも同様のシミュレーションをおこない上記のfGPA算法の最適互換性について比較検討した。

検証 1

上述のとおり、LGごとにGPを対応させてGPAを算定すると100点満点の原成績順位とGPA順位が一致しなくなる。その様子はすでに60名、あるいは600名のシミュレーションにおいて明示されている（半田、2006a、2008）。だが、それらはシミュレーションを1回実施した結果であった。別の条件で反復した場合も同様の結果になることは十分に推察できるが、事実としてそれを確認してはいない。そこでここでは設定集団をさらに増やして1000名にしたうえで、乱数生成による成績得点を変化させながら、10回のシミュレーションを実施し、その結果を検証する。

方法

架空の学生1000名が4年間学修したことを想定し、70～79科目（各2単位として140～158単位）を履修し、各々の成績を得た状況を設定する。1000名は各100名からなる7群、50名からなる2群、200名からなる1群に分け、各群の学生の履修科目数は同等にして、取りうる素点範囲を100名群は7様に、50名群は2様に、200名群は1様に設定した。100名群については素点範囲[1] 100～80点（素点幅20の成績上位群）、[2] 90～70点（同20の中位群）、[3] 80～60点（同20の下位群）、[4] 100～70点（同30の上位群）、[5] 90～60点（同30の下位群）、[6] 95～85点（同100の上位群）、[7] 75～65点（同10の下位群）、50名群について[8] 100～90点（同10の上位群）、[9] 70～60点（同10の下位群）、200名群について[10] 100～60点（同40の全範囲群）であった。

素点幅を全体で100～60点にしたのは現在運用中のGPA制度の大勢で採っている合格素点範囲に準じたためである。この基準では60点未満は不可となり、GPは0になる。GPA算定ではこれを含める点に特徴がある。不合格科目がGPAスコアに損失をもたらすことで学生に不合格回避を動機づけ、むやみな過剰履修を避けさせる仕掛けになっている。したがって、不合格科目がある場合、それを含めて算定しない原成績の順位とGPAの順位は当然違ってくることになる。だが、この点はここで問題にしている算法上の過誤による不公正の発生とは異質のことである。だからこのことが影響しないよう、当シミュレーションで取りうる成績評点の範囲は合格圏内に収まるよう設定した。

手続き

1000名分の全科目の100点満点整数値による成績評点をコンピュータの乱数発生で定めた。その素点からLGを介してGPを求め、各科目の単位数を2にして各人のGPAを算出した。素点とLGとGPの関係は多くの大学に認められる規程に沿い100～90をSで4、80点台をAで3、70点台をBで2、60点台をCで1とした。次に各人の素点平均点とGPAの順位を各々求め比較した。この手続きを乱数発生の条件を変化させながら、10回（以下、回をセッションと呼ぶ）実施した。

結果

現在、大勢で用いられている方法で求めたGPAの順位と原成績順位について10セッションにわたるシミュレーションでの比較の結果、GPA順位が原成績順位と異なった学生は全セッション平均で1000名中969名、順位攪乱の発生率は平均96.9%、レンジ97.7～96.2%であった。順位が10ランク以上変動した学生の割合は平均50.6%に及び、レンジ54.0～47.9%であった。最大の順位変動は97ランクであった。半田（2008）が600名集団で実施したシミュレーションでは、順位攪乱の発生率が94.3%、10ランク以上変動したケースは43.6%、最大変動は88ランクであった。今回の結果は一層大きな攪乱になったことがわかる。これは対象集団の大きさが増せば、その分、一定の合格域における個々の成績の数が増し、格差が小刻みになることによる。1000名というと大規模大学の1学部1学年、あるいは小規模大学の1学部の規模である。したがって、全学での比較ということになれば、ここにみた値はさらに変動が強調されよう。

上記のようにセッション間ではとくに大差のない安

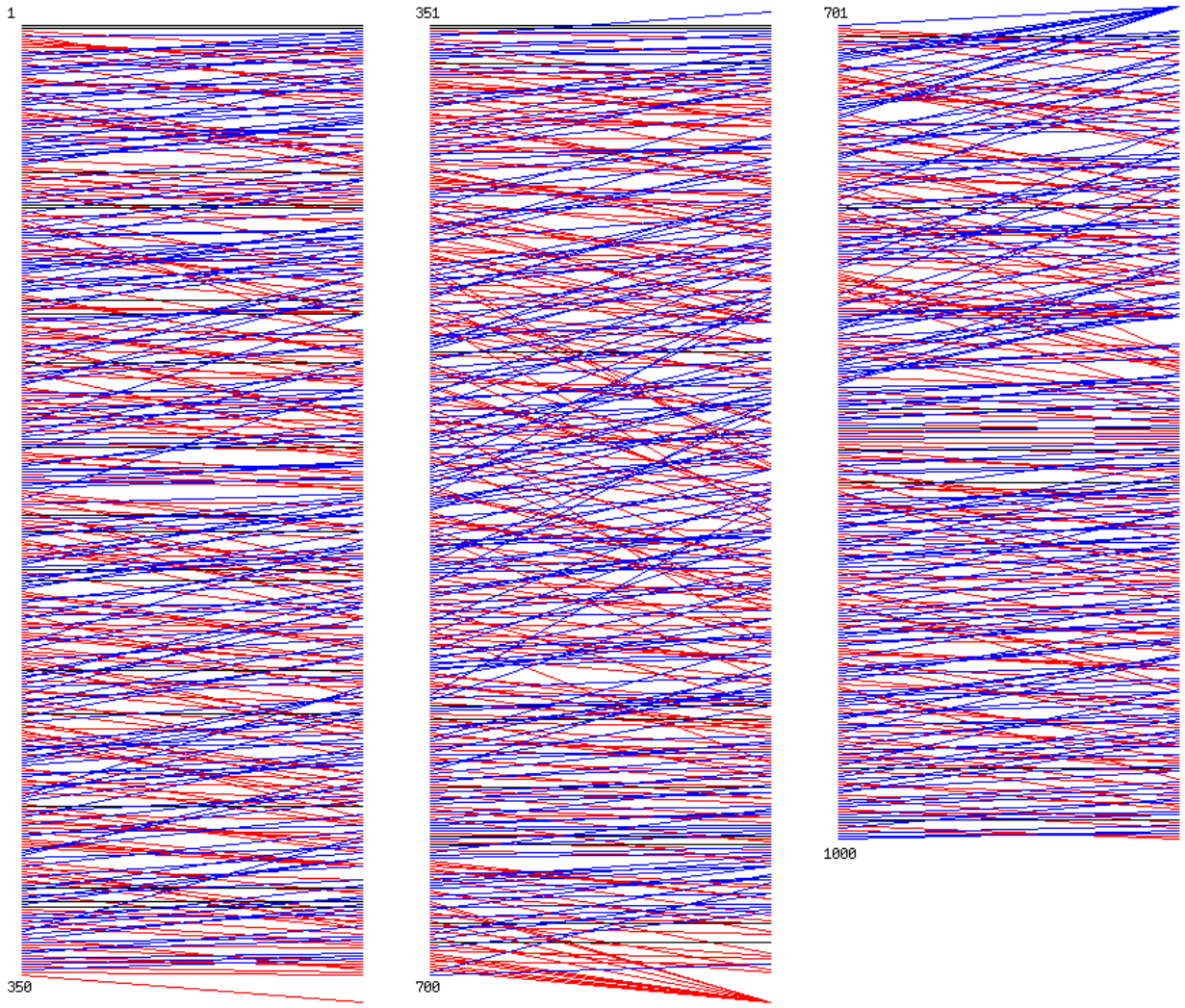

Figure1　原成績順位と GPA 順位の変動の様子

定した結果が導かれた。そのため、順位変動全体の様態はセッション 1 の結果で代表させ Figure 1 にその様子を示した。この図は各線分が各学生を表し、左端が原成績順位、右端が GPA 順位、順位上位から左 2 列は 350 名ずつ、右 1 例は 300 名で計 1000 名全体の成績ランキングを表出している（列を越えて変動した場合は表記上の便宜から列外の一点に集約表示した）。右上がり線分は原成績の順位よりも GPA での順位のほうが上昇した学生、右下がり線分は下降した学生、上下動がなく順位が一致した例は全体の 3.8% しかなかった。

この図に表された順位変動を一瞥すれば、これが攪乱以外の何ものでもないことは十分了解できよう。このような攪乱は決してシミュレーション上の架空の絵図ではない。半田（2011）は 2004 年の時点で GPA 制度を運用していた国内 47 大学におこなった調査で 100 点満点での原成績評価をおこなっていた大学が 96% に及んでいた。このことからすると。こうした攪乱は、日本で GPA 制度を運用している大学の 9 割方で生じている事態ということになる。

検証 2

原成績と GPA の順位間の齟齬を解決するには上述したように、GP を規定するにあたり LG を介さず、直接原成績の値を GP に線形変換する fGPA の方法をとればよい。

fGPA の算法には複数の方法がある。すでにそれら異なる算法で GPA 制度が運用されている事例もある。たとえば、100 点満点の原成績 RS から直接 GP を線形変換する方法として、GP=(RS-X)/10 という算定式がある。この X について、多くの GPA 制度で各 LG に対応させている GP（たとえば A であれば 3.0）が当該の LG 区間(たとえば､A であれば 80 点台の区間)のどの点を代表した値と考えるかという点で、これを区間の最大値におくなら X は 60（以下、X=60 法)、

最小値におくなら X は 50（以下、X=50 法）、中点におくなら X は 55（以下、X=55 法）あるいは 54.5 といった値になる [*2]。

これらは原成績順位が GPA 順位でも保持されるという点では、どの方法でも同等で、成績順位攪乱の問題は解消される。実際のケースとしては、X=55 法を用いている事例に同志社女子大学や静岡大学、お茶の水女子大学、54.5 を用いている例に宮崎大学、X=50 法を用いている例に西南女学院大学や徳島大学工学部（英 ,2003）がある。

ただし、すでに成績攪乱の問題を抱えている GPA 制度から、その問題解決をはかった GPA 制度に移行するという現実的な課題を考えたときには、それまでおこなわれてきた算定値との間の差異を、少なくとも学部単位や全学的な平均値といった大域的にみた場合の代表値においてはできるだけ最小限になるような互換性の高い方法をとる必要がある。また、海外とくに米国で全般化している GPA 指標との通用性という観点からも、それらの算定に比較して明らかに割り増しされたり、割り引かれたりすることがない高い互換性が求められる。

すでに LG 区間の素点代表値を上記の最大値、最小値、中点の 3 通りにおいた場合での互換性については半田（2008）が 600 名集団のシミュレーションにおいて比較検討している。その結果、区間代表値を最大値付近においた X=60 法では LG を介した従前の GPA 算法に比較して GPA 値の集団平均が 0.55 小さくなり、現実的には通用性において学生に不利益が生じることが認められた、反対に、区間代表値を最小値付近においた X=50 法では従前の GPA 算法に比較して GPA 値の平均が 0.45 大きくなり、明らかな割り増し評価になるという不都合が認められた。これらに対して区間代表値を中点付近においた X=55 の方法では、従前の GPA 算法と比較した GPA 値平均の差異は 0.05 にとどまり、高い互換性が認められた。この検証ではあらたに 1000 名集団で条件を変えつつ 10 回のシミュレーションをおこない、以上の結果の再現性を検証する。

Table1 fGPA 算法別の LG 媒介法との GPA 平均値差異

	(X=55)-LG 媒介法	(X=50)-LG 媒介法	(X=60)-LG 媒介法	直接法-LG 媒介法
session 1	-0.039	0.460	-0.540	-0.046
session 2	-0.037	0.463	-0.537	-0.047
session 3	-0.038	0.461	-0.538	-0.048
session 4	-0.037	0.463	-0.537	-0.046
session 5	-0.037	0.463	-0.537	-0.048
session 6	-0.036	0.464	-0.536	-0.048
session 7	-0.036	0.464	-0.536	-0.048
session 8	-0.035	0.465	-0.535	-0.044
session 9	-0.034	0.466	-0.534	-0.044
session 10	-0.037	0.463	-0.537	-0.043
平均	-0.037	0.463	-0.537	-0.046

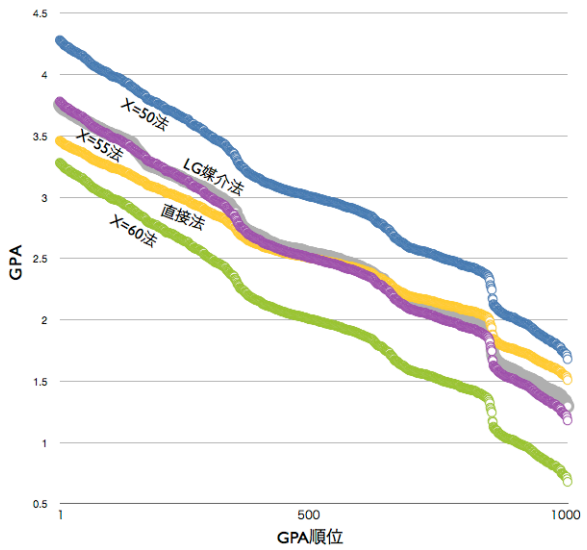

Figure 2　算法別に GPA 値を降順ソートした順位分布

また、これらの算法に加えて、従前の LG と GP の対応関係を考慮せず、原成績の合格域である 100 ～ 60 点をそのまま GP の 4.0 ～ 1.0 に線形変換する方法も原成績順位と GPA 順位の攪乱を防ぐ方法として考えられる、そのため、この算法も加え、結果を同様に比較検討する。

方法

想定学生集団を 1000 名にした 10 セッションのシミュレーションを実施した。その方法は検証 1 と同様であった。

手続き

1000 名の成績評点の生成は検証 1 と同様におこなった。その素点からつぎの 5 つ方法でそれぞれ GPA を求めた。(1) 検証 1 でおこなった現状多くの大学で採られている素点区間と LG、および GP の関係を規定し GPA を算定する方法、以下 LG 媒介法と呼ぶ。(2) 上述の X=55 法、(3) 同 X=60 法、(4) 同 X=50 法、(5) 原成績合格域 100 ～ 60 点を直接 GP の 4.0 ～ 1.0 に線形変換する方法、これは GP=(RS-60)×0.075+1 で求められる。以下これを直接法と呼ぶ。

結果

LG 媒介法は成績攪乱を起こす問題をもつ。だが現在、大勢で用いられている算法であるため、その問題解決にあたる移行算法については、少なくてもその全体平均値に表れるような大域総括的な値については格差が最小に留まるような最適互換性を保証できることが望ましい。したがって、シミュレーション結果については 1000 名集団における GPA 値の平均値に関しての LG 媒介法との比較が検証の焦点になる。その結果を示したのが Table 1 である。

10 セッションのシミュレーション間の差異はどの算法も概ね小数点第 3 位における変動にとどまり、安定的な結果を示した。そのため 10 セッションの平均値で全体の結果を語ることができる。それをみると、半田（2008）の 600 名集団でみた結果を差異の大きさも含めて追認した。すなわち、X=50 法では LG 媒介法よりも 0.46 程度の無視できない増大となり、反対に X=60 法では 0.54 程度の明らかな減少となった。これらに対して X=55 法での同値は 0.04 程度に留まり、明白に高い互換性を確認できた。また、あらたに算定した直接法でも LG 媒介法との平均値差異は 0.05 程度となり、X=55 法と並んで高い互換性が示唆された。

つぎに 1000 データすべての値について、その分布の様態を比較した。これは 5 つの算法別に GPA 値を降順に並べ替え、縦軸に GPA 値、横軸に 1 ～ 1000 位の順位をとって布置したグラフ（Figure 2）によって確認できる。グラフ表記上、比較の主対象となる LG 媒介法は重なりでみえにくくなるが、最も背後に薄いグレーで描出してある。結果は一目瞭然だが、上記の平均値で確認したように、X=50 法では値の全域において LG 媒介法よりも 0.5 程度高い値になっていることがわかる。反対に X=60 法では全域において LG 媒介法よりも 0.5 程度低い値になっている。これらに対し、紫色で描出した X=55 法は値の全域にわたりほとんど LG 媒介法の曲線に重なっており、平均値に示された高い互換性を確認することができた。

また、平均値では X=55 法と並んで LG 媒介法との差異が小さかった直接法は GPA 順位の中程（500 位前後）では LG 媒介法、X=55 法との重なりが認められるものの、順位が高くなるほど LG 媒介法から離れ、GPA 値が逓減して X=60 法の算定値に近づき、逆に順位が高くなるほど LG 媒介法の値から逓増して末尾では X=50 法の算定値に近づくことがわかった。平均値で LG 媒介法との差異がわずかであったのはこの両極方向での増減が相殺されたためであったことがわかる。換言すれば、直接法を用いた場合は、原成績が高いケースほど GPA 値が割り引かれ、原成績が合格最低点に近づくほど GPA 値が割り増されるわけで、はなはだ具合の悪い調整が施されることになる。このことから直接法について上記の平均値で示唆された LG 媒介法との高い互換性は否定された。

以上より、素点原成績と GPA 順位の齟齬をきたさない fGPA 算法のうち、素点から 55 を引き、10 で除する方法が従前の問題含みの GPA 算法からの移行に際し、最適互換性があることがあらためて確かめられた。また従前の合格域において GPA 値がとりうる範囲（4.0 ～ 1.0）が一致し、集団平均値で見比べても従前の算定値から大きく逸脱することがなく、一見したところ互換性が高くみえた直接法は個々の算定値の全体を見わたすと、成績の高い方向と低い方向で異なる方向の重みづけがなされ、根拠のない調整、すなわち不公正をもたらす算法であることを明らかにすることができた。

検証 3

検証 1、2 でおこなったシミュレーションでは学生がとる成績得点は方法に示したように、全体の成績分布が概ね合格域の中点を境に左右に大きな偏りが出ないよう設定した。ところが、多くの大学の成績評定の実際においては、全学や学部単位といった大域的なくくりでの成績分布をみると、LG でいえば A（優）の度数が最も多く、分布が合格域の中点よりも高い方向に偏っていることが稀ではない。たとえば、最近では IR（Institutional Research）の動きが活発化してきていることに伴い、学部や全学規模での大域的な成績分布などが公表されるケースもめずらしくなくなってきている。その代表的なところを比較の便宜上、LG 表記（秀～不可など）を S ～ D という表現に合わせてみれば、たとえば同志社大学（2006）の一年次生 2005 年度実績をみると、概数で S が 15%、A が 27%、B が 25%、C が 17%、D が 13%、その他が 3% となっている。A、B がほぼ同率だが、わずかに A の割合が多く最大値になっている。信州大学の共通教育科目（647 科目）の 2008 年度前期実績（矢部、2009）では、S が 15.0%、A が 34.6%、B が 28.3%、C が 17.5%、D が 4.7% で、A が最多になっている。この報告ではクラス規模別にも集計されている。200 名未満、40 名以上のクラスではほぼこの全

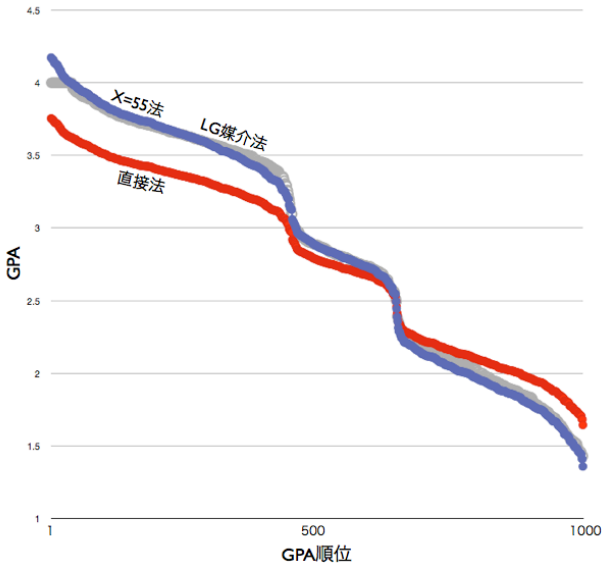

Figure 3　算法別に GPA 値を降順ソートした順位分布

体の割合に沿っているが、200 名を超えると A の割合が 42.0% に増し、20 名未満になると S の割合も増す（S が 27.5%、A が 39.4%）ことが認められている。そこでこうした現実的にみられる高い成績評価の方向への偏りに則したシミュレーションをおこなって fGPA の、とくに X=55 法と直接法の結果について再度、比較検討する。

方法

想定学生集団を 1000 名にした 10 セッションのシミュレーションをおこなった。方法は検証 1 と同様であった。

手続き

1000 名の成績評点の生成は検証 1 と同様におこなった。検証 2 で検討した X=50 法と X=60 法については、LG 媒介法との互換性が低いことが明白となったため除き、LG 媒介法、X=55 法、直接法の 3 通りを比較した。

結果

本検証の目的にそくし成績の高い方向に偏らせた分布は 100 点満点の素点平均値でみると、10 セッションのシミュレーションで平均 83.3 点、レンジ 80.6 ～ 85.7 であった。検証 2 では平均 80.3 点（合格域の中点は 80.0 点）であったからセッションにより程度は異なるが全体に意図どおり成績の高い方向に偏った分布を形成できた。ちなみに LG の割合でみると、10 セッションでの S のレンジは 42.2 ～ 26.4%、A のレンジは 28.8 ～ 24.5% であった。

10 セッションにわたるシミュレーションの結果、GPA 平均値について、X=55 法と LG 媒介法との差異は平均 0.019、レンジ 0.013 ～ 0.037、これに対して直接法と LG 媒介法との差異は平均 0.101、レンジ 0.052 ～ 0.157 で、すべてのセッションにおいて X=55 法の方が差異が小さかった。Figure 3 はセッション中、LG 媒介法との差異がセッション全体の平均値付近であったセッション 6 について、Figure 2 と同様 1000 データすべての値について、3 つの算法別に GPA 値を降順に並べ替え、縦軸に GPA 値、横軸に 1 ～ 1000 位の順位をとって布置したグラフである。

これをみると X=55 法は検証 2 での結果と同様、LG 媒介法とほぼ重なっていることが確認できる。他方、直接法は順位の両端にいくほど LG 媒介法との乖離が大きくなり、とくに高成績ほどその傾向が強く現れ、順位でみると 450 位付近から差異が大きくなることが認められた。

以上から高評価方向に偏りがちな成績評価の状況にあっては、一見互換性が高くみえる直接法は一層、互換性を低下させることが確かめられ、とりわけ高い成績域において広範にわたり不利な GPA 値が算定されてしまうことが明らかにされた。他方、X=55 法は分布の形状によらず従前の方法との間の互換性が成績順位の全域において高く維持されることがあらためて確認された。

考察

わが国における GPA 制度の導入と運用についてのこれまでの経緯はおおよそ 3 相にわけてつかむことができる。第 1 相は一部の大学で地道に運用されていた時期である。「地道に」とは学内的には定着していても、とくにその運用を対外的にアピールすることなく、しかも比較的早い時期から長年にわたり使われてきていたという意味である。「一部の」とは指折り数える程度の数の私学を指す。例をあげれば国際基督教大学や東京神学大学、名古屋商科大学などで、これらの大学での GPA 運用はすでに 30 年以上の実績をみている。

1 相から 2 相への画期は 2000 年であった。この年の 2 年前、大学審議会が『21 世紀の大学像と今後の改革方策について ― 競争的環境の中で個性が輝く大学』という標題の答申を出し、そのなかで厳格な成績評価の実施を話題にし、その具体的な例示として

GPA 制度をあげたことが転機を導いたとみることができる。あわせてその時期、1993 年に設置された青森公立大学が GPA 制度を導入してその値を基準にした退学勧告を実施したことが社会的にも話題になり、大学関係者の頭に GPA の文字が引っかかるようになった。実際、半田（2012）は 2004 年に GPA 運用中の 47 大学に質問紙調査を実施した結果として、2000 年以降に同制度を導入した大学数が増大したことを明らかにし、その割合は調査対象大学数の 85% を占めていたことをみている。

この第 2 相は全国の大学に GPA 制度の運用が広まっていった時期で、国内全大学のほぼ半数が（まずは制度上）運用に至る期間であった。2009 年に文部科学省が 752 大学に実施し、100% の回答率を得た調査結果では国立大学法人で 65%、私学の 47% が GPA 指標を用いている実態が明らかになり、運用大学が概ね半数に至る状況が確認された。現在は様子をうかがってきた未導入大学にも相次ぎ導入の動きが出てきている。冒頭でも触れたように、GPA は米国の大学では全般化し、日本以外のアジアの大学でもわが国以上に浸透している成績評価指標となっている。それだけに、国際的な学生流動における通用性の面でも実務上の要請を充たす必要がでてきている。

こうしたことから、2010 年を分水嶺として、今や GPA 制度は導入の時期から質的に様相を変え、その活用を本格化する第 3 相に入ったとみることができる。活用という点では一橋大学や北九州大学のように、一定の GPA 値を卒業要件にするなど積極的な活用例も現れ出している。だが、大多数の大学での用途は主として学生に対する個別の学修指導や奨学金や授業料免除対象者の選定基準のひとつとして用いる程度に留まっており、それがもつ機能性はまだ十分に発揮されていない現状がみられる（文部科学省、2010）。

活用が積極化していない原因のひとつに、これが厳格な成績評価の文脈下でその活用を含めた具体策として提起されてきた制度であるにもかかわらず、多くの大学で実際になされてきた原成績評価と GPA 指標による評価の順位が齟齬をきたし、実際は厳格厳正な成績評価どころか、著しく不公正な評価に結果するという問題を抱えていることがあげられる。この問題を解決しないまま、一定の GPA 値を進級や卒業の要件にしたり院試の選抜や退学勧告の発動基準に用いたりするなら、もともと評価された成績では要件を充たしているのに、不備な GPA 算定で未充足になったり、その逆が生じるなどの不都合が発生することになる。したがって、第 3 相での第一の課題は GPA 指標の機能特性を十全に発揮させてこれを学修全般に活用していくために、原成績を的確に反映する算法に転換して問題を解決すること、あるいは解決した算法での GPA 制度を導入することに向けられることになろう。

その算法にはいくつかの方法がある。GPA を学修における基準指標として幅広く活用していくには、これまで運用してきたいわば概括的指標としての GPA から、それとできるだけ互換性を保ちながらも厳格な成績評価指標として原成績に律動した GPA に移行できることが望ましい。具体的にはわが国において最も多くの大学で用いられてきている 5 段階の LG に 4.0 から 0.0 までの GP を対応させて算定する GPA を基準にして、それと十分に高い互換性がある方法を採る必要がある。

本検証ではその方法が 100 点満点の原成績から 55 を引き 10 で除して GP を求める functional GPA の算法であることがあらためて確認された。とくに現実的に成績評価が高い方向に偏倚しがちな大学の成績評価にあっても、この算法は従前の算法との間で十分な互換性が維持できることも明らかにされた。

ところで、原成績から 55 を差し引く functional GPA の方法では GP の最大値が原成績 100 点のときで 4.5、合格域の最小値は原成績が 60 点のときで 0.5 になる。一方、従前の多くの GPA の算定では GP の最大値は 4、合格域の最小値は 1 である。ここに相違があるにもかかわらず、両者の互換性が諸法のなかでも最適になるのはなぜだろうか。とくに当検証でみたように、従前の方法にあわせて 4.0 ～ 1.0 の GP にそのまま原成績 100 ～ 60 点を線形変換する直接法よりも互換性が高くなるのはなぜだろうか。

従前の方法では各 LG に対応した一定の原成績区間に対して GP が付与される。よって GP 値は各成績区間の代表値とみなせる。したがってこの値を区間の最大値や最小値におくなら、それぞれの区間の他方の端点では代表値との差異を最大化させることになる。そのため区間代表値としての機能を果たしがたくなる。必然的に代表値は区間の中点付近におくことが適切になる。すると GP 最大値の 4.0 を素点原成績に点対応させてみれば、100 ～ 90 点の区間の中点にあたる 95 点におくことが至当になる。そうなると素点が 100 点のときはもともと従前の方法でも想定上の GP は 4.5 になっていることがわかる。他方の合格域最下位の LG 範囲についてもこれと同様のことがいえる。

これに対して直接法では素点の 100 ～ 60 点を GP

の 4.0 ～ 1.0 に変換するため、素点最大値の端点に近づくにしたがい GP が割り引かれ、反対に素点合格域の最小値に接近するほど GP が割り増されることになる。その結果、本検証でみたように成績尺度の両端点域で互換性を失することになる。

最後に、100 点満点の原成績と LG 媒介法による GPA 値の順位齟齬の問題解決として、100 点満点による評価をせずに、LG 評価そのものを原成績にするという手段をとることについて言及しておく。その際、LG の 5 段階評価ではいささか大まかにすぎるため、米国流に適宜、各 LG に + や - の評価も加えて 10 段階以上の評価尺度に切り替えるということも考えられよう。これは LG を媒介するのではなく、LG そのものが原成績になるのだから、それと GP 値との齟齬の問題は生じなくなり、米国型の GPA 制度を運用することになる。

そもそも 100 点満点での成績評価といってもその尺度を十分に活かし得た評価ができていたかといえば、それを疑問視する声は常に稀ではなかった。だが、100 点満点評価はこれまでわが国の大方の大学の常識としていわば教学文化的に根付いてきたことである。しかも、これは連続量での評価を許容しているから、100 点というよりも完全な達成度を 100% として、そこへの相対量で評価することを可能にしている。そのため実質的には 10 段階評価でも 5 段階評価でも個々の科目の必要に応じて比率変換して運用することに便利な尺度であった。また、一方で昨今では複数の評価手段を用いた多元的な評価が推奨されている。そのこともあって、一度の試験やレポートによる評価だけでなく、種々の観点からの評価を総合して最終評価を出すことも多くなっている。その場合は実際上、満点が 100 を簡単に超えるようなこともある。だが、満点 100% の標準スケールはその場合でも変換して考えやすい尺度として高い通用性をもっている（半田、2009）。

すでにこうした教学環境がある中で、簡単にこれまでの方法を捨て他国の方法を取り入れるなら、国際的には後発、追従に甘んじるだけである。ここにある問題はわが国でとってきた従前の方法にひと手間を加えるだけで、国際的な観点に立ってみてもかえってよりすぐれた方法、すなわち個々の授業科目で採ってきた丁寧な評価方法を活かすかたちで厳格厳正な評価指標を提起できる機会として捉えうるものである。100 点満点での原成績評定は米国の大学では馴染みが薄いが、アジアの大学では十分に馴染んでいる評価方法である。その点からも、これを基盤にした GPA の運用は質の高い成績評価制度として国際的なイニシアティブをとりうる性質のものとしてある。わが国においてあらたな段階に入っている GPA 制度における問題解決は国際的な通用性の観点からも、戦略的な含みをもって推進することができるものとしてある。今後は同制度について諸外国との間の比較調整を進め、そのなかで最適な方法とこの指標の一層の活用法を構築していくことが課題になろう。

注

*1　本論で言及するアジアの代表的な大学とは、半田（2011）で調査対象となった日本を除く韓国、インド、中国、台湾、タイなど 9 ヵ国の 44 大学を指している。

*2　X を 55 にする場合と 54.5 にする場合の差異はほとんどないが、素点評価について小数を許容した連続量評価にしている場合は前者、整数値のみの評価にしている場合は後者にするほうが従前の大方の GPA 算法との互換性がより高くなる（半田、2012）。

参考文献

中央教育審議会（2008）「学士課程教育の構築に向けて」答申 .

大学審議会（1998）「21 世紀の大学像と今後の改革方策について―競争的環境の中で個性が輝く大学」答申 .

同志社大学（2006）「2006 年度 文部科学省「特色ある大学教育支援プログラム (特色 GP)」採択『情報環境の整備と成績評価の厳格化 - 学修支援システム DUET と GPA 得点分布公表 -』概要」同志社大学 .

英崇夫 (2003)「工学部におけるアウトカムズ評価 — GPA 評価とプレゼンテーション評価」高等教育情報センター編 2003『成績評価の厳格化と学習支援システム』地域科学研究会に所収 .

半田智久（2006a）「GPA 制度 : カテゴリー錯誤の問題と解決」大学教育学会誌 28,117-125.

半田智久（2006b）「GPA 制度に対する関心と導入の状況」静岡大学教育研究 ,2,1-9.

半田智久（2008）「機能する GPA とは何か」静岡大学教育研究 4,1-30.

半田智久（2009）「成績の絶対的相対評価を支援するスマートスコアについて」大学教育学会第 31 回大会発表要旨集録 , 200-201,2009.

半田智久（2011）「GPA 制度に関する国際調査研究」高等教育研究 ,14,287-305,2011.

半田智久 (2012)『GPA 制度の研究 : functional GPA に向けて』大学教育出版 .

文部科学省（2010）「大学における教育内容等の改革状況について（平成 20 年度）」文部科学省高等教育局大学振興課大学改革推進室 .

矢部正之（2009）「成績評価分布を用いた Community-based FD の概要 (報告)」信州大学全学教育機構 .

2012 年 2 月 20 日　受稿

2012 年 3 月 16 日　受理