

学生による記述課題の相互評価と教師評価との相関

後 藤 善 友

A Comparative Analysis of Peer Assessment
and Teacher Assessment of Students Reports

Yoshitomo GOTO

はじめに

eラーニングの普及により、学生が提出した課題を他の学生が評価する相互評価が比較的容易に実施できるようになってきた。従来の対面授業においても学生による発表とそれに続くディスカッションという形式で相互評価的な取り組みが行われてきたが、どうしても時間がかかる活動となってしまうため、少人数の授業でなければ実施が出来ずまた何度も実施することは困難であった。しかし近年の主要なLMS（Learning Management System）の殆どは学習者の相互評価の機能を有しているため、対面授業のように受講者数による制限されることなく容易に学生相互評価の実施を可能にしている。

学生の相互評価が盛んになるにしたいが、その評価結果をどのように取り扱うかという研究が進められている。学生による相互評価の場合、評価基準や評価特性が評価者毎に異なるため、甘い評価ばかりする評価者や、まじめに評価を行わない評価者等がいることがあり、相互評価の結果をそのまま受け入れられない場合がある。これに対し、例えば植野ら[1]は項目応答理論の一つである Graded Item Response model により評価者の評価特性を考慮し信頼性

の高い評価値を得ている。また藤原ら[2]は別の評価特性モデルを用いて評点の補正を行い、やはり評点の信頼性を高めることに成功している。

本研究では、簡単な記述問題に対する学生相互評価において、学生の評価基準や評価特性を教師の評価に近づける方法を検討する。具体的には学生の相互評価結果を学生にフィードバックすることにより評価特性や評価基準のズレを意識してもらい、それによってどれくらい評価に変化が生じるのか、またどのくらい教師の評価と一致するのかを検証する。

方法

相互評価活動は、講義中に前回の講義内容に関する設問に400字程度の回答をする記述問題において実施した（「(教科)理科」30人）。各相互評価において各学生は他学生3～5人分の回答を10段階で評価した。評価結果は集計され、教師による評価とともに学生に提示された。評価を開始する前の注意事項として、教師や他の学生と著しく異なる評価をした場合は説明を求められることがあること・場合により評価者の評価を下げることもあることを確認した。

フィードバックによる相互評価の変化

フィードバックが無い場合の相互評価

どの相互評価前にも評価基準や評価の注意は行うが、初回の相互評価は前回の結果が無いために評価のフィードバックの無い相互評価が行われる。

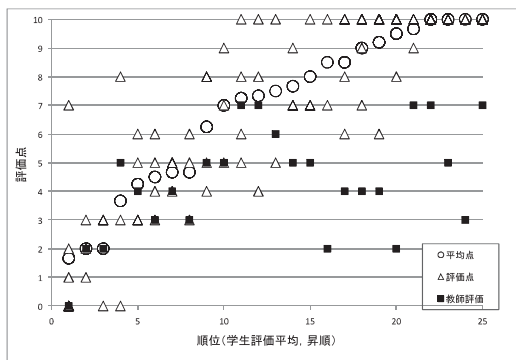


図1 フィードバックが無い場合の学生評価の分布と学生評価平均

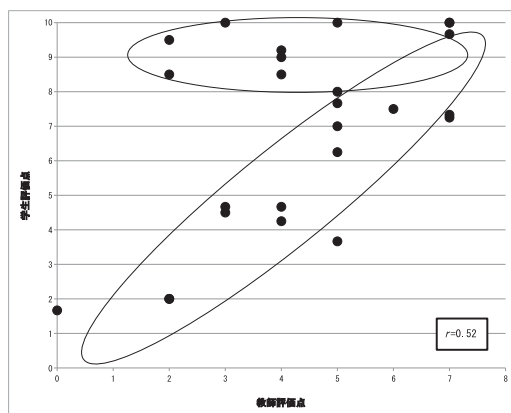


図2 フィードバックが無い場合の教師評価に対する学生評価平均の分布

フィードバックの無い相互評価の結果を図1に示す。学生評価には2つの特徴が見られる。1つは評価のバラツキが大きいことである。これはフィードバックが無いため事前に説明された評価基準の解釈が各学生毎に違うことによると思われる。また評価課題を全て満点で評価す

る学生もいることから、評価活動に対する態度にも違いがあるようである。もう1つの特徴は教師評価より学生評価の方が全体的に高いことである。これは他の学生の課題を甘く採点すれば自分も甘く採点してもらえるという期待による結果ではないかと考えられる[3]

教師評価と学生評価の関係を図2に示す。教師の評価と学生評価平均には大きなバラツキが見られる。しかし、弱い相関も見られる($r=0.52$)。評価点の分布には大きく2つのグループが見られ、ほぼ10点(満点)の評価平均を得ている学生グループを除けば、かなり教師の評価に近い結果になっている。このことから、どの課題にもほぼ10点(満点)の評価をする極端な学生以外の評価平均は、かなり教師評価に近い値であることが予想できる。

フィードバックがある場合の相互評価

相互評価の結果を学生にフィードバックする目的は、学生自身の評価と他の学生の評価結果を比較することで各自の評価を見直し、それにより全体の評価基準のバラツキを小さくすることである。相互評価の結果は、学生本人の課題についての他学生の評価平均とあわせて、他学生の課題に対する自分の評価と他学生評価平均(相互評価平均)が集計され、各学生に伝えられる。

このフィードバックを行った後に別の課題について相互評価を行った結果を図3に示す。図1と比較すると、広い範囲に分布していた最高評価が無くなっている。このことから、どの課題に対しても高い評価を出す学生が少なくなっていることがわかる。また学生評価のバラツキが小さくなっていることから、各学生の評価基準がフィードバック無しの場合より均質化されてきていることもわかる。そして全体的な傾向として、フィードバックにより学生評価平均と教師評価が近づいていることがわかる。

図4にフィードバックが有る場合の学生評価と教師評価の関係を示す。フィードバックの無い場合(図2)に比べ明らかに教師評価との相関が高くなっていることが分る。相関係数は $r=0.81$ であり、フィードバックが無い場合(r

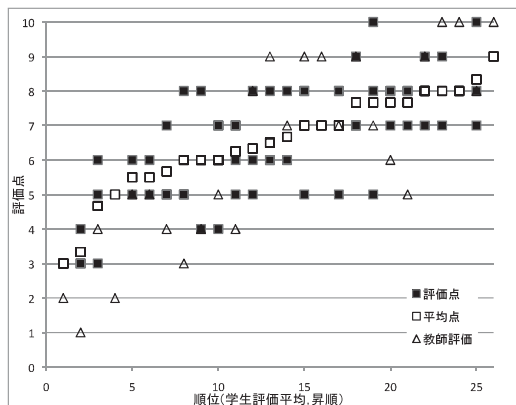


図3 フィードバックの有る場合の学生評価の分布と学生評価平均

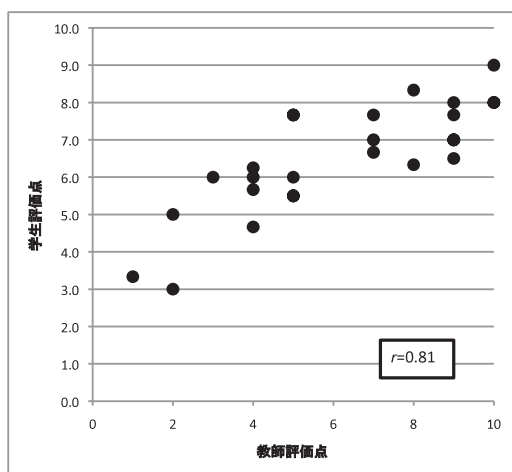


図4 フィードバックが有る場合の教師評価に対する学生評価平均の分布

= 0.52) と比較すると教師評価と学生評価が近づく様子が確認出来る。

学生による課題の順位付けの特性

教師の評価基準と学生の評価基準がどの程度に一致しているのかを判断するため、学生と教師の評価結果についてケンドールの順位相関係数を用いて比較した。

教師はすべての課題に評価点をつけるが、各学生は無作為に割り当てられた数件の課題だけに評価点をつける。この評価点から各学生による課題の順位付けを決められるが、これと対応

する課題についての教師による順位を比較した。

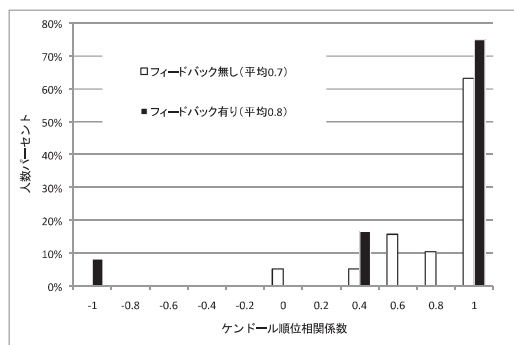


図5 フィードバックの有無による順位相関係数の分布変化

図5にフィードバックの有無による順位相関係数の分布変化を示す。フィードバックによる教師との順位相関が上昇している傾向が見られるが、大きな変化はみられない。しかし教師との順位相関が0.8を超えるような学生が多い(>70%)という傾向がフィードバックの有無にかかわらずみられる。これは大部分の学生の課題評価特性が教師のものに近いということであり、つまり順位付けに限って言えば、学生の評価はわりと信頼できることを示している。

まとめ

簡単な記述課題について学生相互評価を実施し、教師評価との関連について調査した。相互評価と教師評価の結果を学生にフィードバックすることで、学生相互評価の平均値は教師評価と高い相関を示すようになった(0.52 → 0.81)。また学生評価と教師評価の順位相関は、フィードバックの有無に大きな影響を受けず、高い相関を示していた(0.7 → 0.8)。

フィードバックによる学生評価と教師評価の接近は、学生が「どの程度の採点基準で教師評価に近づくか」を意識することで採点基準の均質化が進んだためだと考えられる。一方、順位相関がフィードバックの有無に関わらず高い値を示しているのは、もともと学生の評価特性が教師と近く、学生による課題の善し悪しの判断

は教師と大差なく行われていることを意味している。

今後、学生に対するフィードバックと、順位付け情報を組み合わせ、相互評価活動の頻度を高めることで、相互評価の結果はさらに高い精度で教師評価に一致することが期待できる。

参考文献

- [1] 植野真臣, 岡本敏雄, 永岡慶三, ピア・アセスメントにおける項目応答理論, 電子情報通信学会技術研究報告, ET2006 - 2, pp.7 - 12, 2006
- [2] 藤原康宏, 大西仁, 加藤浩, 形式的評価における相互評価支援システムの利用について, 電子情報通信技術研究報告, ET2006 - 33, pp.65 - 70, 2006
- [3] 藤原康宏, 大西仁, 加藤浩, 公平な相互評価のための評価支援システムの開発と評価 学習成果物を相互評価する場合に評価者の選択で生じる「お互い様効果」, 日本教育工学会論文誌, 31 (2), pp.125 - 134, 2007