

協調学習における学習者間評価の支援に関する研究

藤原 康宏

博士（学術）

総合研究大学院大学
文化科学研究科
メディア社会文化専攻

平成19年度
(2007)

目 次

第 1 章 序論	1
1.1. 研究の背景	2
1.2. 学習の改善を目的とした評価	5
1.2.1. 形成的評価	5
1.2.2. 学習場面と評価が一体化した学習環境	6
1.2.3. コミュニティ内評価	8
1.3. 学習者間の相互評価の課題	9
1.3.1. 信頼性と妥当性	9
1.3.2. 評価の公平性に関する問題	10
1.3.3. 相互評価の効果	12
1.4. 本研究の目的と構成	13
第 2 章 評価者の選択が評価の公平性に与える影響	15
2.1. はじめに	16
2.2. 実験方法	16
2.2.1. 実験の概要	16
2.2.2. 評価者の決定方法	17
2.2.3. 科目の目標及び課題	20
2.3. 結果	21
2.3.1. お互い様条件と非お互い様条件の比較	21
2.3.2. 教員の評価との比較	23
2.3.3. 自由記述項目	26
2.3.4. 別クラスでの実験	26
2.3.5. 考察	27
2.4. まとめ	27

第3章 相互評価データにおける	
評価者の評価特性補正方法	29
3.1. はじめに	30
3.2. 補正方法	30
3.2.1. 評価特性のモデル化	31
3.2.2. 評価特性モデルを使った評点の補正方法	34
3.3. パラメータの計算方法及び精度	34
3.3.1. パラメータの計算方法及び精度	34
3.3.2. 計算例	35
3.4. 適用例	37
3.4.1. データ	37
3.4.2. 結果	37
3.5. まとめ	39
第4章 相互評価支援システムの開発と評価	41
4.1 はじめに	42
4.2 システム概要	42
4.2.1. システム構成	42
4.2.2. 動作環境	45
4.2.3. データベース	45
4.3. 相互評価支援システムの主な機能	46
4.3.1. 提出機能	46
4.3.2. 相互評価機能	49
4.3.3. 結果表示機能	54
4.4. 相互評価支援システムの評価	57
4.4.1 システム改善のための評価	57
4.4.2 学習者による評価	59
4.4.3 教員による評価	62
4.5. まとめ	65

第5章 相互評価支援システムを導入した実践の効果	67
5.1. はじめに	68
5.2. 実践の概要	68
5.2.1. 授業科目の概要	68
5.2.2. 指導計画（プレゼンテーションソフト）	68
5.2.3. 課題	69
5.2.4. 相互評価の方法	70
5.2.5. システムの運用	72
5.3. 結果	76
5.3.1. 実践中の他者の成果物の評価	76
5.3.2. 事前評価と事後評価の比較	77
5.3.3. 事後アンケート	78
5.3.4. 評価者特性モデルによる評価能力の変化	79
5.3.5. 評価人数	82
5.4. まとめ	83
第6章 結論	85
6.1. 本研究のまとめ	86
6.2. 議論	88
6.3. 今後の課題	89

参考文献

本論文に関する研究発表

謝辞

目 次

- 図 2.1 相互評価実施システムの画面例 (p.18)
- 図 2.2 実験で使ったコンピュータ教室 (p.19)
- 図 2.3 学習者の行った評価の分布 (p.23)
- 図 2.4 教員の評価との比較 (p.24)
- 図 3.1 モデルとパラメータ との関係 (p.33)
- 図 3.2 モデルとパラメータ との関係 (p.33)
- 図 3.3 評点の補正の例 (p.34)
- 図 3.4 推定された評価特性曲線と実測値 (p.36)
- 図 3.5 推定されたパラメータの分布 (p.38)
- 図 3.6 推定された値と実測値との差 (p.38)
- 図 4.1 システム構成 (p.43)
- 図 4.2 システムを使った相互評価の流れ (p.44)
- 図 4.3 メインメニュー (p.45)
- 図 4.4 課題設定画面 (p.48)
- 図 4.5 成果物提出画面 (p.49)
- 図 4.6 相互評価方法設定画面 (p.52)
- 図 4.7 相互評価実施画面 (p.53)
- 図 4.8 結果表示設定画面 (p.55)
- 図 4.9 結果表示画面 (p.56)
- 図 4.10 学習者が行った評価と教員の評価との相関の分布 (p.62)
- 図 5.1 課題提出画面 (p.73)
- 図 5.2 課題評価画面 (p.74)
- 図 5.3 フィードバック画面 (補正表示群)(p.75)
- 図 5.4 アンケート結果 (改善の参考になった評価の割合)(p.77)
- 図 5.5 実践前の各学習者の評価特性パラメータの分布 (p.80)
- 図 5.6 実践後の各学習者の評価特性パラメータの分布 (p.80)
- 図 5.7 アンケートの結果 (毎回評価したい人数)(p.82)
- 図 5.8 アンケートの結果 (毎回評価してもらいたい人数)(p.82)

表 目 次

表 2.1	各学習者が担当した成果物 (p.19)
表 2.2	学習者ごとの評価結果 (p.22)
表 2.3	学習者が受けた評価と教員による評価 (p.25)
表 2.4	コメントの集計 (p.26)
表 4.1	ユーザ管理テーブル (p.46)
表 4.2	クラス管理テーブル (p.46)
表 4.3	課題管理テーブル (p.47)
表 4.4	提出管理テーブル (p.47)
表 4.5	お互い様効果を考慮した割り当て (p.50)
表 4.6	評価方法管理テーブルおよび評価シート管理テーブル (p.51)
表 4.7	評価データ管理テーブル (p.51)
表 4.8	結果表示方法管理テーブル (p.54)
表 4.9	結果データ管理テーブル (p.54)
表 4.10	機能と質問項目の対応 (p.57)
表 4.11	最も肯定的な選択肢を選んだ割合 (p.58)
表 4.12	相互評価に関するアンケート結果 (p.59)
表 4.13	学習者に対するアンケートの結果 (p.61)
表 4.14	管理機能一覧 (p.63)
表 4.15	管理機能における各機能の評価 (5 段階) (p.64)
表 4.16	教員が相互評価に対して持っている印象 (5 段階) (p.65)
表 5.1	授業内容 (p.68)
表 5.2	授業内容と使用する機能 (p.69)
表 5.3	群ごとの事前事後の課題と結果表示 (p.71)
表 5.4	評価者の割り当て (p.72)
表 5.5	他の学習者の成果物に対して行った評価の推移 (p.77)
表 5.6	アンケートの結果 (評価者特性の補正に関する項目) (p.78)
表 5.7	実践前後での誤差の変化 (p.79)

第 1 章 序論

1.1. 研究の背景

本研究は、レポートや制作物等の成果物の作成を通じて学習する場面において、学級等の学びの場を共有する学習コミュニティのメンバー同士が、お互いの成果物を相互に評価することによって、学習者が自身の成果物の改善点に気付くことを支援する手法を提案する。本論文では、学習の結果として作成する作文、レポート、作品等のことを成果物と呼ぶ。

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盛んになっている。その背景の1つとして学力や評価に対する考え方の変化がある。Wiggins ら(1989)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評価であるべきとするオーセンティック・アセスメントの考え方を提唱した。オーセンティック・アセスメントの成立に関した実践としては、Coalition of Essential School の教育評価改革(Sizer, 1984)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(岡本, 2007)。このような背景から、パフォーマンス評価やポートフォリオ評価など、ペーパーテストに替わる評価手法が新しい評価として、実践されている(Brown & Hudson, 1998)。我が国では、総合的な学習の時間において、成果物に対して自己評価、相互評価を評価に取り入れた様々な実践が報告されている(例えば、小島・寺崎, 2001; 国立教育政策研究所, 2003)。以前から、学習者の観察、成果物の評価、相互評価などペーパーテスト以外の方法による評価も行われることはあったが、ペーパーテストを補完する役割として行われることが多かった。しかし、ペーパーテストで測定が困難である場合や、学習の場面と切り離れた評価を行うことが適切とはいえない場合が多い。例えば、従来の教科型の学習のように体系的な知識の獲得を目指した学習ではなく、プロジェクト学習のような問題解決志向の学習場面では、狭義の知識獲得を目指したものではないため、その学習場面での状況と切り離して、学習者の個人内の知識のみを評価の対象とするのではなく、文脈の中で総合的に評価されるべきである(加藤ほか, 2006a)。梶田(1994, pp.72-84)は、問題解決力・思考力や関心・意欲・態度といった以前の方法では評価が難しかった能力の評価を「内面的なものの評価」として整理した。また、学力観の変化に伴って、評価の観点や考え方を見直す必要性が議論されてきている(例えば、野嶋, 2002)。

高等教育においても様々な場面で、成果物の相互評価の実践が行われている。プレゼンテーションスキルを題材とした実践(例えば,大倉ほか,2004;山口ほか,2007),外国語教育でコミュニケーションを題材とした実践(例えば,大倉・村尾,2003),情報教育を題材とした実践(例えば,柴田・小川,2001),プロジェクト学習での実践(例えば,松本,2000;Sluijsmans et al.,2001;Akahori & Kim,2003)などが行われており,それぞれ教育効果をあげたと報告されている。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(例えば,Cousins & Whitmore,1988;大塚,2002;鹿毛,2004)。

近年の教育評価の研究では,学習者の一面を客観的に正確に測定することを重視する考え方から,多面的に総合的に学習者を捉えるアセスメントのという考え方に変化してきており(例えば,梶田,1993),情報通信技術の革新を背景にした様々なテストが開発されている。池田(2000)は,テスト技術の発展史を,「主観評定(19世紀以前)」、「客観測定(20世紀)」、「力量査定(21世紀)」の3つの時代に分類している。1990年代には,マルチメディアを出題や回答に用いたテスト(例えばFujihara et al.,1994),コンピュータネットワークを利用したテストが開発された。池田は,客観測定が辿り着いた先は,コンピュータネットワークとマルチメディア技術を利用した人間の多面的力量の査定であり,力量査定時代においては,e-テストが普及し,教授と一体化されたシステムの開発が必要であると述べている。Bundersonら(1989)は,コンピュータテストの発達を4世代に分類し,将来の第4世代のテストは,「知的測定の世代」になると予測している。知的測定によって,オンラインテストで収集した学習者の情報を,従来の数理的な手法だけではなく,人工知能のアプローチで,解釈し,診断することが可能になるとしている。現在では,人工知能の研究は,データマイニングや複雑統計学などの分野に広がり,このような新しい技術を学習評価への応用が試みられている(例えば,植野,2000)。

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,課題などの成果物を提出させ,学習過程に評価を埋め込むことが行われている。ここでの学習コミ

コミュニティは学級を中心とした教室などの現実世界の空間に限定されず、コンピュータネットワークを介しても形成される。植野(2005)は、学習者に提出させる成果物を教師だけでなく、学習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。

先に述べたように多くの分野で相互評価が成功している事例があるが、それらの実践で、相互評価の何が学習効果につながっているかは、ほとんど言及されていない。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(Liu & Sadler, 2003; Figl et al., 2006)。コンピュータを用いた協調学習支援システムに、単に課題提出機能や相互評価等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。

大塚(2002)は、学習コミュニティを、「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」と捉えている。学習コミュニティが評価コミュニティとして適切に機能するためには、同様の条件が必要であると考えられる。目標については、評価目標、評価の観点を適切に設定し、それを共有する試みが行われている。しかし、価値を共有する前提として、学習者が自身の持っている価値を忠実に他者への評価という形で表現でき、他の学習者を公平に扱うことが必要であると考えられる。多くの実践では、学習者が評価者となったときに、適切な評価をできない場合は想定されておらず、学習者が他の学習者を評価する能力について、確認していない場合が多い。また、相互評価を行う際には、学習コミュニティの他の全てのメンバーを評価することは、時間や労力の問題から困難であり、その場合、評価対象を選択する必要が出てくる。しかし、その選択方法によって、公平な評価が行われるかどうか問題となる。

橋本(1971)は、教育評価の目的を「指導目的」、「学習目的」、「管理目的」、「研究目的」に分類している。本論文では、学習者が弱点を把握することを助けるため等に行う学習者の立場から見た、学習の改善のために行う「学習目的」に焦点をあてて論じる。まず、コミュニティ内の相互評価に関係する基礎的事項として、形成的評価、学習

場面と評価が一体化した学習環境，コミュニティ内評価を取り上げ，その研究動向を述べる．次に，コミュニティ内評価の1つである相互評価の課題と研究動向について述べ，最後に，本研究の目的と構成について述べる．

1.2. 学習の改善を目的とした評価

1.2.1. 形成的評価

(1) 形成的評価の定義と PDS モデル

形成的評価とは，学習の途上において行われる評価であり，学習活動が終了したあとに行われる総括的評価と区別される．形成的評価，総括的評価という語は，Scriven(1967)によって用いられたのが最初であるといわれている．Scriven は，カリキュラム開発の途上でその改善のために行う小きざみな各種の評価を総称して形成的評価と呼び，その全体としての価値や採用の可否を最終的に決定するような評価を総括的評価と呼んだ(東ほか, 1988, p.208)．その後，Bloom ら(1971)は，形成的評価はカリキュラム開発だけではなく，教授，学習過程におけるそれらの改善のために重要であると主張した．形成的評価という語は，例えば，鈴木(2002)のように教材開発途上での評価を表すこともあるが，本論文では，教授・学習過程において，教授・学習活動を改善する目的で，学習者を対象として行う評価とする．

生産現場の作業の科学から発生したといわれるマネジメントサイクル PDS(持本, 1991)に，教授・学習活動を当てはめると，Plan は教授・学習計画, Do は教授・学習活動，See は評価にあたる．ここで評価の目的を，教授・学習計画の改善，学習過程の改善のために必要な情報をフィードバックすることに置くとすると，それまでの総括的評価で用いられてきたテスト理論の教育測定の枠組みをそのまま導入することは困難であり，新しい評価ツールが必要となる．Bloom ら(1971)は，形成的テストを提案し，完全習得学習と呼ばれる教授モデルに組み込むことを主張した．完全習得学習については次項で述べる．近年では，PDS のように，評価を他のフェーズと独立して捉えるのではなく，目標と評価，あるいは，教授・学習活動と評価を，それぞれ関連付ける考え方(市川, 1995；鹿毛, 2000)が提案されている．このような考え方は，情報通信技術の発達により，後述する学習場面と評価が一体化した学習環境が提供されるようになってきている(加藤ほか, 2006b)．

(2) 完全習得学習とタキソノミー

Carroll(1963)は、学習の達成度は、学習に必要な時間に対して、学習に費やされた時間の割合で表されるというモデルを提唱した。このモデルでは、学習者の能力は、学習課題を達成するために必要な時間という形で表される。すなわち、学習時間が十分に与えられれば、すべての学習者が学習課題を達成できることを意味している。Bloom(1971, chapter 3)は、Carrollのモデルを基盤とし、Bruner(1966)、Glaser(1968)、Goodlad & Anderson(1959)、Morrison(1926)、Skinner(1954)、Suppes(1966)の知見を基に完全習得学習の考え方と実現方法を示した。完全習得学習の考え方は、梶田ら(1973)によって、日本語に訳され我が国の教育評価研究に大きな影響を与えた。

完全習得学習は、学習過程の改善を目的として、きめ細かく行われる形成的テストとその結果による補充指導の繰り返しで行われる。形成的テストを作成するためには、学習単元を内容と行動の観点から分析を行う。Bloomら(1971, chapter 6)は行動の分析のために、タキソノミー(教育目標の分類学)(Bloom, 1956)のカテゴリーを利用し、目標細目表を作成することを提案している。形成的テストでは、目標細目表に記載された重要な要素を全て出題し、学ばなければならない点を学習者に示すことが重要とされている。

梶田(1994b, pp.168-169)は、タキソノミーや目標細目表を適用する場合の注意点として、タキソノミーや目標細目表は、児童・生徒の現状や成長の実態を分析的に把握し、位置づけるという評価の目的のためのものであり、目標に対応させて教育活動を構成していくためのものであることを指摘している。また、Bloomのタキソノミーをそのまま適用するのではなく、教育目標の特性に合わせたタキソノミーを作成すべきとも述べている。

以上のように、学習の改善を目的とした形成的評価やそのための学習理論が開発されてきたが、学力観や学習形態の変化に合わせた評価方法が必要である。従来の形成的評価において重要視されていた教育目標を適切に記述することは、教授・学習活動と評価が一体化した環境でも重要であり、後述する相互評価においては、学習コミュニティ内で、学習者が評価の観点を共有するために必要である。

1.2.2. 学習場面と評価が一体化した学習環境

(1) WBT

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。WBTは、通常LMS(Learning Management System)で管理されている。LMSは、学習者の進捗状況を管理し、学習者に適切な教材を示す機能を持つ。Okamotoら(2002)はカリキュラムの流れ、学習者の状態からコースを組み立てる機能をもったRAPSODYを開発した。中原・西森(2004)は、オープンソースで汎用性があり、複数の属性の異なるユーザが使用するコミュニケーションツールとしての側面を重視したeXCampusを開発した。Ueno(2004)は、蓄積された学習履歴データを、データマイニング、テキストマイニングの技法で分析する機能を持ったLMSであるSAMURAIを開発した。WBTでは、個別に学習を進めることから、他の学習者の存在を意識することが少ないため孤独感を感じ、学習意欲を持続することが困難な場合がある。教師は1人で多数の学習者に対応する必要があることから、個々の学習者と十分なコミュニケーションを図ることは難しい。そのため、運用の際には、学習者間のコミュニケーションを支援する機能が必要となってくる。

(2) 協調学習

前項で例にあげたLMSはすべてコミュニケーションツールとして機能を有している。最近では、CSCL(Computer Supported Collaborative Learning)は、LMSに統合され、様々な協調学習への支援機能を持つシステムの研究が活発になってきている(植野, 2005)。

本論文では、学習者が学習コミュニティの中で、お互いの学習を助け合いながら学ぶ形態の学習のことを協調学習とする。このような学習をコンピュータで支援することをCSCLと呼び、Koschomann(1996)は、CSCLの出現までの学習支援システムの研究の変遷を、学習理論の観点から行動主義、認知的構成主義、社会的構成主義へ変化してきていると述べている。協調学習は、ピア・チュータリング(Goodlad & Hirst, 1989)や観察学習(Bandura, 1971)など多くの学習理論でその有効性が説明されている。

創造性、論理的思考能力、想像力、表現力の育成が重視される学力観においては、協調学習のように問題解決を通して、学習者が主体的に学ぶことができる環境を提供することが重要である。その際の評価は、従来のペーパーテストではなく、レポートなどの成果物が対象となる。協調学習における評価は、社会的構成主義の考え方に立つと、教師が1人で行うよりも、目標や価値を共有する学習コミュニティのメンバー

が評価に関与することが適切である。協調学習において、学習コミュニティのメンバーが評価することによって、学習意欲が向上したり、学習コミュニティが活発になる等の効果も実証されている（例えば、Sitthiworachart & Joy, 2003; Chinn, 2005; Derntl, 2006）。梶田(1994a, pp.83-84)は、作文やレポート等の評価について、主観的恣意的な評価になる危険性を指摘し、評価の視点を明確にもち、時々複数人で同じ作品や作文を評価することを提案している。また、Ellis & Hafner(2005)は、協調学習における評価では、学習者は、評価対象の学習者と同じ立場に立って評価することができるため、教師には気付にくい点を見つけることができると指摘している。

1.2.3. コミュニティ内評価

(1) 相互評価と自己評価

教育評価における相互評価とは、「学級集団などにおいて、生徒同士に、互いに級友を評価させる方法」(東ほか, 1988, pp.377-378)をいう。教育評価を評価の主体で分類すると、教員による評価、学習者による評価に分けられる。学習者による評価は、相互評価と自己評価があり、それらは密接な関係にある。

自己評価の利点は、自分が考えていたことや不安や気持ちは自分自身のみが知ることができることであり、自身の学習成果を振り返ることは成果物の改善に有用である。しかし、自己評価では、心理的に自己防衛が働き、自己の欠点を直視しない学習者が存在する。このような学習者に対しては、自己評価の効果は限定的である。また、学習者が過大評価する傾向にあるか、過小評価する傾向にあるかは、学習者が行った評価結果のみからは判断できず、評価対象を詳しく見る必要があることから、フィードバックを行うまでに、時間と手間を要する。

相互評価は、自己評価と比較すると、客観的に評価することができ、複数の学習者を評価したり、他の学習者が行った評価を見ることで、他者を評価することを学ぶことができる。また、他の学習者を評価することは、自らを見直す機会となり、評価すること自体が自己へのフィードバックにつながる。

学習者が他の学習者の行った評価に納得して、はじめて相互評価の教育的効果が期待できることから、学習者が納得できるように相互評価を計画することが必要である。学習者が、相互評価の結果を納得して受け入れるためには、評価の公平性の確保や評価の観点への同意が必要である。これらを支援する手法については次項で述べる。

(2) 相互評価支援システム

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(2002)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。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(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。高木ら(2007)は、学習者が問題作成する場面で、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。

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。次節では、学習コミュニティのメンバーが行う評価によって学習効果をあげるための課題とそれに関する研究動向について述べる。

1.3. 学習者間の相互評価の課題

1.3.1. 信頼性と妥当性

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・労力的に困難になることから、評価すべき相手を選択する必要が生じる。このような場合に、評価結果を学習者が受け入れることができるための条件として、信頼性、妥当性について考えてみる。

相互評価における信頼性は、同じものを評価した時に同じ評価結果をつけることができること(一貫性)、評価者と非評価者の間の人間関係などが本来評価したい特性以外の要素が評価に入っていないこと(評価への専念)、評価が甘い、厳しい等の評価者の評価の特性による悪影響がでないこと(公平性)が重要である。相互評価における妥当性は、教育目標全体を代表するような評価の観点を設定し、それを学習者が共有することが重要である。学習者が作成した成果物を評価する場合、作成段階で評価項目を予め提示することで、学習集団が同じ目標を共有することを助け、それを目標に成果物を作成していくものと思われる。学習者が評価項目に納得していることが、評

価結果に納得するための前提であると思われる。

評価の一貫性については、他者を評価する能力と評価に対する責任について考える必要がある。他者を評価する能力は、他の人が行う評価をモデルにし、繰り返し訓練することで身につけられると考えられる。先行研究では、大学生の行う評価は、教員の行う評価と大きな差がなく、しばしば教員の評価よりも適切であることが知られているが（例えば、Arnold, 1981; Orpen, 1982; Falchikov, 1986）、日本の大学生については、他者評価の能力は必ずしも高いとはいえない（例えば、藤原ほか, 2006）。また、責任を持って評価させるためには、学習コミュニティのメンバーとしての役割や責任を理解させる必要がある。

評価への専念については、相互評価に特有の問題がある。評価をする側が評価を受けることになる。このことによって、評価者が自分もよく評価されたいために、評価が甘くなっていく可能性も考えられ、お互いに評価しあうことが評価に影響を与えるのであれば、その影響を考慮した評価者と被評価者の組み合わせを検討する必要がある。

公平性に関しては、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）、たまたま甘い評価者にあたったのか、厳しい評価者にあたったのかによって、学習者間に不公平が生じることが考えられる。その結果として、評価を納得して受け入れることが阻害される。学習者が、他の学習者が行った評価を納得して受け入れるためには、評価者の特性を考慮して評価結果を修正したものを学習者に示す必要がある。次節では、相互評価における公平性について述べる。

1.3.2. 評価の公平性に関する問題

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。まず、不公平が生じる要因として、個々の評価者の特性と評価する学習者が、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について検討する。次に、個々の評価者の評価特性の差によっ

て生じる問題について検討する。

(1) 評価者の選択

評価者を選択する必要がある場合には、学習者を適当なサイズの小集団に分割して、その小集団内で相互に全ての学習者を評価することが行われる場合が多い。協調学習においては、小集団の構成が学習効果に大きな影響を及ぼすといわれており(稲葉・豊田, 1999)、様々なグループ構成法が提案されてきた(例えば、稲葉ほか, 2000; Hishina et al., 2005; Alfonseca et al., 2006; Kitagaki et al., 2007)。グループ構成について、協調学習における学習効果の観点からは議論されてきたが、グループ内での相互評価について、公平性の観点からはあまり議論されてこなかった。グループ内でお互いに評価しあうことが、評価結果に影響を与えることについて考慮する必要がある。

Trivers(1971)は、互惠的利他主義として、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、互惠性に基づく利他的行動について説明した。相互評価において互惠的利他性が働けば、評価が甘くなり、それは容易に取り除けないことが予想される。

学習者が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を調べる必要がある。相互評価において、お互いに評価するかどうか、評価結果に影響を与えるのであれば、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、お互い評価しあわないように割り当てる必要がある。

(2) 個々の学習者の評価特性

全ての学習者が他の全ての学習者を評価する場合は、ほぼ公平であるといえる。しかし、評価者を選択しなければならない状況では、たまたま甘い評価者にあたったのか、厳しい評価者にあたったのかによって、学習者に不公平感を与え、評価結果を納得して受け入れることが阻害されると考えられる。したがって、公平な評価を行うためには評価者の評価特性を考慮した評価結果を修正したものを学習者に示す必要がある。

植野ら(2006)は、e-ラーニングで得られた大量の多段階評価の相互評価のデータを項目応答理論(芝, 1991)の1つである Graded Item Response Model(Samejima, 1969)

に評価者の特性パラメータを加えるという拡張をして、レポートの評価を行う方法を提案している。それぞれのレポートに対して同一尺度上で合理的な評価を行うことができ、それぞれの評価者の特性として、評価基準の厳しさ、評価の一貫性を求めることができることが特徴としてあげられる。

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、評価が甘くなる可能性がある。お互い様に評価しあう組み合わせを避けた場合、1人の学習者ができるだけ多くの評価をした場合でも、データの約半数は欠測値となる。一般的に、1人の学習者が行うことができる評価人数は限られており、その結果、1人の学習者が受ける評価数も多くはなく、このような欠測値の多いデータでも、容易に適用できる評価方法が必要である。

1.3.3. 相互評価の効果

相互評価と学習効果の関係を調べるために、継続した相互評価の実践研究が必要である。ここでは、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を紹介する。

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。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、先に述べた「同じ価値、同じ目標」(大塚, 2002)を共有する試みが行われている。

生田目(2004)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。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みることによる効果、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。

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、学習者の他者評価の能力や学習者が行った他者評価の妥当性について検討していないことも多い。また、相互評価を一度だけ行った事例は多いが、それだけでは、効果はあまり期待できない。本研究では、相互評価を継続的な実施し、その効果について検証する。

1.4. 本研究の目的と構成

本研究では、成果物の作成を通じて学習する場面において、学級等の学びの場を共有する学習コミュニティのメンバー同士が、お互いの成果物を相互に評価することによって、学習者に自身の成果物の改善点を気付かせるために学習者間相互評価を効果的に行う方法を提案する。学習者間相互評価は先に述べたとおり、それ自体に教育的効果が期待されるが、評価の結果を学習者が納得して受け入れることが前提となる。本研究では、学習者が評価を納得して受け入れることに影響を与える評価の公平性に注目し、公平性の高い評価対象選択法と公平性の高い評価結果のフィードバック法について提案する。

第1章では、本研究の背景、目的、および相互評価の必要性と課題について述べた。

第2章では、評価者を選択する必要がある場合に、公平な評価者の選択方法を考えるために、互恵的期待が相互評価に及ぼす影響について調べた実験の結果を考察する。すなわち、評価を行う学習者が、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、評価にどのような変化が見られるかについて検討する。

第3章では、評価者を選択する必要がある場合に、個々の評価者が持つ評価特性を考慮したフィードバックを行う手法を提案する。提案手法の特徴として、項目応答理論と比較して計算が容易にできること、多くの欠測値を持つ相互評価データから個々の評価者の評価特性を推定できることがあげられる。推定された個々の評価者の評価特性を用いて、実際に行った評価の補正するアルゴリズムを提案し、実データに当てはめて、その有効性を検討する。

第4章では、第2章及び第3章の知見を組み込んだ相互評価を支援するシステムについて述べる。実際に、学習活動に相互評価を導入するためには、学習者、教員ともに負担を最小限にすることが求められる。開発した相互評価支援システムでは、公平性の高いフィードバックを行うために、評価者選択アルゴリズムと個人差補正アルゴリズムを持っている。相互評価を導入した講義で学習者による評価を行い、また、相互評価を導入した実践を行った経験のある教員による評価の結果を検証する。

第5章では、第4章で述べた相互評価システムを、継続的に使用することによる学習効果について述べる。プレゼンテーションの作品を制作させ、相互評価を5回実施した。評価の補正值を表示するグループと表示しないグループに分けて実施した。そ

の結果，システムを利用した継続的な相互評価の実施による教育的効果が示された．

第6章では，本研究の知見をまとめ，結論，総合的な考察，今後の課題を示す．

第2章 評価者の選択が評価の公平性に与える影響

2.1. はじめに

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。

本章では、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について検討する。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、評価する成果物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について実験し、考察する。

2.2. 実験方法

2.2.1. 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成果物を、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成果物を評価させた。

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、電子ファイルでの成果物受領機能、学習者間での成果物の評価実施機能を持っている。相互評価実施画面を図 2.1 に示す。学習者は、指定された順番に成果物を評価する。なお、一度評価を決定すると、前の成果物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、同じクラスのもので顔見知りである。図 2.2 に示す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きず、相談せずに評価をさせた。小規模の集団で評価を行っており、学習者は成果物を見れば作成者の予想が付き、また、学習者に責任感をもって評価させるため、学習者の名前を表示した。

学習者は、それぞれ 6 名の成果物を評価した。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成果物が評価され(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成果物が評価されること、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。

学習者は、それぞれ 6 名の成果物を評価した。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成果物が評価され(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成果物が評価されること、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。

2.2.2. 評価者の決定方法

学習者は、表 2.1 に示す学習者が提出した成果物を評価する。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成果物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成果物の評価は行わない。例えば、学習者 1 は、学習者 13, 14, 15, 2, 3, 4 の順に、提出された成果物を評価し、それ以外の成果物は評価しない。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。

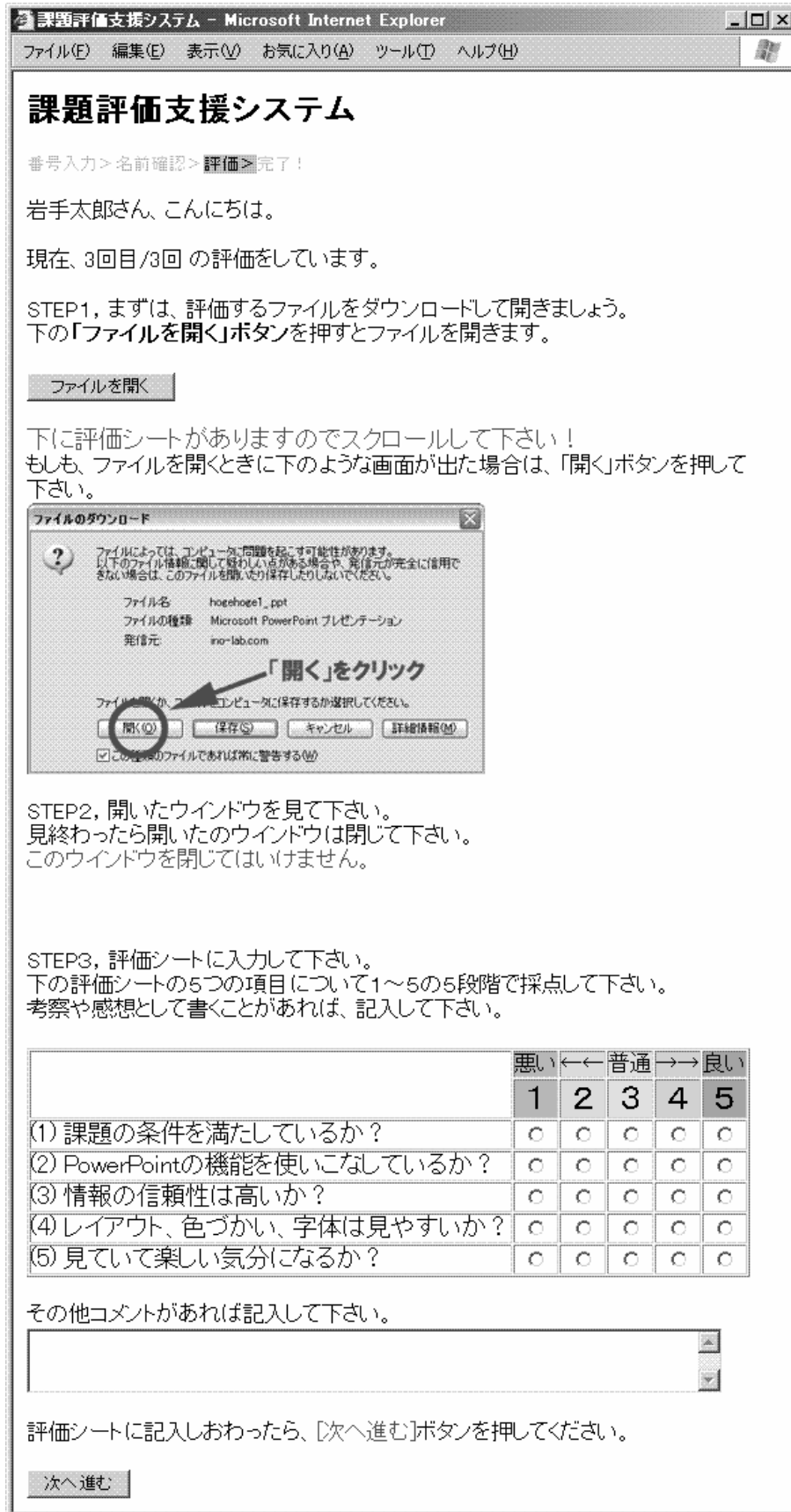
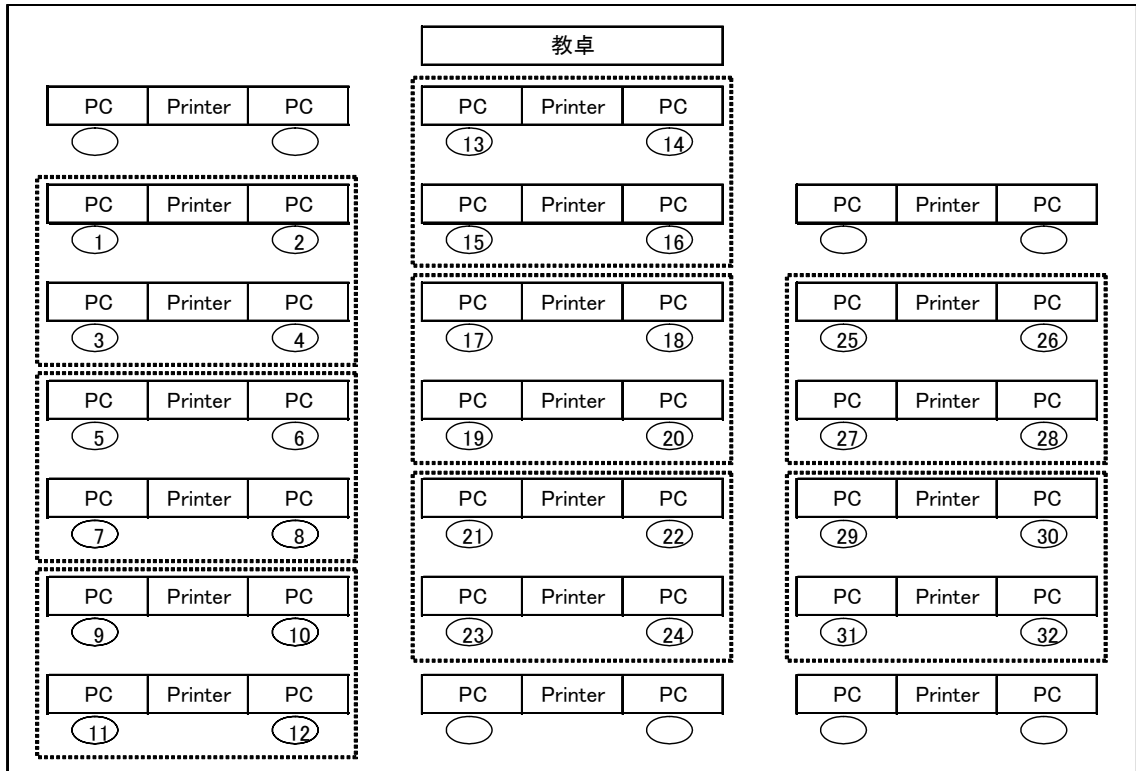


図 2.1 : 相互評価実施システムの画面例



は、机を表す。
 は、条件1でのグループを表す。
 ○ は、椅子を表し、数字は着席した学習者の番号を表す。

図 2.2 : 実験で使ったコンピュータ教室

表 2.1 : 各学習者が担当した成果物

	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																																						

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成果物を担当させた。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成果物を評価する。

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。学習者が評価する成果物の順番は、例えば、ある学習者の成果物が常に最初に評価されることがないように、個々の成果物が評価される順序を考慮して決定した。

2.2.3. 科目の目標及び課題

「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。課題作成前に、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、アニメーションの設定の仕方などについて説明した。良いプレゼンテーション資料の例を示して、プレゼンテーションソフトの効果的な利用法についても解説した。

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れるようになる」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

1 番目の項目は、与えられた条件にあったプレゼンテーションになっているかについて評価させた。通常、プレゼンテーションには、目的、時間の制限があることから、スライドの枚数や入れるべき内容（例えば、名前、タイトルなど）を評価させた。2

番目の項目は、単にプレゼンテーションソフトの機能を使っているのではなく、聴衆の立場から、描画ツールやアニメーションが効果的に使われているか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。学習者は、既に W.W.W 上の情報の信頼性について学習しており、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、例を挙げながら改めて説明した。4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、色の組み合わ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。5番目の項目については、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。

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、操作方法を知った上で、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、各学習者自らが実際に課題を行って、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。

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。評価結果は、成果物の改善の参考にするために、学習者にフィードバックすることを伝えた。また、相互評価の結果は、課題作成途中での評価として、成績評価に反映させた。

2.3. 結果

2.3.1. 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を表 2.2 に、学習者が行った評価の分布を図 2.3 に示す。それぞれの学習者がつけた評価の平均は、25点満点中お互い様条件では 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、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 ($F(1,30)=0.023, p>.10$)。

表 2.2 : 学習者ごとの評価結果

学習者番号	学習者が行った評価の平均			教員が行った評価の平均			学習者と 教員の比較 C-F
	お互い様 条件(A)	非お互い様 条件(B)	差(C) A-B	Aと同じ成果 物の評価(D)	Bと同じ成果 物の評価(E)	差(F) D-E	
1	25.00	25.00	0.00	15.33	19.00	-3.67	3.67
2	25.00	21.67	3.33	18.00	20.00	-2.00	5.33
3	24.33	21.67	2.67	16.33	20.00	-3.67	6.33
4	24.33	22.00	2.33	16.33	19.00	-2.67	5.00
5	25.00	24.00	1.00	22.67	20.67	2.00	-1.00
6	25.00	23.67	1.33	21.33	20.67	0.67	0.67
7	25.00	24.33	0.67	20.33	22.67	-2.33	3.00
8	24.33	25.00	-0.67	20.67	20.67	0.00	-0.67
9	23.00	21.00	2.00	21.67	19.33	2.33	-0.33
10	22.67	23.67	-1.00	19.00	18.67	0.33	-1.33
11	22.67	22.00	0.67	20.33	18.67	1.67	-1.00
12	24.00	20.00	4.00	21.00	17.67	3.33	0.67
13	18.67	21.67	-3.00	20.00	19.67	0.33	-3.33
14	23.33	25.00	-1.67	17.00	21.67	-4.67	3.00
15	24.67	24.67	0.00	20.00	23.33	-3.33	3.33
16	24.67	23.00	1.67	19.00	21.67	-2.67	4.33
17	25.00	21.00	4.00	20.67	21.00	-0.33	4.33
18	24.67	24.00	0.67	24.67	19.33	5.33	-4.67
19	20.67	22.67	-2.00	21.00	20.00	1.00	-3.00
20	23.33	23.33	0.00	20.67	16.67	4.00	-4.00
21	24.33	19.33	5.00	18.67	16.33	2.33	2.67
22	19.00	19.00	0.00	19.00	15.33	3.67	-3.67
23	20.00	18.33	1.67	18.00	17.00	1.00	0.67
24	20.00	20.00	0.00	19.33	18.33	1.00	-1.00
25	25.00	25.00	0.00	21.67	20.67	1.00	-1.00
26	25.00	25.00	0.00	21.67	22.67	-1.00	1.00
27	20.33	19.00	1.33	20.00	21.33	-1.33	2.67
28	23.00	25.00	-2.00	19.67	21.67	-2.00	0.00
29	21.33	22.67	-1.33	19.33	21.00	-1.67	0.33
30	24.33	25.00	-0.67	21.00	21.67	-0.67	0.00
31	22.33	22.67	-0.33	19.67	18.67	1.00	-1.33
32	23.00	19.67	3.33	21.00	20.00	1.00	2.33
平均	23.22	22.50	0.72	19.84	19.84	0.00	0.72
標準偏差	1.95	2.12	1.92	1.93	1.91	2.46	2.89

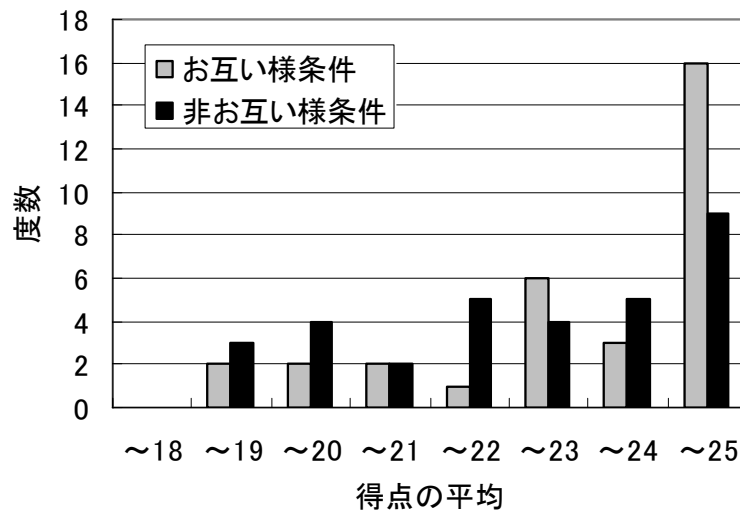


図 2.3：学習者の行った評価の分布

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 2.3 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。

表 2.2 では、学習者が行った評価と、同じ成果物に対して教員が行った評価との比較についても示した。表中の(C)欄は、非お互い様条件(B)で示した成果物よりもお互い様条件(A)で評価した成果物に対して、どの程度甘い評価をつけたかを示している。それぞれの条件で評価した成果物を教員が評価した結果が、(D)欄、(E)欄である。教員についても、(D)欄、(E)欄の差 (F 欄) を求め、その値と学習者の(C)欄と比較することで、評価対象となった成果物の出来具合を考慮して、個々の評価者ごとの各条件で行った評価を比較できる。つまり、「学習者と教員の比較」の欄の数値が大きいほど、教員と比較して、お互い様条件で甘い評価を行っていることになる。この値が 2 より大きい学習者が 12 名おり、この学習者はお互い様条件で、甘い評価を行う傾向があることが分かる。

2.3.2. 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成果物を採点した結果と、学習者の行った評価との関係を調べた。学習者が受けた評価とその成果物に対する教員の評価

を表 2.3 にその散布図を図 2.4 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。お互い様条件の評価が教員の評価と相関が低かった原因の一つに天井効果があげられる。しかし、天井効果自体は、お互いに評価しあうことから生じているため、天井効果が出ていることが、お互いに評価しあうことが適切な評価を阻害するという結果を弱めるものではない。

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。学習者 17 の成果物の評価については、担当教員と担当しない教員の差があったが、それ以外の成果物については、ほぼ同様の評価結果となった。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)

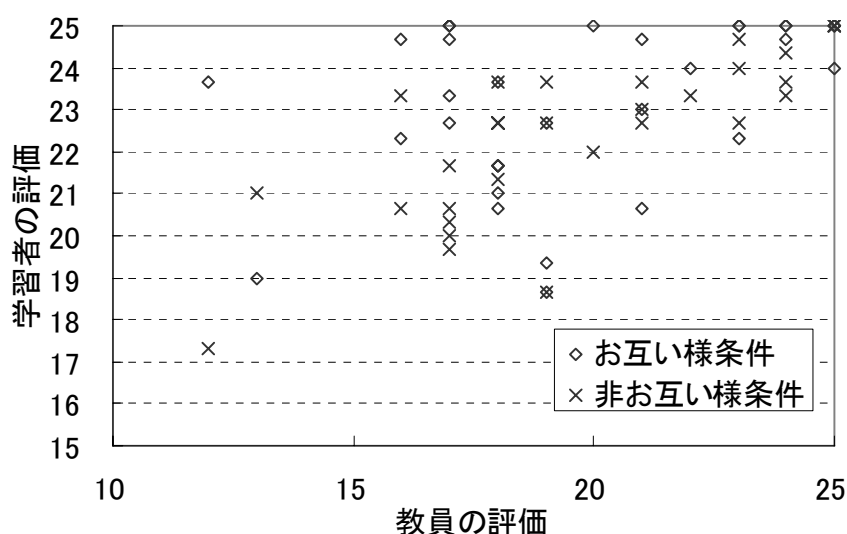


図 2.4：教員の評価との比較

表 2.3 : 学習者が受けた評価と教員による評価

学習者番号	学習者が受けた評価の平均			教員が行った評価		
	お互い様 条件(A)	非お互い様 条件(B)	差(C) A-B	担当教員 による評価	別の教員 A による評価	別の教員 B による評価
1	25.00	22.00	3.00	20.00	19.00	16.00
2	23.67	17.33	6.33	12.00	15.00	12.00
3	25.00	20.33	4.67	17.00	18.00	17.00
4	25.00	20.67	4.33	17.00	14.00	17.00
5	24.67	20.00	4.67	17.00	21.00	19.00
6	24.67	23.67	1.00	21.00	20.00	20.00
7	25.00	23.33	1.67	24.00	19.00	19.00
8	25.00	22.67	2.33	23.00	23.00	22.00
9	22.67	21.67	1.00	17.00	22.00	20.00
10	24.00	25.00	-1.00	25.00	24.00	20.00
11	23.00	23.00	0.00	21.00	22.00	20.00
12	22.67	22.67	0.00	19.00	23.00	19.00
13	24.67	20.67	4.00	16.00	21.00	16.00
14	25.00	25.00	0.00	25.00	21.00	20.00
15	22.33	23.33	-1.00	16.00	20.00	19.00
16	19.33	18.67	0.67	19.00	18.00	18.00
17	25.00	25.00	0.00	25.00	22.00	19.00
18	19.00	21.00	-2.00	13.00	15.00	12.00
19	24.67	24.33	0.33	24.00	23.00	21.00
20	25.00	25.00	0.00	25.00	22.00	23.00
21	18.67	23.67	-5.00	19.00	20.00	20.00
22	20.67	22.67	-2.00	18.00	19.00	16.00
23	20.67	22.67	-2.00	21.00	21.00	20.00
24	23.33	19.67	3.67	17.00	18.00	14.00
25	21.67	23.67	-2.00	18.00	20.00	16.00
26	21.67	21.33	0.33	18.00	21.00	16.00
27	25.00	24.67	0.33	23.00	21.00	20.00
28	25.00	23.67	1.33	24.00	21.00	20.00
29	22.33	24.00	-1.67	23.00	19.00	16.00
30	23.67	22.67	1.00	18.00	21.00	19.00
31	24.00	23.33	0.67	22.00	21.00	19.00
32	21.00	22.67	-1.67	18.00	17.00	13.00
平均	23.22	22.50	0.72	19.84	20.03	18.06
標準偏差	1.97	1.93	2.46	3.57	2.40	2.73

2.3.3. 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 2.4 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 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。

表 2.4：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

2.3.4. 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(Fujihara et al., 2006)。学習者は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 20.6、非お互い様条件 19.4 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 ($t(15)=2.33, p<.05$)。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

者によって差があって不公平」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, 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。

2.3.5. 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。相互評価において互恵的利他性が働けば, 評価が甘くなり, それは容易に取り除けないことが予想される。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。

2.4. まとめ

本章では, お互いに評価しあう場合とそうでない場合の評価を比較する実験とその結果について述べた。

実験では, 評価対象からも自分が提出した成果物を評価される場合と, 評価対象からは自分が提出した成果物は評価されない場合での評価の違いを調べた。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, 評価点が甘くなる場合があることがわかり, この現象のことをお互い様効果と呼ぶこととした。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。

お互いに評価しあわないことで、短所をより適切に指摘できることから、形成的評価として相互評価を導入する際には、お互い様効果のでない組み合わせにすることは、重要である。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り付けについて考慮する必要がある。実際に、相互評価の実践において、お互い様効果の出ない評価者の割り当てを行うためには、お互い様効果を考慮して自動的に評価者を決定するアルゴリズムと支援ツールが必要である。

第3章 相互評価データにおける評価者の評価特性補正方法

3.1. はじめに

評価者によって、評定の甘さや厳しさなど個人差があるといわれている（渡部ほか, 1998）。全ての学習者が他の全ての学習者を評価する場合は、ほぼ公平であるといえる。しかし、評価者を選択しなければならない状況では、たまたま甘い評価者にあたったのか、厳しい評価者にあたったのかによって、学習者に不公平感を与え、評価結果を納得して受け入れることが阻害されると考えられる。したがって、公平な評価を行うためには評価者の評価特性を考慮したフィードバックを行うことが必要である。

植野ら(2006)は、e-ラーニングで得られた大量の多段階評価の相互評価のデータを項目応答理論の一つである Graded Item Response Model(Samejima, 1969) に評価者の特性パラメータを加えるという拡張をして、成果物の評価を行う方法を提案している。それぞれの成果物に対して同一尺度上で合理的な評価を行うことができ、それぞれの評価者の特性として、評価基準の厳しさ、評価の一貫性を求めることができることが特徴としてあげられる。

前章で述べたように、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場合、評価が甘くなる傾向があり（お互い様効果）、お互い様効果を避けるような組み合わせで相互評価を行うと、1人の学習者ができるだけ多くの評価をした場合でも、データの半分は欠測値となる。一般的に、1人の学習者が行うことができる評価人数は限られており、その結果、1人の学習者が受ける評価数も多くはなく、このような欠測値の多い場合でも、適用できる補正方法が必要である。

本章では、項目応答理論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を提案する。提案したモデルを実データに適用しその精度を示すとともに、教員の行う評価と比較することで補正方法の妥当性を示す。

提案する手法は植野ら(2006)などが行っているような項目応答理論を相互評価に当てはめたものではなく、項目応答理論のメタファを用いた評価者の特性を反映したロジスティック回帰による簡易法を提案するものであり、項目応答理論よりも計算が非常に簡易であることが特徴である。

3.2. 補正方法

ここでは、評価者が実際につけた評価値を、その評価者の評価特性に基づいて補正

する方法を提案する。まず、同じ成果物を評価した評価者の評価との関係から、評価者の特性を推定する。次に評価者の特性と平均的な評価との差を使って、実際に行った評価の補正を行う。

3.2.1. 評価特性のモデル化

評価者の評価特性を、項目応答理論（芝, 1991）でよく用いられるモデルの1つである2パラメータ・ロジスティックモデルを元に表現する。項目応答理論は、テスト項目への学習者の応答と、現実には観測されない想定尺度上の学習者の能力を結びつける数理モデルである。大量の正誤データから、それぞれのテスト項目の特性を表すパラメータを求めることによって、学習者の推定された能力から、各テスト項目に正答する確率を求めることができる。2パラメータ・ロジスティックモデルでは、 θ_j の能力を持った学習者 j が、項目 i に対して X の反応をする確率は、式(1)で表される。

$$P(X | \theta_j) = \frac{\exp\{Da_i(\theta_j - b_i)\}}{1 + \exp\{Da_i(\theta_j - b_i)\}} \quad (1)$$

ただし、 a_i, b_i はテスト項目の特性、 D は定数

式(1)では、個々のテスト項目の特性を、項目識別力 a_i 、項目困難度 b_i の2つのパラメータで表している。項目困難度はその数値が大きいほど項目が難しいことを意味し、項目識別力は値が大きいほど、項目困難度の付近での項目応答曲線の傾きが大きいことを意味し、その付近での学習者の能力を識別する力が大きいといえる。

ここで、項目反応理論のメタファとして、以下のようなモデルを考える。ただし、項目応答理論では潜在変数モデルを仮定しているのに対し、本提案では、潜在変数を計算することなく、簡易的に以下のような回帰モデルにより代替するものである。項目応答理論におけるテスト項目が持つ特性を評価者が持つ特性に置き換える。学習者 j の成果物の真のできぐあいを θ_j とすると、平均的な評価者が、学習者 j の成果物に対して、式(2)のように評価者の特性に基づいて、 $R_0(\theta_j)$ ($0 \leq R_0(\theta_j) \leq 1$) の評価を行うものと仮定する。

$$R_0(\theta_j) = \frac{\exp\{Da_0(\theta_j - b_0)\}}{1 + \exp\{Da_0(\theta_j - b_0)\}} \quad (2)$$

これを θ_j について解くと、式(3)を得る。

$$\theta_j = \frac{1}{Da_0} \log \frac{R_0(\theta_j)}{1 - R_0(\theta_j)} + b_0 \quad (3)$$

評価者 i が学習者 j の成果物に対して行う評価 $R_i(\theta_j)$ は、式(2)と同様にして式(4)で表されると仮定する。

$$R_i(\theta_j) = \frac{\exp\{Da_i(\theta_j - b_i)\}}{1 + \exp\{Da_i(\theta_j - b_i)\}} \quad (4)$$

式(2),(4)における a_0, a_i は評価に差をつける度合いを、 b_0, b_i は評価の甘さの度合いを表す。式(4)に対して、式(3)を代入すると式(5)を得る。

$$R_i(\theta_j) = \frac{1}{1 + \exp\{-Da_i(b_i - b_0)\} \left\{ \frac{1 - R_0(\theta_j)}{R_0(\theta_j)} \right\}^{\frac{a_i}{a_0}}} \quad (5)$$

評価者 i が成果物 j に対して、 $R_i(\theta_j)$ の評価を行ったとき、 $R_0(\theta_j)$ を同じ成果物 j を評価した他の評価者の評価の平均を \bar{x}_j と置くことで、平均的な評価と評価者 i との評価の関係を表すことができる。式(5)に対して、パラメータを計算が容易になるように、 $D=1.7, a_0=1, \alpha_i=a_i, \beta_i=b_i - b_0$ と置くと、 $R_i(\theta_j)$ は、 \bar{x}_j とパラメータ α_i, β_i から式(6)のように表される。以降、この式を評価者特性モデルという。

$$f(\bar{x}_j; \alpha_i, \beta_i) = \frac{1}{1 + \exp(-1.7\alpha_i\beta_i) \left(\frac{1 - \bar{x}_j}{\bar{x}_j}\right)^{\alpha_i}} \quad (6)$$

このモデルと2つのパラメータの特性をグラフにしたものを図3.1, 図3.2に示す。 α_i は評価者 i の相対的な差をつける度合い(値が大きいほど差をつける)、 β_i は相対的な評点の厳しさ(値が大きいほど甘い評価をつける)を表していると考えられる。

なお、(6)式は項目応答理論のメタファを用いているが、項目応答理論そのものではないため、項目応答理論に見られる理論的特性をそのまま適用することはできない。しかし、本研究は、評価者の評価特性を集団との関係から推定し、その差異を用いて、実際行った評価を補正することが目的であり、(6)式を用いることにより簡単な計算で

評価者の特性を集団の中で相対的に表現することができる。

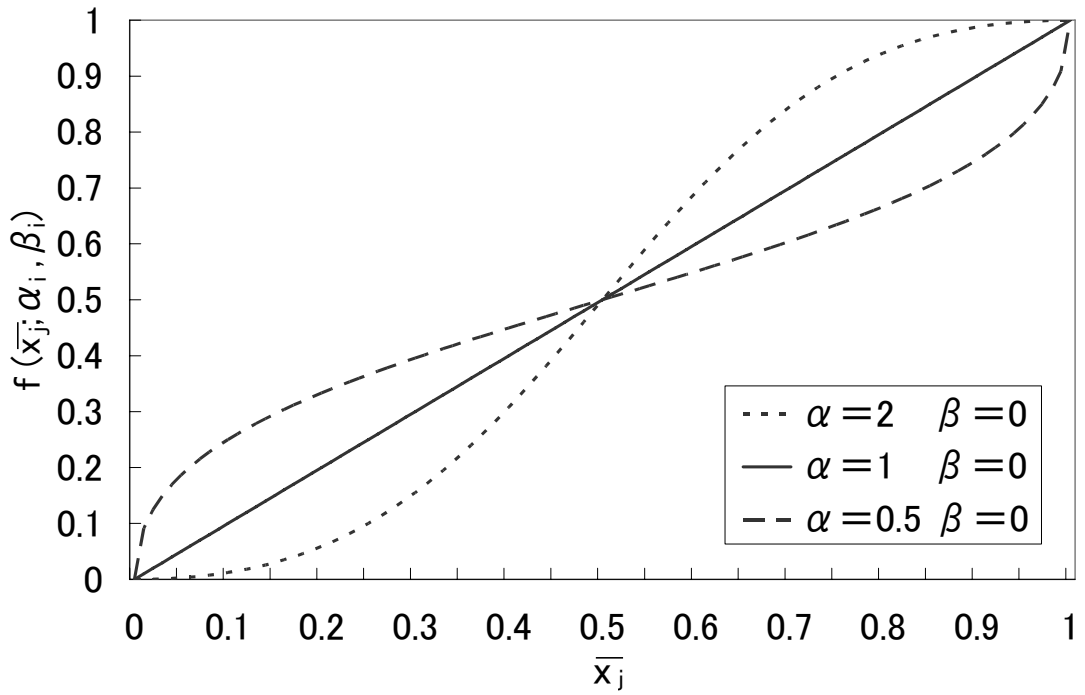


図 3.1 : モデルとパラメータ との関係

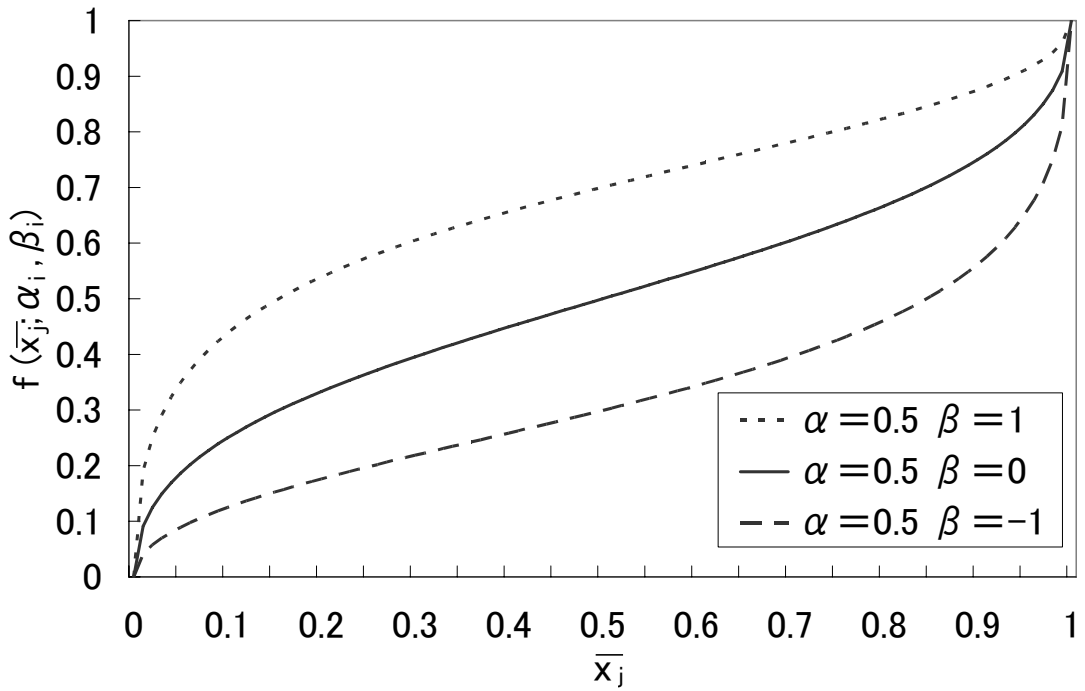


図 3.2 : モデルとパラメータ との関係

3.2.2. 評価特性モデルを使った評点の補正方法

次に式(6)から求められる評価者の特性を用いて、実際に行った評価を補正する。同じ成果物を評価した人がつけた点数の平均 \bar{x}_j を基準にして行う。 \bar{x}_j という評価を得られる成果物に対して、式(6)を用いて i が行うと予測される評価と、平均的な評価者が行う評価とを比較し、その差を実際に i が行った評価から減じる。平均的な特性を持つ評価者の特性パラメータを α_0, β_0 とすると、評価者 i が j の成果物に対して実際に行った評価 y_{ij} に式(7)のように補正する。

$$y_{ij} - \{f(\bar{x}_j; \alpha_i, \beta_i) - f(\bar{x}_j; \alpha_0, \beta_0)\} \quad (7)$$

補正の例を図 3.3 に示す。実線で表される評価者 i の特性曲線と点線で表される平均的な評価との差を補正量として、実際に評価者 i の行った評価を補正する。

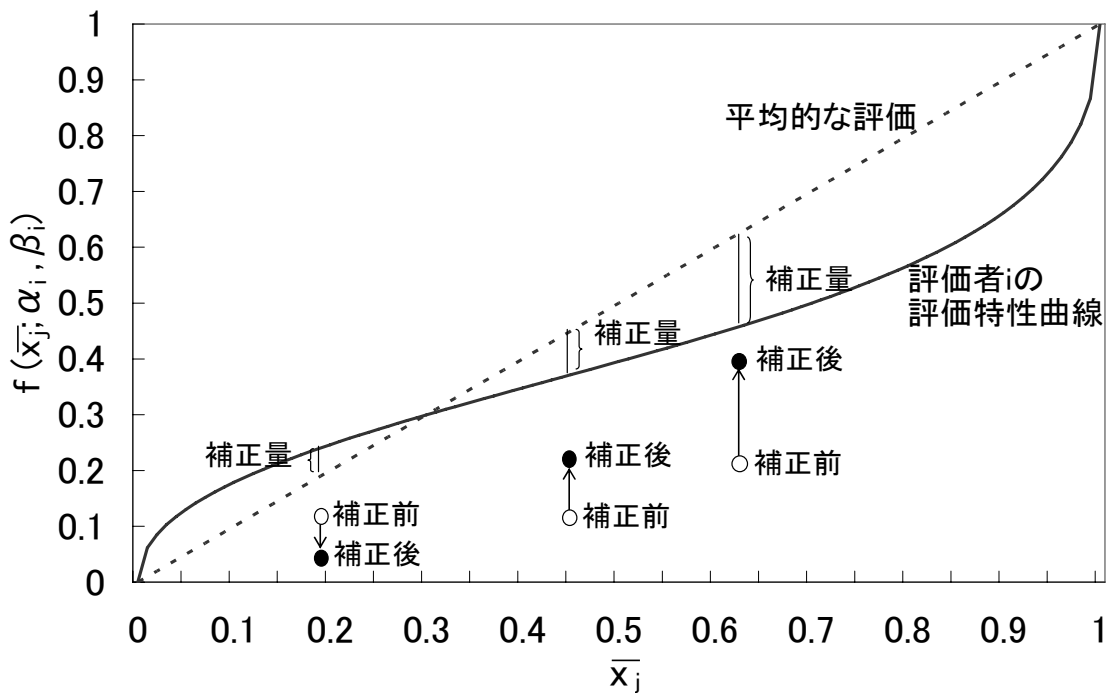


図 3.3 : 評点の補正の例

3.3. パラメータの計算方法及び精度

3.3.1. パラメータの計算方法及び精度

評価者の特性を表すパラメータの計算方法の例について述べる。パラメータは Levenberg-Marquardt 法 (Marquardt, 1963) を用いて推定した。Levenberg-

Marquardt法は、最急降下法と逆 Hesse 法を組み合わせ、非線形最小二乗問題で使用される解法である。当てはめたいモデルを $y = y(x;a)$ 、とすると評価関数 χ^2 は、

$$\chi^2(a) = \sum \left[\frac{y_i - y(x_i;a)}{\sigma_i} \right]^2 \quad (8)$$

となる。パラメータの初期値を a として次の手順で計算した (Press et al., 1988)。

STEP1 $\chi^2(a)$ を計算する。

STEP2 χ^2 の初期値として $\chi^2 = 0.001$ を代入する。

STEP3 連立方程式(9)(k はパラメータ)を解いて a_k を求め $\chi^2(a + \Delta a)$ を計算する。

(χ^2 の減少量が 0.1 以下の場合終了する)

$$\alpha_{kl} = \sum \frac{1}{\sigma_i^2} \left[\frac{\partial y(x_i;a)}{\partial a_k} \frac{\partial y(x_i;a)}{\partial a_l} \right] \quad (9)$$

STEP4 $\chi^2(a + \Delta a) > \chi^2(a)$ ならば、 Δa を 10 倍し、STEP3 に戻る。

STEP5 $\chi^2(a + \Delta a) < \chi^2(a)$ ならば、 Δa を 1/10 倍し、 a を $a + \Delta a$ に更新し STEP3 に戻る。

3.3.2. 計算例

A 大学文学部 1 年生 31 名に対して行った「コンピュータ応用」の講義での成果物の相互評価を例にあげる。この成果物は、与えられたデータを題材にして表計算ソフトを使って加工する課題で、表計算ソフトの使い方の講義が終わった後に中間レポートとして課した。評価項目は、以下のとおりで、それぞれ 0~10 の 11 段階で評価させた。R() の値の範囲は [0,1] であるため、得られた評価値を x とすると、 $R() = x/10$ として計算した。

- ・ 課題は条件を満たしているか？
- ・ 表の入力内容 (計算式) は適切か？
- ・ 完成した表のレイアウトは見やすいか？
- ・ 適切なデータ、グラフの種類を選んでいるか？
- ・ 完成したグラフは見やすいか？

1 人の学習者が 15 人の評価を行った。相互評価は、お互い様効果 (藤原ほか, 2007)

を考慮して、評価した相手からは、評価されないような組み合わせで行った。

評価項目によって評価の特性が異なることも考えられるが、今回は5つの評価項目が表計算ソフトの扱いに対して同じ種類の目標を扱っていると考えて、評価項目ごとではなく、まとめてパラメータ推定を行った。平均的な評価の特性は、同じ成果物を評価した他の14人の評価の平均を用いて計算を行った。ある評価者の行った成果物に対する評価と、推定されたパラメータから予測される評価特性曲線を図3.4に示した。

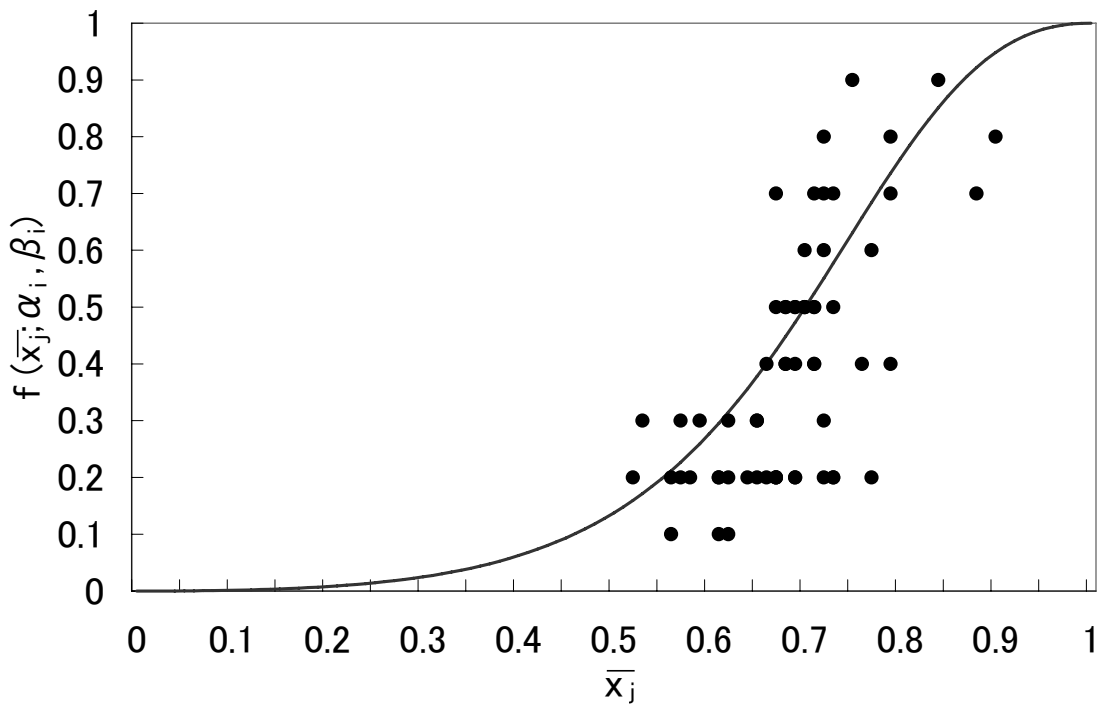


図3.4：推定された評価特性曲線と実測値

この評価者の推定されたパラメータは、 $\alpha_i = 2.16$ 、 $\beta_i = -0.430$ であり、平均より厳しい評価をしていることが分かる。推定されたパラメータを用いて、モデルから求められた推定値と実測値の差を調べるために、学習者ごとに平均二乗誤差(式10)を計算した。

$$e_i = \sqrt{\frac{1}{n} \sum_{j=1}^n \{y_{ij} - f(\bar{x}_j; \alpha_i, \beta_i)\}^2} \quad (10)$$

ただし、 n は行った評価の回数

評価者全体の平均二乗誤差の平均は 0.183 であった。これは、11 段階の評価にすると平均して約 2.0 ずれることになる。なお、図 4 に示した評価者の平均二乗誤差は 0.256 であった。評価者全体の約 1 割にあたる 4 名が平均して 11 段階に換算すると 3.0 以上ずれていた。このように当てはまりが悪かった学習者に対しては、「3.4.2 結果」で述べるような対処が考えられる。

3.4. 適用例

3.4.1. データ

B 大学 1 年生 45 名に対して行った「コンピュータ入門」の講義での相互評価のデータを用いて、補正の効果について検証した。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、グループワークを交えながら 14 回分の授業時間で作成した。相互評価は、15 回目の授業で一度だけ行い、1 人の学習者あたり、6 人分の課題を、次の 5 つの観点でそれぞれ 1~5 の 5 段階で評価させた。

- ・ 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
- ・ プレゼンテーションソフトの機能を適切に使っているか
- ・ レイアウト、色使い、字体は見やすいか
- ・ 全体として内容はうまくまとめられているか

評価にあたっては、評価した相手からは評価されず、また、グループワークを同じ班で行ったものを評価しない組み合わせとした。

3.4.2. 結果

学習者の推定されたパラメータの散布図を図 3.5 に示す。の値について 1.0 を境に見てみると、 >1.0 の者は全体的に差をつけた結果、 <1.0 の者よりも厳しい評価をつける傾向があることが分かる。図 3.5 の左下には離れた点が 3 つ存在する。これについて評価データを調べたところ、いずれもほとんどの成果物について高い評価をつけ、差をほとんどつけていないことが分かった。

次に、推定されたパラメータを用いて、モデルから求められた推定値と実測値の差を調べた。それぞれの学習者ごとの平均二乗誤差を式(10)を用いて計算し、その分布

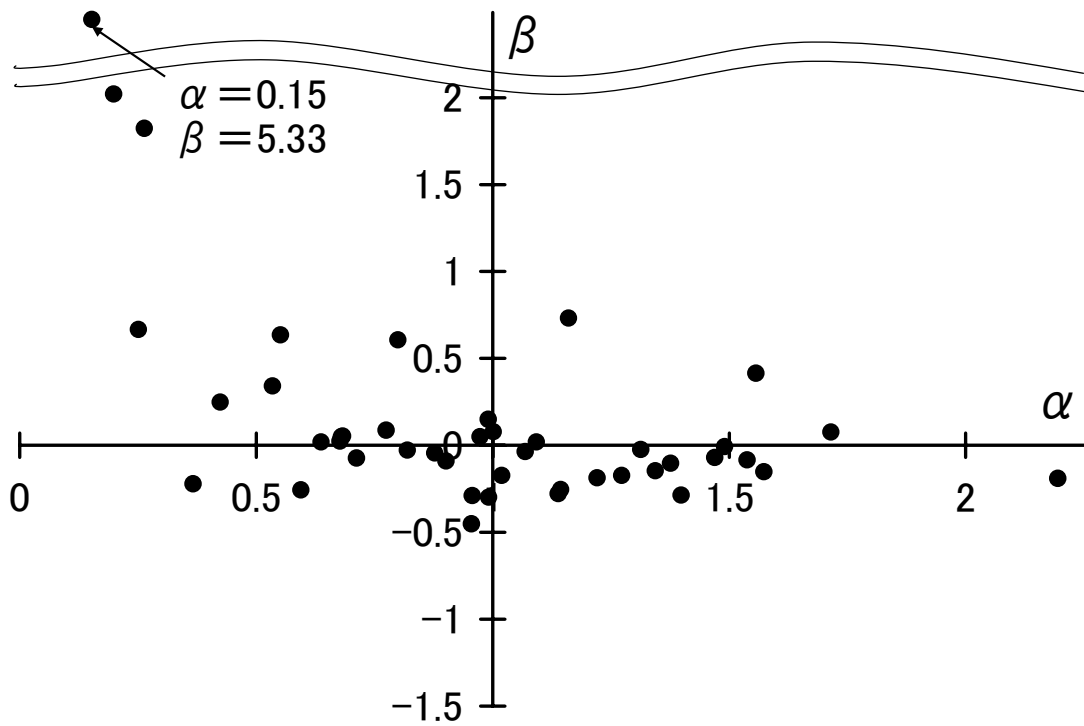


図 3.5 : 推定されたパラメータの分布

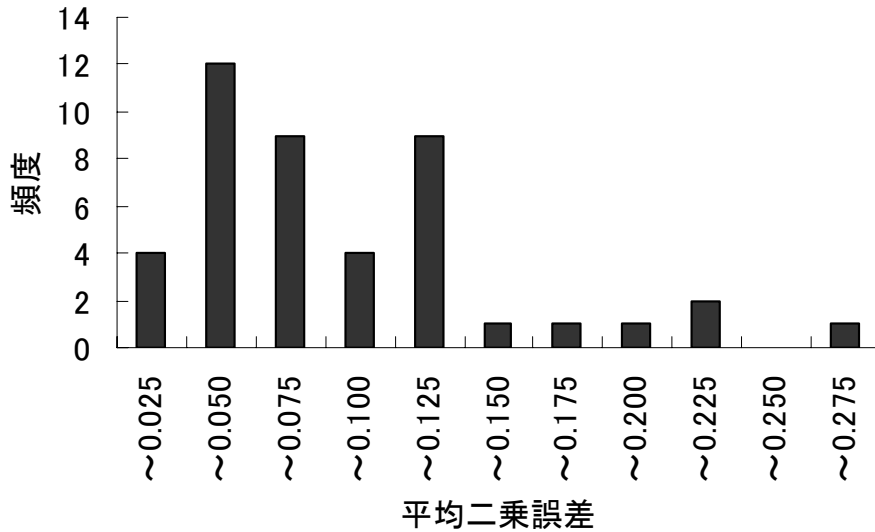


図 3.6 : 推定された値と実測値との差

を図 3.6 に示した．平均二乗誤差が 0.05 の時には，5 段階に変換すると 1 評価項目あたり 0.2 ずれ，同様に 0.125 の時は，0.5 ずれる計算になる．平均誤差が 0.2 を超えた 4 名を調べると，例えば画像の使い方にこだわった評価をするなど，他の者とは異なる

った評価を行う傾向が見られた。全体を見ると、実用に耐える推定の精度であった。このように集団とは異なった評価をするなどの理由で、推定の精度が一定値より低い学習者に対しては、補正の計算からは除外する、学習者に対して他の学習者の評価と離れていることを知らせる、教師にこのような学習者がいることを通知し、当該学習者がどのような評価を行っているか教師がチェックし指導するなどの対処が考えられる。

次に式(7)を用いて、個人差を補正した評価値を計算した。補正前の学習者がつけた評点と教員がつけた評点との Spearman の順位相関係数（以下、相関係数）は 0.303 であったが、補正後の評点と教員がつけた評点との相関係数は 0.427 となり、適切に補正が行われていることが分かった。

この結果を検証するために、別の教員に今回の成果物の評価を依頼し、補正の前後で比較した。2人の教員間の相関係数は 0.566 であった。追加で依頼した教員がつけた評点と補正前の学習者の評点との相関係数は 0.129 であったのに対して、補正後は 0.235 となり、最初の教員と同様の結果が確認された。

3.5. まとめ

評価者によっては、評価が甘かったり、厳しかったり、差をつける評価を行ったり、ほとんど差をつけなかったりするなどの特性がある。評価者を選択する必要がある場合、担当した評価者による評価の特性を考慮しないで、評価値をそのままフィードバックした場合、学習者間に不公平が生じる。

本章では、多数の欠測値を含む相互評価のデータにおいて、評価者の評価特性を考慮するために、個々の評価者の評価特性をモデル化し、その特性を考慮して、実際の評価値を補正する方法を提案した。提案手法では、項目応答理論のメタファを用い、評価特性を評点の厳しさと差をつける度合いをあらわす2つのパラメータで表現した。また、他の評価者が同じ成果物について行った評価の平均と評価特性から推定された評価特性曲線を用いて、実際の評価値を補正した。

次に、提案手法を、実データに適用し、補正が適切であることを確認した。提案手法は、評価者の評価特性にばらつきがある場合に有効である。評価者の評価特性から、適切に評価していない学習者に対して指導を行うこと、評価対象の学習者のフィードバック画面に適切でない評価者による評価結果を掲載しないこと等ができ、より適切

なフィードバックを行うことができる。これらの特徴は、特に形成的評価として、相互評価を行う際に重要である。実際に、相互評価の実践において、このようなフィードバックを行うためには、評価者の評価特性を考慮したフィードバックを容易に行うことができる相互評価支援システムを開発することが必要である。

第4章 相互評価支援システムの開発と評価

4.1 はじめに

第3章までに、相互評価においては、公平性の高い評価方法が必要であることを述べた。第2章では、評価者選択の際に生じる問題を、第3章では、個々の評価者の評価特性の差から生じる問題について指摘した。本章では、公平な相互評価を行うための相互評価支援システムの開発について述べる。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する機能、フィードバック時に評価者の評価特性に応じて評価データを補正する機能が特徴である。

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。

4.2. システム概要

4.2.1. システム構成

今回開発した相互評価支援システムは、成果物の提出、相互評価、評価結果の確認を Web 上から行うことができる。本システムの構成を図 4.1 に、システムを使った相互評価の流れを図 4.2 に示す。学習者を対象とした機能と課題の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。本システムを利用した学習の流れは、課題作成、課題提出、他の学習者の評価、結果の確認、課題修正であり、必要に応じて修正後の課題に対して再度相互評価を行う。

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。メインメニューの画面例を図 4.3 に示す。課題名を選択してから、実行したいボタンを押すことで、各機能を使用することができる。提出機能は、電子ファイルとして作成された成果物をシステムにアップロードし、登録することができる。相互評価機能は、課題を提出した学習者間で、評価相手の成果物を表示し、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成果物の評価をできない場合は、それぞれの学習者がどの成果物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、提出した成果物が受けた評価と、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。

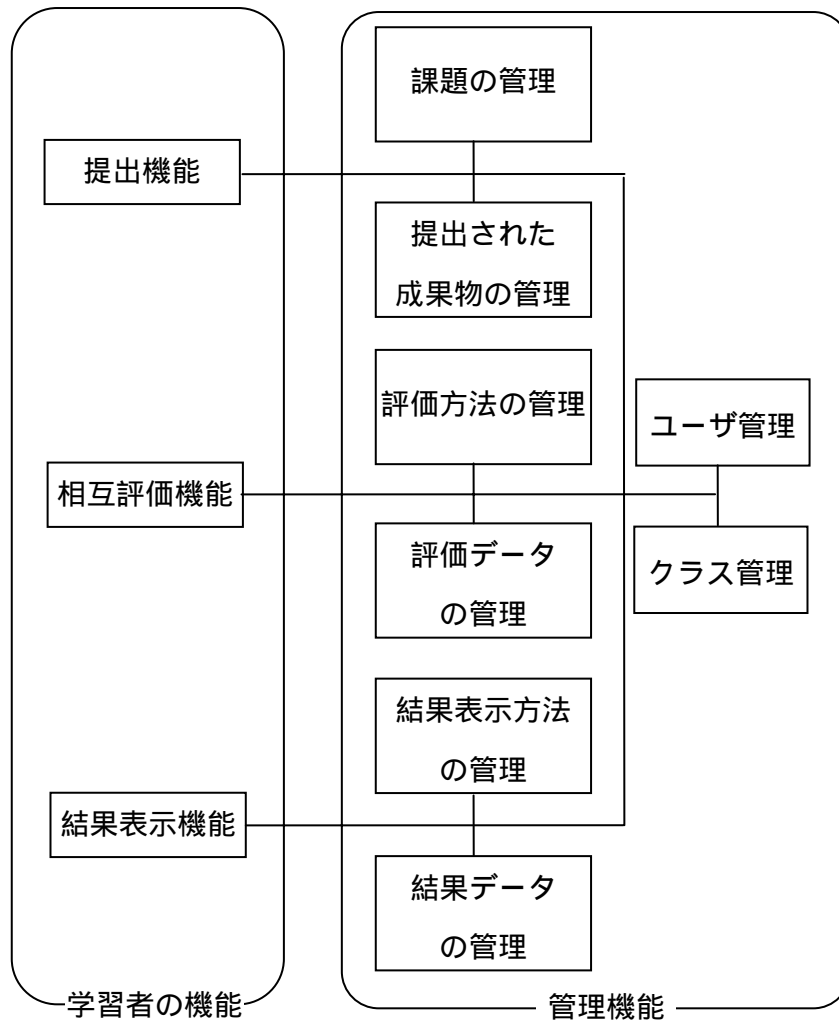
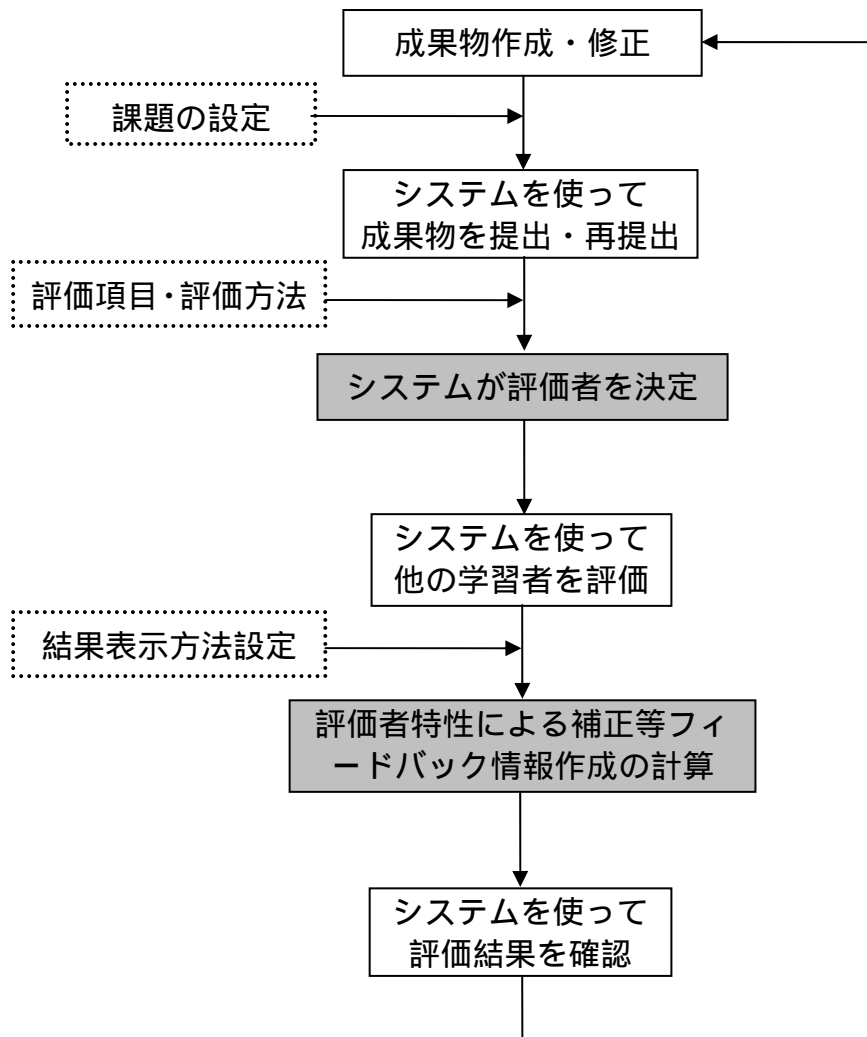


図 4.1 : システム構成

管理機能は、学習者の管理、クラスの管理、成果物の管理、提出された成果物の管理、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。



背景が網掛けの部分にはシステムの機能を表す。

図 4.2 : システムを使った相互評価の流れ

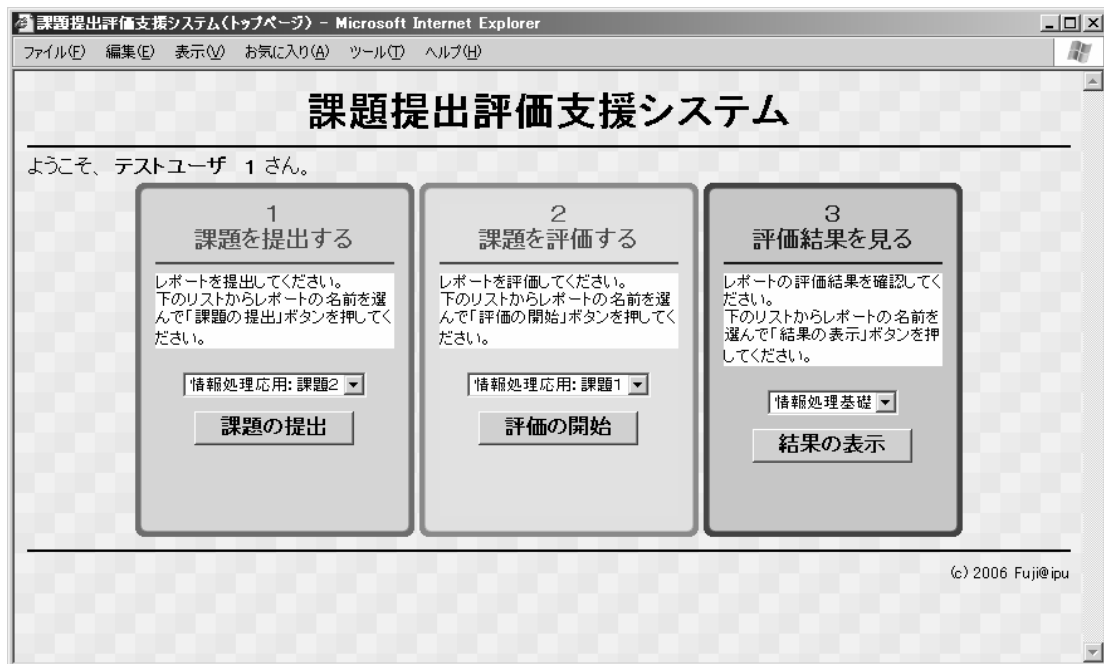


図 4.3 : メインメニュー

4.2.2. 動作環境

データベースを用いたサーバ/クライアントシステムとして開発した．サーバのソフトウェア環境は，以下のとおりである．

- ・ OS : Microsoft Windows Server2003 SP1
- ・ Web サーバ : Apache 2.0
- ・ スクリプト言語 : PHP 4.4
- ・ データベース : PostgreSQL8.0

クライアントは，インターネットに接続された Web ブラウザから利用できる．ただし，PHP スクリプトの実行，ファイルのダウンロード，新しいウインドウを開くことができる環境に限られる．

4.2.3. データベース

本システムは，図 4.1 に示した管理機能それぞれに対応したテーブルを持っている．ユーザおよびクラス（授業）を管理するテーブルを表 4.1，表 4.2 に示す．ユーザは，ユーザ名とパスワードで認証され，それぞれのユーザは複数のクラス（授業）に所属することができる．クラスは，該当する授業を担当する者（以下，授業担当者と呼ぶ）

が管理を行う。授業担当者は、ユーザを授業に登録することができ、ユーザ登録が終わっていない学習者のユーザ登録をすることもできる。

表 4.1：ユーザ管理テーブル

カラム名	説明
userid	ID
username	ユーザ名
password	パスワード
name	本名
class	登録されているクラス
lastlogin	最終ログイン日時
lastIP	最終アクセス IP
account	ユーザの属性（学習者, 授業担当者）

表 4.2：クラス管理テーブル

カラム名	説明
classid	ID
name	授業名
staff	授業担当者名
comment	授業の説明

4.3. 相互評価支援システムの主な機能

4.3.1. 提出機能

(1) 成果物の管理

授業担当者は、学習者に課した課題について、表 4.3 に示す情報を、システムを使って登録する（図 4.4）。成果物は、いずれかのクラスに対応させる必要がある。表 4.4 に示したテーブルで、個々の学習者の成果物の状態や提出されたファイルの情報は管理される。個々の成果物は、成果物の状態（「提出不許可」、「提出期間前」、「未定出」、「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。

表 4.3 : 課題管理テーブル

カラム名	説明
reportid	ID
classid	対応するクラスの ID
rep_name	課題の名前
start_date	提出受付日時
end_date	提出締切日時
max_files	提出ファイル数の上限
mac_size	提出ファイルのサイズの上限
update	提出後の差し替えの可否
file_suffix	提出ファイル形式の制限
commentT2L	提出画面で表示するコメント
commentL2T	教員宛のコメント入力の有無
commentL2L	他の学生宛のコメント入力の有無

表 4.4 : 提出成果物管理テーブル

カラム名	説明
user_reportid	ID
userid	対応するユーザの ID
reportid	対応する課題の ID
status	成果物提出の状況
page	何番目のファイルか
file_oid	提出されたファイルへのポインタ
file_name	提出されたファイルの名前
date	提出日時
type	MIME タイプ
comment2T	学習者が入力した教師宛のコメント
comment2L	学習者が入力した他の学生宛のコメント



図 4.4 : 課題設定画面

(2) 成果物の提出

学習者がメインメニューで、課題の提出を選ぶと、提出期間中であれば、成果物の提出画面（図 4.5）が現れる。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。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。

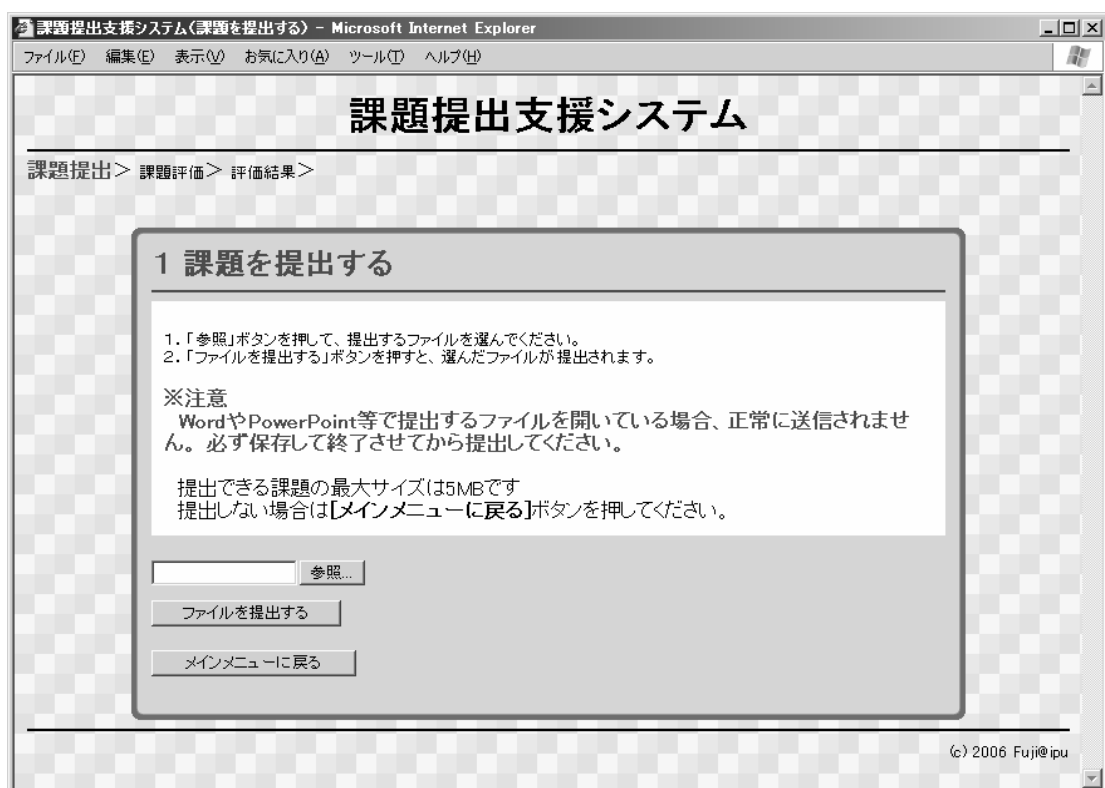


図 4.5：成果物提出画面

4.3.2. 相互評価機能

(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、すべての成果物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入力する。なお、フィードバック時に、評価者の個人差を除去することを考

慮して、すべての学習者が同じ数の成果物を評価するようになっている。



「お互い様効果」を除去する、すなわち評価した相手から、評価されない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 $\lfloor (n-1)/2 \rfloor$ 人分の成果物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人が評価する成果物の数を 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 人分の成果物を評価する。 $n=5$ の時の割り当ての例を表 4.5 に示す。行は評価者の学習者番号、列は評価される学習者の番号を表し、表中の「」がついた成果物が評価される。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。

表 4.5：お互い様効果を考慮した割り当て

$m=2, n=5$ の場合	$m=3, n=5$ の場合																																																																								
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr> <tr><th>1</th><td></td><td></td><td></td><td></td><td></td></tr> <tr><th>2</th><td></td><td></td><td></td><td></td><td></td></tr> <tr><th>3</th><td></td><td></td><td></td><td></td><td></td></tr> <tr><th>4</th><td></td><td></td><td></td><td></td><td></td></tr> <tr><th>5</th><td></td><td></td><td></td><td></td><td></td></tr> </table>		1	2	3	4	5	1						2						3						4						5						<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr> <tr><th>1</th><td></td><td></td><td style="background-color: #cccccc;"></td><td style="background-color: #cccccc;"></td><td></td></tr> <tr><th>2</th><td></td><td></td><td></td><td style="background-color: #cccccc;"></td><td style="background-color: #cccccc;"></td></tr> <tr><th>3</th><td style="background-color: #cccccc;"></td><td></td><td></td><td></td><td style="background-color: #cccccc;"></td></tr> <tr><th>4</th><td style="background-color: #cccccc;"></td><td style="background-color: #cccccc;"></td><td></td><td></td><td></td></tr> <tr><th>5</th><td></td><td style="background-color: #cccccc;"></td><td style="background-color: #cccccc;"></td><td></td><td></td></tr> </table>		1	2	3	4	5	1						2						3						4						5					
	1	2	3	4	5																																																																				
1																																																																									
2																																																																									
3																																																																									
4																																																																									
5																																																																									
	1	2	3	4	5																																																																				
1																																																																									
2																																																																									
3																																																																									
4																																																																									
5																																																																									
お互いに評価しあう部分はない	 の部分はお互いに評価しあう																																																																								

(2) 相互評価の方法および評価データの管理

授業担当者は、評価画面を作るために必要な情報、評価者決定のために必要な情報として表 4.6 に示す情報を、システムを使って登録する(図 4.6)。評価者の割り当ては、前項のアルゴリズムにしたがって自動的に行われる。学習者が入力した評価の結果は、表 4.7 に示したテーブルで、管理される。

(3) 成果物の評価

相互評価を実施する画面を図 4.7 に示す。学習者は、授業担当者の設定に基づいて

表 4.6：評価方法管理テーブルおよび評価シート管理テーブル

カラム名	説明
eval_confid	ID
reportid	対応する成果物の ID
start_date	評価開始日時
end_date	評価終了日時
method	割り当て方法
number	1人あたりの評価人数
mutuality	お互いさま効果除去の有無
group	グループ数

カラム名	説明
eval_itemid	ID
eval_confid	対応する評価方法の ID
order	質問の順序
type	回答方法
time	提示時間
question	質問文，指示内容
number	段階数または最大値

評価方法は，評価方法の実施方法と評価シートを構成する情報の2つのテーブルから構成される

表 4.7：評価データ管理テーブル

カラム名	説明
user_evalid	ID
userid	評価を行ったユーザの ID
eval_confid	対応する評価方法の ID
status	状況
order	順番
time	提示時間の制限（必要な場合）
eval_userid	評価対象のユーザの ID
res1	項目 1 への回答
res2	項目 2 への回答
⋮	⋮
resN	項目 N への回答

選ばれた成果物を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成果物を見て，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。

課題提出支援システム(管理画面) - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

課題提出支援システム(管理画面)

トップページ > 評価方法設定

レポート課題設定	受講生別 提出レポート	評価方法設定	受講生別 評価データ	結果表示設定	結果ファイル出力
----------	----------------	---------------	---------------	--------	----------

評価方法設定

提出されたレポートファイルの相互評価の方法に関する設定を行います。

設定するレポートの選択

設定を行うレポートのクラス名、レポート名を選択して、設定ボタンを押してください。
他のレポートの相互評価方法を元で作成する場合は、複製元レポート名を選択して複製ボタンを押してください。

授業名	コンピュータ入門	レポートを課す授業の名前 リストがない場合はメニューに戻って作成してください
レポート名	課題3:プレゼンテーション	コンピュータ入門で課すレポートの名前 リストがない場合はレポート課題設定画面で作成してください
複製元レポート	クラス名(複製時選択) レポート名(複製時選択)	他のレポートの相互評価方法を複製する時に指定してください

レポート「課題3:プレゼンテーション」の設定

相互評価実施期間	評価開始: 2006年 2月 6日 10時 30分 締め切り: 2006年 2月 6日 11時 30分	年は西暦で記入してください
評価者決定アルゴリズム	<input checked="" type="radio"/> お互い様効果除去 <input type="radio"/> グループ <input type="button" value="グループ設定"/> <input type="radio"/> ランダム <input type="button" value="個別に設定"/> <input type="button" value="割付結果表示"/>	お互い様効果除去: 相互に評価しあわないように自動的に設定します グループ: グループの中でお互いに評価させるように設定します ランダム: 評価者を無作為に設定します 個別に設定: 評価者と被評価者を一人ひとり手動で設定します
評価人数	6人分	1人の評価者が評価する人数
項目数	6項目	学習者が評価する設問数
コメント表示	<input checked="" type="radio"/> 無 <input type="radio"/> 有	ファイル提出時に受講生が入力したコメントの表示
項目1	設問: _____ 回答形式: 段階 5段階 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
項目2	設問: _____ 回答形式: 段階 5段階 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
項目3	設問: _____ 回答形式: 段階 5段階 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
項目4	設問: _____ 回答形式: 段階 5段階 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
項目5	設問: _____ 回答形式: 段階 5段階 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
項目6	設問: コメント 回答形式: 自由記入 8行 理由記述欄: なし	[回答形式]で[自由記入]以外を選んだ場合は、その理由を記述させる権を設定できます
教員による評価	<input checked="" type="radio"/> 無 <input type="radio"/> 有	現在ログインしているIDを使って、提出ファイルを評価することができます

図 4.6 : 相互評価方法設定画面



図 4.7 : 相互評価実施画面

4.3.3. 結果表示機能

(1) 評価者の評価特性に応じて相互評価データの補正

結果表示画面では、個々の評価者の評価特性、評価特性に応じて相互評価データを補正した値を表示することができる。評価者の評価特性の計算及び補正のアルゴリズムは、第3章で示した手法で行った。

(2) フィードバックの管理

授業担当者は、提出された成果物のフィードバック画面について、表4.8に示す情報を、システムを使って登録する(図4.8)。フィードバック画面に必要な場合は、逐次計算され、表4.9に示すテーブルに記録される。

授業担当者は、提出された成果物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。

表4.8：結果表示方法管理テーブル

カラム名	説明
result_confid	ID
eval_confid	該当する評価方法のID
start_date	公開開始日時
end_date	公開終了日時
regression	評価者の個人差補正方法
teacher	教員の評価の表示
record	評価の履歴の表示

表4.9：結果データ管理テーブル

カラム名	説明
user_resultid	ID
userid	該当するユーザのID
result_confid	該当するフィードバック方法のID
	評価特性の推定値
	評価特性の推定値
rev1	回答1に対する補正值
⋮	⋮
revn	回答nに対する補正值



図 4.8 : 結果表示設定画面

(3) フィードバック情報の表示

学習者は、自分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 4.9 に示す。学習者の受けた評価の値とその平均、教員の評価、学習者が他の人を評価した結果が表示される。評価特性に応じて相互評価データを補正した値を同時に表示することもできる。

自分の行った評価を他の学習者が行った評価と比較して、見直すことができる。なお、評価を確認するときに、評価をする際に使ったファイルを見ることもできる。



図 4.9 : 結果表示画面

4.4. 相互評価支援システムの評価

4.4.1 システム改善のための評価

(1)実施方法

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。予め与えたデータを、表計算ソフトを使って加工する課題を出し、システムを使って課題を提出し、相互評価を行った。1週間後に結果を閲覧させた。システムに関して、機能ごとにアンケートで行い、問題が指摘された箇所について、意見を聞いた。アンケートの項目は、システムに関するものと相互評価に関する項目から構成した。システムに関するものとしては、それぞれの機能に対して、次の項目について尋ねた。機能と質問項目の対応を表4.10に示す。

- ・ 操作に関する説明は分かったか（4段階：すぐに分かった，少し考えたら分かった，考えても分からない箇所があった，全く分からなかった）
- ・ 操作は使いやすかったか（5段階：使いやすかった，どちらかといえば使いやすかった，どちらともいえない，どちらかといえば使いにくかった，使いにくかった）
- ・ レイアウトは見やすかったか（5段階：見やすかった，どちらかといえば見やすかった，どちらともいえない，どちらかといえば見づらかった，見づらかった）
- ・ 何をすればよいか分からなくなることはあったか（2段階：あった，なかった）
- ・ 予期しない挙動をすることはあったか（2段階：あった，なかった）

表4.10：機能と質問項目の対応

	メインメニュー	課題の提出	評価の実施	結果の確認
操作に関する説明	1	1	1	1
操作の使いやすさ	1	1	2（成果物を見る操作，評価する操作）	1
レイアウトの見やすさ	1	1	1	2（自分への評価，他人に対して行った評価）
何をするか不明な点	1	1	1	1
予期しない挙動	1	1	1	1

数値は項目数を表す，1項目の場合はその機能全体に対する評価

(2)結果

システムに関するアンケートの結果を表 4.11 に示す。選択式項目 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 件であった。最も肯定的な選択肢以外が選ばれた 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、他の学習者の意見も聞いた。その結果、意見が一致したインターフェースに関する 2 点を修正した。修正した点以外に、評価画面でのファイルの表示が別ウインドウになるのが見づらいとの意見もあった。そのため、システムのウインドウに組み込むことも検討したが、クライアントの動作環境によって使うことができる場合とできない場合があることが予想されたため、現在の仕様は変更せずに、オプションとして対応することとした。

表 4.11：最も肯定的な選択肢を選んだ割合

	メインメニュー	課題の提出	評価の実施	結果の確認
操作に関する説明	80%	70%	50%	70%
操作の使いやすさ	90%	80%	30%	50%
			90%	
レイアウトの見やすさ	80%	80%	90%	80%
				40%
何をすることが不明な点	90%	90%	80%	100%
予想しない挙動	100%	100%	90%	100%

相互評価に関するアンケートの結果を表 4.12 に示す。数字はそれぞれの選択肢の度数で、5 が最も肯定的な回答である。相互評価することに関しては肯定的であったが、人につけた評価を確認することに関してはどちらともいえないという結果となった。次回の評価では、評価結果を表示する画面で、評価を行った答案に対して他の学習者が行った評価を表示するようにした。

表 4.12：相互評価に関するアンケート結果

質問項目	1	2	3	4	5	平均	標準偏差
評価を確認することは勉強になる	0	0	0	3	7	4.7	0.48
自分の評価をあとで確認したい	0	0	0	2	8	4.8	0.42
人の評価することは, 勉強になる	0	1	0	1	8	4.6	0.97
自分がつけた評価をあとで確認したい	2	2	1	2	3	3.2	1.62

4.4.2 学習者による評価

(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人の学習者の成果物を評価させ, 評価結果をフィードバックした。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, それぞれのスライドに何を記述するかは予め指定してある。例えば, 1枚目のスライドでは, タイトルのスライドとし, 指定された項目を記入すること, タイトルにワードアートを使用すること, 背景に白紙以外を指定することを要件とした。同様に9枚のスライド全てに条件を設定した。相互評価では, 次の5つの評価項目に対して, 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。

- ・ 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

1番目の項目は, 与えられた条件にあったプレゼンテーションになっているか, 主にソフトウェアの操作の観点から評価させた。2番目の項目は, アンケートデータを加工して作成した表やグラフが, プレゼンテーションの目的にあっているかどうかについて評価させた。3番目の項目は, 単にプレゼンテーションソフトの機能を使っているのではなく, 聴衆の立場から, 描画ツールやアニメーションが効果的に使われているかについて評価させた。4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、色の組み合わ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。5番目の項目については、成果物全体からアンケートの結果を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。

表計算やプレゼンテーションソフトウェアの操作方法を知り、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、第2章の実践と同様に、実際に課題をやってみて、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。

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。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。

(2) アンケートの結果

アンケート項目と集計結果を表4.13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4.10)、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でき)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)、成果物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。

他者の成果物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、各成果物について、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、その学生の成果物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、他者の成果物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成果物に対して行った評価と、同じ成果物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「他の人が同じ成果物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。

表 4.13：学習者に対するアンケートの結果

質問項目	平均	標準偏差
システムについて		
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
相互評価について		
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成果物を評価することは、勉強になる	4.10	0.93
他の人が同じ成果物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

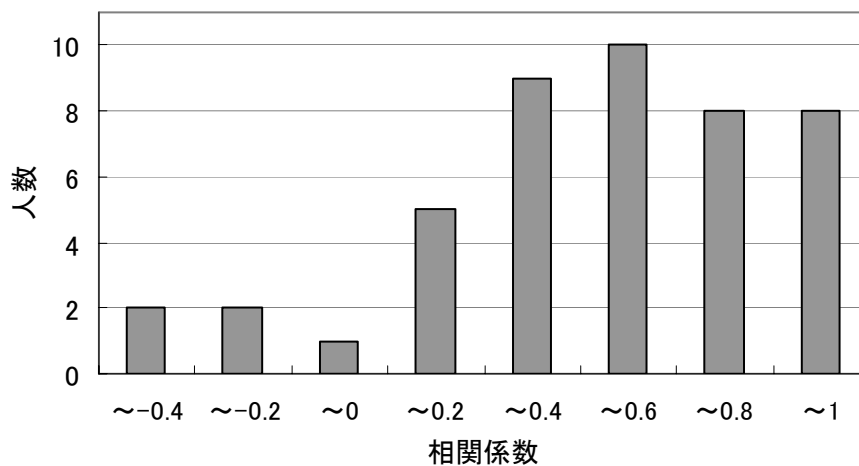


図 4.10：学習者が行った評価と教員の評価との相関の分布

4.4.3. 教員による評価

(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、紙で提出された成果物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。

(2) 結果

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。

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(表 4.14)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 4.15 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、「成果物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。

表 4.14：管理機能一覧

機能	設定項目
課題提出画面	成果物名, 受付期間, 提出後の差し替え, ファイル形式の制限, 提出画面コメント, 教員宛コメント入力欄, 他の学生宛コメント入力欄
課題提出状況の確認	ユーザ名(学籍番号), 氏名, 提出状況, 提出ファイル名, 提出ファイルの表示, ファイルの削除, ファイルの追加, 最終提出日時, 教員宛コメント, 他の学生宛のコメント, 提出状況の設定, 提出状況の一括設定, 全ファイル一括削除, 全コメント一括削除
相互評価実施に関する設定	評価実施期間, 評価者の自動割り当て, グループ内評価の設定, お互い様効果の除去, 提出時のコメントの表示, 評価項目の設定, 評価理由のコメント欄の設定, 教員による評価の有無
相互評価実施状況の確認	ユーザ名(学籍番号), 氏名, 実施状況, 評点の表示, 評点の設定, コメントの表示, コメントの設定, 実施状況の一括設定, 評価項目の表示
結果のフィードバックに関する設定	結果閲覧期間, 平均値の表示, クラスの平均の表示, 教員の評価の表示, 個人差を補正した結果の表示, 他の学習者に対する評価の履歴の表示
結果分析用ファイル作成に関する設定	ラベル(見出し), 出力対象の指定, 学習者の属性, 評点(素点), 記述式項目, 教員の評価, 補正後の評点, 補正用の評価特性パラメータ

表 4.15：管理機能における各機能の評価（5段階）

機能	項目数	平均	分散
課題提出画面の設定	7	4.60	0.66
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	4.57	0.74
結果のフィードバックに関する設定	5	4.63	0.56
結果分析用ファイル作成に関する設定	8	4.77	0.50

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 4.16 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学習者に評価集計した結果を正確に知らせることを考えると、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がえた。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。

表 4.16：教員が相互評価に対して持っている印象（5段階）

質問項目	評価前	評価後
他の学生の成果物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てみたい	4.67	4.83

評価後欄の記号は Wilcoxon の符号付き順位検定の結果(*:5%有意,+:10%有意傾向)

4.5. まとめ

実際に、学習活動に相互評価を導入するためには、学習者、教員ともに負担を最小限にすることが求められる。本章では、公平なフィードバックを行うために、お互い様効果を除去して合理的に評価すべき相手を選択できる機能と、個々の評価者の評価特性に基づいた評価値の補正機能を持った相互評価支援システムの開発と評価について述べた。

本システムは、学習者向けの機能としては、電子ファイルでの課題提出機能、相互評価機能、結果表示機能から構成される。相互評価の際には、お互い様効果を考慮して、自動的に評価者を決定することができる。結果の表示では、個々の学習者の評価特性とそれに基づいた評価の補正值を表示することができる。

システムのユーザビリティ、システムを使った相互評価の印象について、学習者による評価を行った。また、相互評価を導入した授業の経験がある教員に対しても、同様の評価を行った。学習者による評価から、相互評価について積極的に受け入れることが分かり、教員による評価から、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することが重要であることが分かった。学習者による評価、教員による評価で、容易に相互評価を実施することができ、より公平な相互評価を支援できることから、本システムは、有用であることが分かった。

また、システムを利用した実践では、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することが分かった。今回は、授業の最終回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が、形成的評価として行う場合は、適切な評価ができない学習者がさらに増えることも考えられる。本システムの教

育的効果を調べるためには、形成的評価として継続的に相互評価を行い、学習者の他者評価の能力が向上するのか、学習の改善を行うことができるのかについて、検討を行う必要がある。

第5章 相互評価支援システムを導入した実践の効果

5.1. はじめに

本章では、相互評価支援システムを継続的に使用することによる教育的効果を調べるための実践について述べる。システムのフィードバック画面で評価者特性に基づいて補正した評価値等を表示させ、提案した補正方法によって学習者の評価能力が向上するかどうかについて検証した。

5.2. 実践の概要

5.2.1. 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5.1に示す。このうち、プレゼンテーションソフトでの課題でシステムを利用した実践を行った。

授業の履修者は、文学部1年生19名であったが、以下の計画及び結果では、単位取得者16名を対象として述べる。なお、不合格となった3名は、いずれもプレゼンテーションソフトの前の回までに、既に出席が常ではなかった者であり、不合格の理由は今回の実践と関係しない。

表 5.1：授業内容

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

5.2.2. 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5.2に示す。第6回の授業から3時間をかけて課題の制作を行い、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時間をかけて課題の改善を行い、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。

表5.2：授業内容と使用する機能

回	授業内容	提出	評価	結果
6	PowerPointの使い方(60分) 制作(30分)			
7	制作(90分)			
8	制作(30分) システムの使用法・事前評価(60分)	制作	(事前) 制作	
9	画像の引用(30分) 改善(60分)	改善		制作
10	アニメーション(30分) 改善(60分)	改善	改善	改善
11	見やすいスライド(30分) 改善(30分)	改善	改善	改善
12	分かりやすいスライド(30分) 改善〔完成〕(30分)	完成版	改善	改善
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)		(事後)	

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。

「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す。改善された成果物は、再び提出され、再評価を受ける。

()内はそれぞれの授業での配分時間を表す。(11,12,14回の授業時間は60分、13回の授業時間は30分。改善 ~ の時間にはシステム使用時間を含む)

5.2.3. 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。課題でのスライド数は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、色の組み合わ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、成果物中の文章などの了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、操作方法を知った上で、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、課題制作、評価、課題の改善のサイクルを繰り返しながら、学習させた。

5.2.4. 相互評価の方法

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、お互い様効果の影響が確認されるか、フィードバックの際に個人差を補正した結果を表示することによる効果はあるかについて検討を行う。

学習者を第1回の授業でのコンピュータの知識についてのアンケートをもとに4群に分け、表5.3にあげた条件で学習させた。相互評価実施前後での他者への評価の違

いを比較するために、昨年度の学習者が作成した課題を評価させた。6人分の課題を2セット用意し、群によって事前・事後で使用するセットを変えた。それぞれのセットには、昨年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。フィードバックの際に個人差を補正した結果を表示することによる影響をみるために、群によって補正した結果を表示の有無を設定した。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。

表 5.3：群ごとの事前事後の課題と結果表示

	事前評価： 課題セット	結果表示画面での 個人差補正の有無	事後評価： 課題セット
A群	セット	表示なし	セット
B群	セット	表示なし	セット
C群	セット	表示あり	セット
D群	セット	表示あり	セット

5回の相互評価の評価対象は、お互い様効果を考慮しない場合とする場合の両方について行った。評価の割り当てを表5.4に示す。お互い様効果を考慮しない評価対象者としては、4人で1グループを組ませ、自分以外の3人の成果物を毎回評価させる(表4.4で背景が濃い網掛け部分)。お互い様効果を考慮した評価対象としては、グループ以外の学習者3名を評価させる(表5.4で背景が薄い網掛けの部分)。次回の相互評価では、その続きから3人を選び、候補者全員の評価を行った後は最初に戻る。

表 5.4 : 評価者の割り当て

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		1	2	3						1	2	3				
2	3		1	2							1	2	3			
3	2	3		1								1	2	3		
4	1	2	3										1	2	3	
5						1	2	3						1	2	3
6	3				3		1	2							1	2
7	2	3			2	3		1								1
8	1	2	3		1	2	3									
9		1	2	3						1	2	3				
10			1	2	3				3		1	2				
11				1	2	3			2	3		1				
12					1	2	3		1	2	3					
13						1	2	3						1	2	3
14							1	2	3				3		1	2
15								1	2	3			2	3		1
16									1	2	3		1	2	3	

行が評価者，列が評価対象を表す．数字は評価する順序
背景が濃い部分はお互いに評価しあい，薄い部分はお互いに評価しあわない

5.2.5. システムの運用

第4章で評価を行ったシステムは，繰り返し相互評価を行うことを前提にしたものではなかったため，今回の実践に合わせて，システムの改訂を行った．ここでは，今回の実践でのシステムの運用方法について述べる．

(1) 課題の提出

学習者がメインメニューで，課題の提出を選ぶと，成果物提出期間中であれば，図5.1のような課題の提出画面があらわれる．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，再提出できるようにした．ファイルを提出するときに，課題を評価する他の学習者に対するコメント（課題の進捗状況，アドバイスを欲しい点など）を入力させた．

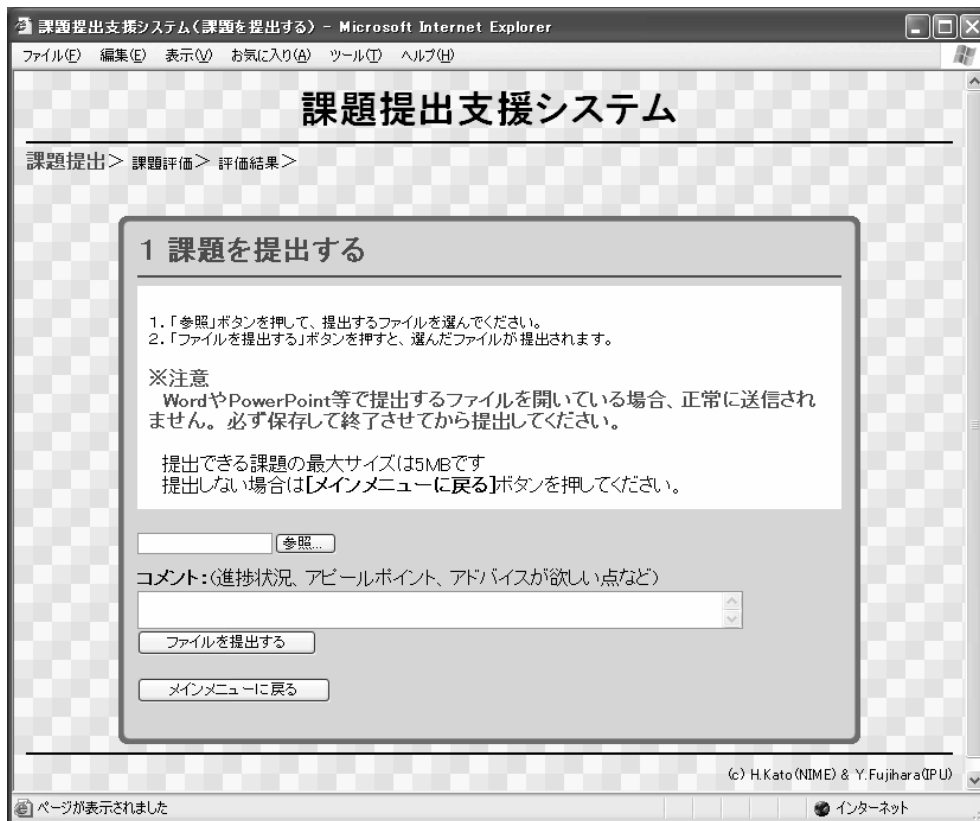


図 5.1 : 課題提出画面

(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、評価すべき相手を自動的に割り当てる機能を持っている。今回は、「2.4 相互評価の方法」で示したように割り当てを行った。1人の学習者が評価する課題は6つとし、うち3つは所属するグループ内、残り3つは「お互いさま効果」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ように自動的に割り当てを行った。

成果物を評価する画面を図 5.2 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、その評価結果を図 4.2 の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、課題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こともできる。

課題提出支援システム(課題を評価する) - Microsoft Internet Explorer

STEP3
 下の評価シートの縦軸に書かれてある5つの項目について、
 1～5の5段階で採点して下さい。
 考察や感想などのコメントがある場合は、評価シートの下にある
 コメント欄に記入して下さい。無い場合は未記入で構いません。

※注意
 評価シートの入力もれが無いように注意してください。
 コメント欄に入力できる最大文字数は250文字です。

作成者のコメント
 進捗状況: 半分、何でもいいのでアドバイスください

評価シート

	悪い ←	普通	→	良い	
	1	2	3	4	5
(1)レイアウト、色づかい、字体は見やすいか？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
理由:	<input type="text"/>				
(2)表現は分かりやすいか？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
理由:	<input type="text"/>				
(3)情報の信頼性は高いか？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
理由:	<input type="text"/>				
(4)見ていて楽しい気分になるか？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
理由:	<input type="text"/>				
(5)全体としてうまくまとめられているか？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
理由:	<input type="text"/>				

その他コメントがあれば記入して下さい。

評価シートに記入しおわったら、【次へ進む】ボタンを押してください。

ページが表示されました インターネット

図 5.2 : 課題評価画面

(3) 結果表示

フィードバック画面では、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、他者に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できるようにした。学習者は、図 5.3 に示すように自分が受けた評価を確認することができる。学習者の受けた評価の値とその平均、C 群・D 群の学習者には補正後の評価値が表示される。さらに、C 群・D 群では、学習者の評価の特徴をあらゆるパラメータを偏差値に変換して表示させた。学習者が他の学習者の作成した課題に対して行った評価についても、同様に確認することができる。

課題提出支援システム(評価結果を見る) - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

課題提出支援システム

課題提出 > 課題評価 > 評価結果 >

3 評価結果を見る

あなたのファイルをみた人たちの評価結果が下に表示されます。
自分が提出したファイルを見る場合は[ファイルを開く]ボタンをクリックしてください。

点数の目安
点数は5段階評価になっており、数値が多いほど「良い」と評価されたこととなります。

評価段階 悪い ← 普通 → 良い

評価段階	1	2	3	4	5
------	---	---	---	---	---

評価結果
あなたのレポートに対する評価結果です。
数値は評価者の個人差(甘い、厳しいなど)を補正した点数です。()内は、補正する前の元の点数です。

	(1) レイアウト、色づかい、字体	(2) 表現の分かりやすさ	(3) 情報の信頼性	(4) インパクト(楽しさ)	(5) 全体のまとめ	平均
1人目	4.6(5) テーマと色使いがよく合っていた	4.6(5) 読みやすい文章だった	4.4(5) 詳しく書かれていた	3.7(4)	4.7(5) アニメーションの効果がでていた	4.4 (4.8)
2人目	3.4(3) 字が大きくてよ	3.5(3) うまくまとめられていた	3.5(3) 顔写真や本の写真が付いているので	3.3(3) アニメーションもたくさんありおもしろい	4.3(4) 文章構成が良	3.6 (3.2)
3人目	2.8(3) ホラーのように暗かった。	3.8(4) わかりやすかった。	4.8(5) 高いと感じた。	4.7(5) 山田悠介を知っているので楽しいというかた	4.8(5) 山田悠介についてよくわ	4.2 (4.4)
平均(素点)	3.7	4.0	4.3	4.0	4.7	4.1
平均(補正後)	3.6	4.0	4.2	3.9	4.6	4.1
(参考)クラス全員の提出されたレポートの平均(補正後)です。						
クラス平均	3.9	3.8	3.9	3.8	3.9	3.9

あなたが評価したレポート
あなたの行った評価と他の人が同じレポートにつけた評価を見ることができます。
()内は評価者の個人差を補正した点数です。
()内は他の人が同じレポートに対して行った評価の平均です。
評価したレポートを確認するには、表の右端の「参照」ボタンを押してください。

	(1) レイアウト、色づかい、字体	(2) 表現の分かりやすさ	(3) 情報の信頼性	(4) インパクト(楽しさ)	(5) 全体のまとめ	平均	評価したレポート
1番目 あなたの評価 (他の人の評価) [補正後の評価]	4 (3.4) [4.3]	2 (3.4) [2.3]	3 (3.7) [3.3]	4 (3.7) [4.3]	4 (3.3) [4.2]	3.4 (3.5) [3.7]	<input type="button" value="参照"/>
2番目 あなたの評価 (他の人の評価) [補正後の評価]	1 (3.1) [1.2]	1 (3.1) [1.2]	4 (2.6) [4.0]	3 (2.7) [3.1]	2 (3.0) [2.2]	2.2 (3.1) [2.4]	<input type="button" value="参照"/>
3番目 あなたの評価 (他の人の評価) [補正後の評価]	4 (2.7) [4.1]	4 (2.6) [4.0]	3 (3.0) [3.2]	3 (3.4) [3.3]	4 (3.4) [4.3]	3.6 (3.0) [3.8]	<input type="button" value="参照"/>

あなたの評価の傾向 (数値の説明)
甘さ指標(偏差値): -0.13(47.1), 極端さ指標(偏差値): 0.72(49.3)

図 5.3: フィードバック画面 (補正表示群)

5.3. 結果

5.3.1. 実践中の他者の成果物の評価

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。その結果を表 5.5 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。1 項目あたりの評価の平均は、お互いに評価しあう場合は 4.13 ,お互いに評価しない場合は 3.87 となり($t(15)=2.33, p<.05$)、先行研究と同様に、お互い様効果が確認された。

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 5.5 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。

形成的評価として相互評価をおこなったため、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 5.5 に示す短所を具体的に指摘しているコメントの数が、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、課題の改善の参考になった評価の

割合」を尋ねた結果、図 5.4 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。

表 5.5：他の学習者の成果物に対して行った評価の推移

		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

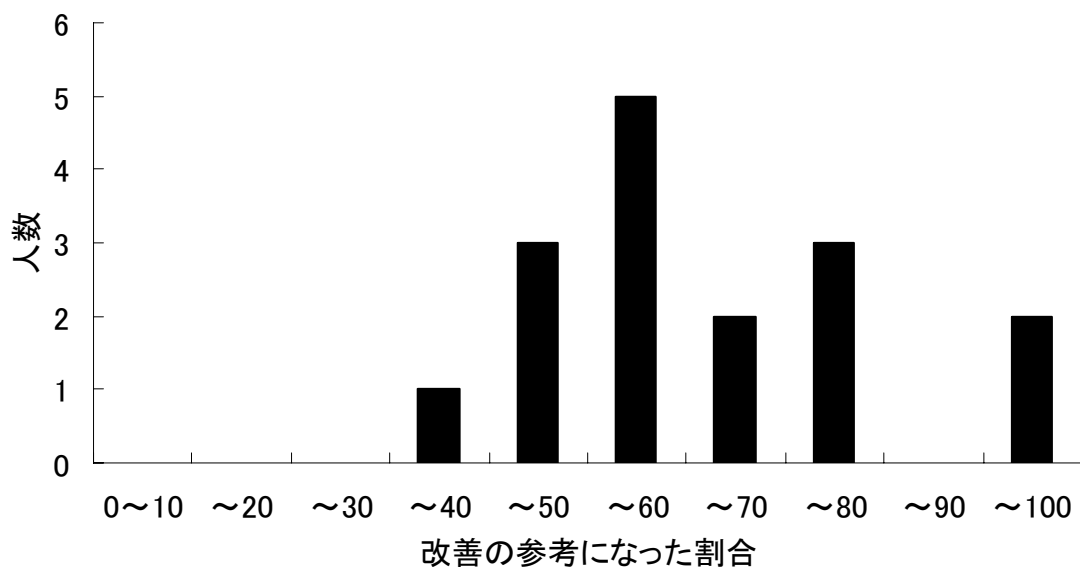


図 5.4：アンケート結果（改善の参考になった評価の割合）

5.3.2. 事前評価と事後評価の比較

相互評価実施前後で、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 が 3.43，セッ

ト が 3.43) で、事後評価は、3.72 (セット が 3.75, セット が 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。実施時期と補正值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: $F(1,14)=0.137, p>.10$)。

次に、教員が行った評価と学習者が行った評価の関係について調べた。Spearman の順位相関係数は、事前評価では補正表示群 0.260, 非表示群 0.284 であったのが、事後評価では補正表示群 0.431, 補正非表示群 0.192 となった。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。

5.3.3. 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 5.6 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。また、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。「学生間での評価を導入した授業を受けたい」についてのみ、有意な差が見られた ($U=14.0, p<.05$)。

表 5.6 : アンケートの結果 (評価者特性の補正に関する項目)

質問項目	補正あり	補正なし	検定
補正前の評価は、おおむね納得できた	4.13	4.50	
補正後の評価は、おおむね納得できた	4.25	-	
パラメータ の意味は理解できた	3.75	3.25	
パラメータ の意味は理解できた	3.88	3.38	
自分の評価の特徴に気がついた	4.25	3.38	
評価者によって評価に差がある	4.25	4.38	
評価者によって差があって不公平	3.00	2.25	
評価者の差を補正するほうが望ましい	3.75	2.38	*
学生間での評価を導入した授業を受けたい	3.88	3.00	

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

アンケートから、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いった補正表示群では、補正することが望ましいと回答した者が多く見られた。補正のパラメータの意味は理解できたかという問いに、2以下と回答した者が3名見られた。自分の評価の特徴が分かったかという問いには、パラメータの意味が理解できなかった3名を除いた平均は3.92であった。補正表示群では、補正後の得点には納得できたという項目の平均は4.25であり、補正值は高く評価されていたが、補正前と補正後でWilcoxonの符号付順位検定の結果、有意な差はなかった($Z=.577, p>.10$)。

5.3.4. 評価者特性モデルによる評価能力の変化

個々の評価者の評価特性をフィードバック画面で使用したことによる、評価能力の変化について述べる。初回と最終回での平均二乗誤差の変化を表5.7に、初回、最終回の評価のそれぞれで推定された学習者のパラメータを図5.5、図5.6に示す。

表5.7：実践前後での誤差の変化

条件	学習者番号	実践前	実践後	差
補正表示あり	1	.087	.151	.064
	2	.161	.142	-.019
	3	.030	.055	.024
	4	.182	.189	.006
	5	.091	.037	-.054
	6	.149	.091	-.058
	7	.032	.101	.069
	8	.197	.071	-.126
	平均	.116	.105	-.012
補正表示なし	9	.236	.186	-.050
	10	.172	.190	.018
	11	.157	.226	.069
	12	.068	.122	.053
	13	.112	.134	.022
	14	.172	.165	-.007
	15	.072	.141	.069
	16	.079	.024	-.055
	平均	.134	.149	.015
	全体の平均	.125	.127	.002

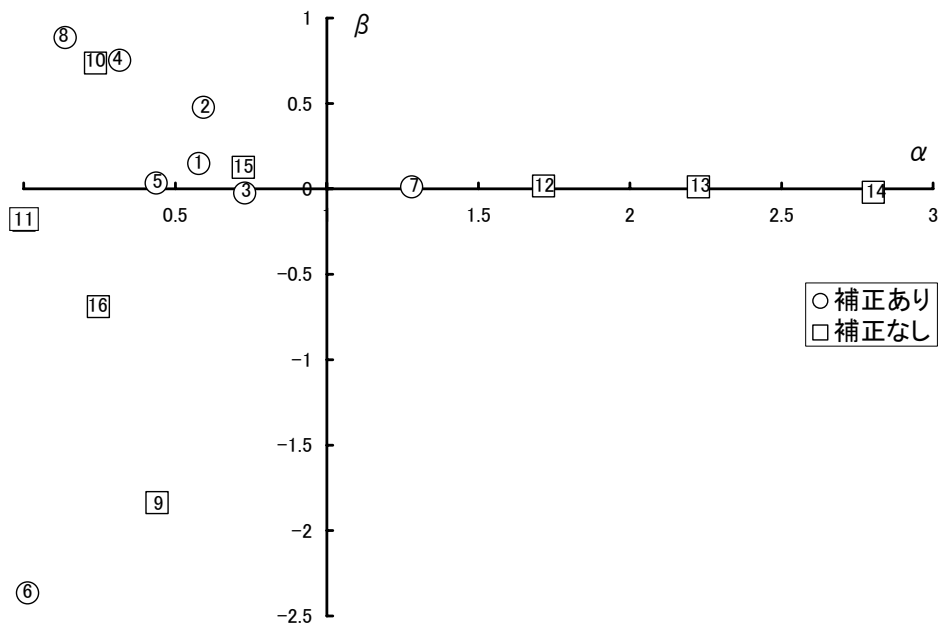


図 5.5 : 実践前の各学習者の評価特性パラメータの分布

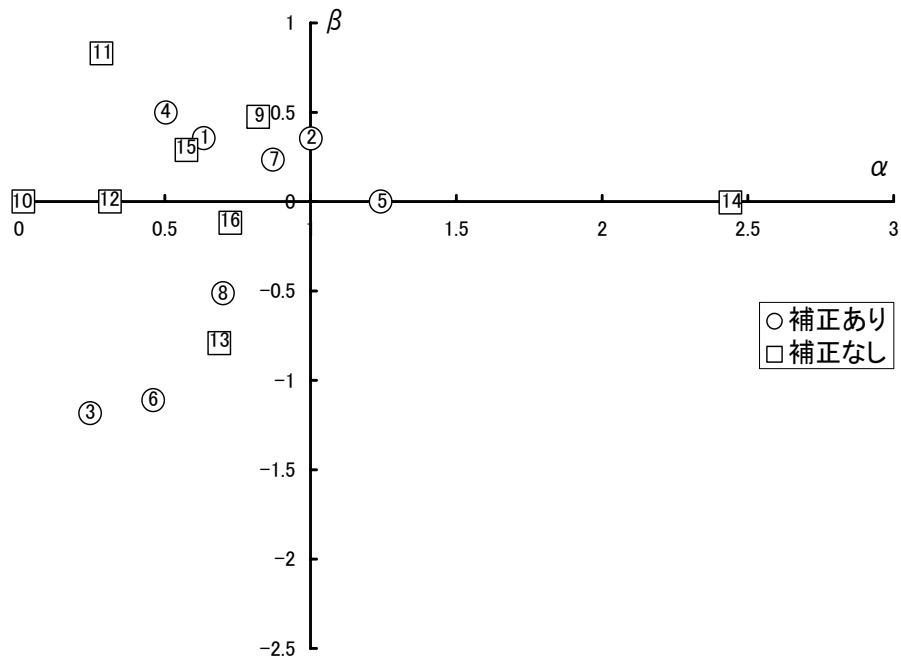


図 5.6 : 実践後の各学習者の評価特性パラメータの分布

図 5.5 から、初回の評価では、評価に差をつけることができず が低い学習者が多いことがわかる。図 5.6 から、最終回の評価では、補正表示群の学習者は が向上していることが読み取れる。次に、フィードバック画面で補正表示を行った学習者のうち、実践前にモデルへの当てはまりがよくなかった学習者について取り上げる。学習者 8 は、実践前の平均二乗誤差は 0.197 であったのが、実践後は 0.071 に向上した。実践前の は 0.012 であり、評価に差をつけずに、他の学習者と比べてかなり甘い評価をする傾向にあった。実践後の は 0.460 に上昇し、実践前よりも評価に差をつけることができるようになった。学習者 6 についても、学習者 8 と同様に評価に差をつけることができるようになったことで、平均二乗誤差は減少した。学習者 2 については、実践前と実践後の平均二乗誤差からはほとんど変化が見られなかったが、学習者 2 は、学習者 6,8 と同様に評価に差をつけるようになった。学習者 4 については、実践前後で平均二乗誤差も変化せず、パラメータの変化もほとんど見られなかった。この学習者は、実践後のアンケートで、パラメータの理解（ と それぞれに対する回答の平均）については、2（あまり理解できなかった）、自分の評価の特徴に気がついたかについては 3（どちらともいえない）と回答していた。このような学習者については、教師が介入してパラメータの意味など評価特性を理解させるなどの対処が必要である。

補正表示群、補正非表示群ともに、他人の評価を見せたが、教員の評価との相関係数、パラメータ、最小二乗誤差のいずれにおいても、補正表示群の方が補正非表示群よりも優れていた。適切な評価ができるようになるためには、単に他人が行った評価の結果を見せるだけでは十分でないことが示唆された。

実践後のアンケートのパラメータの理解、自分の評価の特徴への気づきと、実践前後での評価の変化について述べる。実践後のアンケートで、パラメータの理解について、4 以上と回答した者と 4 未満と回答した者に分けて評価の変化を調べた。4 以上と回答した学習者 2,5,6,8 は や平均二乗誤差の変化から 評価に差をつけるようになり、評価の傾向が他の評価者に近くなった。また、アンケートで自分の評価の特徴への気づきについても同時に 4 以上と回答していた。4 未満と回答した学習者 1,3,4,7 は、 や平均二乗誤差の変化から、それ以外の学習者と比較すると効果が低かった。パラメータの理解が 4 以上の者は、平均二乗誤差が平均 0.064 下がり、4 未満の者は 0.041 上がったことから、単に補正值を表示するだけでは不十分で、自分の評価特性を理解することが必要であることが示唆された。

5.3.5. 評価人数

1回当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，毎回何人分の課題を見てみたいか，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 5.7 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かについてのアンケートの結果を図 5.8 に示す。

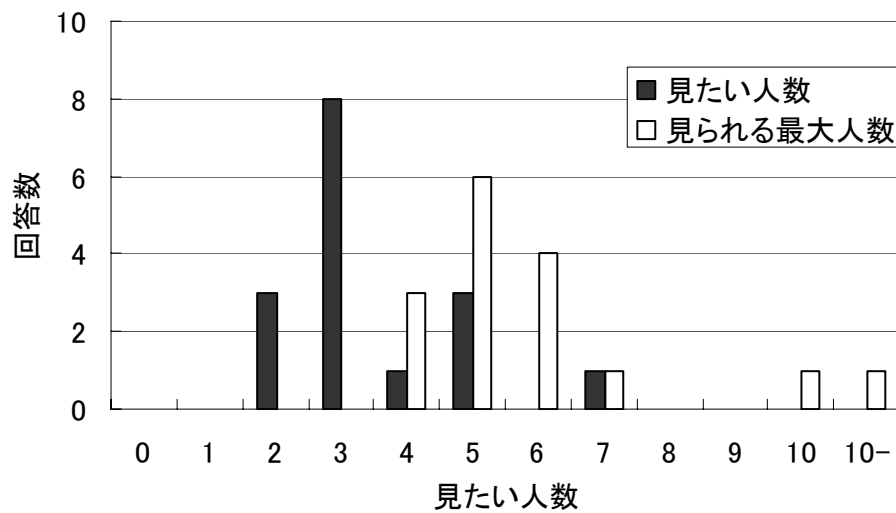


図 5.7：アンケートの結果（毎回評価したい人数）

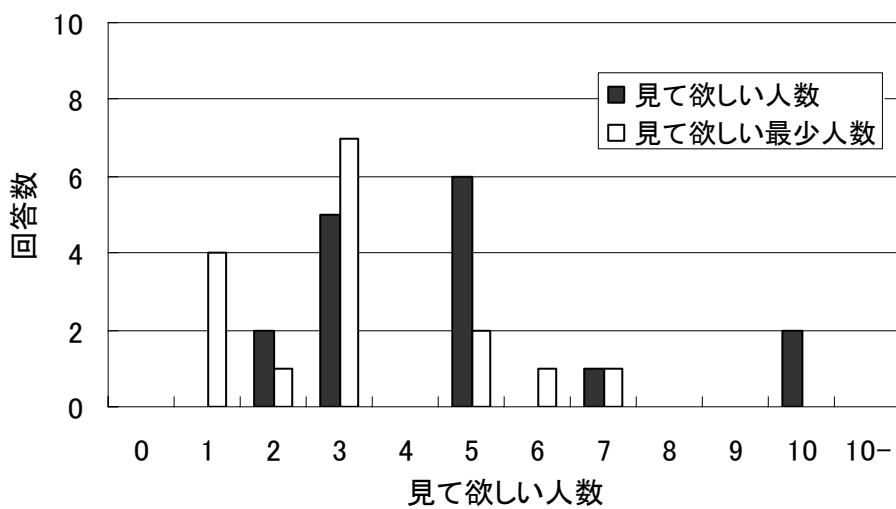


図 5.8：アンケートの結果（毎回評価してもらいたい人数）

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欲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、受ける評価は共に 6 人で行った。先述した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。

5.4. まとめ

形成的評価では、改善を目的とした評価が行われることが必要であるが、先述したとおり学習者の評価の能力が十分ではない可能性がある点について留意する必要がある。本章では、これら 2 つの視点から、システムを継続的に利用した相互評価を導入した実践について述べた。

連続 9 回行った授業の中で、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、課題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、他者の成果物の長所や短所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、他者の成果物の長所や短所を具体的に指摘できるようになり、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。

評価者特性モデルにおける評価特性パラメータ、平均二乗誤差を用いて、評価能力の変化について分析した。個人差補正を行った値を表示したグループの方が、そうでないグループと比較して、より適切な評価を行うことができるようになった。このことから、適切な評価を行うためには、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示唆された。

これまでに、相互評価の実践研究では、継続的に相互評価を行ったものは多くなかった。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。

第 6 章 結論

6.1 本研究のまとめ

本研究では、成果物の作成を通じて学習する場面において、学級等の学びの場を共有する学習コミュニティのメンバー同士が、お互いの成果物を相互に評価することによって、学習者に自身の成果物の改善点を気付かせるために学習者間相互評価を効果的に行う方法を提案した。特に、学習者がフィードバックを納得して受け入れるために、学習者に評価の方法や結果に対する公平性を担保した評価手法を提案した。

第1章では、最初に、近年盛んになってきている学習者間相互評価の背景と研究動向を概観した。次に、相互評価の問題点として、評価の公平性についてあげた。評価が公平に行われていると学習者が感じるかどうかは、他の学習者からの評価を受け入れることを前提とした相互評価の学習効果に影響する。しかしこれまで、相互評価における公平性を問題にした研究はほとんどなかった。公平性と関係して、すべての学習者が他のすべての学習者を評価できない場合に、起こりうる2つの問題を提起した。1つめの問題は、評価を行う学習者が、評価対象となる学習者から評価を受けるかどうかで、行う評価が変わってくる可能性があることである。2つめの問題は、様々な評価特性を持つ学習者からの評価結果をそのままフィードバックすることは、たまたま甘い評価者にあたったのか、厳しい評価者にあたったかによって、学習者間に不公平が生じることである。相互評価による学習効果をあげるためには、これらの問題点を解決し、学習者が他者評価をする能力を身に付けることが重要である。

第2章では、評価者を選択する必要がある場合に、公平な評価者の選択方法を考えるために、評価を行う学習者が、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、評価にどのような変化が見られるかについて述べた。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、評価点が甘くなる場合があること（お互い様効果）がわかった。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、また、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同等で、適切に評価行うことがわかった。

第3章では、評価者を選択する必要がある場合に、個々の評価者が持つ評価特性を考慮したフィードバックを行う手法について述べた。提案した手法では、項目応答理論のメタファを用い、評価特性を評点の厳しさと評点に差をつける度合いを表す2つのパラメータで表現した。推定した評価特性を用いて評価値を補正するアルゴリズムを提案した。次に、提案手法を、実データに適用し、補正が適切であることを確認し

た。提案手法は、評価者の評価特性にばらつきがある場合に有効である。評価者の評価特性から、適切に評価していない学習者を教師に通知して、指導を行うこと等ができ、より適切なフィードバックを行うことができる。これらの特徴は、特に形成的評価として、相互評価を行う際に重要である。

第4章では、公平なフィードバックを行うために、お互い様効果を除去する機能と、個々の評価者の評価特性に基づいた評価値の補正機能を持った相互評価支援システムの開発と評価について述べた。実際に、学習活動に相互評価を導入するためには、学習者、教員ともに負担を最小限にすることが求められる。本システムは、学習者向けの機能としては、電子ファイルでの課題提出機能、相互評価機能、結果表示機能から構成される。相互評価の際には、お互い様効果を考慮して、自動的に評価者を決定することができる。結果の表示では、個々の学習者の評価特性とそれに基づいた評価の補正值を表示することができる。システムのユーザビリティ、システムを使った相互評価の印象について、学習者による評価を行った。また、相互評価を導入した授業の経験がある教員に対しても、同様の評価を行った。学習者による評価から、相互評価について積極的に受け入れることが分かり、教員による評価から、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することが重要であることが分かった。学習者による評価、教員による評価ともに、本システムで容易に相互評価を実施することができ、その有効性が示唆された。また、システムを利用した実践では、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することが分かった。今回は、授業の最終回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が、形成的評価として行う場合は、適切な評価ができない学習者がさらに増えることも考えられる。形成的評価では、学習者に学習過程で評価の観点を理解させることが必要である。例えば、評価特性パラメータや最小二乗誤差等から、理解度が低い学習者を検出することで、このような学習者を見つけて、早期に教師が介入することができる。

第5章ではシステムを継続的に利用した相互評価を導入した実践について述べた。実践では、1つの課題に対して相互評価を5回行い、課題の改善と学習者の評価の能力の変化について分析した。相互評価を繰り返した結果、他者の成果物の長所や短所を具体的に指摘できるようになり、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示唆された。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。

今後、新しいが学力観に基づいた実践や e-ラーニングでの協調学習の場面で、学習者間相互評価はますます重要になると予想される。その際に、学習者が納得して受け入れられる相互評価を容易に行うことは、その協調学習の成否の大きな要因になると考えられる。本研究で得られた知見を生かすことにより、協調学習による学習効果を向上させることができる。

6.2. 議論

第3章で提案した補正方法は、評価者の特性を平均的な評価者の特性との差を用いて行っているため、補正後の評価は平均的な評価に近づく。平均的な評価特性から離れた評価を行う評価者がいた場合、その評価が合理的であったとしても、その評価者の補正後の評価は平均的な評価に近づく。合理的な根拠に基づいた評価であれば、評価者の評価方法を直す必要はない。教師が、平均二乗誤差や評価者特性パラメータでこのような評価を見つけ、授業中に取り上げることにより、学習コミュニティの評価能力の向上につなげることもできると考えられる。また、このような評価者に対してでも、他の評価者との評価の違いをフィードバック画面で表示し、個性的な評価をしていることを知らせることで、他の学習者の行った評価から学ぶ機会を得ることになり、有益である。

第3章で提案した評価者特性モデルを実データでの検証の結果、モデルへの当てはまりが悪い評価者も見られた。例えば、他の不適切な基準で評価をしたり、ランダムに評価をつけたり、全員に同じ評価をつけるなど不真面目に評価を行っていたり、評価者の中での評価基準が安定していないなどの理由が考えられる。これらの評価者は、平均二乗誤差の数値などを見ることで、容易に見つけることができる。該当する評価者全て、もしくは教員が不適切な評価であると判断した評価者の行った評価については、該当する評価者を評価特性パラメータの推定から外したり、該当する評価者の評価を被評価者の画面に表示しないようにしたりするなどフィードバックに悪影響を与えない処理を行うことも考えられる。該当する評価者への対処方法としては、システムが該当する評価者に対して、他の評価者の評価と離れていることを提示することや、教師がシステムを使ってこのような評価者を検出し、該当する評価者がどのような評

価を行っているかチェックした上で、個別に指導することが考えられる。他の評価者との評価の特徴の違いについてフィードバックすることで、次回以降の評価ではより適切な評価ができるようになると考えられる。

第4章で開発した相互評価支援システムの評価において、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。第1章で述べたように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、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、評価の観点で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が期待される。

第5章で取り上げた相互評価の実践では、分析のためにそれぞれの評価値をつけた根拠を書かせることにした。形成的評価のために行う相互評価では、根拠を書かせることは、適切なフィードバックを行うために、次の理由から特に重要である。

成果物作成者にとって、フィードバック時に示された評価値を補完して、学習の改善につなげることができる

評価者にとって、自身の行った評価を見直す時に、他人が行った評価の根拠が分かり、次回の評価をする際の参考にすることができる

教員にとって、評価者特性モデルの当てはまりが悪い者や個性的な評価を行う者の評価の根拠が分かり、介入して指導することが容易になる。

本論文では数値データの処理について扱ったが、記述式の回答と対にしてフィードバックすることで、学習効果が期待される。

6.3. 今後の課題

第2章でお互い様効果の影響、第3章で評価者の個人差補正アルゴリズムの効果に

ついて、情報処理入門科目を例に検証した。本論文で取り上げなかった芸術分野のように評価基準にある程度のばらつきが許容される領域等、様々な分野での検証が望まれる。また、評価項目・評価基準の影響、通信ネットワークを使った対面でない学習形態など様々な条件でも、同様の結論ができるか検証することが望まれる。

第4章の実践で学習者が他者に行った評価では、適切な評価をすることができない学習者も見られ、大学生でも他の学習者を評価する能力を育成することの必要性が明らかになった。第5章で行った実践で行ったフィードバックは、学習者の評価の特性を数値で表現したが、パラメータの意味が理解できない学習者もみられた。単に補正値を表示するだけではなく、評価特性の解釈をコメントで表示したり、視覚的に分かりやすいように2つのパラメータを x - y 平面に配置したグラフで表示することによって、学習者が自分の評価特性を理解できるようなインターフェースが有効かもしれない。また、前節で述べたように、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も必要である。例えば、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。

本研究では、相互評価について述べたが、評価を通じてより積極的なコミュニケーションをとることも有効であると考えられる。例えば、評価者に自分が受けた評価や改善案についてコメントを求めたり、自分が受けた評価が適切であったかを逆に評価することなど、学習者と評価者との間でコミュニケーションが取れる仕組みを取り入れることで、学習コミュニティ内で学習者同士が協力することで、お互いに学習効果を高めあうことができる。改善のためのコミュニケーションの手段としての相互評価の結果を材料として協調的な学習を促進させることも今後の課題としてあげられる。

参考文献

- Akahori, K., & Kim, S. M. (2003) "Peer evaluation using the Web and comparison of meta-cognition between experts and novices," Proceedings of World Conference on Educational Multimedia, Hypermedia, & Telecommunications (ED-MEDIA 2003), pp.1484-1487.
- Alfonseca, E., Carro, R., Martin, E., Ortigosa, A. & Paredes, P. (2006) "The impact of learning styles on student grouping for collaborative learning: a case study," User Modeling and User-Adapted Interaction, vol.16, no.3-4, pp.377-401.
- 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, pp.7-8.
- Arnold, L. (1981) "Use of Peer evaluation in the assessment of medical students," Journal of Medical Education, no.56, pp.35-42.
- 東洋, 梅本堯夫, 芝祐順, 梶田叡一 (1988) "現代教育評価事典," 金子書房, 東京.
- Bandura, A. (1971) "Social learning theory," General Learning Press, New York.
- Bloom, B. S. (1956) "Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain," Mc Kay, New York.
- Bloom, B. S., Hastings, J. T. & Madaus G. F. (1971) "Handbook on formative and summative evaluation of student learning," McGraw-Hill, Columbus.
- Brown, J. D., & Hudson, T. (1998) "The alternatives in language assessment," TESOL Quarterly, vol.32, pp.653-675.
- Bruner, J. S. (1966) "Toward a theory of instruction," Harvard University Press, Cambridge.
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989) "The four generations of computerized educational measurement," In R. L. Linn

- Ed., "Educational measurement," pp.367-407, Collier Macmillan, London.
- Carroll, J. B. (1963) "A model of school learning," Teachers College Record, no.64, pp.723-733.
- Chinn, D. (2005) "Understanding algorithms: Peer assessment in the algorithms course," Proceedings of 10th annual SIGCSE conference on Innovation and technology in computer science education, pp.69-73.
- Cousis, J. B. & Whitemore, E. (1998) "Framing participatory evaluation," New Directions for Evaluation, no.80, pp.5-23.
- Derntl, M. (2006) "A Generic View on Online Peer-Evaluation," Proceedings of World Conference on Educational Multimedia, Hypermedia, & Telecommunications (ED-MEDIA 2006), pp.1809-1816.
- Edelson D. C., O'Neill D. K., Gomez L. M., & D'Amico, L (1995) "A design for effective support of inquiry and collaboration," Proceedings of CSCL'95, pp. 107-111.
- Ellis, T. J., & Hafner, W. (2005) "Peer evaluations of collaborative learning experiences conveyed through an asynchronous learning network," Proceedings of 38th Annual Hawaii International Conference on System Sciences, p.5b.
- 遠藤貴広 (2007) "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.
- Falchikov, N. (1986) "Product comparisons and process benefits of peer group and self assessments," Assessment and Evaluation in Higher Education, vol.11, pp.146-166.
- Figl, K., Bauer, C., Mangler, J., & Motschnig, R. (2006) "Online versus face-to-face peer team reviews," Proceedings of 36th ASEE/IEEE Frontiers in Education Conference, pp.7-12(T1H).
- Fujihara, Y., Shibata S., Yoshikawa Y., & Nagaoka, K. (1994) "Development of multimedia test system," 2nd International Interactive

- Multimedia Symposium, pp. 147-151.
- 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, pp.65-70 .
- Fujihara Y., Ohnishi H., & 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 2006), pp.2816-2822.
- 藤原康宏, 大西仁, 加藤浩 (2007) “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.123-132.
- 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, vol.19, no.4, pp.206-211.
- Glaser, R. (1968) “Adapting the elementary school curriculum to individual performance,” Proceedings of the 1967 Invitational Conference on Testing Problems, Educational Testing Service, pp.3-36.
- Goodlad, J. I. & Anderson, R. H. (1959) “The nongraded elementary school,” Harcourt Brace, New York.
- Goodlad, S., & Hirst, B. (1989) “Peer tutoring: A guide to learning by teaching,” Kogan page, London.
- 橋本重治 (1971) “学習評価の研究,” 図書文化社, 東京.
- Hishina M., Okada R., & Suzuki K. (2005) “Group formation for Web-based collaborative learning with personality information,” vol.4, no.3, pp.351-364.
- 市川伸一 (1995) “学習と教育の心理学,” 岩波書店, 東京.
- 池田央 (2000) “アセスメント技術からみたテスト法の過去と未来,” 日本教育工学会論文誌, vol.24, no.1, pp.3-14.
- 稲葉晶子, 豊田順一 (1999) “CSCL の背景と研究動向,” 教育システム情報学会誌, vol.16, no.3, pp.166-175.

- 稲葉晶子, テップチャイサップニティ, 池田満, 溝口理一郎, 豊田 順一
(2000) “学習理論に基づく協調学習グループ構成のための学習目的オントロジー,” 電子情報通信学会論文誌(D-I), vol.83, no.6, pp.569-579.
- 鹿毛雅治 (2000) “学びの場で経験される評価”, 長尾彰夫, 浜田寿美男編, “教育評価を考える,” ミネルヴァ書房, 京都.
- 鹿毛雅治 (2004) “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.
- 加藤浩, 山下淳, 藤原康宏, 鈴木栄幸 (2006a) “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 30, pp.179-180.
- 加藤浩, 山下淳, 市丸俊亮, 鈴木栄幸 (2006b) “モバイル相互評価端末の設計思想と概要,” 日本教育工学会第 22 回全国大会講演論文集, pp.31-34.
- 梶田叡一, 渋谷憲一, 藤田恵璽訳 (1973) “教育評価法ハンドブック : 教科学習の形成的評価と総括的評価,” 第一法規出版, 東京.
(Bloom, B. S., Hastings, J. T. & Madaus G. F. (1971) “Handbook on formative and summative evaluation of student learning,” McGraw-Hill, Columbus.)
- 梶田叡一 (1993) “教育評価の現代的動向と内面への着目,” 学習評価研究, no.15, pp.28-35.
- 梶田叡一(1994a) “教育における評価の理論 - 学力観・評価観の転換 - ,” 金子書房, 東京.
- 梶田叡一 (1994b) “教育における評価の理論 II - 学校学習とブルーム理論 - ,” 金子書房, 東京.
- Kitagaki, I., Hikita, A., Takeya M. & Fujihara Y. (2007) “Development of An Algorithm for Groupware Modeling for A Collaborative E-learning,” International Journal of Computers, Communications and Control, vol.2, no.1, pp.66-73.
- 小島宏, 寺崎千秋編 (2001) “総合的な学習の評価計画と評価技法,” 明治図書, 東京.
- 国立教育政策研究所編 (2003) “総合的な学習の時間実践事例集,” ぎょうせい, 東京.

- Koschmann, T. (1996) "CSCL: Theory and practice," LEA Publishers, Florence.
- Liu J., & Sadler R.W. (2003) "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.
- Marquardt, D. W. (1963) "An algorithm for least-squares estimation of non-linear parameters," *Journal on Applied mathematics*, vol.11, no.2, pp.431-441 .
- 松本重男 (2000) "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," *日本教育工学会論文誌*, vol.24, no.1, pp.93-98.
- 持本志行 (1991) "現代の経営：戦略的経営管理," 中央経済社, 東京.
- Morrison, H. C. (1926) "The practice of teaching in the secondary school," University of Chicago Press, Chicago.
- 中原淳, 西森年寿, 杉本圭優, 浦嶋憲明, 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*, vol.24 no.Suppl., pp.97-102.
- 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 (2002) "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.
- 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.
- 生田目康子 (2004) "ピア・レビューをともなうグループ学習の評価 - 一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.
- 野嶋栄一郎 (2002) "学力観に伴って変わる教育測定観," 野嶋栄一郎編, "教育実践を記述する," 金子書房, 東京, pp.1-16.
- 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*, vol.17, no.3, pp.263-274.
- Okamoto, T., Seki, K., Kayama, M., & Cristea, A. I. (2002) "The integrated

- e-learning system – RAPSODY based on distance ecology model and its practice,” *Educational Technology and Society*, vol.5, no.4, pp.124-136.
- 大倉孝昭, 村尾敏彦 (2003) “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 C A L L 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.
- 大倉孝昭, 高村博正, 奥田アレックス・H (2004) “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 .
- 大塚雄作 (2002) “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.
- Orpen, C. (1982) “Student versus lecturer assessment of learning: a research note”, *Higher Education*, 11, pp.567-572.
- Press H. W., Teukolsky A. S., Vetterling T. W. & Flannery P. B.(1988) “Numerical recipes in C,” Cambridge University Press, Cambridge.
- (丹慶勝市, 奥村晴彦, 佐藤俊郎, 小林誠訳 (1993) “ニューメリカルレシピ・イン・シー,” 技術評論社, 東京.)
- Samejima F. (1969) “Estimation of Latent ability using a response pattern of graded scores,” *Psychometric Monograph*, 17, pp.1-100.
- Scriven, M. (1967) “The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation*,” American Educational Research Association monograph series on evaluation, 1, Rand McNally, Chicago.
- 芝祐順 (1991) “項目反応理論 基礎と応用,” 東京大学出版会, 東京.
- 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*, vol.25, no. Suppl., pp.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*, no.23, pp.31-36.
- Skinner, B. F. (1954) “The science of learning and the art of teaching,” *Harvard Educational Review*, vol.24, no.2, 86-97.
- Sitthiworachart, J., & Joy, M. (2003) “Web-Based Peer Assessment in

- Learning Computer Programming,” Proceedings of Third IEEE International Conference on Advanced Learning Technologies, pp.180-184.
- Sizer, T. (1984) “Horace's Compromise: The Dilemma of the American High School Boston”, Houghton Mifflin, Boston.
- Sluijsmans, D. M. A., Moerkerke, G., van Merriënboer, J. J. G., & Dochy, F. J. R. C. (2001) “Peer assessment in problem based learning,” Studies in Educational Evaluation, no.27, pp.153-173.
- Suppes, P. (1966) “The uses of computers in education,” Scientific American, vol.215, no.51, pp.206-221.
- 鈴木克明 (2002) “教材設計マニュアル - 独学を支援するために - ,” 北大路書房, 京都, pp.113-143.
- 高木正則, 田中充, 勅使河原可海 (2007) “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 .
- Trivers R. (1971) “The evolution of reciprocal altruism,” Quarterly Review of Biology, no.46, pp.35-57.
- 植野真臣 (2000) “ベイズ・アプローチによるグラフィカル・テスト理論,” 日本教育工学会論文誌,” vol.24, no.1, pp.35-52.
- Ueno M. (2004) “Data mining and text mining technologies for collaborative learning in LMS SAMURAI,” Proceedings of IEEE Conference (Computer Science), pp.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, no.44, pp.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, ET-2006-2, pp.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, no.28, pp.143-164.
- Wiggins, G. (1989) “A true test: Toward more authentic and equitable assessment,” Phi Delta Kappan, vol.70, pp.703-713.

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 (2007)
“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 - ,” 工学教育, vol.55, no.1, pp.41-46.

吉田宏史, 高橋岳之, 竹田尚彦 (2004) “自由記述を重視した成果物相互評価
システム,” 情報教育シンポジウム SSS2004, pp.159-162.

本論文に関する研究発表

学術論文

- 藤原康宏,大西仁,加藤浩 (2007) 学習者間の相互評価に関する研究の動向と課題,メディア教育研究,4(1),pp.77-85.〔第1章〕
- 藤原康宏,大西仁,加藤浩 (2007) 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.123-132.〔第2章,第4章〕
- 藤原康宏,大西仁,加藤浩 (2007) 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,31(3),印刷中.〔第3章〕

国際会議

- Fujihara Y., Ohnishi H., & 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 2006), pp.2816-2822.〔第5章〕

国内学会口頭発表

- 藤原康宏,加藤浩 (2005) 学習コミュニティメンバーによる相互評価のための評価者選択方法,電子情報通信学会技術研究報告,ET2004-118,pp.97-100.〔第2章〕
- 藤原康宏,大西仁,加藤浩 (2006) 相互評価のための課題提出評価支援システムの開発,電子情報通信学会技術研究報告,ET2005-94,pp.35-40.〔第4章〕
- 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,pp.65-70.〔第3章,第5章〕

謝辞

本研究を進めるにあたり終始あたたかいご指導と激励を賜りました総合研究大学院大学文化科学研究科メディア社会文化専攻 大西 仁准教授 ,加藤 浩教授に心より感謝と敬意の意を申し上げます .

学位論文審査の主査として ,多くの貴重なご助言を賜りました総合研究大学院大学文化科学研究科メディア社会文化専攻 黒須 正明教授に ,深くお礼申し上げます .

お互い様効果の調査のための実験ツールのプログラミングでは ,熊本大学大学院 井ノ上 憲次氏 (当時 ,岩手県立大学ソフトウェア情報学部)に ,相互評価支援システムの画面設計では ,元岩手県立大学ソフトウェア情報学部 佐藤 晶一氏に協力いただきましたことに ,感謝申し上げます . 学習者の課題の評価では ,ご多忙中にもかかわらず丁寧に採点をしていただいた岩手県立大学ソフトウェア情報学部 伊藤 久祥講師 , 児玉 英一朗講師 , 市川 尚助手に ,感謝申し上げます . また ,アンケート調査に快くご協力いただいた受講生の皆様及びシステムの評価に協力いただいた先生方にお礼申し上げます .

東京大学総合教育研究センター 望月 俊男客員准教授 , 青山大学ヒューマン・イノベーション研究センター 松田 岳士助手 , 東京工業大学教育工学開発センター赤堀研究室の学生の皆様には ,多くのご助言をいただきましたことに ,深く感謝いたします . 学会活動を通して ,ご意見・ご助言をいただいた多くの先生方に感謝申し上げます . また ,総合研究大学院大学文化科学研究科メディア社会文化専攻の皆様には大変お世話になりました . 深く感謝いたします .

熊本大学大学院社会文化科学研究科教授システム学専攻 鈴木 克明教授 (当時 ,岩手県立大学ソフトウェア情報学部)には日々の教育 , 研究に対して数々のご教示と激励いただきましたことに深く感謝いたします . 故沼田 俊昭先生 , 元岩手県立大学アドミッション・オフィス 大塚 剋佳教授 , 岩手県立大学社会福祉学部 佐々木 民夫教授 (当時 ,岩手県立大学アドミッション・オフィス長) , 岩手県立大学ソフトウェア情報学部 船生 豊教授 , 村山 優子教授には ,研究の面だけでなく ,日頃から様々な面でご指導とご助言をいただき ,深く感謝いたします .

神戸大学教育学部の学生時代から研究の楽しさと難しさを教えてくださり、また、学位論文審査をお引き受けいただきました電気通信大学大学院情報システム学研究科 植野 真臣准教授に深くお礼申し上げます。

早稲田大学人間科学学術院 永岡慶三教授には大学学部学生時代より学位論文審査に至るまで、永年にわたり多大なるご指導をいただいてまいりました。ここに深く感謝いたします。

最後に、これまで私をあたたく応援してくれた両親，妻 一恵，娘 百花に心からの謝意を記します。

平成 19 年 9 月