

継続的な学習者間評価を導入した情報教育の実践

藤原 康宏^{†1} 大西 仁^{†2,†3} 加藤 浩^{†2,†3}

情報教育では、実践的な課題やグループワークの場面で、学習者間の相互評価を導入した実践がよく行われている。しかし、相互評価の研究では、学習のプロセスで繰り返し相互評価を行う事例は少なく、また相互評価と学習効果の関係について客観的なデータに基づいた報告は少ない。本研究では、大学生を対象としたプレゼンテーションソフトの活用法の演習で、相互評価支援システムを継続的に利用した実践を行い、成果物の改善に役立っているのか、他者の成果物への評価能力が向上するかどうかについて検証した。実践では、1つの課題に対して相互評価を5回行い、成果物の改善と学習者の評価の変化について分析した。相互評価を繰り返した結果、他者の成果物の短所や改善点を具体的に指摘できるようになり、評価者の個人差を補正した評価結果を表示することでより適切な評価をすることができることが示された。実践が終了した後にに行ったアンケートでは、多くの学習者が相互評価は有意義であると回答し、また、評価者の個人差を補正することが望ましいと考えていることが分かった。

A Practice of Repetition Peer Assessment in ICT Education

YASUHIRO FUJIHARA,^{†1} HITOSHI OHNISHI^{†2,†3}
and HIROSHI KATO^{†2,†3}

In ICT education, peer assessment is well known as the effective evaluation way for project based learning, group work and so on. However the reports on the effects of repetition peer assessment practices from quantitative data are rarely seen. We practiced peer evaluation five times on application of presentation software in a university class to show learning effects of peer assessment. As a result of the practice, it was found that the evaluators received feedback with revised score improved the evaluation ability, but the evaluator received only evaluation value before revision did not improved. Because the above-mentioned result, it was shown that proposed revision method was effective to an impartial evaluation and the improvement of the evaluation ability.

1. はじめに

近年、学習者が同じ学習コミュニティに属する他の学習者の成果物を評価する相互評価がさかんになっている。相互評価が注目される背景の1つとして、真正な学習という考え方がある。Wigginsは、標準テストでは断片化した知識を文脈と切り離して評価している本物でない評価になっているのに対して、大人が現実世界で直面する問題解決場面をシミュレートした課題での評価であるべきとする真正な学習の考え方を提唱した¹⁾。真正な学習の成立に関わった実践としては、SizerのCoalition of Essential Schoolの教育評価改革²⁾があり、そこでは、真正の評価として、学習発表会における卒業審査と、ポートフォリオが用いられていた³⁾。このような背景から、学習者間評価、パフォーマンス評価やポートフォリオ評価など、ペーパーテストに代わる評価手法が新しい評価として、注目されている⁴⁾。

高等学校における普通教科「情報」や大学などにおける情報処理入門科目において、学習すべき内容は、コンピュータやソフトウェアの操作など、形式知だけにとどまらない。情報機器はつねに新しいものが生産されるため、ある時期に覚えた操作だけでは役立たないため、生徒、学生が実社会に出てからも生かされるような、総合力としての情報教育が求められている⁵⁾。このような観点から、実践的な課題を導入した教育や学習者間の共同作業を重視した実践が行われてきており、真正な学習となることが望ましい。

相互評価の利点として、「学習者間の評価は、最も自然な評価体系であり、評価結果を受け入れやすくし、学習者の内省を引き出し、結果として知識の内化や深化を導くことや、学習動機の向上が期待される」といわれている⁶⁾。情報教育やコミュニケーションの分野では、多くの相互評価の教育実践が行われている。たとえば、プレゼンテーションスキルを題材とした実践^{7),8)}、外国語教育でコミュニケーションを題材とした実践⁹⁾、情報教育を題材とした実践¹⁰⁾、プロジェクト学習での実践¹¹⁾⁻¹³⁾などが行われており、それぞれ教育効果をあげたことが報告されている。従来の評価は、テストなどにより教師が学習者を一方的に評価するものであったが、教師と学習者の間にある学習コミュニティ内で、評価を行っていくことが有効であるといわれている¹⁴⁾⁻¹⁷⁾。多くの分野で相互評価が成功している事例がある

†1 岩手県立大学

Iwate Prefectural University

†2 メディア教育開発センター

National Institute of Multimedia Education

†3 総合研究大学院大学

The Graduate University for Advanced Studies

が、継続的に相互評価を行い、相互評価と学習効果の関係について言及されている研究はほとんどない。

本研究では、相互評価による教育的効果として、内省によって学習者が成果物の改善点に気づき、その改善を行うことができるかどうかに関心をあてる。形成的評価として、相互評価支援システムを使用した相互評価を継続的に行うことにより、成果物の改善に役立っているのか、他者の成果物への評価能力が向上するかどうかについて検証した。

相互評価が効果をあげる前提として、学習者が他の学習者を適切に評価できることがあげられるが、相互評価をあまり経験していない学習者が、最初から適切に評価できない場合も多いと思われる。また、成果物の改善には、フィードバックされる評価結果が適切である必要がある。しかし、評価者によって、厳しい評価をしたり、極端に差をつける傾向があったりするなど、評価者の特性が反映されたものになっており、評価結果をそのまま提示するだけで、十分な効果があるとはいえない可能性がある。特に、相互評価の経験の少ない学習者の評価については注意が必要である。藤原らは、項目応答理論のメタファを用いた評価モデルによって、評価値の補正アルゴリズムを提案し、実データに適用して、補正を行うことにより妥当な評価値に近づくことを確認した¹⁸⁾。本研究では、このアルゴリズムを使用して、評価者の評価特性を考慮した評価結果を補正したものを学習者に示し、学習成果物の改善や他者の成果物への評価能力に与える影響についても考察する。

2. 関連研究

本章では、特に情報教育やコミュニケーションに関係する事例を紹介する。

相互評価にコンピュータを利用する場合と、そうでない場合を比較した研究が行われており、対面で相互評価を行った場合の方が評価後の学習者同士のコミュニケーションを促進するといわれている^{19),20)}。コンピュータを用いた協調学習支援システムに、単に提出機能や相互評価などを付加するだけでは、対面で行う相互評価と同様の教育的効果が得られないことも考えられる。コンピュータを使って相互評価を行う際に、どのような支援を行うかについて考える必要がある。

コンピュータを用いた様々な相互評価ツールが開発、実践され、その教育的効果が報告されている。中原らは、電子掲示板に、任意にポイントを加点することで相互評価を行うことができるシステムを開発した²¹⁾。アンケートの結果から、学習者はシステムを積極的に利用し、好意的に受け入れられたと述べられている。布施らは、ビデオオンデマンドとウェブデータベースを利用した相互・自己評価システムを開発し、プレゼンテーションの授業で、

学習者の意欲が向上したと述べている²²⁾。高木らは、学習者が問題作成する場で、作成した問題を数人のグループの中で相互評価するシステムを開発した。システム運用の結果、学生同士の対話が増え、学習意欲が向上したと述べている²³⁾。これらのシステムの評価をはじめとする相互評価の実践では、学習者が積極的に学習に取り組んだり、学習意欲が向上したりすることがあげられている。しかし、学習効果をあげるための要因については、ほとんど議論されていない。

相互評価が学習効果をあげるための要素について言及した実践例について述べる。天野・下村は、共同での作品制作の中間段階で、Webベースの相互評価システム²⁴⁾を用いた実践を行い、シンプルな評価観点を学習コミュニティで共有することで、作品制作途中での意見の交流が活発になったと述べている²⁵⁾。学習コミュニティは、「何らかの部分で、同じ価値、同じ目標をもった、人の集合・つながり」と考えられ¹⁵⁾、この実践では、学習者に評価の観点を意識させることを通じて、「同じ価値、同じ目標」を共有する試みが行われている。生田目は、プログラミングの授業で、グループ学習を導入し、グループ内での相互評価を行わせた²⁶⁾。その結果、プログラミングやフローチャートの理解度が向上と述べている。事後に行ったアンケート結果に対して、因子分析を行った結果、教えあうことによるグループ学習の効果、他者の良いプログラムを見ることによる効果、他者のフローチャートを見ることより誤りを発見する効果の3つの因子が抽出された。藤原らは、プレゼンテーションソフトウェアの利用法の授業で相互評価の実践を行った²⁷⁾。その結果、学習者がお互いに評価しあう場合は、そうでない場合に比べて、高い評価値をつけることが分かり、この現象を「お互い様効果」と名づけた。また、評価する相手に評価されない場合の方が、より適正な評価を行うことも分かった。評価者を選択する必要がある場合、お互い様効果を考慮した評価者の割付けについて考慮する必要がある。

相互評価の実践は多く行われているが、学習効果については、主観的なデータをもとにした報告が多く、他者の成果物を評価する能力や学習者が行った他者の成果物に対する評価の妥当性について検討していないことも多い。また、相互評価を1度だけ行った事例は多いが、形成的評価として相互評価を実施する場合、それだけでは高い効果は期待できないと思われる。

3. 継続的な学習者間評価を導入した実践

今回の実践では、継続的にシステムを使った教育的効果として、成果物の改善に役に立つか、他者の成果物を評価する能力が身につくかどうか検討を行う、また、フィードバックの

表 1 授業内容
Table 1 Contents of the class.

授業内容	配当時間
オリエンテーション	第 1 回
パソコンの基本的な使い方	第 1 回
e-mail	第 2 回
www ブラウザ	第 2~3 回
ドローツール	第 4~5 回
プレゼンテーションソフト(本実践)	第 6~14 回
ワードプロセッサ	第 11~15 回

表 2 授業内容と使用するシステムの機能
Table 2 Contents of each lesson.

回	授業内容	提出	評価	結果表示
6	PowerPoint の使い方 (60 分) 制作① (30 分)			
7	制作② (90 分)			
8	制作③ (30 分) システムの使用方法・事前評価 (60 分)	制作③	制作③	
9	画像の引用 (30 分) 改善① (60 分)	改善①		制作③
10	アニメーション (30 分) 改善② (60 分)	改善②	改善①	改善①
11	見やすいスライド (30 分) 改善③ (30 分)	改善③	改善②	改善②
12	分かりやすいスライド (30 分) 改善④ [完成] (30 分)	完成版	改善③	改善③
13	完成版の評価・結果確認 (30 分)		完成版	完成版
14	事後評価, 発表会 (60 分)			

※ 授業内容欄の斜体は、講義及びその内容についての演習を表す。

() 内はそれぞれの授業での配当時間を表し、改善①~④の時間にはシステム使用時間を含む。

際に個人差を補正した結果を表示することによる影響について考察を行う。

実践では、1 つの課題に対して相互評価を 5 回行い、実践中の学習者の評価の変化と成果物の改善について分析した。また、相互評価の実践の前後で、以前の受講生が作成した成果物を評価させ、他者の成果物への評価が変化するかどうか調べた。さらに、実践後に本実践について、アンケートを行った。

3.1 実践内容

3.1.1 授業科目の概要

A 大学文学部 1 年生を対象に行っている「情報処理基礎」(2 単位)で、相互評価を取り入れた授業実践を行った。コンピュータの基本的な操作を通じて、情報処理の基本について学ぶもので、授業は演習を中心にコンピュータ演習室で行った。授業内容を表 1 に示す。このうち、プレゼンテーションソフトの課題で相互評価支援システムを使った相互評価を導入した。

授業の履修者は、文学部 1 年生 19 人であったが、以下の計画および結果では、単位取得者 16 人を対象として述べる。なお、不合格となった 3 人は、いずれもプレゼンテーションソフトの前の回までに、すでに出席がつかねではなかった者であり、不合格の理由は今回の実践内容とは関係しない。

3.1.2 指導計画(プレゼンテーションソフト)

プレゼンテーションソフトについての授業では、プレゼンテーションソフトの操作方法、プレゼンテーションソフトを利用したスライドの作成方法、成果物の改善のプロセス、他者

の成果物を評価する方法について学習した。ここでは、9 時間をかけて、演習を中心に授業を行った。各時間の授業内容を表 2 に示す。表中の「改善」とは前時に作成した成果物を相互評価の結果をふまえて改善することをさしており、改善された成果物は、再び提出され、再度評価を受ける。第 6 回の授業から 3 時間をかけて制作し、システムを用いて成果物を提出させた。第 6 回の授業では、ソフトウェアの操作として、基本的操作、レイアウト、字体などの変更方法などについて説明した。第 9 回の授業からは、相互評価を行いながら 4 時間をかけて成果物の改善を行い、第 13 回の授業で最終的な評価を行い、第 14 回の授業で発表会を行った。なお、11, 12, 14 回の授業では、プレゼンテーションの内容にあてた授業時間は 60 分、13 回の授業では 30 分であった。

今回の実践では、単にプレゼンテーションソフトの使用法ではなく、相互評価を行うこと

で、受け手を意識させて成果物を作成させることができた。情報教育においては、情報発信について学ばせる場面が多く、このような場合に相互評価を導入することは意義があることであると思われる。また、評価者として他人の成果物を評価することは、情報活用能力の前提となる情報を正しく理解するための練習方法として有効である。

3.1.3 課題

Microsoft Office PowerPoint を使用して、自分の趣味や興味があることについて紹介するプレゼンテーション資料を作成させ、システムを使用して学習者間で評価させた。スライド数は5~9枚で、クリップアート、インターネット上の素材は自由に使ってよいこととした。

課題を出す際に、次の5つの評価項目を提示し、それぞれの項目について1~5(5が最も良い)の5段階で行った。それぞれの評価項目についての説明は行ったが、1~5をつけるための基準は指示しなかった。

- レイアウト、色づかい、字体は見やすいか?
- 表現は分かりやすいか?
- 情報の信頼性は高いか?
- 見ていて楽しい気分になるか?
- 全体としてうまくまとめられているか?

1番目の項目は、スライドの視覚的な印象と文字の読みやすさについて評価させた。プレゼンテーションソフトウェアの使い方について説明する際に、色の組合せや文字の量などについて、具体例を示しながら説明を行った。2番目の項目は、成果物の中の文章などの理解性について評価させた。3番目の項目は、プレゼンテーションの信頼性として、成果物に書かれていることの根拠について評価させた。4番目の項目は、成果物全体から受ける印象について、注意を引くような構成やデザインになっているかどうか評価させた。5番目の項目は、成果物全体から伝えたいテーマについて伝わってくるような構成になっているかどうか評価させた。

プレゼンテーションソフトウェアの操作を評価するのではなく、操作方法を知ったうえで、効果的なプレゼンテーション資料を作成できることを評価することを目指して、評価項目を設定した。これらの項目は、実際に課題を行い、他の学習者の成果物を見ることによって、身につく目標であると思われる。本実践で、制作、評価、改善のサイクルを繰り返しながら、学習させた。

表3 群ごとに使用した課題と結果表示
Table 3 Conditions of each group.

	相互評価実践前に 使用した課題セット	結果表示画面での 個人差補正の有無	相互評価実践後に 使用した課題セット
A群	セット①	表示なし	セット②
B群	セット②	表示なし	セット①
C群	セット①	表示あり	セット②
D群	セット②	表示あり	セット①

3.2 グループごとの評価内容

第1回の授業で行ったコンピュータの知識についてのアンケートをもとに学習者を4群に分け、表3にあげた条件で学習させた。それぞれの群の学習者数は、4人である。

3.2.1 実践中の評価の変化と成果物の改善

評価者の個人差を補正した結果を表示することによる影響について調べるために、A、B群の学習者には、フィードバック画面で補正した結果を表示し、C、D群の学習者には表示しなかった。以下、補正結果を表示しないA、B群をまとめて補正非表示群、C、D群をまとめて補正表示群とよぶ。

3.2.2 別集団の成果物の評価

相互評価の実施の前後での他者の成果物に対する評価の違いを比較するために、昨年度の受講生が今回の実践と同じテーマで作成した成果物を、今回の実践と同じ評価項目で評価させた。実践の前後に、すべての群で6人分の成果物を評価させた。実践の前後で異なる成果物を評価させ、かつ実践の前後で比較するために、A、C群には、実践前にはセット①、実践後にはセット②の成果物を評価させ、B、D群には、セット②、セット①の順で評価させた。なお、セット①とセット②には重複する成果物はなく、それぞれのセットには、昨年度の評価が良かったものから悪かったものまで混ぜ、2つのセット間での差が大きくなるようにした。すべての群で、実践前、実践後ともに、後述する相互評価支援システムの相互評価機能を使用して評価させた。

3.3 相互評価支援システム

今回の実践では、筆者らが開発した相互評価支援システム²⁷⁾を利用して、webを利用して、成果物の提出、相互評価、評価結果の確認を行った。このシステムの構成を図1に示

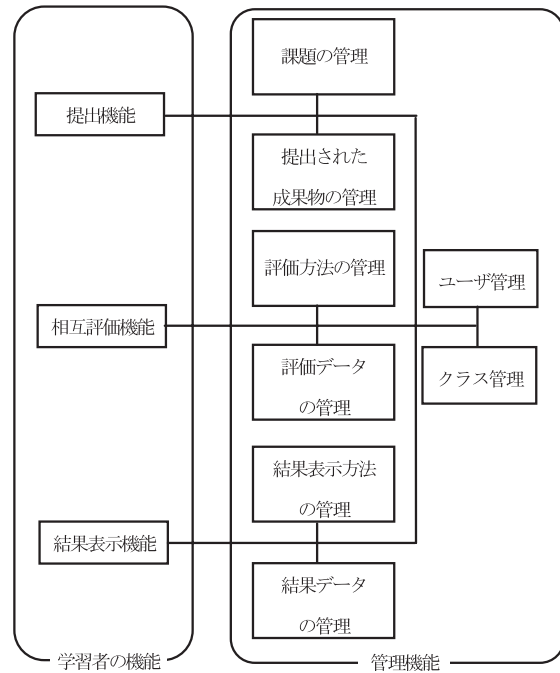


図 1 相互評価支援システムの構成

Fig. 1 Construction of peer assessment support system.

す．ここでは，学習者が使用する提出機能，相互評価機能，結果表示機能について簡単に述べる．

3.3.1 成果物の提出

学習者は，作成した成果物の電子ファイルを，Web ブラウザ上から提出する．提出ファイルは，パワーポイントで作成した 1 ファイルに限定した．提出期間中であれば，回数の制限なく，提出したファイルを差し替えられるようにした．ファイルを提出するとき，評価を担当する他の学習者に対するコメント（進捗状況，アドバイスを欲しい点など）を入力させた．

3.3.2 相互評価

本システムは，すべての学習者が，すべての成果物を評価できない場合に，評価すべき

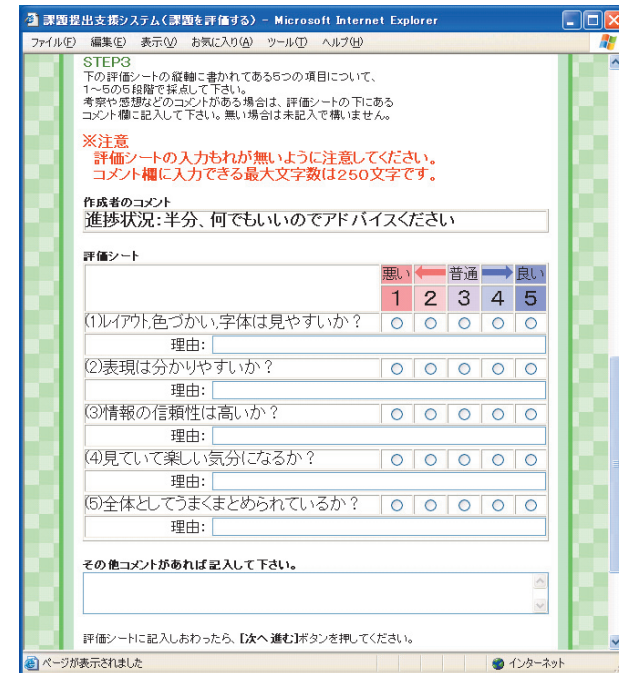


図 2 評価画面

Fig. 2 Screenshot of "peer evaluation."

相手を自動的に割り当てる機能を持っている．以前に，同じ科目で行ったプレゼンテーション資料を作成する場面での相互評価で，前述の「お互い様効果」が観測されたため²⁷⁾，今回の実践では，「お互い様効果」を除去する，すなわち評価した相手から，評価されることがないようにした．具体的には，1 人が評価するレポートの数を m とし，学習者番号を $1, 2, \dots, i, \dots, n$ とすると，学習者 i は，学習者番号 $i - m$ ($i - m < 0$ のときは $i - m + n$) から連続して m 人分のレポートを評価するようにした．今回は，1 人の学習者が 6 人分の評価を行うようにした．成果物を評価する画面を図 2 に示す．学習者は，まず，別ウインドウに表示された評価対象となる成果物を見て，その評価結果を採点フォームに入力する．各評価項目には 5 段階で採点し，その点数をつけた理由を記入させる．評価をする際には，成果物を提出する際に入力したコメントを見ることができ，それに基づいたコメントを書く

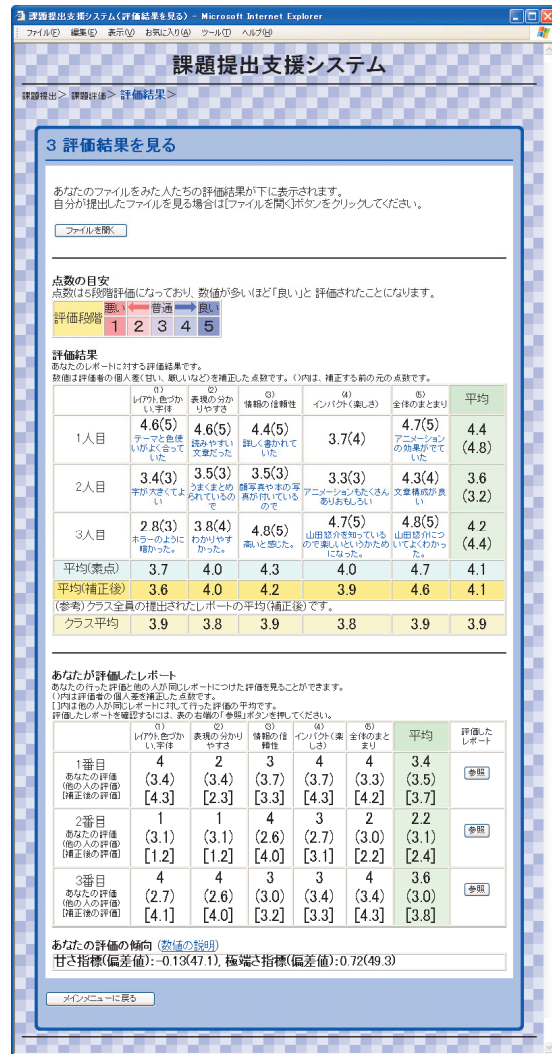


図3 フィードバック画面(補正表示群)
Fig.3 Screenshot of "results display."

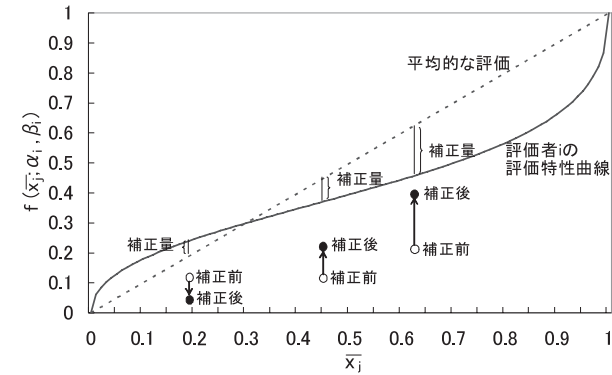


図4 Example of a revision of an evaluated score in considering of the evaluator's characteristic.
横軸は j の提出した成果物を評価した他の評価者の評価の平均を、縦軸は評価値を表す

こともできる。

3.3.3 結果表示

補正表示群では、評価者の評価特性を補正した数値として、学習者が受けた評価に対する補正後の評価と補正された評価の平均、他者の成果物に対して行った評価に対する補正後の評価と評価の特徴を表す評価特性パラメータを表示した(図3)。評価者の評価特性の補正には、項目応答理論²⁸⁾のメタファを用いた評価モデルにより各評価者の評価特性を推定し、平均的な評価特性の評価者による評価値との差を用いて、評価値の補正を行う方法¹⁸⁾を使用した。評点の補正例を図4に示す。評価者の評価特性曲線は、評点に差をつける度合いを表す α 、評点の厳しさの度合いを表す β の2つのパラメータを用いて表している。評価特性曲線は、 α の値が大きいほど変局点付近での傾きが急になり、 β の値が大きいほど全体的に低い評価値つけることになる。比較的少数のデータや欠損値の多いデータでも適用できることが特長である。

4. 結果

4.1 実践中の評価の変化と成果物の改善

まず、他の学習者に対して行う評価が、相互評価の回数を重ねることや評価値を補正した結果を提示することによって、甘くなったり、厳しくなったりするのかを調べた。その

表 4 他の学習者の成果物に対して行った評価の推移
Table 4 Changing the evaluation of other learners' products.

	1回目	2回目	3回目	4回目	5回目	平均	
評価値の平均	補正表示群	19.56	20.53	20.19	20.08	19.87	20.05
	補正非表示群	20.10	21.03	20.79	20.65	20.51	20.62
	平均	19.83	20.78	20.49	20.37	20.19	20.46
評価値の分散	補正表示群	0.39	0.45	0.54	0.65	0.74	0.56
	補正非表示群	0.38	0.42	0.50	0.60	0.66	0.51
	平均	0.39	0.44	0.52	0.63	0.70	0.53
改善に有益なコメント数の平均	補正表示群	0.63	0.63	1.00	1.63	1.50	1.08
	補正非表示群	0.75	0.75	0.88	1.50	1.63	1.10
	平均	0.69	0.69	0.94	1.56	1.56	1.09

結果を表 4 に示す。授業の回数と補正表示の有無が評価に及ぼす影響を 2 元配置による分散分析を行った結果、いずれの主効果も有意ではなかった（補正表示： $F(1, 14) = 0.003$, $p > .10$, 回数： $F(4, 14) = 1.573$, $p > .10$, 交互作用： $F(4, 14) = 1.402$, $p > .10$ ）。つまり、繰り返し評価を行ったり、補正した結果を表示したりすることによって、評価が甘くなったり、厳しくなったりすることはなかった。

次に、他の学習者に対して行った評価の差のつけ方（メリハリのつけ方）を調べるために、相互評価の回数を重ねることによって分散が変化するかどうかについて調べた。その結果を表 4 に示す。授業の回数と補正表示の有無が分散に及ぼす影響を 2 元配置による分散分析を行った結果、授業回数の主効果と交互作用が有意であり（授業回数： $F(4, 14) = 23.374$, $p < .01$, 交互作用： $F(4, 14) = 3.382$, $p < .05$ ）、補正表示の主効果は有意ではなかった（ $F(1, 14) = 0.694$, $p > .10$ ）。つまり、繰り返し評価を行うことによって、評価に差をつけるようになり、交互作用について調べた結果、回数を重ねるごとに大きくなる割合は補正表示群の方が大きく、補正表示機能によって評価により差をつけることができるようになった。

形式的評価として相互評価を行ったため、成果物の改善のために有益であると思われる短所を具体的に指摘しているコメント数の変化を調べた。その結果を表 4 に示す。短所や改善点を具体的に指摘し、改善に有益なコメントの数が、相互評価の回数を重ねることや評価値を補正した結果を提示することによって、増加するかどうかについて調べた。授

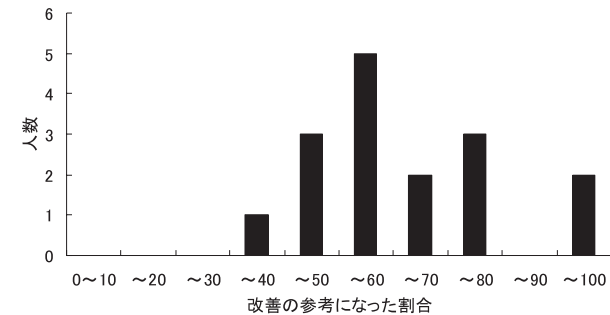


図 5 改善の参考になった評価の割合
Fig. 5 Ratio of the useful evaluation to improve.

業の回数と補正表示の有無がコメント数に及ぼす影響を 2 元配置による分散分析を行った結果、授業回数について主効果が有意であり（ $F(4, 14) = 7.799$, $p < .01$ ）、補正表示の有無と交互作用は有意ではなかった（補正表示： $F(1, 14) = 0.069$, $p > .10$, 交互作用： $F(4, 14) = 1.571$, $p > .10$ ）。つまり、補正表示の有無に関係なく、繰り返し評価を行うことによって、短所を具体的に指摘しているコメントが増えたことが分かった。実践後に行ったアンケートで、「自分が受けた評価の中で、改善の参考になった評価の割合」を尋ねた結果、図 5 に示すように、平均 67% の評価が参考となったと答えた。また、指摘されたコメントを反映して成果物の改善が行われているか調べたところ、82% の短所の指摘について、何らかの修正が行われていた。このことから、多くの学習者がコメントを参考に、成果物の改善を行ったことが推察された。

4.2 別集団の成果物の評価

相互評価の実施の前後で、他の学習者の成果物に対する評価がどのように変わるかを、昨年度の受講生が作成した成果物を使用して調べた。事前評価における 1 項目あたりにつけた平均は、3.43（セット ①：3.43, セット ②：3.43）で、事後評価は、3.72（セット ①：3.75, セット ②：3.68）となり、相互評価事後に行った評価の方が、評価が高い傾向となった。実施時期と補正値の表示の有無が評価に及ぼす影響を 2 元配置による分散分析を行った結果、実施時期についての主効果が有意であり（ $F(1, 14) = 12.270$, $p < .05$ ）、補正の表示の有無と交互作用は有意ではなかった（補正表示 $F(1, 14) = 1.134$, $p > .10$, 交互作用： $F(1, 14) = 0.137$, $p > .10$ ）。

評価対象となった成果物は、昨年度に成績評価のために同じ評価項目で授業担当教員が評

表 5 事後アンケートの結果
Table 5 Result of the questionnaire.

質問項目	補正あり	補正なし	検定
1. 他人の成果物を見ることは有意義だった	3.50	4.00	
2. 評価結果を確認することは有意義だった	4.13	4.25	
3. 自分の行った評価と同じ成果物に対して他人が行った評価とを比べることは有意義だった	3.88	3.38	
4. 学生間での評価を導入した授業を受けたい	3.88	3.00	
5. 補正前の評価は、おおむね納得できた	4.13	4.50	
6. 補正後の評価は、おおむね納得できた	4.25	-	
7. 自分の評価の特徴に気がついた	4.25	3.38	
8. パラメータ α (評点に差をつける度合い) の意味は理解できた	3.75	3.25	
9. パラメータ β (評点の厳しさの度合い) の意味は理解できた	3.88	3.38	
10. 評価者によって評価に差がある	4.25	4.38	
11. 評価者によって差があつて不公平である	3.00	2.25	
12. 評価者の差を補正するほうが望ましい	3.75	2.38	*

※検定欄の記号は Mann-Whitney の U 検定の結果(*:5%有意)

価を行った。教員が行った評価と今回の受講生が行った評価の関係について調べた¹⁸⁾。それによると、事前・事後でのそれぞれの群の Spearman の順位相関係数を比較した結果、「補正結果を表示することにより、他人の行った評価をより強く意識して、適切な評価をつけられるようになることが示唆された」と考察されている。

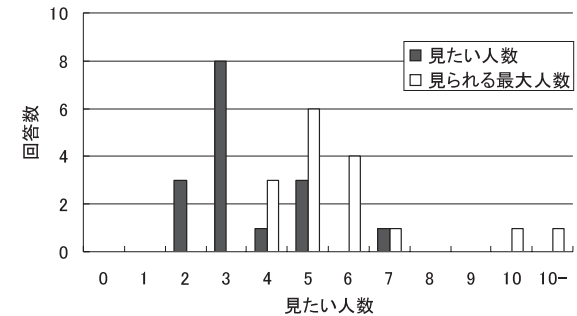


図 6 評価したい人数
Fig. 6 Number of reports to evaluate.

4.3 事後アンケート

実践の終了後に行ったアンケートの結果を表 5 に示す。補正非表示群にも、補正結果を表示したフィードバック画面を提示して補正についても尋ねた。表 5 の項目 4~12 は、文献 18) を再掲したものである。

補正表示群と補正非表示群の回答について、Mann-Whitney の U 検定を行った。「12. 評価者の差を補正するほうが望ましい」についてのみ、有意な差が見られた ($U = 14.0$, $p < .05$)。文献 18) によると、「学習者は、評価者によって評価が甘かったり、厳しかったりすることを認識している。また、その差を不公平であると受け取る者も一定数見られた。実践で毎回補正を見て慣れてきた補正表示群では、補正することが望ましいと回答した者が多く見られた」と考察されている。

今回の実践での相互評価は、他人の成果物の評価、評価結果の確認、他人の評価との比較について、補正表示群と非表示群ともに、肯定的にとらえられていた。

4.4 評価人数

1 回あたりの相互評価を行う人数について検証するために、事後アンケートで評価人数について尋ねた。今回の課題、評価方法の場合に、毎回何人分の成果物を見てみたいか、最大で何人見ることができるかについてのアンケートの結果を図 6 に、毎回何人に見てもらいたいか、最小でも何人に見てもらいたいについてのアンケートの結果を図 7 に示す。

毎回評価してもらいたい人数の平均が 4.8 人であったのに対して、評価したい人数は 3.5 人となり、68%の学習者が評価したい人数よりも評価してほしい人数を多くあげた。今回は、1 回の講義で、1 人の学習者が行う評価、受ける評価はともに 6 人で行った。先述した

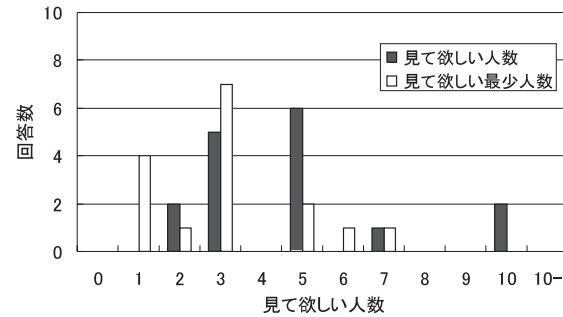


図 7 評価してもらいたい人数

Fig. 7 Number of reports to be evaluated.

改善のためのコメント数などから、評価人数としては十分であったと思われる。学習者の印象としては、6人はやや多く、4人程度を希望する者が多くいた。評価人数を減らしたときに、有益な評価の数や評価特性の補正に与える影響について、今後検討を行う必要がある。

5. まとめ

連続9回行った演習授業の中で、1つの課題に対して相互評価を繰り返し5回行い、成果物を改善していった。結果表示画面で、個人差補正を行った値を表示するグループとそうでないグループに分けて、成果物の改善と学習者の評価の能力の変化について分析した。

実践によって、他者の成果物の短所や改善点を具体的に指摘できるようになり、多くの学習者が相互評価でのコメントを活かした改善を行うことができた。また、評価者の評価特性の個人差の補正表示を行うことで、より適切な評価をすることができることができた。実践が終了した後に行ったアンケートでは、多くの学習者が相互評価の教育的効果を認識し、また、評価者の個人差を補正することが望ましいことが示唆された。適切な評価ができるようになるためには、他人が行った評価の結果を、そのまま見せるだけでは十分でないことが分かった。

今回使用した相互評価支援システムの課題として、適切な評価を行わなかった学習者に対して、教員が介入するための機能があげられる。たとえば、評価が適切でない学習者について、当該学習者とその評価を教師に自動的に報告する機能、当該学習者が行った評価を他の学習者に見せない、またはフィードバック情報の計算に含めない機能、当該学習者に他の評価者との評価の違いを指摘する機能などが考えられる。教員が適切に介入することにより、

相互評価における効果をあげることができると思われる。また、協調学習の場面で相互評価を行う場合は、学習コミュニティの一員ではない教員からの評価が最も適切であるとは限らない。教員が、評価を通じて学習者の考えを知るとは、学習者や学習コミュニティを理解する助けになると思われる。

情報教育に相互評価を導入する意義について述べる。今回の実践では、単にプレゼンテーションソフトの使用法ではなく、相互評価を行うことで、受け手を意識させて成果物を作成させることができた。情報教育においては、情報発信について学ばせる場面が多く、このような場合に相互評価を導入することは意味がある。また、評価者として他人の成果物を評価することは、情報活用能力の前提となる情報を正しく理解するための練習方法として有効であり、今後、多くの実践が期待される。

これまでの相互評価の実践研究では、成果物の改善のプロセスに、学習者間相互評価を継続的に行い、その教育的効果を検討した研究はなかった。継続的に学習者間評価を実施することで、他者の成果物の評価を適切にできるようになることを実証したことは、重要である。

参考文献

- 1) Wiggins, G.: A true test: Toward more authentic and equitable assessment, *Phi Delta Kappan*, Vol.70, pp.703-713 (1989).
- 2) Sizer, T.: *Horace's Compromise: The Dilemma of the American High School* Boston, Houghton Mifflin (1984).
- 3) 遠藤貴広: 米国エッセンシャル・スクール連盟における「逆向き計画」による学校改革: セイヤー中・高等学校の実践を例に, 京都大学大学院教育学研究科紀要, Vol.53, pp.220-232 (2007).
- 4) Brown, J.D. and Hudson, T.: The alternatives in language assessment, *TESOL Quarterly*, Vol.32, pp.653-675 (1998).
- 5) 大岩 元, 橘 孝博, 半田 亨, 久野 靖, 辰巳丈夫: 情報科教育法, オーム社 (2001).
- 6) 植野真臣: 先端的 e-learning の理論と実践, 教育心理学年報, No.44, pp.126-137 (2005).
- 7) 大倉孝昭, 高村博正, 奥田アレックス・H: 英語による効果的なプレゼンテーションと相互評価—連続性・発展性の視点から, 教育福祉研究, No.30, pp.20-30 (2004).
- 8) 山口顕司, 大塚 茂, 森田慎一, 松本 至, 矢壁正樹, 早水庸隆, 大塚宏一: 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, 工学教育, Vol.55, No.1, pp.41-46 (2007).
- 9) 大倉孝昭, 村尾敏彦: 映画の音声と解答を同期させ相互評価により聴解力を高める CALL システムの開発と学習者評価, 日本教育工学会論文誌, Vol.27, No.3, pp.315-324 (2003).

- 10) 柴田好章, 小川 亮: 相互評価システムの開発と大学情報科目における利用, 日本教育工学会論文誌, Vol.25, No.Suppl., pp.33-38 (2001).
- 11) 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, 日本教育工学会論文誌, Vol.24, No.1, pp.93-98 (2000).
- 12) Sluijsmans, D.M.A., Moerkerke, G., van Merriënboer, J.J.G. and Dochy, F.J.R.C.: Peer assessment in problem based learning, *Studies in Educational Evaluation*, No.27, pp.153-173 (2001).
- 13) Akahori, K. and Kim, S.M.: Peer evaluation using the Web and comparison of meta-cognition between experts and novices, *Proc. World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp.1484-1487 (2003).
- 14) Cousins, J.B. and Whitmore, E.: Framing participatory evaluation, *New Directions for Evaluation*, No.80, pp.5-23 (1998).
- 15) 大塚雄作: 高等教育における評価の諸要素とその機能—改善指向の評価文化の形成に向けて, 大学評価, No.1, pp.27-66 (2002).
- 16) 鹿毛雅治: 教育評価再考—実践的視座からの展望, 心理学評論, Vol.47, No.3, pp.300-317 (2004).
- 17) 加藤 浩, 山下 淳, 藤原康宏, 鈴木栄幸: 社会構成主義から見た相互評価の意義, 日本科学教育学会年会論文集, No.30, pp.179-180 (2006).
- 18) 藤原康宏, 大西 仁, 加藤 浩: 多数の欠測値を含む相互評価データにおける評価者の評価特性補正方法, 日本教育工学会論文誌, Vol.31, No.3, pp.373-381 (2007).
- 19) Liu, J. and Sadler, R.W.: The effect and affect of peer review in electronic versus traditional modes on L2 writing, *Journal of English for Academic Purposes*, Vol.2, No.3, pp.193-227 (2003).
- 20) Figl, K., Bauer, C., Mangler, J. and Motschnig, R.: Online versus face-to-face peer team reviews, *Proc. 36th ASEE/IEEE Frontiers in Education Conference*, pp.7-12 (T1H) (2006).
- 21) 中原 淳, 浦嶋憲明, 西森年寿, 鈴木真理子, 今井 靖, 山際耕英, 永田智子: 相互評価機能を実装した電子掲示板の開発と評価, 日本教育工学雑誌, Vol.26, No.Suppl., pp.33-38 (2002).
- 22) 布施雅彦, 湊 淳, 小澤 哲: ビデオオンデマンドとウェブデータベースを利用した相互・自己評価システムの開発—高専における問題解決学習の事例, 教育システム情報学会誌, Vol.19, No.4, pp.206-211 (2002).
- 23) 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型 WBT システム, 情報処理学会論文誌, Vol.48, No.3, pp.1532-1545 (2007).
- 24) 下村 勉, 天野昌和, 須首野仁志: 学習成果の改善を図る Web ベース相互評価システムの開発と活用, 三重大学教育実践総合センター紀要, No.23, pp.31-36 (2003).
- 25) 天野昌和, 下村 勉: Web ベース相互評価システムによる学習集団内での評価観点の共有, 日本教育工学会第 19 回全国大会講演論文集, pp.7-8 (2003).

- 26) 生田目康子: ピア・レビューをとまなうグループ学習の評価—斉型プログラミング授業への適用, 情報処理学会論文誌, Vol.45, No.9, pp.2226-2235 (2004).
- 27) 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, Vol.31, No.2, pp.125-134 (2007).
- 28) 芝 祐順: 項目反応理論—基礎と応用, 東京大学出版会 (1991).

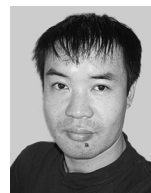
(平成 19 年 12 月 10 日受付)

(平成 20 年 7 月 1 日採録)



藤原 康宏 (正会員)

平成 5 年神戸大学教育学部卒業。平成 7 年同大学大学院教育学研究科修士課程修了。平成 19 年総合研究大学院大学文化科学研究科博士課程修了。博士 (学術)。平成 10 年岩手県立大ソフトウェア情報学部講師。教育工学の研究に従事。日本教育工学会, 電子情報通信学会, 教育システム情報学会, 日本行動計量学会, IEEE, ACM 各会員。



大西 仁 (正会員)

平成 2 年東京工業大学工学部卒業。平成 7 年同大学院総合理工学研究科博士課程修了。博士 (学術)。同年放送教育開発センター助手。メディア教育開発センター助手を経て, 現在, メディア教育開発センター准教授, および総合研究大学院大学文化科学研究科准教授併任。認知科学の研究に従事。電子情報通信学会, 日本認知科学会, 日本ヴァーチャルリアリティ学会, The Cognitive Science Society 各会員。



加藤 浩（正会員）

昭和 58 年慶應義塾大学大学院工学研究科修士課程修了。同年日本電気入社。平成 11 年東京工業大社会理工学研究科博士課程修了。博士（工学）。平成 12 年メディア教育開発センター助教授。平成 12 年東京工業大大学院社会理工学研究科助教授連携併任。平成 13 年総研大文化科学研究科助教授併任。現在、メディア教育開発センター教授、および総研大文化科学研究科教授併任。教育工学の研究に従事。日本教育工学会、日本科学教育学会、電子情報通信学会、日本認知科学会、ヒューマンインタフェース学会、日本テスト学会、American Educational Research Association 各会員。
