

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」—[†]

藤原康宏^{*1,*3}・大西仁^{*2,*3}・加藤浩^{*2,*3}

岩手県立大学ソフトウェア情報学部^{*1}・メディア教育開発センター^{*2}

総合研究大学院大学文化科学研究科^{*3}

近年の教育評価の研究では、学習の場面と独立した評価ではなく、学習の場面に埋め込まれた評価が試みられている。その方法の1つとして、学習者同士が評価を行うことが有用であることが知られている。相互に学習コミュニティメンバー全員の評価をすることは、メンバーの人数が多くなるにつれて困難になるため、評価すべき相手を選択する必要性が生じる。その選択方法を考えるために、評価する学習者が、評価対象となっている学習者からも評価されるか否かにより評価が変化するかについて実験を行った。その結果、お互いに評価しあう方が甘い評価を行う傾向があり（「お互い様効果」）、お互いに評価しあわない方が教員の行った評価に近いことが分かった。そこで「お互い様効果」を除去し合理的に評価すべき相手を選択し、相互評価を容易に実施できるツールを開発し、その評価を行った。学生と教員による評価の結果、相互評価をさせる場合に有効であることがわかった。

キーワード：相互評価，教育評価，お互い様効果，評価ツール

1. はじめに

近年の教育評価の研究では、情報通信技術の革新を背景にして、学習の場面と独立した評価ではなく、学習の場面に埋め込まれた評価が試みられている。池田(1992)は、未来のテストとして、授業、学習(訓練)、評価の3者の区別がない一体化したテストを予見しており、その後に開発されたWBT(Web Based Training)では、学習過程に評価を組み込んだ多くの教材が開発

され、実践されている(例えば、中原・西森 2004, UENO 2004, 藤原ほか 2005)。

e-ラーニングの普及によって、電子会議室での議論や協調学習が行われるようになった(例えば、EDELSON *et al.* 1995, 中原ほか 2000, 緒方ほか 2000)。このような学習場面では、狭義の知識獲得にとどまらず、学習者が学習活動を通して実践的能力を獲得することを想定しており、従来のような形式のテストではなく、課題などの提出物を課して、学習過程に評価を埋め込むことが行われている。植野(2005)は、この提出物を教師だけでなく、コミュニティ内における学習者同士で行うことの利点として「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」と述べている。

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。柴田・小川(2001)は、情報関連科目の作品制作で、Webベースの相互評価システムを用いてお互いに評価させた後に、その結果にもとづいて評価レポートを作成させた。その結果、他者のコメントに基づいて多くの学生

2007年2月6日受理

[†] Yasuhiro FUJIHARA^{*1,*3}, Hitoshi OHNISHI^{*2,*3} and Hiroshi KATO^{*2,*3}: An Evaluation Support System for Impartial Peer Evaluation - "Reciprocal Effect" Occurred by Selecting Evaluators in Case of Peer Evaluation of Learning Outcome -

^{*1} Iwate Prefectural University, 152-52, Takizawa-Sugo Takizawa-mura Iwate, 020-0193 Japan

^{*2} National Institute of Multimedia Education, 2-12, Wakaba, Mihama-ku, Chiba-shi Chiba, 261-0014 Japan

^{*3} The Graduate University for Advanced Studies, 2-12, Wakaba, Mihama-ku, Chiba-shi, Chiba, 261-0014 Japan

が作品の改良を行うことができた」と述べている。布施ら(2002)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。天野・下村(2003)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(下村ほか 2003)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。

しかし、実際に学習者間で評価を行う際に、相互にすべての学習者を評価することは、学習者の数が増えるにつれて時間的・労力的に困難になっていき、評価すべき相手を選択する必要がある。学習者が、他の学習者による評価を受け入れるには、評価が公平に行われることが重要である。全ての学習者が他のすべての学習者を評価する場合は、ほぼ公平であるといえるが、評価者を選択しなければならない状況では、評価者の割り当て方によっては不公平が生じる。評価者によっても評定の甘さや厳しさなど個人差があるといわれており(渡部ほか 1998)、個々の評価者の特性についても考慮する必要があるが、ここでは、不公平が生じる要因として、評価する学習者が、評価対象となっている学習者からも評価されるか否かによって生じる可能性を検討する。なぜなら、この要因は、評価者の選択方法に関するものであり、個々の評価者の評価特性を知る以前に考慮できるからである。

本研究では、評価者を選択しなければならない状況で、公平性の高い相互評価を行うために、評価対象となっている学習者からも評価されることが評価に与える影響を調べ、その影響を低減することを目的とする。

本研究では、お互いに評価しあう場合(以下、お互い様条件)とお互いに評価しあわない場合(以下、非お互い様条件)では、どちらがより適切な評価をするか実験した。そして、その結果、お互いに評価しあわない場合の方が、評価がより厳しくなり、かつ教師の評価に近くなることが明らかになった。

次に、実験結果をもとに、適切なフィードバックを行うための相互評価支援システムを開発した。本システムは、相互評価において評価すべき相手を選択せざるを得ない場合に、お互いに評価しあうことによる評価の変動を最小限に抑える評価者決定、レポートの提出、相互評価の実施、結果のフィードバックの機能から構成される。システムの実用性を調べるために、学習者に対しては、システムの使いやすさ、システムを

使った相互評価の意義、得られた相互評価の結果と教員の評価の比較を、教員に対しては、システムの使いやすさ、授業担当者が設定する項目の妥当性、システムを使った相互評価の意義について、評価を行った。評価の結果、お互いに評価しあわないようにすることで、より公平で適切な評価結果がフィードバックされることができるとの示唆を得た。

2. 実 験

学習者間で評価を行う際、評価する学習者が、評価対象となっている学習者からも評価されるか否かにより評価が変化する可能性がある。評価する学習者が、評価対象となっている学習者からも評価される場合の方が、そうでない場合より、お互いの評価が甘くなることは、経験的に予想される。しかし、お互いに評価しあうと、評価するレポートの短所を見落としやすくなるかもしれないが、長所を見出しやすくなる可能性もあり、お互いに評価しあうのとしあわないのとでは、どちらの評価がより適切であるかは明らかではない。そこで本実験では、学習者がお互いに評価しあうか否かにより評価値が変化するか、どちらの評価値が適切であるかに加え、評価の仕方がどのように変化するかを検討した。

2.1. 方 法

2.1.1. 実験の概要

A大学文学部1年生を対象に行っている情報処理入門科目「コンピュータ入門」の2004年7月20日に行った授業の中で、あらかじめ電子ファイルで提出させたレポートを、受講者32人に相互評価実施システムを利用して他の学習者のレポートを評価させた。

相互評価実施システムは、本実験のために開発したものであり、電子ファイルでのレポート受領機能、学習者間でのレポートの評価実施機能を持っている。学習者は、指定された順番にレポートを評価する。なお、一度評価を決定すると、前のレポートに戻って再評価することはできない。

この科目の受講者は、同じクラスの者で顔見知りである。コンピュータ教室で一斉にシステムを使用して、評価させた。隣の学習者とは離れており、画面を覗き見ることでせず、相談せずに評価をさせた。

学習者は、それぞれ6名のレポートを評価した。6名のうち3名については評価をした相手からも自分が提出したレポートが評価され(お互い様条件)、残りの3名については評価した相手からは評価されない(非お互い様条件)。学習者には、それぞれの条件で評価を開始

する前に、お互い様条件では評価した相手からも自分が提出したレポートが評価されること、非お互い様条件では評価した相手からは評価されないことを説明した。

2.1.2. 評価者の決定方法

学習者は、表1に示す学習者が提出したレポートを評価する。1列目が評価を行う学習者の学習者番号、1行目が評価対象となるレポートを提出した学習者の学習者番号を表している。学習者番号は、座席の並びから通し番号で付けた。表中の数字は、それぞれの学習者が評価を行う順番を表しており、数字の入っていない学習者が提出したレポートの評価は行わない。例えば、学習者1は、学習者13, 14, 15, 2, 3, 4の順に、提出されたレポートを評価し、それ以外のレポートは評価しない。薄い網掛けはお互い様条件を、濃い網掛けは非お互い様条件を表している。

お互い様条件では、4名1組のグループを8グループ作成し、同じグループに属する他の3人について評価する。グループは座席の近い者で構成し、自分以外のグループメンバーのレポートを担当させた。非お互い様条件では、それぞれの学習者番号を*i*とすると、*i*+12 (32を超える場合は32を減じる、以下同様)、*i*+13、*i*+14の3人の離れた座席に座っている学習者のレポートを評価する。

学習者番号が偶数の者は先にお互い様条件から、奇数の者は先に非お互い様条件から評価を行った。学習者が評価するレポートの順番は、例えば、ある学習者のレポートが常に最初に評価されることがないように、個々のレポートが評価される順序を考慮して決定した。

2.1.3. 科目の目標及びレポートの課題

「コンピュータ入門」は、1年生を対象にした科目

であり、1) コンピュータに関する基本的な操作技能の習得、2) 主なアプリケーションの利用方法の習得、3) アプリケーションを使用した表現力の育成を目標としている。Microsoft Office PowerPoint を使用して、3時間をかけて、「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、相互評価実施システムを使用して学習者間で評価させた。ソフトウェアの操作については、課題作成前に修得させ、相互評価の結果は成績評価に反映させた。

レポート課題を出す際に、「与えられた条件の中で見る人を意識したプレゼンテーション資料を作るようになる」という目標を提示し、その評価項目として、次の5つを示した。

- ・課題の条件（スライド数など）を満たしているか
- ・プレゼンテーションソフトの機能を適切に使いこなしているか
- ・情報の信頼性は高いか
- ・レイアウト、色づかい、字体は見やすいか
- ・見ていて楽しい気分になるか

評価は、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対する詳細な評価基準は指示しなかった。

2.2. 結 果

2.2.1. お互い様条件と非お互い様条件の比較

学習者が行ったそれぞれの評価項目の評価値を合計し、それぞれの条件で学習者が行った評価の分布を図1に示す。それぞれの学習者がつけた評価の平均の平均はお互い様条件では23.22、非お互い様条件では22.50となった。条件と実験を行う順序の要因が評価に及ぼす影響を2元配置による分散分析により検定を行った。結果、条件について主効果が有意であり($F(1,30)=4.344$, $p<.05$)、実験の順序($F(1,30)=0.808$, $p>.10$)およびそれらの交互作用は有意ではなかった($F(1,30)=0.023$, $p>.10$)。

学習者が他者につけた評価の全体での平均値は22.9（1項目あたり4.58）と高く、天井効果が見られた。図1から、特にお互い様条件でこの傾向が強いことがわかる。したがって、お互い様条件と非お互い様条件との差は見かけの平均値の差よりも大きいと考えられる。

2.2.2. 教員の評価との比較

お互い様条件と非お互い様条件ではどちらが適正な評価を行っているかを検討するために、教員が成績評価の目的で同じ評価項目でレポートを採点した結果と、学習者の行った評価との関係を調べた。その散布図を

表1 各学習者が担当したレポート

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1																																
2																																
3																																
4																																
5																																
6																																
7																																
8																																
9																																
10																																
11																																
12																																
13																																
14																																
15																																
16																																
17																																
18																																
19																																
20																																
21																																
22																																
23																																
24																																
25																																
26																																
27																																
28																																
29																																
30																																
31																																
32																																

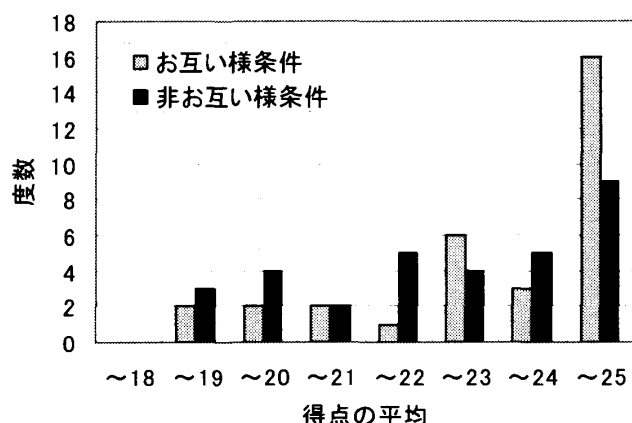


図1 学習者の行った評価の分布

図2に示す。教員の評価とお互い様条件での学習者の評価の Spearman の順位相関係数(以下、相関係数)は0.439, 教員の評価と非お互い様条件での学習者の評価の相関係数は0.804であり有意な相関が見られた(お互い様条件: $t(30)=2.68$ $p<.05$, 非お互い様条件: $t(30)=7.42$ $p<.01$). そして、両者の相関係数には有意な差があった($\chi^2(1)=5.93$ $p<.05$). すなわち、非お互い様条件での評価の方がお互い様条件と比べると、教員の行った評価に近い結果となった。ただし、相関係数の値の差が大きくなった原因として、お互い様条件では天井効果の影響で差がつきにくかったことも影響していると考えられる。

担当教員の評価の妥当性を検証するために、別のクラスの同じ科目の授業を、同じテキストを使って行っている教員2名に評価を依頼した。前述の教員と依頼した教員の行った評価の相関係数は、それぞれ0.561, 0.666, 追加評価を行った教員間の評価の相関係数は0.751であり、3名の教員が行った評価の相関は高かった。2名の教員の評価と学習者の評価との相関を調べた結果、授業を担当した教員の評価と同様に、非お互い様条件での学習者の評価の方が、いずれの教員の評

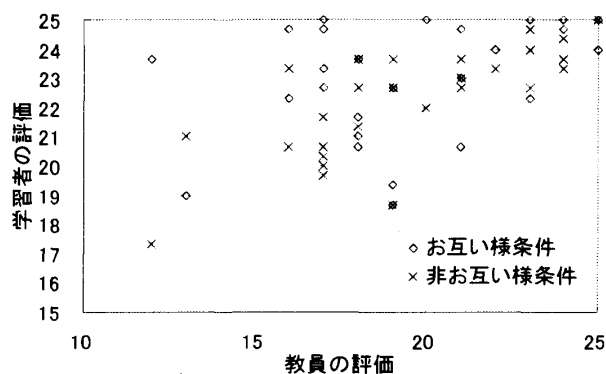


図2 教員の評価との比較

価に近いことが分かった。(1人目の教員:お互い様条件での相関係数0.268, 非お互い様条件での相関係数0.510, 2人目の教員:お互い様条件での相関係数0.383, 非お互い様条件での相関係数0.600)

2.2.3. 自由記述項目

学習者に評価させる際に同時に記入させたコメントを分類し、集計したものを表2に示す。入力されたコメントの数は、お互い様条件、非お互い様条件で差がなかった($\chi^2(1)=0.40$ $p>.10$)。コメントの内容を肯定的なもの、批判的なものに分類したところ、お互い様条件、非お互い様条件の間で差が生じた($\chi^2(1)=5.98$ $p<.05$)。例えば「絵が動く所がおもしろくいろいろな種類の絵を使っていて良かったです。」など、肯定的なコメントのうち、良かった点を具体的に指摘しているコメントは、どちらの条件でも半数以上だった。逆に、例えば「改行の位置が変だと思った。」など、批判的なコメントは、非お互い様条件の方が多く指摘され、その内容は教員による評価とすべて一致していた。また、批判的なコメントは、どちらの条件でもほとんどが問題点を具体的に指摘していた。今回の評価項目とは直接関係ない自己紹介の内容に関するコメント(例えば「私も〇〇が好きでよく見に行きます」)は非お互い様条件の方がやや多く見られた。

2.2.4. 別クラスでの実験

同様の結果が別の集団に対しても得られるかどうかについて評価した。2006年6月6日に、同じ科目で、同様の課題及び評価項目で実験を行った(FUJIIHARA *et al.* 2006)。受講生は16名で、お互い様条件、非お互い様条件をそれぞれ3名分の評価をさせた。

その結果、それぞれの条件でつけられた評価値の平均は、お互い様条件20.63, 非お互い様条件19.36であり、お互いに評価しあう方が、評価が甘くなることが確認された。また、教員の評価との相関を調べたところ、お互い様条件0.180, 非お互い様条件0.676となり、非お互い様条件の方が適切に評価されていた。

評価結果のフィードバック後に行った相互評価についてのアンケート(5段階、5が当てはまる)の回答の平均値は、「評価者によって評価に差がある」が4.31, 「評価者によって差があつて不公平」が3.06, 「親しい友達の課題は、公平な評価は難しい」が3.69であり、

表2 コメントの集計

	コメントなし	コメントあり	肯定的(具体的)	批判的(具体的)	内容
お互い様条件	14	91	79 (48)	7 (6)	13
非お互い様条件	15	90	66 (43)	18 (17)	17

学習者にとって相互評価をすることの難しさがうかがわれた。公平性についての問題は、4.4.で述べる。

2.3. 考 察

実験の結果、学習者間で評価をする時に、評価する相手も評価者を評価する場合は、そうでない場合に比べて、評価値が甘くなる場合があることがわかった。教員の評価と比べることで、どちらの条件が、より適正な評価を行っているかを調べたところ、評価する相手に評価されない場合の方が、より適正であることもわかった。さらに、自由記述のコメントを分析した結果、肯定的なコメントにおいては条件による差はなかったが、否定的なコメントは、評価する相手に評価される場合には、抑制されるということもわかった。

これらの実験の結果は相手に高い評価をすることで、互恵的に自分にも高い評価をしてもらいたいという期待があるため起きたのではないかと考えられる。そこで、ここではこの現象を「お互い様効果」と呼ぶことにする。TRIVERS (1971) は、互恵的利他主義として、社会生物学の観点から、他者に対して返報が期待できる状況で、互恵性に基づく利他的行動について説明した。もし、お互い様効果が互恵的利他主義から起こるものであれば、短期間で修正することは困難であると予想される。相互評価において、全員が相互評価することが困難で、相手を選択しなければならない場合には、自分が評価した相手からは評価されないような組み合わせにすることが望ましいことがわかった。

3. 相互評価支援システムの開発

3.1. システム概要

評価すべき相手を選択する必要がある場合に、適切なフィードバックを行うために、評価すべき相手を自動的に選択し、相互評価を容易に行うことができるシステムを開発した。本システムは、PHP と PostgreSQL を用いて開発し、Microsoft Internet Explorer6.0で動作を確認した。

今回開発した相互評価支援システムは、レポートの提出、相互評価、評価結果の確認を Web 上から行うことができる。学習者を対象とした機能とレポート出題者を対象とした機能（以下、管理機能と呼ぶ）に分けることができる（図3）。

学習者を対象とした機能は、提出機能、相互評価機能、結果表示機能の3つから構成される。提出機能は、電子ファイルとして作成されたレポートをシステムにアップロードし、登録することができる。相互評価機

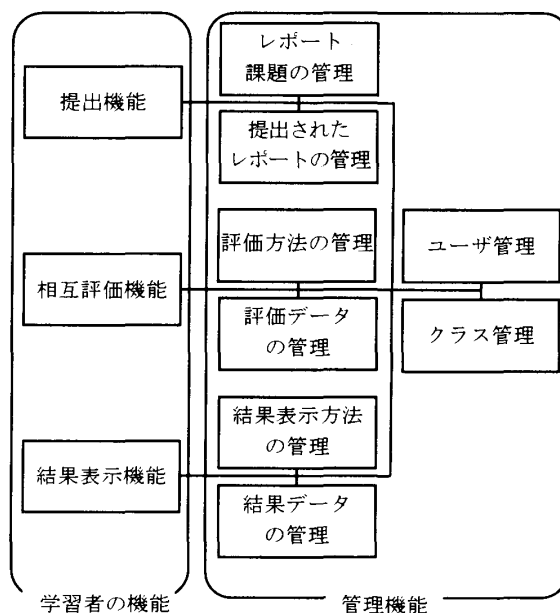


図3 システム構成

能は、課題を提出した学習者間で、評価相手のレポートを表示し、評価を行うことができる。それぞれの学習者が他のすべての学習者のレポートの評価をできない場合は、それぞれの学習者がどのレポートを評価するかを自動的に設定することができる。結果表示機能では、提出したレポートが受けた評価と、学習者が他の学習者に行った評価を確認することができる。

管理機能は、学習者の管理、クラス管理、レポート課題の管理、提出されたレポートの管理、評価方法の管理、評価データの管理、結果表示方法の管理、結果データの管理からなる。

3.2. 提出機能

3.2.1. レポートの管理

授業担当者は、学習者に課したレポートについて、レポートの名前、提出期間、提出ファイルの制限に関する情報を、システムを使って登録する。個々のレポートは、レポートの状態（「提出不許可」、「提出期間前」、「未提出」、「提出済」、「提出受付終了（提出確定）」、のいずれか）、提出されたファイルについての情報、学習者が入力したコメントの情報と共に管理される。

3.2.2. レポートの提出

学習者がメインメニューで、課題の提出を選べると、レポート提出期間中であれば、課題の提出画面が現れる。学習者は、作成したレポートの電子ファイルを、Web ブラウザ上から提出する。あらかじめ設定した提出可能なファイルの形式、ファイル数の上限、ファイルのサイズの上限、提出した後のファイルの差し替え

の可否に基づいて、提出ファイルは扱われる。課題提出時にコメントを入力させ、その内容を相互評価の際に表示させることもできる。

3.3. 相互評価機能

3.3.1. 評価者自動割り当てアルゴリズム

本システムは、すべての学習者が、すべてのレポートを評価できない場合に、評価すべき相手を自動的に選択できる機能を持っている。その際、小グループに分けてその中で評価させるかどうか、小グループに分けない場合は、「お互い様効果」を除去するかどうかというオプションを選択できる。

「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するため、全評価者数を n とすると、1人当たり最大で $(n-1)/2$ 人分のレポートを評価することができる。それを超える場合は、超えた人数に応じて、全学習者が同程度にお互いに評価しあうように設定する。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 $i-m$ ($i-m < 0$ の時は $i-m+n$) から連続して m 人分のレポートを評価する。 $n=5$ の時の割り当ての例を表3に示す。1列目は評価者の学習者番号、1行目は評価される学習者の番号を表し、表中の「○」がついたレポートが評価される。この場合、 $m=2$ の時はお互い様効果が除去されるが、 $m=3$ の時は、お互い様効果を完全に除去できる上限を超えているので、表の網掛け部分に対してお互いに評価しあうことになる。

3.3.2. 評価方法の設定

授業担当者は、まず評価画面を作るために必要な情報として、評価項目、それぞれの評価項目への回答方法などについてシステムを使って登録する。次に、相互評価期間、割り当て方法、学習者1人当たりの評価人数、お互い様効果を除去するかどうかなどを設定する。評価者の割り当ては、前述のアルゴリズムにしたがって自動的に実行される。

3.3.3. レポートの評価

相互評価を実施する画面を図4に示す。学習者は、

表3 お互い様効果を考慮した割り当て

$m=2, n=5$ の場合

	1	2	3	4	5
1				○	○
2	○				○
3	○	○			
4		○	○		
5			○	○	

お互いに評価しあう部分はない

$m=3, n=5$ の場合

	1	2	3	4	5
1					○
2	○				
3		○			
4			○		
5				○	

の部分はお互いに評価しあう

授業担当者の設定に基づいて選ばれたレポートを1つずつ評価する。まず、別ウインドウに表示された評価対象となるレポートを見て、評価結果を入力する。評価項目は、 n 段階評価ボタン、数字入力フィールド、自由記述のフィールドのうち該当するものが表示される。なんらかの原因で、評価途中で終了した場合、前回の続きから評価を行うことができる。

3.4. 結果表示機能

3.4.1. フィードバックの設定

授業担当者は、提出されたレポートのフィードバック画面に関して、閲覧可能期間、教員による評価の表示の有無、他者に行った評価の表示の有無、平均値など表示する統計量の設定を行う。全員が評価を終了していない時であっても、その時点で受けている評価を見られるように設定することもできる。

3.4.2. フィードバック情報の表示

学習者が受けた評価と他の学習者に行った評価を確認する画面を図5に示す。学習者の受けた評価の値と

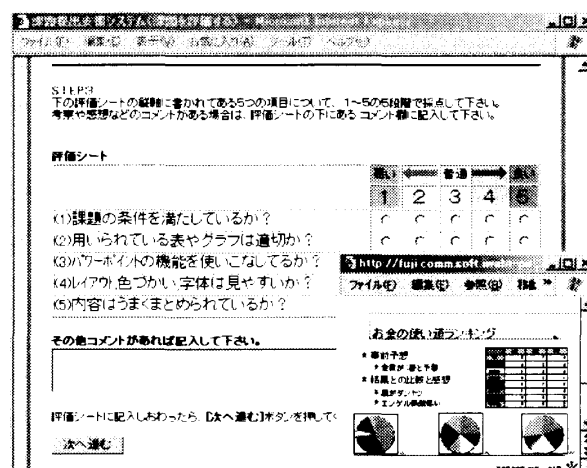


図4 相互評価画面

評価者	1	2	3	4	5	平均	コメント
1 日目	4	4	3	1	2	2.8	文字が少し見づらい
2 日目	5	4	4	4	3	4.0	すごくきれいな仕上がりだと思えます。
3 日目	4	2	4	3	4	3.4	最後に1つ誤字なところがありました。
平均	4.3	3.3	3.7	2.7	3.0	3.4	
クラス平均	3.6	3.5	3.5	3.3	3.4	3.4	

図5 結果表示画面

その平均などと、学習者が他の人を評価した結果が表示される。自分の行った評価を他の学習者がつけた評価と比較して、見直すことができる。なお、評価を確認するときに、評価の際に使ったファイルを見ることができる。

4. 相互評価支援システムの評価

4.1. ユーザビリティに関する評価

2006年1月24日に、システム改善のための評価として、A 大学1年生10名にシステムを使用させた。表計算ソフトを使って、指定したデータを加工する課題を出した。課題はシステムを使って提出させ、相互評価を行い、結果を確認させた。システムの機能ごとにアンケートを行い、問題が指摘された箇所について、意見を聞いた。

システムのユーザビリティに関する選択式項目に対して得られた220件の回答のうち、それぞれの選択肢の中で最も肯定的なものが選ばれたのは173件であり、残り47件についてインタビューを行い、インタビューで指摘された内容について、他の人の意見も聞いてみた。その結果、意見が一致したインターフェースに関する2点について修正した。

4.2. 学習者による評価

4.2.1. 評価方法

2006年2月6日に、システムを利用して適切な相互評価ができるかどうか評価するために、B 大学1年生45名に情報処理入門科目の最終回の授業でシステムを使用させた。あらかじめ与えていた課題をシステムを使って提出させ、他の6人の受講生のレポートを評価させた。課題は、プレゼンテーションソフトを使用し、アンケートデータをまとめるもので、「課題の条件を満たしているか」、「用いられている表やグラフは適切か」、「パワーポイントの機能を使いこなしているか」、「レイアウト、色使い、字体は見やすいか」、「内容はうまくまとめられているか」の観点で、それぞれ5段階で評価させた。システムの使用後に、システムの使いやすさと相互評価について、アンケートを行った。

4.2.2. アンケートの結果

アンケートは、システムの使いやすさに関するもの7項目、相互評価に関するもの6項目、自由記述項目から構成される。選択式の項目に対しては5段階で回答し、5が最も肯定的を意味する。アンケート項目と集計結果を表4に示す。

システムの使いやすさに関する7項目は、いずれも肯定的な結果となっており、自由記述項目からも、特

表4 アンケートの結果

	質問	平均	標準偏差
システム	課題の提出機能は、使いやすかった	4.02	0.87
	他人の課題を見る操作は、使いやすかった	3.79	1.14
	他人の課題に点数をつける操作は、使いやすかった	4.10	0.76
	結果表示画面は、分かりやすかった	3.88	0.97
	システムを全体的に見ると、使いやすかった	3.88	1.02
	システムを全体的に見ると、操作方法は分かりやすかった	3.74	0.96
	システムを全体的に見ると、レイアウトは見やすかった	4.02	0.75
相互評価	今回のように自分に対してつけられた評価結果を、後で確認したい	3.86	1.12
	自分に対してつけられた評価結果は、納得できた	3.74	1.13
	自分に対してつけられた評価結果は、参考になった	3.95	1.03
	今回のように他人のレポートを評価することは、勉強になる	4.10	0.93
	他の人が同じレポートにつけた評価と比べることは、勉強になる	3.76	1.12
	自分の評価の特徴（点が甘い、辛いなど）に気がついた	3.45	1.29

に問題点は見つからなかった。相互評価や評価の結果に対しても、肯定的に捉えられていた。自由記述項目への回答から、自分に対してつけられた評価に対して疑問に思う学習者も存在した。そこで、それぞれの学習者のつけた評価と教員のつけた評価の相関係数を計算したところ、教員との相関が0.2未満の学習者が10名おり（図6）、この10名がつけた評価をみると、1つの視点（例えばアニメーションの動き）に過度に重視するなど独自の視点で評価している者（2名）、レポートの内容を反映していない評価をつけた者（3名）、欠点を指摘する表現が適切でない者（4名）が見られた。これらの特徴は相関が0.2以上の者には見られなかった。このことが、アンケートで、自分に対してつけられた評価結果が納得できないと回答した者が一定数いた1つの要因となっていると思われる。

他者のレポートを適正に評価する能力とよい課題を遂行する能力の関係を調べるために、各レポートにつ

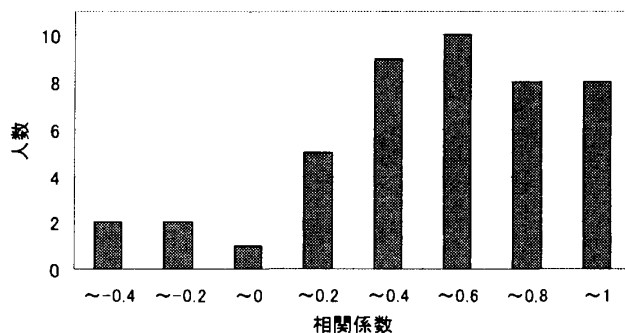


図6 学習者が行った評価と教員の評価との相関の分布

いて、それぞれの学習者が行った評価と教員が行った評価の差と、その学生のレポートに対して教員が行った評価を比較した。その相関係数は-0.10となり、他者のレポートを適正に評価する能力と課題を遂行する能力は、別の能力であることが示唆された。

今回は、学習者が他の学習者のレポートに対して行った評価と、同じレポートに対する別の学習者の評価結果も表示させた。アンケートで「他の人が同じレポートにつけた評価と比べることは勉強になる」の項目に対する平均は3.76であったのに対して、「自分の評価の特徴に気づいた」と回答した者には3.45であった。適切な評価をできない者に、評価のやり方を自ら学ばせるためには、学習者にとって自分の評価の特徴をより理解しやすいように表示する方法などについて検討する必要がある。

4.3. 教員による評価

4.3.1. 評価方法

B大学で5年以上情報処理入門科目を担当している教員6名に、システムの評価を依頼した。今回評価を依頼した6名の教員は、紙で提出されたレポートや学生の発表で相互評価を行った経験があった。

まず、評価前に相互評価に対して持っている印象について、アンケートに回答してもらい、それに基づいてインタビューを行った。次に、学習者による評価で行った提出、相互評価、結果の確認のプロセスを体験してもらい、次に授業担当者が設定する項目を管理者モードで確認してもらった。それぞれの管理画面ごとに、アンケートに記入してもらい、それを元にインタビューを行った。アンケートは、それぞれの管理画面での設定内容の妥当性、教員が授業で使用する際に必要な機能からなる。システムの使用が終わった後、このシステムを使った相互評価について、評価前に行った相互評価に対して持っている印象と同じ項目に回答してもらった。

4.3.2. 結果

(1) 学習者用機能のユーザビリティ

課題提出機能、相互評価機能、結果表示機能の学習者用画面の使いやすさについて5段階で回答してもらった(5が学生にとって使いやすい、1が使いにくいを表す)。それぞれの機能の平均は、4.17、4.83、4.50であり、コンピュータに不慣れな者が使用することを考慮した要望が2点あったが、大きな問題は指摘されなかった。

(2) 授業担当者用機能の評価

次に授業担当者が、課題提出、相互評価、結果表示

の各機能で設定、確認できる項目について、その必要性について5段階で回答してもらった。その結果を表5に示す。「相互評価実施状況の確認」機能の中の「学習者が回答した評価結果を授業担当者が変更する機能」(3.33)を除いて、評価の平均が4以下となる項目はなかった。また、提示した項目以外で、必要とされる項目について尋ねたところ、「結果のフィードバックに関する設定」で評価をした人の名前の表示・非表示を設定できるようにしたて欲しいという意見があったが、それ以外は特に指摘はされなかった。全体としては、管理機能で設定、確認ができる項目については適切であったと考えられる。しかし、例えば、「レポートの提出状況の表記が分かりにくい」など、表現や設定方法が分かりにくい点が7箇所指摘され、それぞれ改善を行った。また、「設定の変更履歴を記録しておきたい」など、より使いやすくするためのインターフェースに関する改善点が5箇所提案され、それについては今後の課題とした。

(3) 相互評価の印象

システムの評価に先立って行った相互評価の印象についてのアンケート(5段階で回答し、5があてはまる、1があてはまらないを表す)の結果を表6に示す。相互評価の教育的効果に関しては、自身の教育実践を踏まえて、全ての教員が肯定的な意見を持っており、機会があれば相互評価を授業に取り入れていきたいと考えていた。しかし、実施にかかる手間や評価結果のフィードバックのことを考えると、相互評価を手軽に行うことは難しいという印象を持っていることがうか

表5 各機能の評価(5段階)

機能	項目数	平均	分散
課題提出画面の設定	7	4.69	0.56
課題提出状況の確認	14	4.67	0.57
相互評価実施に関する設定	8	4.73	0.45
相互評価実施状況の確認	9	4.54	0.77
結果のフィードバックに関する設定	5	4.63	0.49
結果分析用ファイル作成に関する設定	8	4.77	0.50

表6 相互評価に対して持っている印象(5段階)

項目	評価前	評価後
他の学生のレポートを見せることは勉強になる	4.50	4.50
他の学生からの評価を見せることは勉強になる	4.33	4.83 *
実施に教員の手間がかかる	2.83	2.17 *
評価結果のフィードバックが難しい	2.83	2.33 *
教員が評価する手間が省ける	3.00	3.17
学生の評価には信頼性がある	2.83	3.67 *
相互評価を取り入れた授業をしたい	4.67	4.83

記号はWilcoxonの符号付き順位検定の結果を表す。*:5%有意、+:10%有意傾向

がえる。学生の行った評価の信頼性に対しては、相互評価実施前に評価の方法や評価項目を明確に示すなどが必要であるとの意見が多かった。学生の評価結果について、教員の評価より甘くなる傾向にあるが、学生は教員が行う評価よりも他の学生による評価結果の方が真摯に受け止める傾向があるとの意見が聞かれた。

システムの使用後に、同じ項目について尋ねたところ、評価前の印象と比べて2つの項目で統計的に有意な差がみられた。このシステムでは、電子ファイルでの提出、相互評価、フィードバックを管理することができることから相互評価を実施しやすく、学生に評価方法を明確に伝えることができ、それぞれの学生に確実に結果を伝えることができることが評価された。学生の評価の信頼性について、成績評価に用いることを前提としている場合は、できるだけ公平な評価が必要となり、お互い様効果が出ないように自動的に割り当てる機能は有用であるとあるという意見が聞かれた。学習過程で行う課題の改善を目的として相互評価をさせる場合は、課題の改善につながるコメントを必ず入力させるなどが必要で、このシステムは学習過程での課題の改善のための評価でも有効であるという意見を得た。

4.4. 議 論

教員による評価の中で、多くの教員が、相互評価では学習者にとって納得できる評価結果をフィードバックされることが重要であることを指摘した。特に、成績評価など成果に対して評価を行う場合などは、個々の評価者による評価の特性の違いを吸収・是正して、可能な限り公平な評価値を求める手法が必要である。本論文では、お互い様効果に注目して実験及びシステム開発について述べてきたが、評価者の個々の評価特性を平準化し、より適切なフィードバックを行うことが今後の課題である。植野ら(2006)は、e-Learningで得られた大量の相互評価のデータから項目応答理論を利用して、レポートの採点を行う方法を提案している。筆者らは、今回の行ったようなクラス集団など少数のデータを対象とした個人差の補正方法を開発し、シミュレーションでは適切に補正できることを確認している(藤原ほか 2006)。

大学生の行う他者評価と教員の行う評価と差はないことは知られているが(例えば、ORPEN 1982)、今回の実践では、他の学習者を適切に評価する能力が劣る学習者も少数ながら存在した。最終回の授業で相互評価を行ったため、評価の観点が理解できない学習者は少なかった。しかし、形成的評価として行う場合は、適

切な評価ができない未熟な学習者がいることも考えられる。このような場合は、評価の分布から異質な評価をする学習者を検出する機能をシステムに付加して、フィードバックから取り除いたり、教員が介入するなど対処する必要があると思われる。協調学習における相互評価では、他者評価の能力を育成することを期待する場合もあり、本システムを効果的に利用して、学習者の他者評価の能力を向上させることについても今後の課題である。今回開発したシステムに評価者の個々の評価特性を平準化した値をフィードバックする機能を付加し、一連の作品作りの中で6回の相互評価を行わせる実践を行ったところ、他者評価の能力が向上したことが確認された(FUJIHARA *et al.* 2006)。以上から、本システムを利用し、より適切なフィードバックを行うことで評価の能力の育成を助けることができると考えられる。

5. お わ り に

本研究では、学習者間で評価をする時に、お互いに評価しあう場合とお互いに評価しあわない場合では、どちらがより適切な評価をするか実験した。実験の結果、評価する相手も評価者を評価する場合は、そうでない場合に比べて、評価が甘くなる場合があり(お互い様効果)、お互いに評価しあわない場合の方が、教員の評価と相関が高く、より適切であることがわかった。また、お互いに評価しあわない場合の方が、短所をより適切に指摘し、長所の指摘はお互いに評価する場合と比べて劣らないことが分かった。

評価する相手を選択する必要が生じる場合に、合理的に評価すべき相手を選択できる相互評価支援システムを開発した。学習者及び教員に対する評価の結果、ユーザビリティについては、コンピュータの操作に熟達していない者でも使いやすいたことが確認された。また、教員に対する評価から、相互評価では公平な評価が行われることが重要であるとの意見が出された。相互評価による学習効果を高めるためには、教員が行う評価とは異なり、同僚の学習者の行った評価を真摯に受け入れられる状況を作ることが重要であり、できるだけ公平性が高く、適切な評価結果をフィードバックすることが必要であると考えられる。以上のことから、お互い様効果を除去し、より公平な相互評価を支援できる本システムは、有用であることが分かった。

今後は、本システムを学習過程で繰り返し利用する場合を想定し、評価結果のフィードバックの後に、例えば、自分が受けた評価が適切であったかを評価する

ことなど、評価者と被評価者との間でやりとりができるコミュニケーションツールとしての利用についても検討していきたい。

参 考 文 献

- 天野昌和, 下村勉 (2003) 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 : 7-8
- EDELSON D.C., O'NEILL D.K., GOMEZ L.M. and D'AMICO, L (1995) A design for effective support of inquiry and collaboration. Proceedings of CSCL'95 : 107-111
- 藤原康宏, 大西仁, 永岡慶三 (2005) 情報処理入門科目におけるオンライン個別学習システムを利用した授業実践とその効果. 日本教育工学会論文誌, **29**(Suppl) : 109-112
- 藤原康宏, 大西仁, 加藤浩 (2006) 形成的評価における相互評価支援システムの利用について. 電子情報通信学会技術研究報告, **ET2006-33** : 65-70
- FUJIHARA, Y., OHNISHI, H. and KATO, H. (2006) A practice of peer evaluation in ICT education using a report evaluating support system. Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn): 2816-2822
- 布施雅彦, 湊淳, 小澤哲 (2002) 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, **19**(4) : 206-211
- 池田央 (1992) テストの科学. 日本文化科学社, 東京
- 中原淳, 西森年寿, 杉本圭優, 浦島憲明, 永岡慶三 (2000) 議論を通じた共同的な問題解決を支援する CSCL 環境の開発. 日本教育工学雑誌, **24**(Suppl.) : 97-102
- 中原淳, 西森年寿編著 (2004) e-learning マネージメント. オーム社, 東京
- 緒方宏明, 松浦健二, 矢野米雄 (2000) WWW を利用した開放型グループ学習支援における Knowledge Awareness Map. 教育システム情報学会誌, **17**(3) : 263-274
- ORPEN C. (1982) Student versus lecturer assessment of learning: a research note. Higher Education, **11** : 567-572
- 柴田好章, 小川亮 (2001) 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, **25**(Suppl) : 33-38
- 下村勉, 天野昌和, 須曾野仁志 (2003) 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, **23** : 31-36
- TRIVERS R. (1971) The evolution of reciprocal altruism. The Quarterly Review of Biology, **46** : 35-57
- UENO M. (2004) Data mining and text mining technologies for collaborative learning in LMS SAMURAI. Proceedings of IEEE Conference (Computer Science) : 1052-1053
- 植野真臣 (2005) 先端的 e-learning の理論と実践. 教育心理学年報, **44** : 126-137
- 植野真臣, 岡本敏雄, 永岡慶三 (2006) 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** : 7-12
- 渡部洋, 平由実子, 井上俊哉 (1998) 小論文評価データの解析. 東京大学教育学部紀要, **28** : 143-164

Summary

In recent research of an educational evaluation field, learning involved evaluation is practiced instead of a term-end exam. It is known to executes peer evaluation as a useful evaluation method. However, the more the number of class members increases, the more evaluating all the others becomes difficult. It is necessary to devise a method of selecting the evaluator so that the evaluation is impartial. In this paper, we conducted an experiment to examine whether there is a difference in the evaluation that the learner evaluates one another or not. As a result of the experiment, it was found that in the case of evaluating one another, the evaluation is softer("reciprocal effect") and farther from teacher's evaluation than the case of evaluating not one another. Then, we developed peer evaluation support system to execute the peer evaluation easily that removed "reciprocal effect". As a result of the assessment of our system by student and teacher, it was confirmed that our system is effective in peer evaluation.

KEY WORDS: PEER EVALUATION, EDUCATIONAL EVALUATION, RECIPROCAL EFFECT, EVALUATION TOOL

(Received February 6, 2007)