

グループ発表を対象とした学生相互評価の 妥当性に影響する要因の分析

三浦 元喜^{1,a)}

概要：学生相互評価は、グループ活動の成果物としての発表やレポート等の評価方法として、ひろく利用されている。しかし、学生相互評価では学生が直接評価に関わるため、公正性や妥当性がかならず担保されているとはいえない。我々は、グループ発表を行う実験講義科目を対象として、学生相互評価の妥当性に影響する要因を調査した。分析の結果、教員が学生個人の発表を評価した際の点数と、学生個人が行った相互評価の妥当性（教員評価との順位相関）には、弱い正の相関があることがわかった。また、学生が入力したコメント量の多寡や、学生が別途提出したレポートの点数は、学生相互評価の妥当性との有意な相関はみられなかった。

キーワード：アクティブラーニング、PBL、評価手法、ピアアセスメント

Analysis of Factors Influencing Validity of Peer Assessment Between Students for Group Presentations

MOTOKI MIURA^{1,a)}

1. はじめに

学生相互評価は、講義における学生の成果を評価するための方法として以前から使用されている [1]。学生相互評価を導入する講義としてはアクティブラーニングやPBL型授業との親和性が高く、評価活動の対象としてはグループ活動の成果物としてのプレゼンテーション [2], [3] やレポート [4]、作問学習 [5] の事例が報告されている。しかし、学生によるグループ活動や成果物に対する評価は、公正性や正確さがかならず担保されるとは限らない。Shibaらは、個々の学生の公正な評価のための相互評価手法を提案している [6]。この方法では、信頼ネットワークを維持するために学生グループの再編成を行っている。しかし、学生グループの再編成を行うことで、それまでのグループ活動に

おける継続性が失われるという問題点がある。

我々は再編成を行わない条件の下で、学生相互評価における精度や信頼性を高めるための方法や、それらに影響を与える要因を、実際の講義で収集した相互評価データの分析を通して探索してきた [7], [8]。[7]では、LEGO Mindstorms を利用した創造的なPBL講義において、小テストや演習などプログラミングスキルに関連する個人の成果と、グループによるプレゼンテーションに対する相互評価点の相関を確認した。また、2013年から2018年の6年間にわたる学生相互評価データから、他グループを評価する際、入力初期状態にある平均的な点数よりも高い点数をつける学生ほど、全員の評価をまとめた最終順位との順序相関が高くなっていることがわかった。このことから、相互評価に真剣に取り組んでいるかどうかの指標のひとつとしての、他グループへの相互評価点の平均値の有用性を示した。[8]では、相互評価活動に付随する、発表者へのフィードバックコメント入力を「必須」として実施した場合と、必須ではなく「任意」とした場合とで、相互評価の

¹ 千葉工業大学 工学部 情報通信システム工学科
Chiba Institute of Technology, Narashino, Chiba 275-0016,
Japan

^{a)} motoki.miura@p.chibakoudai.jp

効果がどのように変化するかを調査した。分析の結果、コメント入力を「必須」とした場合について、他グループへの相互評価点の標準偏差と、コメント長の標準偏差のあいだに正の相関が確認できた。また、コメント入力を「任意」とした場合については、他グループへの相互評価点の平均値と、コメント長の平均値とのあいだに正の相関が確認できた。これらの結果は、相互評価におけるコメント入力の指示方法や運用によって、評価者となる学生の振る舞いに変化することを示している。しかし [8] の段階では、最終的な総合順位との順序相関については分析に含めていなかった。そのため、評価の精度や信頼性との関連については未確認であった。

そこで本稿では、[8] で不足していたデータを追加し、より詳細に分析することによって、学生相互評価における精度や信頼性、教員評価との関連をあきらかにする。これらがあきらかになることのメリットとして、学生相互評価における学生の行動原理や特性、傾向の理解がすすみ、相互評価の精度や信頼度を高める一般的で有効な方法の開発につながることを期待できる。

2. 対象講義と分析方法

本章では、今回の調査における対象講義と分析方法について述べる。

2.1 対象講義

今回の調査における対象講義は、著者が所属している大学の2年次後期に開講されている情報通信工学基礎実験とした。この科目は実験・必修科目である。今回は2023年度のデータを対象とした。この実験科目では、学生は5つのテーマ（UNIXの基礎、ネットワークプログラミング、アナログ信号とデジタル信号、トランジスタ回路、受動回路）にそれぞれ2週間ずつ取り組む。そして、学生は最後に取り組んだテーマに関してプレゼンテーションを準備し、最終週にグループで発表する。1グループの人数は4～5名であった。

学生は、対面方式の発表会に参加しながら、図1に示す学生相互評価のWeb入力フォームから、各グループの発表（自分たちのグループを除く）に対して点数とコメントを入力した。入力された点数とコメントは、入力後すぐにサーバに送信された。学生は発表会終了までは、点数とコメントを何度でも修正することができた。また、自分たちのグループの発表に対するコメントも、参照すればすぐに見られる状態であった。

教員は、すべての学生の発表について、グループ単位ではなく学生ごとに20点満点での採点を行った。こちらはWeb入力フォームではなくExcelシートで発表会終了後に収集した。

登録学生数は125名、グループの数は30であった。グ

ループの数が多いため、同じ時間に2つの教室に分かれて発表会を実施した。担当教員5名も3名と2名に分かれ、片方の教室のみで採点を行った。

グループに与えられた発表時間は10分で、その後質疑の時間が3分設けられた。学生からの質問を促すため、良い質問は加点の対象とすることと、一度も質問しなかった学生の発表点は減点することを周知した。また、すべての発表に対してWebフォームからコメントを入力することを指示し、未入力は減点の対象となることを伝えた。前の8件と後ろの7件のあいだに、15分の休憩を挟んだ。

図1 学生相互評価のフォーム（グループに対して点数とコメントを入力可能。点数の初期値は-1。教員は本フォームを使用せず、個人に対して20点満点で採点を行った。）

2.2 リサーチクエスチョン

本研究でのリサーチクエスチョンは、以下の通りである。

- (1) RQ1: 学生が入力した相互評価と、教員による評価には相関があるか？
- (2) RQ2: 学生が入力した相互評価と、学生が入力した他のグループに対するコメント量には相関があるか？
- (3) RQ3: 学生が入力した相互評価の精度（教員評価との順位相関）と、学生のレポート点には相関があるか？
また、教員による学生個人の評価との相関はあるか？

RQ1については、学生の相互評価の妥当性について確認するために行った。RQ2については、発表会当日の相互評価を真剣に行っている学生のほうが、コメントをきちんと入力していると考えられ、相互評価の精度との相関が高い可能性があるため設定した。RQ3については、普段から課題やレポートをしっかり出している学生のほうが、相互評

価活動に真剣に取り組む傾向が高いと考えられるため設定した。また、発表会における学生の主体性や態度との影響を調べるため設定した。

2.3 分析の方法

相関表や散布図行列を作成し、変数間の相関を確認した。相関については、無相関の検定の結果も参照した。

3. 分析結果

本章では、今回の調査における分析結果について述べる。以下、グループを「班」と短縮表記する場合がある。

3.1 教員評価と学生評価との相関

最初に、教員による班の評価と、学生による班の評価に大きな差異があるかどうかを確認する。仮に学生による班の評価が、教員による班の評価と大きくかけ離れていた場合、学生相互評価活動そのものの信頼性が低下すると考えたためである。

図2に、教員による班の評価と学生による班の評価との相関と散布図（散布図行列）を示す。散布図行列の対角成分にある棒グラフはデータの分布を示しており、 r は相関係数、 p は無相関の検定を行った際の検定統計量（ t ）の両側確率を示している。また、 p 値の右の*の数は、 $p < 0.05$ 、 $p < 0.01$ 、 $p < 0.001$ の3つの不等式のうち満たしている式の数を示している。なお、教員は学生個人単位での評価をしており、班単位での評価はしていない。そのため、学生が所属していた班のメンバーに与えられた評価点の平均を、班の評価値とした。

図2から、Pearsonの相関係数は0.66であり、学生による評価と教員による評価は高い相関があることがわかった。また、平均値を順位に変換したSpearmanの相関係数は0.70であった。

この結果から、学生による班の評価は教員による評価と類似性があり、学生による評価そのものの妥当性が確認できた。ただし学生が入力する評価値は平均(3)よりも高い点数が多く、分布が偏っている。そのため、学生の入力値に対しては分布の偏りを防ぐため、平均や標準偏差の幅を指定するといった制限を加えることも検討すべきである。

3.2 学生→班評価におけるコメント文字数の影響

全体として、学生による班評価の妥当性は確認できたが、学生によっては真剣に相互評価をしていなかったり、適当に点数を入力していたりする可能性が考えられる。また著者らが以前行った研究[8]では、コメント文字数の標準偏差と学生→班の評価値の標準偏差*1に弱い相関があることが確認できている。加えて、コメント文字数の標準偏

*1 学生→班の評価値の標準偏差が大きいほうが、学生はより積極的に相互評価に取り組んだと考えられる。

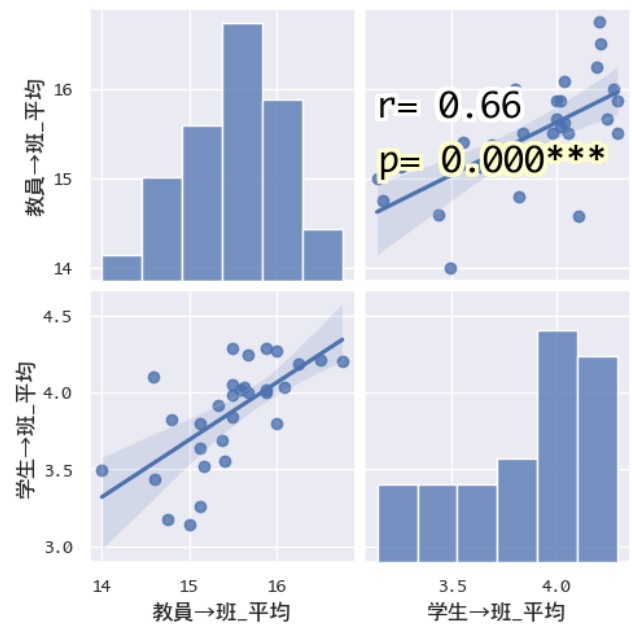


図2 教員→班評価と学生→班評価との相関と散布図（平均点）

差は、文字数の平均と相関が高い（今回のデータについては $r = 0.75$ ）こともわかっている。そこで、コメント文字数によって学生を4つのグループに分けたときに、教員評価との相関に変化がみられるかどうかを確認する。

最初に学生のコメント文字数を調べたところ平均値は101.1、標準偏差は50.1、中央値は97.6であった。欠席者や一部欠損値があった学生を除く116名を対象とし、29人ずつ4つのグループに分けて分析した。

図3に、教員→班評価とコメント文字数でグループ分けした学生→班評価の散布図行列を示す。学生グループG1（平均150.4文字）～G4（平均60.6文字）を比較したが、どのグループも教員→班評価と顕著な差はみられず、相互の相関も高い状態であった。順位相関についても調査したが、傾向は同じであった。このことから、教員→班評価の妥当性を仮定した場合、コメント文字数の多寡と学生→班評価の妥当性については明確な関連は確認できなかった。

3.3 学生→班評価における、学生レポート点との関係

3.2の分析において、コメント文字数と、学生→班評価の妥当性についての関連がみられなかったため、その他の確認可能な指標として、学生が提出したレポート点との相関がみられないか確認することにした。また、3.2の分析では図3は学生を4つのグループに分けて分析を行ったが、今回は個人単位での関連を調査した。

ここで分析に使用したレポートの点数は著者が担当する実験テーマに関するものであり、教員とTA2名が10点満点で採点した。ただし、点数配分は教員を70%、TA2名の平均点を30%とし、それらを足し合わせた点数を使用した。最大値は8.8、最小値は3.2、中央値は7.4であった。

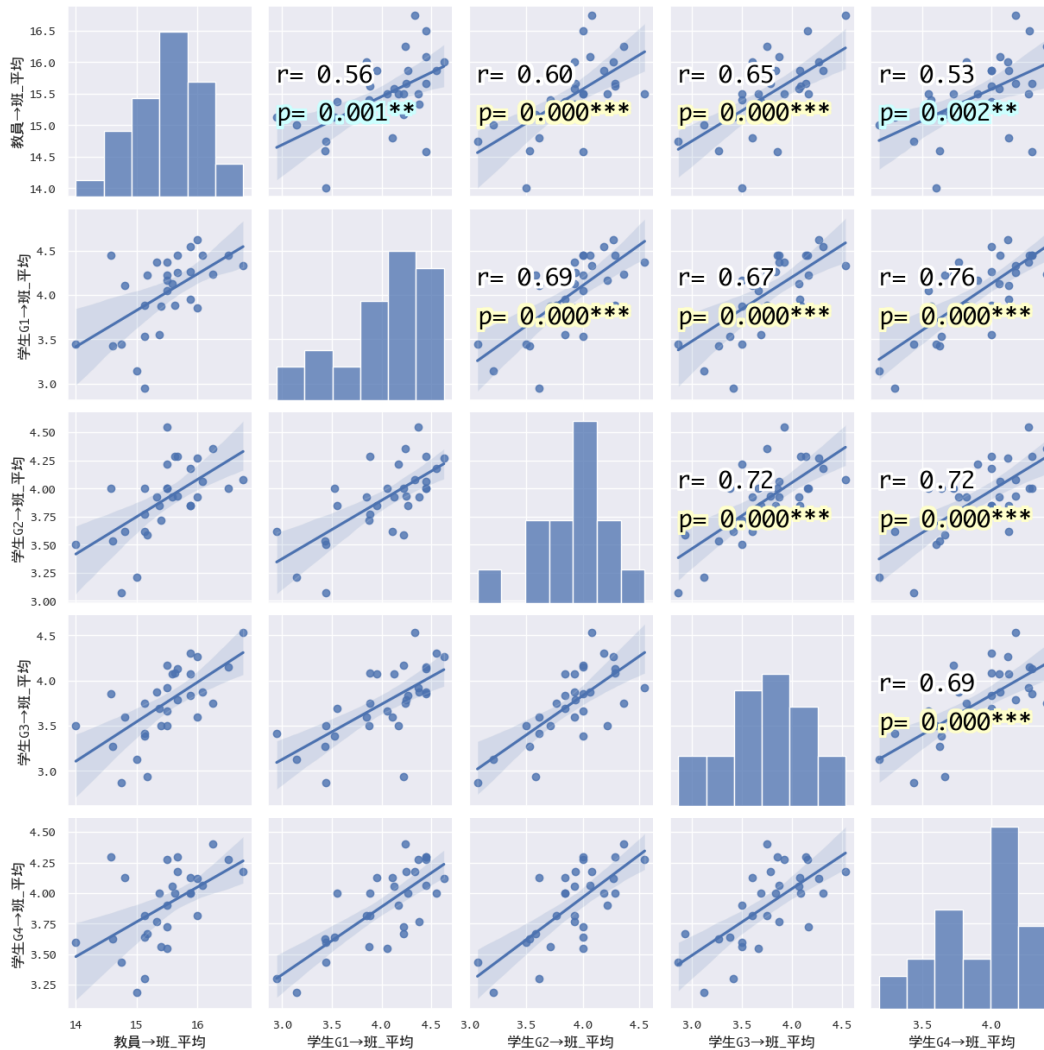


図3 教員→班評価とコメント文字数でグループ分けした学生→班評価の散布図行列

欠席者や一部欠損値があった学生を除く115名を対象として分析を行った。図4に、散布図行列を示す。ここで「教員→個人発表点」は教員が学生の発表に対して個人ごとに採点した点数であり、3.1で使用したものと同一である。また「教員→班との順位相関」は、学生個人が提出した「学生→班」評価と、教員→班評価との順位相関である。この値が+1に近いほど、教員→班評価の妥当性を仮定した場合、当該学生の評価の妥当性が高かったことを示す。

図4からは、「レポート点」と「教員→個人発表点」、「教員→班との順位相関」との相関や有意性は確認できなかった。しかし、「教員→個人発表点」と「教員→班との順位相関」のあいだに弱い正の相関と有意性が確認できた。このことは、発表会の当日、学生が発表する際の姿勢・意欲^{*2}と、他グループに対して相互評価を行う際の態度や熱意には多少の関連があったことを示している。

*2 教員→個人発表点には、発表会当日に一回も質問しなかったことによる減点も反映されている。

4. まとめ

本研究では、グループ発表を行う必修科目を対象とした学生相互評価の妥当性や精度に影響する要因を調査・分析した。今回の調査結果から、確認できたことをまとめると、以下ようになる。

- 学生が入力した相互評価と、教員による評価は高い相関がある。(RQ1)
- 学生が入力したコメント量の多寡は、相互評価の妥当性にはあまり影響しない。(RQ2)
- 学生の提出レポート点と、学生相互評価の質との相関に有意性はない。(RQ3)
- 教員の学生個人に対する評価点と、学生相互評価の質には弱い正の相関がある。(RQ3)

今回の調査では、教員の評価をいわゆる「正解データ」として、学生の評価の妥当性を検討した。ただし、教員の評価が常に正解であるとは限らない。また今回は発表会評価における詳細な観点について、事前に共有していなかつ

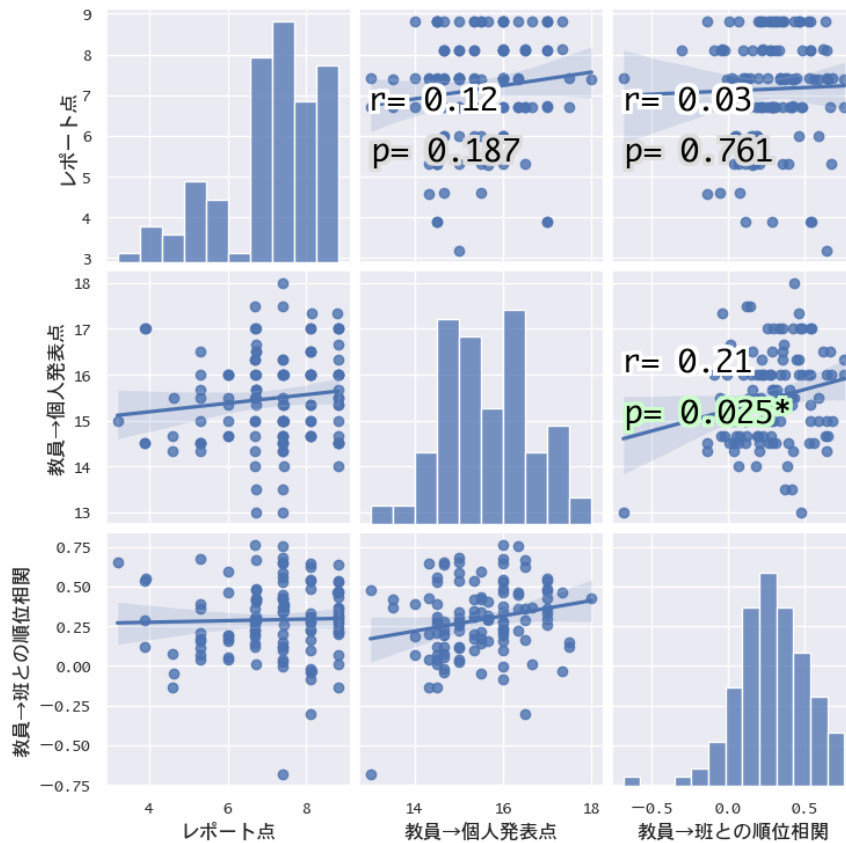


図 4 教員→班評価と学生→班評価との相関と散布図（平均点）

たため、評価軸がぶれていた可能性もある。こうした点を考慮しつつ、今回の結果を踏まえて、学生の相互評価の質や精度を高めるための方策について引き続き試行と分析を続けていきたい。

謝辞

本研究の一部は JSPS 科研費 JP19K03056 および JP22K12319 の支援によるものです。

参考文献

- [1] Topping, K.: Peer assessment between students in colleges and universities, *Review of educational Research*, Vol. 68, No. 3, pp. 249–276 (1998).
- [2] 松本重男: チームでプロジェクト活動を行う科目での教育評価: 学生の相互評価と教員の評価観点, *日本教育工学雑誌*, Vol. 24, No. 1, pp. 93–98 (2000).
- [3] 竹田尚彦, 吉田宏史, 佐合尚子: プレゼンテーション演習における学生間相互評価の分析, *情報処理学会研究報告コンピュータと教育 (CE)*, Vol. 2005, No. 36 (2005-CE-079), pp. 55–62 (2005).
- [4] 河野昭彦, 齊藤博嗣, 佐々木大輔, 平澤一樹, 須田 達, 鶴谷奈津子: 学生の相互評価によるアクティブラーニング型授業, *工学教育*, Vol. 62, No. 6, pp. 6.62–6.67 (2014).
- [5] 高木正則, 田中 充, 勅使河原可海: 学生による問題作成およびその相互評価を可能とする協調学習型 WBT システム, *情報処理学会論文誌*, Vol. 48, No. 3, pp. 1532–1545 (2007).
- [6] Shiba, Y., Umegaki, H. and Sugawara, T.: Fair Assessment of Group Work by Mutual Evaluation with Irresponsible and Collusive Students Using Trust Networks,

PRIMA 2015: Principles and Practice of Multi-Agent Systems, Cham, Springer International Publishing, pp. 528–537 (2015).

- [7] Miura, M.: Analysis of student performance and mutual evaluation activity in creative project-based learning using LEGO mindstorms, *The 41st Annual Conference of Japan Creativity Society, Proceedings of International Session*, pp. 17–20 (2019).
- [8] Miura, M.: Analysis of mutual evaluation activity and student performance in creative project-based learning, *8th International Conference on Knowledge, Information and Creativity Support Systems (KICSS2023)* (2023).