

オンラインレポート相互評価システムの開発と実践

野口峻輔^{†1} 藤村直美^{†2}

MOOC (Massive Open Online Course)などの多人数の受講者がいる学習コースでは、教員による課題の評価が困難なことから、受講者が相互に評価する仕組みである相互評価方式が提案されている。相互評価で適正な評価ができるかどうかを検証するためにオンラインレポート相互評価システムを開発し、本学の講義で運用した。受講者の利用状況のデータの分析や教員による評価との比較をもとに相互評価の有用性を検証する。また受講者に対するアンケートの結果も報告する。検証の結果、相互評価には一定の有用性があることがわかった。

Development and operation of online peer grading system

SHUNSUKE NOGUCHI^{†1} NAOMI FUJIMURA^{†2}

In an online learning course like MOOC (Massive Open Online Course), peer grading is a useful tool for reducing the effort of teacher. In order to verify whether peer grading works well and returns accurate score, we develop online peer grading system and operate it in our university's class. In this paper, we verify it with an analysis of the data of user behavior and comparison between peer and teacher grading score. We also show questionnaire results. As a result, we find the validity in peer grading to some extent.

1. はじめに

大学の講義で学生の成績を評価する方法として、出席、レポート・課題提出、筆記試験などが行われている。この中でレポート・課題提出は、講義の理解度・習得度を測り、また卒業研究等の下地にもなる重要な項目である。本論文では、与えられた課題に対する調べ学習や自分の意見を書くレポート、演習系の授業での作品などをまとめてレポートと呼ぶことにする。本研究は、このレポートの採点を学生同士が行う「相互評価」に注目し、実際に大学の講義で運用するためのシステムを構築し、運用した成果について述べる。

近年、相互評価に関する研究や実践が盛んに行われている。Sadler と Good (2006) は自己評価・相互評価の有用性を統計的に示した[1]。近年、相互評価に注目が集まっている背景には、ICT (Information and Communication Technology) の進歩があげられる。主に米国で、大学の講義をオンライン上に動画やテキストで公開する OCW (Open Course Ware) という取り組みがなされている。代表的なものに MIT の OpenCourseWare[2]、Khan Academy[3]などがある。

OCW に加えて、毎週講義を聞き、課題を提出することで修了証を取得できる MOOC (Massive Open Online Course) というものが出現し、急速に普及しつつある。代表的なものにスタンフォード大学の Coursera[4]や MIT の edX[5]などがある。この MOOC で成績を評価する際に使われているのが相互評価である。MOOC の講義の受講者は多いものだと

数万人にもなる。この多人数の受講者の課題を数人のスタッフが毎週採点していくことは不可能である。そこで相互評価が用いられている。Chris ら(2013)は MOOC での相互評価の精度を上げる手法を提案している[6]。多くの MOOC では他者の学習を評価するだけでなく、受講者同士のコミュニティを作り、積極的な意見交換が行われている。

2. 背景

九州大学では上述の OCW の流れを受け、一部の授業を YouTube[7]や iTunes U[8]などのプラットフォームで公開している。また、LMS (学習管理システム) として Blackboard 社の Blackboard Learn (以後、Blackboard) が導入されている[9]。このシステムでは時間割の管理、授業資料の配布、課題提出などができる。担当教員はこのシステム上で課題を採点でき、学生は教員からのフィードバックを得られる。ただし、学生同士の相互評価機能は実装されていない。

九州大学では、学部1年生向けに全学教育を行っているが、講義によっては学部を問わず受講できる、あるいは全学生が対象であるため、受講生数が非常に多くなることがある。教室のスペースの問題や、スタッフの負担などの理由で受講を断られる学生もいる。

基幹教育院による全学教育の教育課程の全面見直しの一環として、2015年度には「社会基盤としてのネットワーク」という授業が開講予定である。この授業は遠隔講義システムを利用して、大橋キャンパス、伊都キャンパス、箱崎キャンパスなど、複数のキャンパスで開講予定の授業であり、受講学生は400~500人ほどが見込まれている。この多人数の学生が提出する課題を教員一人が採点するのは困難である。

^{†1} 九州大学大学院芸術工学府
Graduate School of Design, Kyushu University

^{†2} 九州大学大学院芸術工学府
Faculty of Design, Kyushu University

3. システム概要

開発したオンラインレポート相互評価システム（以下本システム）の概要を記す。本システムは MOOC の一つである Coursera を参考にして実現している。

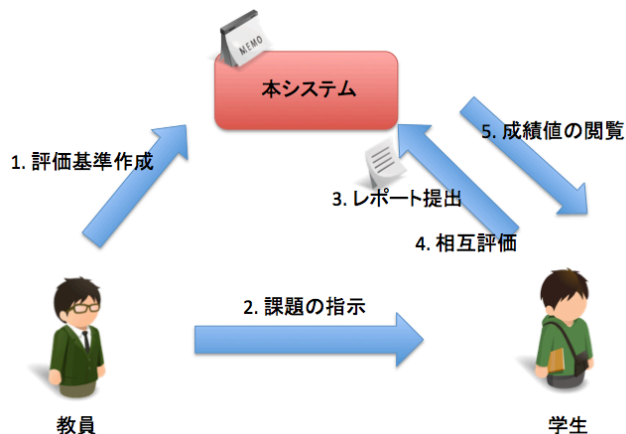


図 1 システム概要図
Figure 1 Abstract View of the System.

図 1 をもとに説明する。

1. 評価基準作成

本システムは大学の講義と連携して使う。教員は授業前に課題の内容・締め切りと評価基準を本システムにアップロードしておく。評価基準は学生が相互評価する際に指針とするものなので重要である。

2. 課題の指示

教員は授業中に学生に課題の内容と評価基準を伝える。

3. レポート提出

課題を指示された学生は、提出締め切りまでに本システムにレポートをアップロードする。提出締め切り前であれば何度でも修正したものを再提出できる。アップロードされたレポートはサーバ上に保存される。また、評価基準を参考にして自身のレポートを評価する「自己評価」もできる。自己評価するかどうかは任意である。

4. 相互評価

提出締め切りを過ぎると、相互評価期間に入る。この期間中に学生は他の学生のレポートを採点する（図 2）。この際、成績値だけでなくコメントも入力できる。利用者（学生）には見えていないが、評価に要した時間も記録している。あまりに短時間で評価をした場合にはその評価は疑わしい。誰のレポートを評価するかはシステムによってランダムに決められる。また、評価は匿名で行う。これは、人間関係等、他の要因が入るのを防ぐためである。

5. 成績値の参照

学生は自身のレポートにつけられた成績を参照できる。コメントを見て次回以降のレポート作成の参考にする、モ

チベーションを上げるといった教育的効果を狙っている。

6. 教員用管理ページ

教員用に管理ページを作成している。このページから課題・評価基準の作成や提出されたレポートを参照できる。また、教員もこのシステムから採点できる。

アンケートをこのシステム上で行った。結果は後で報告する。

図 2 相互評価画面

Figure 2 Peer Grading Screen.

4. 実践

4.1 概要

本システムを九州大学芸術工学部で平成 26 年度後期に開講されている「ネットワークサービスデザイン」で運用した。学部 2 年生対象の授業で、電話やインターネットの基礎知識の習得を目的とし、講義形式で進められる。成績評価は各回の授業後に指示されるレポートの提出をもとにしている。受講者は 66 人であり、2015 年 1 月 16 日現在、9 つの課題が指示され、441 つのレポートが提出されている。

教員は授業時に学生に課題を指示する。学生は一週間以内にレポートを提出する。次の一週間のうちに当該課題の相互評価を行う。図 3 にレポート提出と相互評価の流れを図示した。

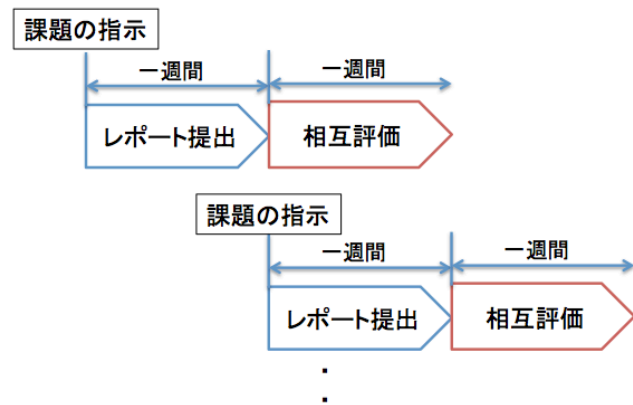


図 3 レポート提出と相互評価のスケジュール

Figure 3 Schedule for assignment

4.2 結果

4.2.1 相互評価の結果

本システム上で得られたレポートの提出数と相互評価数を表 1 に示す。レポート提出数とは当該課題に対して提出されたレポートの数である。相互評価数とは当該課題に対して提出されたレポートを評価した数の総計である。

表 1 相互評価の集計結果

Table 1 Results of peer grading.

	レポート提出数	相互評価数
1 回目	55	250
2 回目	52	177
3 回目	54	241
4 回目	55	197
5 回目	50	200
6 回目	47	194
7 回目	49	214
8 回目	32	78
9 回目	37	42

4.2.2 アンケートの結果

授業日程の半分が過ぎた時点でアンケートを行った。結果を表 2 に示す。右端には「とてもそう思う」「そう思う」の回答の合計をパーセンテージで示した。表 2 の結果以外にも任意でコメントを書いてもらった結果、以下のようない意見があった。

- ・ ユーザーインターフェースに関する要望
- ・ コメントをもらえるので意欲が上がった、等の肯定的な意見
- ・ 不可解な評価があって不快だった、等の否定的な意見

5. 考察

5.1 相互評価の考察

レポート提出数を見てみると、50 前後で安定している。本システムに登録している学生数は 66 人なので提出率は 75%前後となる。一度だけ授業に顔を出してそれ以降来ていない学生や本システムに登録だけして全く使っていない学生もいるので提出率 75%という数字は決して低くない。8 回目のレポート提出数・相互評価数ともに少なくなっているが、システムに問題がありレポートを提出できない状況が数日間あったからだと思われる。なお、第 9 回目は原稿執筆時点で集計の途中である。

相互評価数を見てみると、回によってバラつきはあるが 200 前後になっている。一人あたり約 4 つのレポートを毎回相互評価していることになる。学生には最低 3 つ以上相互評価するように授業中に伝えている。

5.2 アンケートの考察

Q1～Q2 は本システムの使い方やレイアウトに関する設問だったが、概ね肯定的である。コメントでよせられたユーザーインターフェースに関する指摘は次回以降の実践の参考にする。

Q3～Q7 は相互評価することによる学習効果について問うた設問である。「Q3. 相互評価によって自身の課題につけられた成績値は妥当なものだった」と「Q5. 相互評価をすることにより授業内容を振り返る機会になった」には肯定的な回答が多い。Q3 の結果から、相互評価によって成績をつけられることは受け入れられていると考えられる。Q5 の結果から、相互評価が復習のいい機会になっていることが分かる。

しかし、「Q6. 相互評価活動を通して自身の採点能力の向上を実感した」は肯定的な回答が少ない。相互評価の前に採点指導をするべきだったと反省している。さらに、「Q7. コメントが有意義だった」に肯定的な回答をしているのは 33%しかない。本システムにアップロードされたコメントに目を通したが、確かにちゃんと見ていないと思われる相互評価がいくつか見られた。ただ、意外ときちんと見てくれていて嬉しかった、という学生もいた。学生にきちんと相互評価をするインセンティブを与えることが今後の課題になりそうだ。

5.3 相互評価と自己評価の比較

相互評価によって得られた値と自己評価による値を比較した。いずれも共通の評価基準をもとに評価している。相互評価の値には複数人によってつけられた値の平均値を使っている。第 1 回目の結果を図 4 に示す。相関係数には Pearson の積率相関係数を用いた。相関係数は $r = -0.17$ となり、相関はないという結果になった。グラフを見てみると、相互評価では 4 前後になっているのに自己評価で 5 と

表 2 アンケートの集計結果 (N = 33)

Table 2 Results of questionnaire(N = 33).

	とても そう思う	そう思う	どちら でもな い	そう 思わな い	全くそう 思わない	肯定的 (%)
Q1. システムの操作は分かりやすかった	11	21	0	0	0	100
Q2. 課題の表示・レイアウトは適切だった	8	17	5	3	0	76
Q3. 相互評価によって自身の課題につけられた成績値は妥当なものだった	9	17	5	2	0	79
Q4. 相互評価をすることにより授業への意欲が増した	5	13	11	4	0	55
Q5. 相互評価をすることにより授業内容を振り返る機会になった	8	17	6	1	1	76
Q6. 相互評価活動を通して自身の採点能力の向上を実感した	5	9	11	8	0	42
Q7. コメントが有意義だった	4	7	12	7	3	33

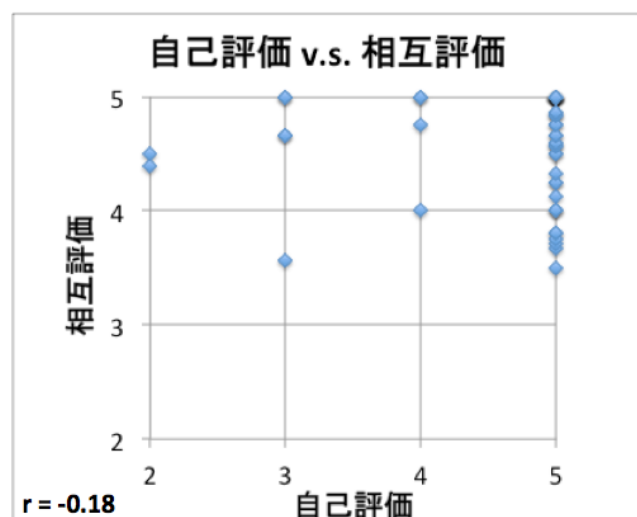


図 5 相互評価と自己評価の比較
Figure 5 Comparison between peer- and self-grade.

している人が多いことが分かる。やはり自身のレポートを客観的に見るのは難しいと考えられる。他の回でも調べてみたが、いずれの回でも自己評価と相互評価の間には相関関係はなかった。

5.4 相互評価と教員による採点との比較

相互評価によって得られた値と教員による採点を比較した。5.3 と同じくいずれも共通の評価基準をもとに評価し、相互評価の値には複数人によってつけられた値の平均値を使っている。第 6 回目の結果を図 6 に示す。

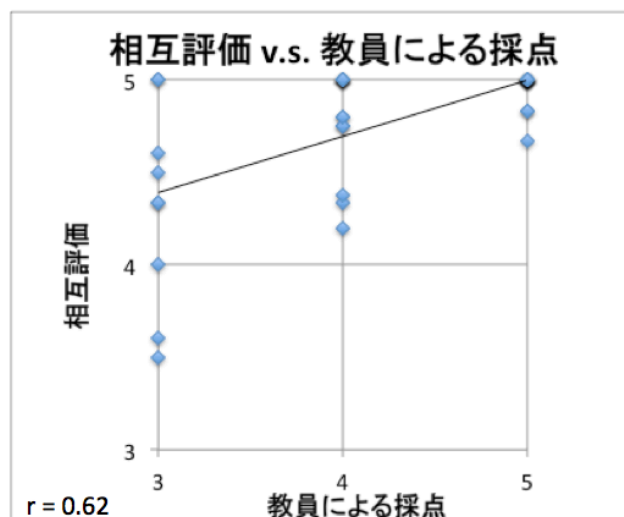


図 6 相互評価と教員による採点との比較
Figure 6 Comparison between peer- and teacher grade.

相関係数は $r = 0.62$ と正の相関が認められた。グラフを見てみると、全体的に上に寄っているのが分かる。言い換えると、相互評価よりも教員による採点のほうが厳しくつけられているということだ。実際に採点されたレポートと成績値を見てみると、教員から見てレポート内容や量に不足を感じて減点しているレポートが多いことがわかった。逆に学生は評価基準に書かれている内容を満たしているので高得点にしているというパターンが見られた。教員が頭の中で考えている評価基準と、学生が受け取った評価基準にズレがあったためこのような傾向が見られたと推測できる。

今回は評価するレポートをランダムで選んだため、レポートによって評価される件数にバラつきがあった。例えば、一人からしか評価されていないレポートと、10人から評価されたレポートがあった。このことにより外れ値が出てきてしまっていた。

6. 今後の展望

今回得られた知見をもとに、本システムを改善・発展させていく。具体的には以下のようなことが挙げられる。

- ・アンケートから得られたユーザインターフェースの改善
- ・学生の相互評価を行うモチベーションの向上
- ・相互評価をさせる前の採点指導
- ・評価時間が極端に短い・コメントがすべて同じ等の不適當と思われる評価を排除

また、本論文では検証できなかった研究課題としては以下のようなことが挙げられる。

- ・信頼できる相互評価のために必要な評価数の検証
- ・学生の成績と評価能力・傾向との関連
- ・評価に要した時間と評価能力との関連
- ・相互評価の精度を高めるための補正アルゴリズムの検証

本システムは 2015 年度前期開講の「社会基盤としてのネットワーク」での運用を予定している。以上のような改善を行い学生によりよい教育環境の提供をしたい。

参考文献

- [1] P. M. Sadler and E. Good. The impact of self-and peer-grading on student learning. Educational assessment, 11(1):1-31, 2006.
- [2] MIT Open Course Ware : <http://ocw.mit.edu/index.htm>
- [3] Khan Academy : <https://www.khanacademy.org/>
- [4] Coursera : <https://www.coursera.org/>
- [5] edX : <https://www.edx.org/>
- [6] Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. Tuned models of peer assessment in MOOCs. Proc. of International Conference on Educational Data Mining, 2013.
- [7] YouTube 九州大学 : <https://www.youtube.com/user/KyushuUniv>
- [8] 九州大学 iTunesU : <http://itunes.icer.kyushu-u.ac.jp/>
- [9] 九州大学 Web 学習システム : <https://bb9.iii.kyushu-u.ac.jp/>