

新潟大学教養教育 GP データにみる成績評価の分布

教育人間科学部 藤村正司

What earns the marks?

Masashi Fujimura (Faculty of Education and Human Sciences)

This survey had two purposes. The first purpose is to draw the Box-plot distributions of GP (grade point) of first year course in Niigata University. The second purpose is to investigate the relationship between the GP distribution and the evaluation methods of professors who are charge of first year course. The followings are major findings. 1) The distribution of GP is so wide among and between the groupings of courses. Especially, there are some outliers in the upper end of distribution in grade A. So, standards of the evaluation should be made clear from the point view of fairness.

2) The grade points depend on the method of evaluation, class size, the characteristics of employment of the professors, to say nothing of the type of courses. Small class and the method of class participation tend to obtain gratifying results. But it is wonder that an assignment of report cannot anticipate the GP distribution. This may mean that because evaluation of student's report relies on the professor's intuition, it is difficult to grade objectively.

Key words : GPA, evaluation method, grading, teaching method, effective instruction

はじめに

この報告のねらいは、新潟大学学生部教務課が管理する教養教育 GP ファイルを利用して成績ランクの分布を描き、シラバスに記載される成績評価の方法との関係を検討することにある。このような定量的な側面から成績評価のあり方を検討するのは、大きく三つの理由がある。

一つは、GP に換算された学生の成績分布それ自体が、評価のあり方や教育方法を見直す上で包括的・客観的な情報を提供するからである。大学における教育方法の改善については、平成 10 年に出された大学審議会答申『21 世紀の大学像と今後の改革方策について』が、成績評価基準の明示と厳格な成績評価の実施を提唱し、各大学の状況に応じて GPA (グレード・ポイント・アベレージ)を活用した取り組みを例示した。GPA とはアメリカの大学で一般に用いられている評価方式で、A (秀)、B (優)、C (良)、D (可)、F (不可) の 5 段階による成績評価であり、100 点満点で A

(100~90)、B (89~80)、C (79~70)、D (69~60)、F (59 以下) に応じて、5 段階もしくは 4 段階評価を行うものである。

本学に限らず、運用は別にして、答申後、成績評価基準の明示と GPA 換算による学生の成績管理システムの構築がはかられている。運用はともかくというのは、答申が留年生の増加を覚悟せよと読めるからである。これについては、かつて賛否両論を含めたメディアの様々な反応やとくに私学関係者からの批判が相次いだ。重要なことは留年生の増加を懸念することよりも、GPA の分布から授業改善や成績評価基準の明確化に関わる情報を得ることだと考える。そのことが、成績評価まで踏み込んだ「答申」の意図だと読んでおきたい。

第二は、答申の背景にあるものだが、今日のゆとり教育のなかでは、学生の基礎学力が保証されにくい状況があることによる。国立大学では 5 教科 7 科目が復活するが、昨年実施した本学学生の高等学校での履修状

況から判断する限り、高校でこれだけ選択制が浸透しては、入試選抜だけで基礎学力が保証されたとみることとはできない（藤村：2002）。また、入学後、医師や教員のライセンスを取得し、国家もしくは自治体の採用試験を受験するものは限られた学生であり、大方は面接だけで企業に就職していく。

こうして大学の入り口でも出口でも学力保証ができていく環境のなかでは、大学内部でしか学生の学力ないしは能力をキチンと評価するところがないといってもよい。それだけに学生の学力や能力をどう評価するか、その重要性はいっそう増してきているようにおもえる。

第三は、成績評価について学生に不信感があることである。上記の学生授業理解度調査では、8割以上の学生が成績評価の判定基準の公表を望んでいる（藤村：2002）。一般に、評価が低いと学生の不信は高まるが、満点を取る学生でも不信をもつことはある。成績評価は、学生と教員の関係のなかで最も非対称な側面である（ローマン：1987）。これが学生の成績評価基準に対する不信の根本にある。言うまでもなく、採点は教員個人に任されていて、採点基準が教員個人の判断に依存する。分野や受講生数にもよるが、評価方法は試験、レポートの回数、出席回数、毎回の課題、予習・復習の程度、授業への参加の程度などを加味して＜総合的＞に評価することになっている。

しかし、そもそも学力ないし知的能力という定性的・直感的なものを数量化しランクづけする試みには、多少ともジレンマがつきまとう。実際、A、B、C、D、不可の線引きをどこで行うか、79点と80点、59点と60点の間に明確な妥協点を設定することは悩ましい作業である。とくにレポートではそうであろう。しかも、評価基準には様々な関係的要因も影響する。同じ教員でも評価基準が異なってくる。筆者の卑近な経験では、若い頃は厳しい採点をしていたが、年をとるとやさしくなった。したがって、キャリアを通せば平均値に近づいていくのかもしれない。

さらに、クラス・サイズの大小が評価に及ぼすことは改めて言うまでもないが、学生の所属によっても評価基準に一貫性が欠けることがある。一般に、知っている学生よりも知らない学生の方が、客観的で公正な

評価を下しやすい。また、担当者によっても成績の妥協点は影響を受ける。複数教官がコースの一部を担当する場合と一人で担当する場合とでは、前者は教員の負担は軽減するが、評価基準は曖昧になりやすい。成績評価の基準には、以上のような条件からバイアスが生まれるといつてよい。

この調査報告では、以上の理由と、個人的な体験から得た若干の仮説めいたものを踏まえて、平成13年度新潟大学教養教育で前後期に開講された1,119科目について、成績評価の方法、GPランクの分布の形状、及び両者の関係について検討し、今後の成績評価のあり方を考える上での基礎的なデータを提供しておきたい。

GPデータの利用は、個別科目名を特定化しないという条件で、本学教養教育委員会の許可を得た。したがって、GPデータの分析にあたっては、大分類＝科目群レベルで集約されたものを用いる。なお、データで示される不可（59点以下）は、シラバスには「科目の目標に及んでいない」と定義されるが（『新潟大学教養科目講義概要』、以下「シラバス」と略）、出席・試験放棄等による評価不能も含まれている。

1. 成績評価の方法と基準

1-1. 科目分布と担当教員の配置状況

まず、平成13年度シラバスから教養教育科目の群別分布と担当教員の配置をみておきたい。授業科目の大分類である群について、前後期通じて開講科目数の多い順にみると、「外国語科目群」543が最も多く全体の48.3%を占める。そのうち、第一外国語の英語が「外国語科目群」の65.6%、全科目では32.7%になる。まさしく、教養教育とは外国語教育であるといつてよい。

次いで多いのが、「自然科学科目群」の146科目（13.3%）、以下「保健体育科目群」が103科目（内、体育実技が92）で9.4%、「人文科学科目群」が94科目で8.6%、「総合科目群」が92科目で8.4%、「社会科学科目群」が51科目で4.5%、「情報処理科目群」は51科目で4.7%、そして「日本語・日本事情科目群」が48科目で4.3%となっている。

担当教官の属性をみると、学内の教官が一人で担当するのは、全体の47.7%で半数以下である。次いで、

非常勤一人が担当するクラスは全科目の 34.6%、学内の複数教官が担当する科目は全体の 9.4%、学内教官と非常勤の組み合わせは 6.4%、複数の非常勤教員が担当する科目は 1.8%となっている。科目群別にみると、「総合科目群」で学内の教員が一人で担当する科目は 3 割程度で、過半数は複数教官が担当している。「外国語科目群」と「保健体育科目」は非常勤依存率が高く、それぞれ 6 割と 4 割になっている。

1-2. 定員と受講生数

GP 分布をみる前に、シラバスに記載される定員と実際の受講生数を比較しておきたい。成績評価のみならず、授業で最も影響を受けるクラス・サイズに対して、定員のしびりがどの程度のものかみるためである。科目群別にみると、「総合科目群」92 科目の内、受講生が定員を上回る授業は 20 科目 (21.7%) にすぎず、8 割近くは定員を下回っていることがわかる。「人文科学科目群」では、定員オーバーは 175 名を最高に 20 科目 (21.3%) あるが、残りは定員以下である。「社会科学科目群」では定員オーバーが 51 科目中 11 科目 (21.6%) で、残りの 40 科目は 237 名を最高に定員を下回っている。

以下、定員オーバーの割合をみると「自然科学科目群」は 15%、「情報処理科目群」は 14.6%、「保健体育科目群」は 11.7%、「日本語・日本事情科目群」は 11.1 (2 科目)、そして「外国語科目群」では 2.7% (最大 48 名オーバー) である。

定員オーバーといっても程度問題だが、シラバスに記載される定員よりも上回っているのは全科目数 1,119 の 22.4%、5 科目に 1 つ程度にすぎない。このことは、外国語科目や体育実技は別にして聴講の際に学生への指導が厳しく行われていることを示している。実際、定員オーバーが 100 名を超える全科目は 7 クラスにすぎない。ところが、定員を 100 名も下回るクラスが 33 クラスある (200 名を下回るのは 6 科目、300 名を下回るクラスは 2 科目)。200 名から 300 名も定員を下回るのは、蓋を開けてみないと学生数の予想ができない、ということなのであろう。

1-3. 成績評価の方法と配点

シラバスに記載される成績評価の方法と配点はどのようなになっているのか。一般的なのは、「試験」、「レポート」、「出席」、そして「平常点」の 4 種類の組み合わせである。どれも記載がない科目は全体の 3.3%であった。他に、出席条件として 3 分の 1 規定、予習、参加重視、課題・宿題を記載する科目もある。ここでは成績評価のもっともポピュラーな「試験」、「レポート」、「出席」、そして「平常点」の分布をみておこう。

まず、全科目のなかでの比率をみると、「試験」を行うと明記してあるのが 70.9%、「出席」が 69.3%で最も多く、「平常点」が 56.4%、そして「レポート」が意外に少なく、32.7%となっている。試験と出席、または平常点が評価の要である。科目群別にみると、「試験」の記載が最も高いのが、「外国語科目」で 95.4%、「社会科学科目群」で 80.4%、「自然科学科目群」で 75.3%となっている。

最も低いのが、「情報処理科目群」で 14.6%、「保健体育科目群」で 21.4%、そして「総合科目群」で 39.1%である。「レポート」について、最も高いのが「情報処理科目群」で 95.8%、「総合科目」で 67.4%であり、低いのは「外国語科目群」の 14.2%となって「試験」の多い科目群と逆になっていることがわかる。

「出席点」の記載状況は、「保健体育科目群」が 94.2%で最も高く、最も低いのは「社会科学科目群」で 39.2%となる。同様に、「平常点」についても、「保健体育科目群」が 92.2%で最も高く、次いで、「外国語科目群」が 72.7%、「情報処理科目群」が 68.8%で比較的高い。逆に、「自然科学科目群」(28.8%) と「社会科学科目群」(15.7%) は低くなっている。

このように科目群別に評価方法の特徴をみると、試験と出席が重視される「外国語科目群」、試験重視の「社会科学科目群」、レポート重視は「情報処理科目群」と「総合科目群」、そして「出席」と「平常点」を重視する「保健体育科目群」となる。参考までに、試験、レポート、出席の 3 点セットが記載される割合が最も高いのは、「自然科学科目群」で全体の 21.2%の科目が該当する。さらに、「自然科学科目群」は上述の三分の一「出席条件」の規定を明記する割合が高い (18.5%)。その他、「参加重視」を記載するのは、「外国語科目」(20.1%) と「人文科学科目群」(18.1%) が高く、「外

国語科目群」については「課題・宿題」を出す割合が25.2%になっている。

2. GPの分布

2-1. 科目群別にみた箱ヒゲ図

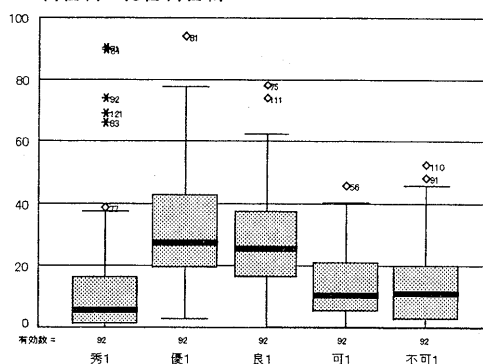
秀、優、良、可、不可の成績分布はどのようなになっているのか、科目群別に箱ヒゲ図で示しておいた。縦軸は、該当科目の受講生合計に対する5つの評価の比率を示している。100人の受講生で「秀」を10人、「優」を20人、「良」を30人、「可」を30人、「不可」を10人につければ、それぞれ10%、20%、30%、30%、10%の分布となる。箱内に科目数の50%が含まれる（箱内の横線が中央値、上下のヒンジが第1・第3四分位数）。これによって、同じ科目群でも教員のGP評価の幅に大きな差異があることがわかる。たとえば、総合科目群で「優」をつけた科目の分布をみると、20%から40%までの間に全総合科目の50%が含まれる。

しかし、ハズレ値（○印）と極値（×印）をみると、70%以上の学生に「秀」（＝科目の目標を越えている）つけている気前の良い科目が4つある。「優」については1科目だけ90%の学生に出していることがわかる。「秀」の中央値は5%程度であるが、高い方にシフトしている教科が少なくない。GP換算により「秀」を設けたことで、「優」を含めて全体に評価が甘くなっていることが読みとれる。「情報処理科目」と「保健体育科目群」も同様の傾向がある。

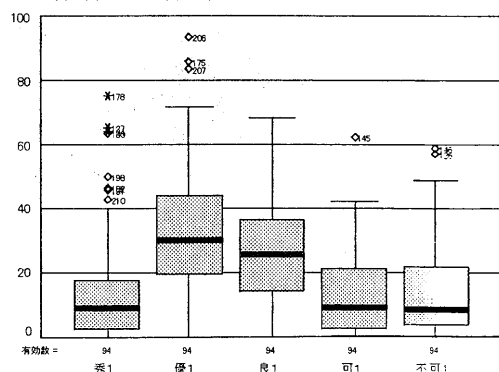
これに対して、ハズレ値と極値を別にすれば、「社会科学科目群」、「自然科学科目群」、「外国語科目群」は科目間のバラツキが比較的小さい。科目群によって評

価にこれだけの幅があることは、公正の点から言って問題があるのかもしれないが、これは大分類である科目群をみたものだから、かなりのバラツキがでるのは当然であろう。しかし、個別には図示しないが、小分類の科目系について箱ひげ図を描いても成績評価に相当のバラツキがある。一つの大きな理由は、教養部解体後、教養教育が各部局の応援によっているため、或いは複数教官が担当するため、一貫した成績基準が設定しにくいこと、やがて所属することになる学生には厳しい評価が下しにくいことがあるのかもしれない。あるいはその逆で、学内教官ほど評価が厳しくなるのかもしれない。科目にもよるが、たとえば、所属の1年次生が対象となる「スタディ・スキル」だけ取り出してみると、「秀」と「優」の割合が高くなっていることがわかる（図参照）。

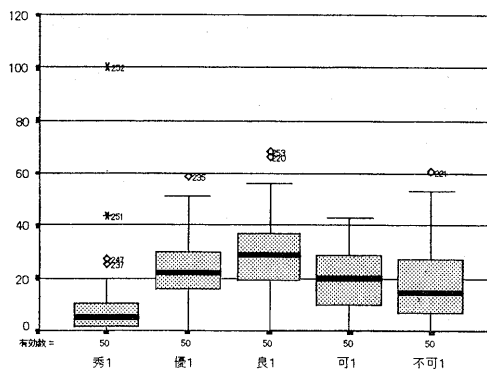
科目群= 総合科目群



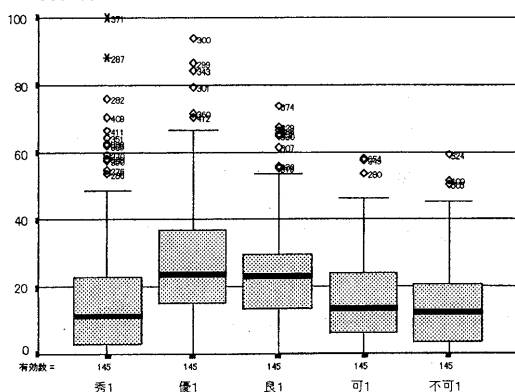
科目群= 人文科学科目群



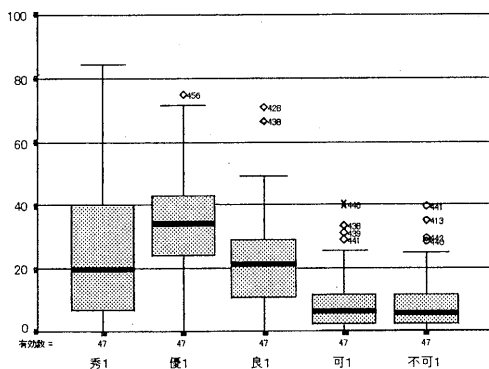
科目群= 社会科学科目群



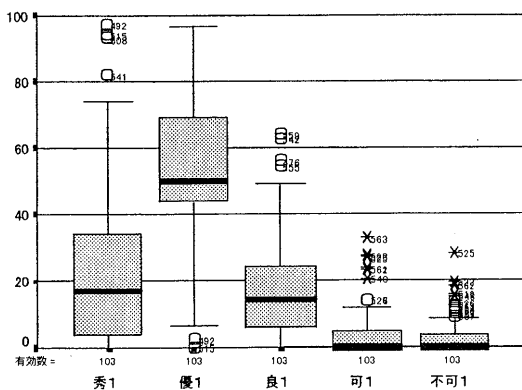
科目群= 自然科学科目群



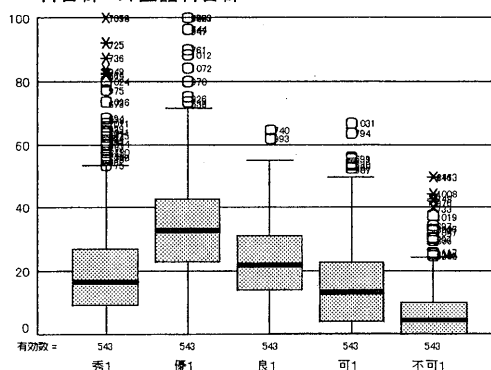
科目群= 情報処理科目群



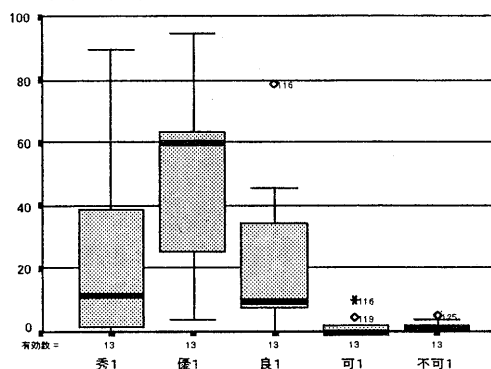
科目群= 保健体育科目群



科目群= 外国語科目群



学系= その他(スタディ・スキルズ科目)



2-2. 成績評価方法と GP の分布

それでは、成績の分布は評価方法とどのような関係にあるのだろうか。シラバスには、上述の通り試験、レポート、出席、平常点などによって<総合的>に評価するとあるが、どのように総合されているのか検討してみたい。変数は、先に見た授業シラバスに試験、レポート、出席点、平常点の記載のあるものについて、

それぞれ記載があれば1, なければ=0のダミーとした。さらに、科目群の影響は、「総合科目群」をベースにしたダミーである(日本語・日本事情は除外)。モデルにはクラス・サイズを示す学生数も入れておいた。

下の表が、秀、優、良、可、不可の比率に対して回帰分析を行った結果である。表には、有意な符号条件のみ記述している。これによって、評価方法は成績ランクによって符号条件が異なることが分かる。

まず、学生数については、受講生が多い科目ほど、評価方法や科目群が同じ場合でも、秀と優が少なくなり良と可が増える傾向にある。マスプロの解消は、採点者の負担の軽減を含めて授業の目標達成を容易にするための前提条件といえる。次いで、評価方法についてみると、客観的な評価方法である「試験」は、可と不可が増えて優が減少する。

表 成績評価と成績の関係 N=1,032					
	秀	優	良	可	不可
学生数	-	-	++	++	
試験		-		++	++
レポート					
出席点		+		-	
平常点		+++		-	-
人文					
社会		-			++
自然		-		++	+
情報	++	-			
体育	+++	+	-	-	-
外国語	+++			+	-
本学教官				+	++
R2	.032	.213	.038	.142	.181

- +...p<5% +++...p<1%, +++...p<0.1%

興味深いのは「レポート」である。「レポート」は、成績ランキングと明確な関係にはない。「レポート」は提出の有無を別にすれば、その内容的な評価は直感的なものだからであろう。成績ランクは、「出席点」や「平常点」と関係する。平常点をみる科目ほど、優が増え可と不可が減少する。つまり、平常点を評価する科目は、出席と併せて学生の動機づけを高めているものと思われる。これは単にシラバスに記載されるだけでなく、実際に教員が毎回チェックしていることを示している。科目群と評価の関係は、評価方法とクラス・サイズが同じでも、いわゆる必修科目ほど「総合科目群」よりも評価が高くなる傾向にある。さらに、本学の教

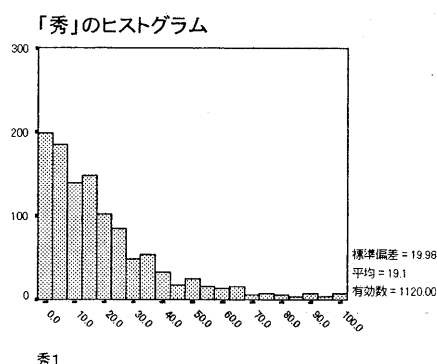
官一人で担当する授業科目は、複数教官もしくは非常勤の場合よりも成績が厳しくなる傾向にある。

おわりに：厳格な成績評価？

この調査では平成13年度前後期に開講された本学教養教育1,119科目をGP換算とシラバスをリンクさせたデータ・ファイルから、GPの分布と評価方法との関係を検討した。明らかになったのは、

(1)従来の4ランク評価から「秀」を加えたGP評価に改正されたことで、評価基準が高い方にシフトする(気前がよい)科目が少なくないことである。平成13年度では、全科目の10%にあたる科目で受講生の5割に「秀」が出ている。

成績評定が高いことは、学生の満足度と動機づけが高まるかもしれないが、科目の目標を超えている「秀」と科目の目標に十分達している「優」との間に明確な妥協点が引きにくくなっている。かつての「良」が「優」になっているのである。それは、大分類の科目群によっても、小分類の系のなかでも評価基準に幅がある。このことはある程度までは仕方がないが、「秀」と「優」の区分、とくに「秀」の評価については、少なくとも同じ学系内で調整が必要であるようにおもう。取得単位の上限についてのキャップ制は実施されているが、成績評価のキャップ制についても配慮すべきである。



(2)成績ランクは、科目群の違いのみならず、成績評価方法やクラス・サイズ、担当教官の属性によっても影響を受けることである。クラス・サイズは少人数制授業の導入により改善されつつあり、科目の目標達成に有益であることが確認された。一方、「試験」は、目標達成の程度をみることができると同時に、可と不可

を認知する客観的指標となっている。

(3)科目の目標を高める評価方法として「平常点」のあり方は示唆するところが大きい。高校まで慣れ親しんだ「平常点」は、いまの学生には高校までの教育との連続性があり、効果的であるといえる。大学の高校化というのではなく、「平常点」が学生の緊張感と動機を高める評価方法だということである。単位数の上限にしばりがかった現在、「平常点」の実施は、5割の教員が実施しているように、可能な環境になっている。

(4)教室の外と内を繋ぐレポートは、今回のデータでは成績ランクとは無関係である。これはレポートが課題の明確化や採点方法などを含めて、評価方法の中で最も主観的・直感的な方法であることによるのかもしれない。この点は、別途検討を要しよう。

以上が平成13年度本学教養教育GPデータから得られた結果の概要である。成績評価が担当教員個人の判断に委ねられていることは改めて言うまでもないし、選抜された学生の成績GPを正規曲線に近づける理由などもととない。

しかし、クラス・サイズの大きい授業、複数教員で担当する授業、少なくとも同じ科目内では学外非常勤に依頼する場合であっても評価基準についてある程度のまとまりのあることは、公正の点から大切である。むろん、採点の統計的な面に夢中になる必要もその余裕もないが、学生の知的能力の発達を促す条件を見出すうえで、定量的なデータを利用することもあるだろう。

国立大学でも大学入学試験の多様化のもと、入試データと在学中の成績データが管理・蓄積されているが、今後はそれを有効に利用することが必要となろう。

参考文献

- R.ピアド・J.ハートレイ,平沢茂訳『大学の教授・学習法』玉川大学出版部,1986.
- J.ローマン,阿部美哉監訳『大学のティーチング』玉川大学出版部,1987.
- W.マッキーチ,高橋靖直訳『大学教授法の実際』玉川大学出版部,1984.
- 藤村正司「大学での学習理解度と高校での履修状況」『大学教育研究年報』新潟大学大学教育開発研究センター,第7号,2002.

<付記>GPデータ・ファイルの作成については、本学学生部教務課管理系の協力を得た。記して謝意を表したい。