

広島大学 高等教育研究開発センター 大学論集  
第 38 集 (2006年度) 2007年 3月発行 : 277-292

## 絶対評価・相対評価が学生の学習行動に与える影響

—大学の専門科目における実験—

松 繁 寿 和・井 川 静 恵



# 絶対評価・相対評価が学生の学習行動に与える影響

—大学の専門科目における実験—

松 繁 寿 和\*  
井 川 静 恵\*\*

## 1. 問題

本稿では、絶対評価と相対評価という評価方法の違いが、大学生の学習行動にどのような影響をもたらすかを、大学の専門科目における実験を通じて検証する。学校における子供たちの学習達成度に関する評価から企業で働く労働者の働き方に関する評価まで、評価は社会のいたるところに存在し、さまざまな問題を抱えている。絶対評価と相対評価のどちらを採用するべきかという問も、常に議論をかもし出す重要なテーマの一つである。

2002年から、学校においてかつての5段階評価による相対評価の仕組みが変更され義務教育において絶対評価による通知票が作成されるようになった。鹿毛雅治(1989)においても主張されているように、相対評価は、児童・生徒の学習意欲を下げるとの意見が重視された。もし全員が同様に努力し、同じようにレベルを上げたとすれば、絶対的尺度で見れば成果を上げたとしても、相対的尺度で見ると評価は変わらないことになる。努力の量およびその結果による高いレベルの達成が評価に反映されないために、学習意欲の向上を阻害する可能性がある。なかでも、低い評価を受ける者たちの学習意欲の向上が難しいという問題が指摘された。

この問題に注目した実証研究もいくつか行われた。たとえば、白井博、袋佑加理、河内京子(2004)は、中学生を対象に絶対評価に変わった通信簿に関する感想や評価、および、相対評価と絶対評価の概念に関する理解度を調べている。また、白井博、袋佑加理、河内京子(2005)では、さらに北海道教育大学の大学生をくわえ評価方法についての意識と理解を調査している。また、鹿毛(1993)は小学生を対象にした調査において、到達度評価が内発的動機づけを高めることを示した。しかし、日本において、評価方法の違いが大学生の学習態度に実際どのような影響を与えるかを計測し分析した研究はほとんど見当たらない。

評価の問題は、企業の人事管理においても重要なテーマであった。従業員の労働意欲を高めるための評価方法や報酬制度のあり方は、労働経済学や組織の経済学の中でも重要な位置を占める分野として近年発達した。Lazear and Rosen(1981)やMalcomson(1984)に代表されるように、従業員の労働意欲は昇進競争によって向上するという始点から研究が始まった。通常、企業の上位のポストは限られており、昇進は数ある候補者のなかで限られた数だけ与えられる。すべての候補者が十分に職責を果たせる能力を持っていても、全員が昇進するわけではない。相対的に成績のよい上位

\* 広島大学高等教育研究開発センター客員研究員／大阪大学大学院国際公共政策研究科教授

\*\* 大阪大学大学院国際公共政策研究科博士後期課程

から順番に選抜されていく。したがって、基本的には企業内のトーナメント競争は相対評価による競争システムであるといえ、経済学の分野で発達した理論は暗黙のうちに相対的評価が前提となっていた。

その後、就業状態や生産に投入された労力量をモニターする困難さやそれに伴う費用の問題が目されるようになるとともに、それらを内包した理論分析と実証的検証が行われだした。ただし、相対評価と絶対評価の問題のうち分析が進められたのは、Gibbons and Murphy (1990), Milgram and Roberts (1992), Janakiraman, Lambert and Larcker (1992) で取り扱われているように、本人の努力以外の要因による成果のぶれの大小と評価のあり方に関する問題であった。

一方、日本経済の低迷が長引くとともに、そこから脱出する手段として企業内人事制度の変革が唱えられだし、成果に基づき個々の従業員の貢献をより厳密に測定・評価し報酬格差を広げようとする「成果主義」を取り入れようとする企業が増加した。個人の貢献の違いがこれまでよりも大きく処遇の差に反映されるようになったのである。

しかし、成果主義は新たな問題を生み出した。企業は組織であり、仕事を進める上でチームワークや協力・協調は欠かせない。仲間を助けることや後輩に技能を伝達する行為は企業内において積極的に進められるものであり、それが企業の競争力を支えていたりする。しかし、相対評価が行われている状況では、このような行為は、自己の相対的優位性を消してしまう行為となってしまう。手を差し伸べることで仲間が失敗を免れれば、仲間は評価を下げなくてすむ。また、後輩を指導し力をつけさせることができれば、その後輩の評価が上がる。すなわち、技能の伝達は自分の競争相手を育て自分の評価を下げてしまうことを意味する。評価が個人の貢献により強く依存し、評価差がより強く報酬に反映される状況では、この問題はより深刻になる。相対評価に基づいたままの評価・処遇格差の拡大は、企業の競争力を減退させる可能性がある。

実は、相対評価が従業員間の協力を阻害する問題は、Barron and Kreps (1999) などの人的資源管理基本的文献でも重要な問題として取り上げられている。しかし、残念ながら実証的検証はほとんどなされていない。

一方、絶対評価にも短所がある。絶対評価においては、ある一定以上の結果を出せば高い評価が得られる。したがって、その水準を上回った段階で満足し、より一層の向上を目指そうとしなくなる傾向がある。このことはすでに高い達成度にいる者に生じやすい<sup>1)</sup>。相対評価においては自分の評価は常に競争相手の成果に依存するために、自分と相手の力量の差や競争に参加するメンバーに関する情報が不十分であると、十分な努力水準を確定しにくい。その不安ゆえに出来るだけ努力をしておこうとする。絶対評価においては、達成度と自分の能力や努力との関係が明確になったとき、最適な努力水準はその関係を通じて解かれ、無駄な努力をしようとはしなくなる。

以上のように相対評価と絶対評価の是非に関する議論は、パースペクティブの発揮に関する問題をめぐってなされてきた。しかし、両者がどのような差を生み出すかを実際に示した研究は少ない。そこで、本稿では、大学の専門科目において試みた実験結果を用いて、評価方法の違いによって学生の学習行動に差が生じるかどうかを実証的に分析する。特に、注目するのは以下の3点である。

- 1) 相対評価と絶対評価の間で努力水準に差が生じるか。
- 2) 相対評価は協力を抑止するか。
- 3) 努力および協力はどのように成績に影響するか。

## 2. データ

分析に使用されるデータは、2002年度に開講された大学の専門科目で行った試験の成績とその際にとられたアンケートからの情報である。科目は2年生以上を対象に設定され、約600人が登録した。しかし、実際に最終試験となる第3回目の試験を受けた人数は200人前後である。1度でも試験を受けた者は約400人、その内訳は、2年生が約47%、3年生が約43%、4年生が約10%となっている。

試験は3回行われ、それぞれ、30点満点となっている。第1回は素点がつけられ、第2、3回は偏差値化された点数がつけられることが知らされ、かつ、偏差値の意味とそれがなぜ相対評価になるか、また、偏差値で成績を出すことと素点をそのまま用いることの違いについて30分ほど説明した。すなわち、第1回は絶対評価が点数化され、第2、3回は相対評価が行われる。

第2、3回目の相対評価は、

$$\text{点数} = \text{偏差値} - 30$$

の式を当てはめて計算され、偏差値60以上は30点に、偏差値30以下は0点とみなし、30点満点に換算される。

アンケートの質問項目は「この試験の準備のために合計何時間勉強しましたか」「あなたの授業のノートを何人の友達にコピーさせましたか」「あなたはこれまで（第2回目と第3回目は前回の中間試験以後）授業を何回欠席しましたか」のように、努力水準や協力水準について時間、人数、回数などを記入させるものと、出身高校（公立・私立）、自宅通学（はい・いいえ）、所属クラブ・サークル（体育会系クラブ・体育会系サークル・文科系クラブ・文科系サークル・所属していない）などについて、あてはまるものを選択する形式のもので構成されている。

## 3. 分析方法

### (1) 平均値の差の検定

第2節の課題1)・課題2)を検証するため、絶対評価（第1回の試験）と相対評価（第2回、第3回の試験）では、努力水準や協力水準に差が生じるかをみる。努力水準として、アンケート質問項目の「試験準備のための勉強時間」「授業欠席回数」の2項目を使用する。協力水準は、アンケート質問項目の「ノートのコピーをさせた人数」「ノートのコピーをさせてもらった人数」「協力して準備した人数」の3項目を分析する。

それぞれの項目について、第1回試験（絶対評価）時のアンケート回答の平均値と第2回試験（相対評価）時のアンケート回答の平均値の差が統計的に有意かどうかを検定する。同様に、第1回試験（絶対評価）時と第3回試験（相対評価）時の平均値の差の検定を行う。

たとえば、努力水準「試験準備のための勉強時間」について、絶対評価時の勉強時間の平均値の方が相対評価時の勉強時間の平均値よりも長いか、短いか、同じであるかを検証するために、勉強時間の平均値の差をとって、検定を行う。統計的に有意な差があれば、絶対評価による試験のための平均勉強時間の方が相対評価による試験のための平均勉強時間よりも長い(短い)ことが示される。

## (2) 努力水準・協力水準が成績に与える影響についてのパネル分析

続いて、課題3)にあげた、努力水準や協力水準が成績に与える影響について分析する。ここでは、個人ごとに3時点のデータが存在するため、パネルデータ(複数個人について複数期間のデータが存在する、横断面時系列データ)として分析することが可能である。

パネル分析の利点として、個人差を考慮できることがあげられる。このためにはパネルデータを固定効果モデルで推定する必要がある。個人差を考慮した固定効果モデルの推定が望ましいか、個人差を考慮する必要がない(個人効果が存在しない)通常のクロスセクション推定が望ましいかについては、Fテストによって統計的に検証し、選択することになる。個人効果が存在しなければ、3期のデータをプールしたデータ(プールド・データ)による最小二乗法での推定が支持される。

固定効果モデルでの推定では、個人を3期間追うため、たとえば3回の試験でずっと高成績であるとか、成績や順位が変動しているといったことが推定結果に反映される。また、個人の能力や資質といった観察できない要因もコントロール可能であり、努力水準や協力水準が成績に与える効果を見るために望ましい精度の高い推定となる。

推定式は次の通りである。

$$\begin{aligned} \text{成績} = f & (\text{努力水準: 試験準備の勉強時間, 欠席回数,} \\ & \text{協力水準: コピーをさせた人数, コピーをさせてもらった人数, 協力して準備した人数,} \\ & \text{その他の属性: 性別, 学年, クラブ活動, 自宅通学, 出身県, 公立高校出身, など}) \\ & + \text{攪乱項} \end{aligned}$$

## 4. 分析結果

### (1) 平均の差の検定結果

課題1), 課題2)に関する結果は、図表1から図表5に示されている。まず、努力水準の差について試みる。努力水準のひとつである「試験のための準備時間」に関しては、絶対評価である第1回と相対評価である第2回では第1回の方が長い、第1回と相対評価である第3回では有意な差が見られない(図表1)。もうひとつの努力水準の項目である「授業欠席回数の変化」については、第1回の方が第2回、第3回よりも多いという結果が示された(図表2)。やや相対評価の方が多くの努力を引き出しそうであるが、その傾向は確定的ではない。

つづいて、協力水準について試みる(図表3)。協力の項目のひとつは、「ノートのコピーさせた人数」である。図表に示されているように、相対評価が行われた第2、3回においてノートをコ

ピーさせた人数は明らかに減少したといえる。協力の第2の項目は「ノートをコピーさせてもらった人数」である（図表4）。絶対評価であった第1回より、相対評価が行われた第2, 3回において人数が減少したことが統計的に明らかとなった。協力の最後の項目は、「準備を行った友達の数」であるが、この数も相対評価（第2回, 第3回）の際に減少したことが統計的に有意となった（図表5）。

図表1：準備状況の変化

		サンプル数	平均値	標準偏差	
準備のための勉強時間	第1回試験	196	4.811	5.434	
	第2回試験	196	3.668	4.250	
	第1回-第2回		1.143	2.769	
	t値	5.781			
	対立仮説	第1回<第2回	第1回≠第2回	第1回>第2回	
	P値	1.000	0.000	0.000	
			サンプル数	平均値	標準偏差
	第1回試験	186	4.788	5.448	
	第3回試験	186	5.019	4.496	
	第1回-第3回		0.231	3.261	
t値	-0.967				
対立仮説	第1回<第3回	第1回≠第3回	第1回>第3回		
P値	0.168	0.335	0.833		

図表2：授業欠席回数の変化

		サンプル数	平均値	標準偏差	
授業欠席回数	第1回試験	188	2.046	2.183	
	第2回試験	188	1.047	1.757	
	第1回-第2回		0.999	1.754	
	t値	7.809			
	対立仮説	第1回<第2回	第1回≠第2回	第1回>第2回	
	P値	1.000	0.000	0.000	
			サンプル数	平均値	標準偏差
	第1回試験	186	2.081	2.195	
	第3回試験	186	1.019	1.212	
	第1回-第3回		1.062	1.933	
t値	7.494				
対立仮説	第1回<第3回	第1回≠第3回	第1回>第3回		
P値	1.000	0.000	0.000		

図表3：ノートをコピーさせた人数

		サンプル数	平均値	標準偏差	
何人にノートを コピーさせたか	第1回試験	196	0.829	1.231	
	第2回試験	196	0.638	0.926	
	第1回-第2回		0.191	1.077	
	t値	2.488			
	対立仮説	第1回<第2回	第1回≠第2回	第1回>第2回	
	P値	0.993	0.014	0.007	
			サンプル数	平均値	標準偏差
	第1回試験	200	0.813	1.224	
	第3回試験	200	0.470	0.838	
	第1回-第3回		0.343	1.293	
t値	3.746				
対立仮説	第1回<第3回	第1回≠第3回	第1回>第3回		
P値	1.000	0.000	0.000		

図表4：ノートをコピーさせてもらった人数

		サンプル数	平均値	標準偏差	
何人にノートを コピーさせても らったか	第1回試験	198	0.558	0.689	
	第2回試験	198	0.444	0.574	
	第1回-第2回		0.114	0.648	
	t値	2.468			
	対立仮説	第1回<第2回	第1回≠第2回	第1回>第2回	
	P値	0.993	0.014	0.007	
			サンプル数	平均値	標準偏差
	第1回試験	202	0.552	0.686	
	第3回試験	202	0.431	0.765	
	第1回-第3回		0.121	0.899	
t値	1.919				
対立仮説	第1回<第3回	第1回≠第3回	第1回>第3回		
P値	0.972	0.057	0.028		

図表5：準備を行った友人の人数

		サンプル数	平均値	標準偏差	
何人の友達と 準備をしたか	第1回試験	197	1.937	1.598	
	第2回試験	197	1.371	1.460	
	第1回-第2回		0.114	1.352	
	t値	5.877			
	対立仮説	第1回<第2回	第1回≠第2回	第1回>第2回	
	P値	1.000	0.000	0.000	
			サンプル数	平均値	標準偏差
	第1回試験	201	1.930	1.601	
	第3回試験	201	1.109	1.516	
	第1回-第3回		0.821	2.003	
t値	5.812				
対立仮説	第1回<第3回	第1回≠第3回	第1回>第3回		
P値	1.000	0.000	0.000		



## (2) パネル分析結果

このような差が、試験の成績にどのような影響を及ぼすかは、図表6に示されている。ここでは、第1回の成績も標準化し受験者別に3回の試験成績をとり、それを被説明変数とし、努力水準に関する変数(準備時間、欠席回数)、協力に関する変数(コピーさせた人数、コピーさせてもらった人数、協力して準備した友達の数)、その他個人属性(学年、その他の変数)に回帰させて分析している。推定結果には重要な変数の結果のみを示した。ただし、パネルデータ・固定効果モデルでの推定では、期間を通じて変化しない変数(学年、公立高校出身、性別など)は推定から落ちることになる。

まず、推定モデルの選択であるが、統計的にはデータをプールした回帰分析よりも、パネルデータを用いた固定効果モデルが支持された(全ての個人効果=0という帰無仮説の検定統計量は、 $F=1.930$ ,  $P$ 値=0.00)。ただし、かならずしもすべての試験を受けたわけではない者も存在するのでアンバランスド・パネル分析を用いている。

固定効果モデルの推定結果では、試験結果に影響をおよぼす変数として、協力して準備した友達の数が残った。係数が正であるから、協力して準備した友達の数が多いほど、成績は良いと解釈される。相対評価が協力関係を抑制するのであれば、それによって各人の成績が下がり、ひいてはクラス全体の成績が下がる可能性が示唆される。

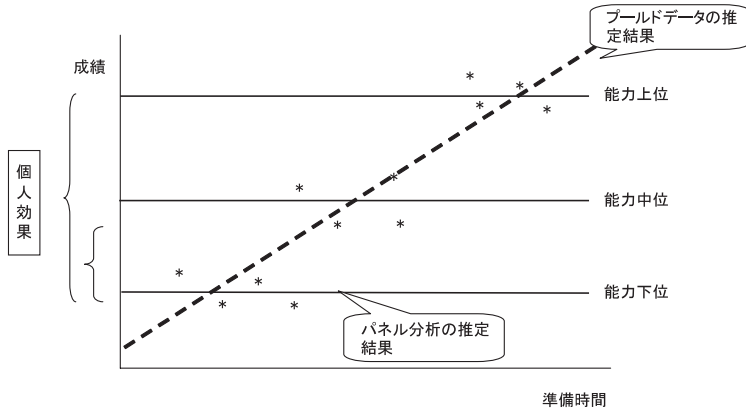
図表6：プールド・データとパネルデータ(固定効果モデル)の推定結果

プールド・データでの推定			パネルデータ・固定効果モデルでの推定		
変 数	係数	P 値	変 数	係数	P 値
準備時間	0.170	0.025	準備時間	0.142	0.402
欠席回数	-0.859	0.000	欠席回数	-0.169	0.538
コピーをさせた人数	0.803	0.022	コピーをさせた人数	-0.032	0.946
コピーをさせてもらった人数	-1.632	0.002	コピーをさせてもらった人数	-0.390	0.586
協力して準備した友達の数	0.124	0.580	協力して準備した友達の数	0.805	0.007
定数項	51.170	0.000	定数項	49.020	0.000
サンプル数	849		すべての $u_i = 0$		
F (5,843)	12.460		F (387,456)	1.930	
P 値	0.000		P 値	0.000	
R2	0.069		サンプル数	849	
自由度修正済みR2	0.063		グループ数	388	
			F (5,456)	1.720	
			P 値	0.129	
			R2		
			within	0.019	
			between	0.038	
			overall	0.020	
			ランダム効果—固定効果モデルに関するHausmanテスト		
			X2	24.320	
			P 値	0.000	

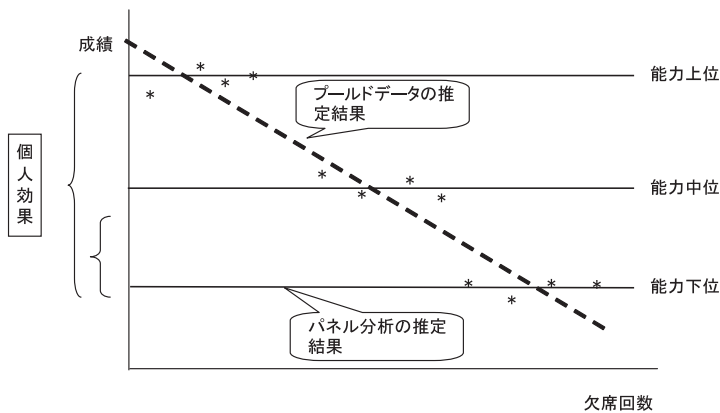
最後に、プールドデータの推定結果とパネルデータ（固定効果モデル）の推定結果を比較する。特に、プールドデータ推定では有意となった変数（「準備時間」「欠席回数」「コピーをさせた人数」「コピーをさせてもらった人数」）が、パネルデータ推定では非有意となり、逆にプールドデータ推定では非有意であった「協力して準備した友達の数」がパネルデータ推定で有意となった。すなわち、プールドデータ推定（最小二乗法による推定）では推定結果にバイアスが生じていた可能性がある。この点について以下の図表を用いて考察する。

「準備時間」は、プールドでは正で有意、パネルでは非有意となった。プールドの結果からは図表7の点線のような右上がりの関係があるように思われるが、パネルの結果をみれば、実際には個人間の差が成績の差を生んでいるのであって、同一個人の勉強時間の差は、一見して感じるほど成績に影響しないと解釈される。

「欠席回数」は、プールドでは負で有意、パネルでは非有意となっている。プールドからは、欠席回数が多いほど成績が悪いという右下がりの関係（図表8の点線）があるように思われるが、パ



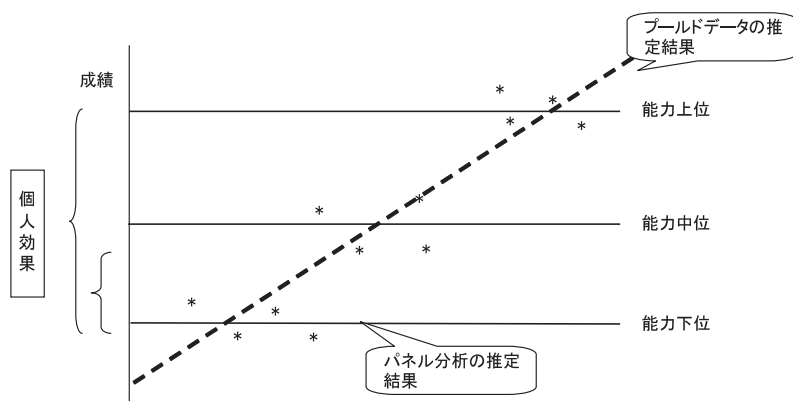
図表7：準備時間



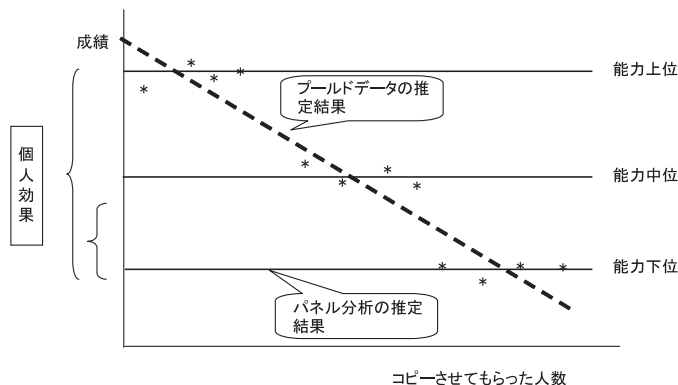
図表8：欠席回数

ネルの結果を考慮すると、個人間の欠席回数之差が成績之差を生んでいる（よくできない者ほど出席しない）という関係が示唆され、同一個人の授業出席傾向の変化は、一見して感じるほど成績に影響しないといえる。

同様に、「コピーさせた人数」は、プールドでは正で有意、パネルでは非有意となった。これについても、プールドデータの推定結果はバイアスを持っており、実際には、よくできる者ほど多くの友達にコピーをさせているのであって、多くのコピーをさせたから成績が上がるわけではないと考えられる（図表9）。「コピーさせてもらった人数」も、プールドでは負で有意、パネルでは非有意となったが、できない者ほど多くの友達からコピーをさせているのであって、多くの者からコピーをさせてもらったからといって成績が下がるわけではない（図表10）。

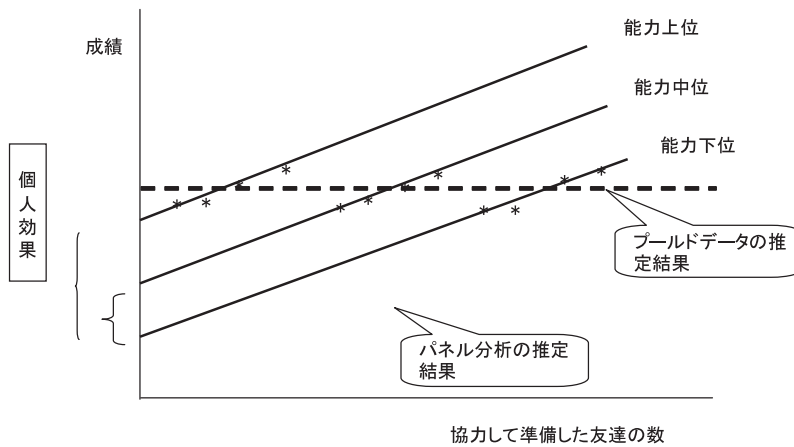


図表9：コピーさせた人数



図表10：コピーさせてもらった人数

以上の3変数とは異なり、「協力して準備した友達の数」はプールドでは非有意、パネルでは正で有意となった。プールドでは「協力して準備した友達の数」と成績の間に関係がないように思われる(図表11の点線)。しかし、実際にはよくできる者ほど友達と協力して準備しない傾向があるが、協力そのものは各人の成績を上げるという実線の関係があることがパネルデータの結果から推察される。



図表11：協力して準備した友達の数

## 5. 結語

本稿では、絶対評価と相対評価という評価方法の違いが、大学生の学習行動にどのような影響をもたらすかを、2002年度に開講された大学の専門科目で行った試験の成績とその際にとられたアンケートからの情報を用いて明らかにした。特に、パネル分析という統計手法を応用することで、個々人の属性や観察されない能力などをコントロールした上で、評価方法の違いがどのような変化をもたらすかを観察した。

分析の結果、まず、試験のための準備時間に関しては、絶対評価を行った場合と相対評価を行った場合では、有意な差が見られないことが分かった。すなわち、努力の水準に関しては相対評価と絶対評価の間にはっきりとした差は観察されなかった。一方、協力は相対評価が行われた場合において明らかに減少することが分かった。ノートのコピーをさせた、またはさせてもらった人数、および、協力して準備を行った友達の数減少は統計的に有意であった。

次に、このような差が、試験の成績にどのような影響を及ぼすかを分析した。試験の成績を被説明変数とし、準備時間、協力に関する変数、学年、その他の変数に回帰させた。重要な発見は主に2つある。まず、統計的にはデータをプールした回帰分析よりも、パネルデータを用いた固定効果モデルが指示されることが分かった。このことは、データをプールした回帰分析やクロスセクショ

ン分析では推定された説明変数の係数がバイアスを受けている可能性があることが分かった。もう一つは、成績に大きな影響をおよぼすのは友達との協力関係であることが分かった。このことと先の分析結果を合わせると、相対評価は成績を上昇させると思われる友達との協力関係を抑制し、結果として全体の成績を下げるという影響を持つことがわかった。先行研究では、相対評価が学習意欲を低下させるという直接的効果が問題視されたが、本稿の分析は、それ以外に相対評価は協力関係の発生を抑制し、協力を通じて学習効果を上げるという筋道も閉ざすことを明らかにした。

ただし、今後対処されるべき課題も多く残る。最も大きな問題は実験の設定である。試験を3回行ったが、それぞれ出題される範囲が異なる。範囲で扱われている内容の性質と難易度によって、行動が変化する可能性は高い。また、試験はかなり時間を置いて実施されている。大学の他の授業における課題等の存在や、クラブ活動の状態によって個人の属性をコントロールしても、試験を行った時期毎になんらかの変化があった可能性がある。すなわち、パネル分析における期別の効果を処理していない。最後に、今回の実験はある大学のある授業という限られた状況で行われたものである。結論に一般性をもたらすには、同様の実験が数多くなされる必要がある。今後の研究の蓄積が期待される。

## 【注】

1) この問題にくわえて、企業ではいくつか別の問題を抱えることになる。まず、技能や成果に関して企業横断的に絶対的水準を設定することが極めて困難である。仕事の世界においては、知識や技能をマニュアル化し体系的に統合した上で難度順に並べるという作業は、極めて膨大な労働と時間の投入を必要とするからである。また、技術が刻々と変化・進歩するがゆえに、たとえある時点で整理できたとしても、すぐさま変更が迫られるような事態が発生する。さらに、企業の競争力の源泉はそもそもマニュアル化できないような知識や技能の中にあるものであったり、他の企業が持っていないような特殊な知識や技能であったりする。したがって、多くの企業は、それら要となる技術を公開しようとはしないであろう。要するに企業間で比較可能となる技術や知識はそもそも限られている。

また、企業の場合、賃金支払い額に関する原資の制約がある。全員がよい成果を上げたからといって彼らへの給与をすべて上げられるわけではない。結局は限られた原資を分け合うことになる。高い評価を受けた者により高い報酬を支払おうとすると、その分、評価の低い者への報酬を削らなければならない。すなわち、評価の尺後は絶対であっても報酬に変換する段階で相対評価になってしまっていることが多い。

## 【参考文献】

白井博，袋佑加理，河内京子（2004）「新通信簿に対する中学生の反応と絶対評価・相対評価の概念の理解」『北海道教育大学紀要（教育科学編）』第55巻，第1号，北海道教育大学，183-194頁。

- 白井博, 袋佑加理, 河内京子 (2005) 「新通信簿に対する中学生と大学生のとらえ方と評価の基本  
概念の理解」『北海道教育大学教育実践総合センター紀要』第6巻, 北海道教育大学教育実践総  
合センター, 99-105頁。
- 鹿毛雅治 (1989) 「教育評価と学習意欲の関連について」『慶應義塾大学大学院社会学研究科紀要』,  
第30号, 慶應義塾大学大学院社会学研究科, 73-80頁。
- 鹿毛雅治 (1993) 「到達度評価が児童の内発的動機づけに及ぼす効果」『教育心理学研究』, 第41巻,  
第4号, 1-11頁。
- Barron, J. and Kreps, D. (1999), *Strategic Human Resources: Frameworks for General Managers*, John Wiley  
and Sons.
- Gibbons, R. and Murphy, K. (1990), "Relative performance evaluation for chief executive officers," *Industrial  
and Labor Review*, 43, Special Issue, pp.30-51.
- Janakiraman, S., Lambert, R. and Larcker, D. (1992), "An empirical investigation of the relative performance  
evaluation hypothesis," *Journal of Accounting Research*, 30, 1, pp.53-69.
- Latham, G. and Locke, E. (1991), "Self-regulation through goal setting," *Organizational Behavior and  
Human Decision Process*, 50, pp.212-247.
- Lazear, E. (1995), "Relative compensation," *Personnel Economics*, MIT, Chapter 2, pp.25-38.
- Lazear, E. and Rosen, S. (1981), "Rank-order tournaments as optimum labor contracts," *Journal of Political  
Economy*, 89, pp.841-864.
- Malcomson, J. (1984), "Work incentives, hierarchy and internal labor market," *Journal of Political Economy*,  
92, pp.486-507.
- Milgram, P. and Roberts, J. (1992), *Economics, Organization and Management*, Englewood Cliffs, N.j.:  
Prentice-Hall.

## **Absolute and Relative Evaluation of University Students and Their Learning Behavior: An Experiment in a Specialized Subject Class**

Hisakazu MATSUSHIGE\*  
Shizue IGAWA\*\*

This paper presents the results of an experiment that investigated the effect of the difference between relative and absolute scoring on students' test preparation activities in a specialized subject class. Among the many aspects that arise when comparing these two scoring methodologies, in this paper, we have selected the following three.

First, we investigate whether there is a difference in the amount of effort that students devote to preparing for a test that is relatively scored in comparison with one that is absolutely scored. As expected, the effort spent in preparing for absolutely scored tests was dependent on the expected difficulty of the examination. In other words, when students expected an easy test, they spent less time and effort preparing for it; on the other hand, when they expected a test that would be difficult to pass, they spent more time and effort studying. Relative scoring, on the other hand, introduces the presence of competitors. Regardless of how well any given student does on such a test, his/her relative position will suffer—and his/her score will decrease—if a competitor demonstrates the same level of achievement. For this reason, the amount of effort that students devote to preparing for relatively scored tests does not depend on the absolute difficulty of the exam itself; rather, it depends on the achievement level of competitors and the pass rate for the test.

Next, we examine whether relative scoring inhibits cooperation among students by observing whether there is any difference in the amount of cooperation at the time of preparing for absolutely versus relatively scored tests. The class-wide level of knowledge—that is, the scores observed on an absolutely scored test of that knowledge—increases as students work toward teaching one another and consequently bolstering their understanding of the course material. However, it is expected that the use of relative scoring would inhibit cooperation since helping friends to prepare for such an examination would adversely affect one's own relative standing.

Finally, based on the above results, we analyze the influence of the effort devoted to preparation and cooperation on the test results by investigating how variations in the observed amount of effort and cooperation affect actual test scores.

The data used for this analysis consists of scores from a series of quizzes that were included in a university course held in 2002 as well as the responses to a survey administered to the students taking the test.

---

\* Professor, Osaka School of International Public Policy, Osaka University

\*\* Doctoral Student, Osaka School of International Public Policy, Osaka University

Enrollment for the course, which was open to second-year and higher students, was about 600 students, while a little more than 400 students actually took the three quizzes of 30 points each. Students were informed in advance that they would receive a raw score for the first quiz, while the second two would be graded on a curve. In other words, the first quiz would be scored absolutely, while the second and third would be marked relatively.

The results of the analysis are as follows. First, preparation time for the first quiz was longer than that for the second quiz, but there was no statistically significant difference in the preparation time between the first and third quizzes. In other words, no prominent difference was observed between the amount of effort that the students devoted toward preparing for absolutely and relatively scored tests. On the other hand, there was a distinct decrease in cooperation as students prepared for the second and third (relatively scored) quizzes. There was a statistically significant reduction in the number of students copying class notes and allowing class notes to be copied as well as in the number of friends studying together.

Subsequently, we analyzed the impact of these differences on the test scores. We performed a regression analysis using the test scores as the explained variables and the preparation time, cooperation-related variables, grade, and other variables as explanatory variables. This analysis indicated that the fixed effect model using panel data was better suited to the data in this study than the regression model, which pooled the data statistically.

In particular, the pooled regression analysis performed in this study yielded results that cannot be explained using merely one argument: the preparation time had a positive correlation with the test score, as did the number of friends that a student allowed to copy his/her class notes. On the other hand, the number of friends from whom a student copied notes had a negative correlation, as did class attendance. The number of friends with whom a student cooperated during preparation was not significant. However, the fixed effect model revealed that differences in individual ability were significant, while the actual effects of other variables estimated when considering those differences revealed great variation. Preparation time, the number of friends that a student allowed to copy his/her class notes, the number of friends from whom a student copied notes, and class attendance were not significant. Only the number of friends with whom a student cooperated during preparation remained as a variable that influenced test scores.

These results suggest the possibility that relative scoring of tests inhibits cooperative relationships with friends, which is considered to be beneficial, and as a result, causes a reduction in the overall scores.