

# GPA の抱える問題とその解決法を探る —心理計量学的観点から—

○脇田貴文（関西大学社会学部）

近年、日本の多くの大学でも学生の成績を表す指標として、Grade Point Average（以下、GPA）が用いられている。GPAは、科目の成績を、秀（A+）：4点、優（A）：3点、良（B）：2点、可（C）：1点、不可（F）：0点で得点化し、その平均を求めたものである。本稿では各科目の得点（0点～4点）をGrade Point（GP）と表記し、その平均をGPAとする。GPAの利用法は、学生が自身の成績を知ることだけでなく、大学側においてもInstitute Research（IR）やカリキュラム評価、入試政策の評価などのアウトカム指標としてGPAが用いられることが多い。このような検討を行った例として日本では大久保（2010）など、海外においては、Koretz（2014）やShaw et al.（2012）などが挙げられる。

しかし、GPAには後藤（2006）や林（2010）でも指摘されているようなさまざまな問題点があり、その解決方法としてfunctional GPA（半田，2011）なども考案されている。稲垣・能上（2013）では、relative GPAとstandard score GPAの比較研究がなされている。

本発表ではGPAが有する問題点のうち、心理計量学的観点から1) 各科目の成績を表す得点は加算して良いのか、2) その平均を求めても良いのかの2点を提起し、その検証、解決方法を探ることを目的とする。

はじめに、科目の成績を表す得点（GP）を加算することが適切であるかという問いに対しては、測定次元がどのようなものかを検討する必要がある。具体的には、一般教養科目と専門科目、語学科目の各成績をすべて足し合わせることは妥当であるといえるのだろうか。これは英語・数学・国語の得点を足しても良いのかという議論と同様である。先述したような利用法を考慮すると、あらゆる分野の履修・修得科目からなる学生の成績が1つの指標で表されることには抵抗がある。例えば学部が専門教育科目のカリキュラム評価をする場合、一般教養科目が含まれるGPAではなく、専門科目のみで求められたGPAを使うことが望まれるのではないだろうか。本発表ではこの点に関して、分野ごとのGPAを算出しその関連を検討する。また、因子分析を用いて測定次元の検討を行う。

次に、各科目のGPの平均が意味を持つのかという問いに関しては、GPがStevensの尺度水準の考え方において、間隔尺度水準にあるのか、比率尺度水準にあるのかという議論になる。この点に関しては、Item Response Theory（項目反応理論；以下IRT）を利用することがその解決につながる可能性がある。GPAの検討にIRTを用いた例としてはYoung（1990）、Stricker et al.（1994）が従前のGPAとIRTに基づくGPAの比較を行っているが、それ以降IRTにも続くGPAに関する検討はほとんどなされていないようである。本発表では、改めてIRTを用いた検討を行う。

## 方法

私立大学A学部の卒業生537名のデータを用いた（これは複数年にわたるデータである）。全履修科目数は609であった。うち200名以上の履修があった科目数は22科目であった。分野別GPAとして、共通教育科目（174科目）、基礎必修科目（7科目）、専門科目（139科目）、英語科目（120科目）それぞれでGPAを求めた。また全科目によるGPAを算出した。

また、200名以上の履修者があった22科目に関して因子分析（最尤法・プロマックス回転）を

行った。この結果をもとに、7科目（専門A群）に関して、多値型IRTモデルの1つである Generalized Partial Credit Model (Muraki, 1992) を用いて項目パラメタの推定を行った。パラメタの推定には PARSCALE (Muraki & Bock, 2003) を用いた。

## 結果と考察

全科目による GPA と分野ごとの相関係数を求めた (Table 1)。一般教養科目と専門科目の相関は.679であった。この値を高いと評価するか低いと評価するかは議論が分かれるところではあるが、GPA のあり方を示す1つのエビデンスとなるだろう。また、200名以上の履修者があった22科目に関して因子分析（最尤法・プロマックス回転を行った）。固有値の減衰状況は、6.64, 1.70, 1.55, 1.17, 1.11, 1.02, …であり、3因子構造であると考えられた。因子間相関を考慮すると、分野を考慮せずに全科目の GPA のみを扱うことは、重要な情報を見落とす可能性があるのではないだろうか。

続いて Table 2 の結果より、試行的に第1因子の専門Aの7科目に関して、IRT分析を行った。各項目パラメタと各評価の割合を Table 3 に示した。そのうち Location パラメタを Figure 1 に示した。さらに、科目ごとの評価に関して詳細に検討するために、Wakita et al. (2012) で提案した方法を用いて、専門A1、専門A3、専門A5の科目ごとに評価間の距離を算出し Figure 2 に示した。なお、詳細は後述するが、category パラメタの s.e. も大きいことが指摘できる。また、専門A2、専門A4、専門A6、専門A7では category パラメタの順序性も崩れているため、このIRT分析結果は参考に留めるべきであろう。

Table 1 分野別 GPA 間の相関

	基礎 必修	専門 科目	外国語	全科目
一般教養	.571	.679	.530	.771
基礎必修		.709	.711	.797
専門科目			.736	.956
外国語				.840

Table 2 22科目の因子分析結果

	F1	F2	F3	$h^2$
専門A1	.870	-.190	-.080	.510
専門A2	.690	-.220	.040	.330
専門A3	.690	-.090	.060	.430
専門A4	.640	-.050	.040	.390
専門A5	.630	-.090	.080	.370
専門A6	.600	.010	.030	.390
専門A7	.590	-.080	.100	.350
専門B1	.560	.110	-.110	.350
専門B2	.510	.070	-.020	.300
専門B3	.330	.150	-.090	.170
専門B4	-.440	.920	.100	.510
一般教養科目1	-.090	.610	-.130	.270
教職科目1	.050	.610	.040	.440
教職科目2	.090	.480	-.060	.280
専門B5	.150	.440	.110	.360
専門B6	.340	.350	-.080	.370
専門B7	.200	.340	.050	.270
専門B8	.190	.280	.020	.190
専門B9	.260	.280	-.020	.230
ゼミ1	.040	-.020	.750	.590
ゼミ2	-.050	.100	.750	.580
ゼミ3	.080	.110	.590	.480
因子間相関	F1	.690	.480	
	F2		.380	

Table 2 および Figure 3 の結果から、最も難易度の高い科目は専門 A7 であることが、難易度が低い科目は専門 A2, 専門 A3 であることが分かる。本稿では具体的な科目名、科目内容を明示できないが、科目内容や科目特性を考慮して検討することで有益な情報が得られるだろう。

Table 3 7 科目の IRT 分析結果及びおよび各評価の割合

	IRT								各評価の割合								
	slope		location		category				不可	可	良	優	秀				
					1	2	3	4									
専門A1	0.806	(0.078)	-0.533	(0.069)	1.120	(0.199)	1.026	(0.142)	-0.124	(0.095)	-2.022	(0.122)	0.06	0.09	<b>0.28</b>	<b>0.43</b>	0.13
専門A2	0.515	(0.052)	-0.953	(0.096)	0.203	(0.373)	1.260	(0.301)	0.346	(0.171)	-1.809	(0.150)	0.05	0.05	0.18	<b>0.44</b>	<b>0.28</b>
専門A3	0.732	(0.086)	-0.960	(0.081)	1.277	(0.318)	0.683	(0.185)	-0.298	(0.124)	-1.661	(0.118)	0.03	0.08	<b>0.22</b>	<b>0.38</b>	<b>0.29</b>
専門A4	0.666	(0.067)	-0.621	(0.088)	0.778	(0.281)	1.142	(0.225)	0.521	(0.139)	-2.441	(0.170)	0.06	0.07	0.20	<b>0.54</b>	0.12
専門A5	0.439	(0.043)	-0.448	(0.094)	1.327	(0.325)	0.595	(0.247)	0.380	(0.195)	-2.302	(0.222)	0.08	0.14	<b>0.22</b>	<b>0.41</b>	0.15
専門A6	0.538	(0.062)	-0.359	(0.097)	0.044	(0.359)	2.179	(0.310)	-0.282	(0.168)	-1.941	(0.216)	0.10	0.06	<b>0.34</b>	<b>0.36</b>	0.15
専門A7	0.523	(0.045)	-0.119	(0.088)	1.288	(0.278)	1.793	(0.201)	-0.019	(0.132)	-3.063	(0.248)	0.07	0.10	<b>0.35</b>	<b>0.42</b>	0.06

括弧内の数値はs.e.を表す。

各評価の割合では、0.2を越えるセルに下線を示した。

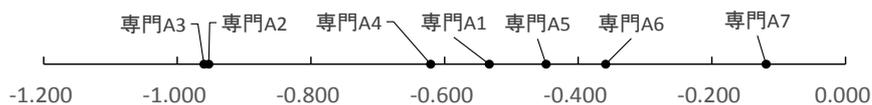


Figure 1 location パラメタのプロット

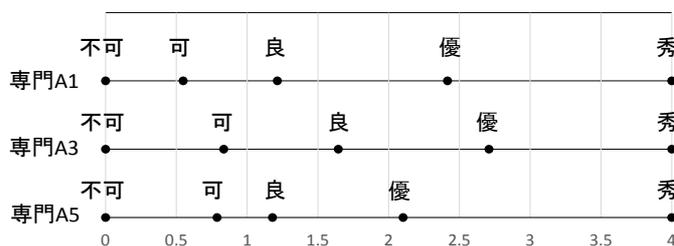


Figure 2 各評価間の距離

Figure 2 の結果から、各評価間の評価の近さを見ることができる。例えば、専門 A5 は可と良が比較的近く、その一方で、優と秀の距離が遠い。これは、可と良ではあまり差がなく、一方で秀をとることが難しい科目であることが分かる。実際のところ、GP は、秀・優・良・可・不可の 5 段階に対して、5 点~0 点の整数を与えたものである。したがって、Figure 2 において、1・2・3 の点に可・良・優の各評価があればこの GP の与え方は妥当であることになる。しかし、科目によって必ずしも整数値の GP を与えることが妥当とはいえないこともあることが分かる。

## まとめ

以上の結果を踏まえ、本稿の結論は 1) 分野ごとの GPA は有益な情報となりうること、2) IRT を利用することにより、科目に関してより詳細な検討ができることが示された。特に、各評価間の距離は、科目の難易度の有利不利の議論（成績のとりやすい科目のみを履修することによるメリット）に有益な情報となるだろう。また、本稿では触れていないが、IRT に基づき推定される  $\theta$  は個人レベルの評価に関しても可能性を広げると考えられる。

最後に、本稿で行った分析は無理をしていることも事実であり、結果の解釈には注意が必要である。以下に箇条書きで GPA の分析の難しさを挙げる。

- ・データ行列に関して、かなり欠測値の多いデータ行列になるため様々な制限が生じる。
- ・カリキュラム上もしくは時間割上、科目 A と科目 B の重複履修がない場合、相関が得られないため、因子分析などに制限が生じる。
- ・上記と同様、履修者数が少ない科目の場合、分析にのらない可能性が高い。
- ・本稿のデータは、すべて初回履修時の成績を用いている。再履修により成績評価が生じた場合、それを上書きするかしないかの判断が難しい。
- ・IRT 分析において、初年時教育の必修科目など不可が 0 人もしくは非常に少ない場合に項目パラメタの推定に影響が大きい。

### 引用文献（一部抜粋）

後藤和雄 (2006). GPA 定義の問題点とその一般化 鳥取大学教育総合センター紀要, **3**, 11-27.

半田智久 (2011). 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 高等教育と学生支援, **2**, 22-30.

林直嗣 (2010) . 大学教育のガバナンスと成績評価基準（上・中・下）＝質保証と GPA 制度＝ 経営志林, **47**.

稲垣麻央・能上慎也 (2013). 科目難易度を考慮した GPA について 2013 年電子情報通信学会総合大会 情報・システム講演論文集 1, 193.

Koretz, D. (2014). Predicting Freshman Grade-Point Average from High-School Test Scores: are There Indications of Score Inflation? (Doctoral dissertation, City University of New York).

大久保 敦 (2010). 大学入学者受け入れにおける高校調査書情報活用の可能性 -高校評定平均値と大学 GPA の分析から- 大学入試研究ジャーナル, **20**, 159-164.

Shaw, E. J., Kobern, J. L., Patterson, B. F., & Mattern, K. D. (2012). The validity of the SAT for predicting cumulative grade point average by college major. New York, NY: College Board.

Stricker, L. J., Rock, D. A., Burton, N. W., Muraki, E., & Jirele, T. J. (1994). Adjusting college grade point average criteria for variations in grading standards: A comparison of methods. *Journal of Applied Psychology*, **79**, 178.

Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the likert scale comparing different numbers of options. *Educational and Psychological Measurement*, **72**, 533-546.

Young, J. W. (1990). Are validity coefficients understated due to correctable defects in the GPA?. *Research in Higher Education*, **31**, 319-325.

**謝辞** 貴重なデータの使用を許可くださった学部の先生方、そして学部との間をつないでくださった森朋子先生に深く感謝申し上げます。