

ICT を活用した学生の主体的相互評価の研究

大橋 信行・佐野村 学・大熊 誠二

(要旨)

学習者間の相互評価には、高い教育効果が認められている。近年は、キャリア教育、授業レポートなど様々な事柄を対象とした取り組みがなされるようになってきた。

今回、インターネットを使った情報通信技術（以下、ICT）を利用した授業において、学生の相互評価をおこなった場合の問題点および妥当性について検討した。

本研究では、グループで作成した演技を YouTube 上で閲覧し相互評価をおこなった。学生間の相互評価を実施する場合には、お互いに評価をおこなうことで評価点が高くなる傾向があり、それを防ぐためには評価対象者と評価者を一致させないなどの対策が必要であることがわかった。また、YouTube 上で演技を閲覧し相互評価をおこなった場合、動画の再生率が低くなることがわかった。評価の正確性を向上させるには、動画の全編視聴の注意喚起と指導の必要性が示唆された。

アンケートによる自由記述をテキストマイニングにより解析した結果、作品の評価の高低や、その作品をどのようなイメージとして受け止めているかがわかりやすく図式化されることがわかった。

(キーワード)

レクリエーション、ICT、相互評価、テキストマイニング、動画視聴

1. はじめに

2000 年頃からおこなわれてきた学習者間の相互評価には、高い教育効果が認められるとされている。当初、相互評価の対象は、プレゼンテーションや成果物（作品、システム等）が中心であったが、近年は、キャリア教育、授業レポートなど様々な事柄を対象とした相互評価の取り組みがなされている。

石橋（2010）は、授業内で学生同士がレポートを相互評価した結果、レポート作成能力が向上したとし、腰山（2004）は、保育士養成課程の学生を対象に、紙芝居を取り入れ授業改善を目的とした相互評価を実施し、教師のみの評価よりも質の高い評価が可能となり、より有効な評価支援情報を提供すると報告している。

授業時の学習効果を向上させる取り組みとして、ビデオカメラを使い学習者を撮影し、

動画を閲覧させ、フィードバックをおこなうという試みがおこなわれてきた。古くは、外国語の演習授業で授業中の様子と学生のパフォーマンスをビデオカメラで撮影する手法(志柿 1997)が示されている。体育の授業では、授業中にビデオ撮影をおこない、映像の確認を行なうことで学生同士の学び合い、動きの認識、技能の向上、動きの改善が見られるという報告(岡本 2015)や、「跳び箱運動」で、動画コンテンツの有用性について検討した結果、自ら課題を明確にさせることで学習効果が上がるという報告(山本 2004)もある。

ここ数年、新型コロナウイルス感染症予防措置の一環により遠隔授業が進められる中、多くの高等教育機関で manaba や LMS を使った e ラーニング、Zoom や Google Classroom を使ったオンライン授業など、ICT を活用した授業方法が選択され実践されているが、レクリエーションの分野においてもオンラインによる学習活動の有効性が報告されており、講義科目に限らず実技・演習科目であってもその効果に期待できることがわかっている(大家 2021、中井ほか 2021)。

実技や演習科目としておこなわれることが多いレクリエーション授業ではあるが、オンラインによる学習活動の有効性が報告されていることから、今回、ICT を利用した授業において、動画を活用した学生の相互評価の問題点および妥当性について検討した。

2. 方法

2. 1. 対象学生

本研究の対象は、レクリエーション・インストラクター資格を取得することのできる演習授業である「レクリエーション指導演習」履修者を対象とした。授業は2コマ開講しており、各曜日の履修者数、授業に参加した対象者数、全てのアンケートに回答した学生数を有効回答数、グループ発表を行ったグループ数を表1に示した。2コマ合計の履修者数は128名であったが、授業に出席した学生は102名で、課題を全て提出し本研究の対象となった学生は96名であった。

表1 履修者数、対象者数、有効回答数、グループ数 (名)

	履修者	対象者	有効回答	グループ
月曜2限	56	41	40	10
水曜4限	72	61	56	15
合計	128	102	96	25

2. 2. 授業内容

「レクリエーション指導演習」では、グループ・ワーク・ゲーム、コンセンサス・ゲーム、ウォークラリー体験と作成、スタンツの作成と評価を授業内容としておこなっているが、本研究では、スタンツの作成と評価を調査対象とした。スタンツとは、野外キャンプでおこなわれることが多い「キャンプファイヤー」などでおこなわれる演技（出し物）のことで通常グループ単位でおこなわれる。授業では、野外キャンプで使えるスタンツを制作し発表することとし、制作と発表はグループでおこなわせた。学生には任意の4から8名でグループを作らせた。

2. 3. 授業の流れ

授業は、演技内容の検討・作成（小道具の作成含む）、動画撮影、動画の公開の順で進めた。その後、YouTubeに公開した動画URLの一覧を配布し、オンラインにて自身の所属グループの動画も含め、全ての動画を閲覧させた（図1）。

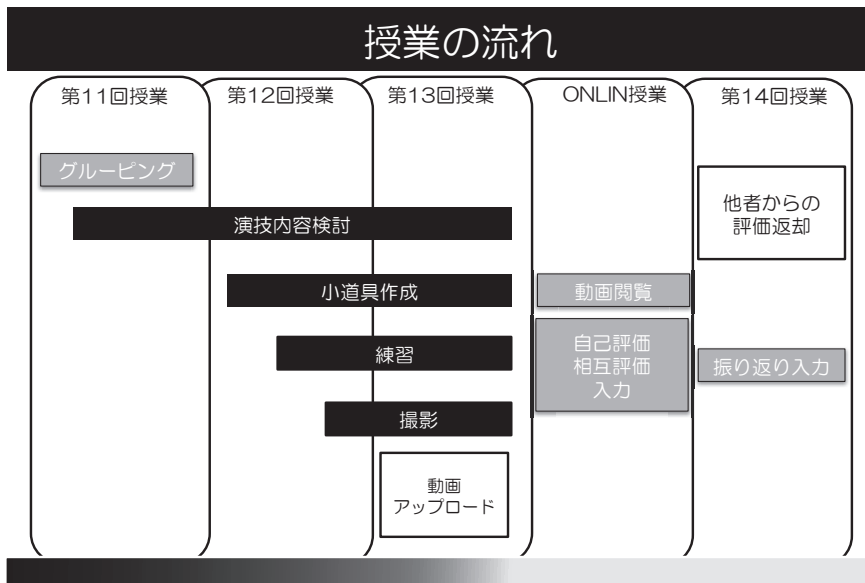


図1 授業の流れ

2. 4. アンケート

2. 4. 1. 種類

自身の所属グループ動画を閲覧後に「自己評価アンケート」を記入させ、自身の所属グループ以外の動画閲覧後は、「相互評価アンケート」を記入させた。その後、他グループからの評価を確認した後、振り返りとして「振り返りアンケート」を記入させた。アンケートは、全て Google Form を利用した。

2. 4. 2. 「自己評価アンケート」「相互評価アンケート」の内容と回答

設問は、「流れがスムーズだった」「理解しやすかった」「話し方について」「声の大きさについて」「話す早さについて」「演者の楽しさが伝わった」「内容がまとまっていた（完成されていた）」「演者が内容を理解していた」「板書／資料／小道具がうまく活用されていた」「総合評価」の10項目とした。今回の演技内容には制限を設けなかったため、内容は様々であった。話をしない演技や小道具を使わない演技もあったことから、全てのグループを公平に分析するため「総合評価」を本研究の評価項目とした。

回答は、「大変良かった」「良かった」「普通」「良くなかった」「全く良くなかった」の5点法とした。結果は、平均±標準偏差（以下、AVG ± SD）で示した。

2. 4. 3. 「振り返りアンケート」の内容と回答

設問は、「流れがスムーズだった」「理解しやすかった」「話し方について」「声の大きさについて」「話す早さについて」「演者の楽しさが伝わった」「内容がまとまっていた（完成されていた）」「演者が内容を理解していた」「板書／資料／小道具がうまく活用されていた」「総合評価」の10項目とした。「振り返りアンケート」も「自己評価アンケート」「相互評価アンケート」同様、「総合評価」を本研究の評価項目とした。

回答は、「全くその通り」「その通り」「そう言われればそうかもしれない」「納得できない」「全く納得できない」の5点法とした。結果は、AVG ± SD で示した。

2. 5. 解析方法

自由記述の解析は、株式会社ユーザーローカルのAIテキストマイニング (<https://keyword.userlocal.jp/>) を使用した。

テキストマイニングは、自由記述で書かれた文字列をデータ化する手法であり、出現頻度、共起頻度などを分析し図式化するものである。本研究では、文章中出现する単語の頻出度をスコア化する「ワードクラウド」、ポジティブな感情の文章とネガティブな感情の文章の存在比を示す「ポジネガ」、重要な文を抜粋して表示する「ダイジェスト」を使用した。「ダイジェスト」は、抜粋する行数を3行・5行・10行から選択できるが、本研究では10行とした。

3. 結果

3. 1. 自己評価アンケート、相互評価アンケート、振り返りアンケート

3. 1. 1. 各アンケートの平均得点

各アンケートの「総合評価」の得点の平均を示したものが表2である。

授業開始時のグループ数は、月曜2限授業が10、水曜4限授業が15であったが、アンケートを提出しない学生や記入漏れのあったものはデータから削除した結果、各授業で1

グループずつ減少し、最終的には月曜2限授業が9、水曜4限授業が14となった。

また、M3とW9はグループメンバーのうち複数名が提出しなかったことから、人数が1名となっている。全体の平均は、自己評価アンケートが 4.0 ± 0.4 、相互評価アンケートが 4.0 ± 0.7 、振り返りアンケートが 3.9 ± 0.5 であった。

本研究では、3つのアンケート全てにもれなく回答した学生のみを解析の対象とした結果、対象人数が1名になったグループや全員が対象外となってしまったグループもあったことから、グループ間の統計解析は行わず、 $AVG \pm SD$ で示した。

表2 各アンケートのグループ毎の得点（総合評価）

履修時間	グループ	人数	自己評価	相互評価	振り返り
月曜 2 限	M1	7	3.0 ± 0.6	3.6 ± 0.8	3.1 ± 0.4
	M2	5	3.8 ± 0.4	4.0 ± 0.8	3.8 ± 0.4
	M3	1	4.0	3.2 ± 0.8	3.0
	M4	5	4.8 ± 0.4	4.7 ± 0.5	4.8 ± 0.4
	M5	4	4.3 ± 0.5	4.4 ± 0.6	3.8 ± 0.5
	M6	3	4.0 ± 0.0	4.0 ± 0.6	4.3 ± 0.6
	M7	4	4.5 ± 0.6	4.1 ± 0.7	4.0 ± 0.0
	M8	3	4.7 ± 0.6	4.2 ± 0.7	4.0 ± 1.0
	M9	8	3.9 ± 0.6	3.8 ± 0.7	3.9 ± 0.8
水曜 4 限	W1	3	3.7 ± 0.6	3.9 ± 0.8	3.0 ± 1.0
	W2	3	3.3 ± 0.6	3.8 ± 0.9	3.7 ± 0.6
	W3	4	4.0 ± 0.0	4.5 ± 0.5	4.5 ± 0.6
	W4	4	3.5 ± 0.6	4.0 ± 0.4	4.0 ± 0.0
	W5	4	3.8 ± 0.5	3.7 ± 0.9	3.5 ± 1.0
	W6	4	4.0 ± 0.0	4.5 ± 0.6	4.0 ± 0.8
	W7	5	2.8 ± 0.4	3.4 ± 0.8	3.6 ± 0.9
	W8	4	4.0 ± 0.0	4.3 ± 0.8	4.0 ± 0.0
	W9	1	4.0	3.5 ± 0.9	4.0
	W10	3	3.7 ± 0.6	3.7 ± 0.8	3.3 ± 0.6
	W11	4	4.5 ± 0.6	4.5 ± 0.6	4.8 ± 0.5
	W12	6	4.3 ± 0.5	3.9 ± 0.6	4.0 ± 0.6
	W13	4	4.5 ± 0.6	4.2 ± 0.8	4.0 ± 0.0
	W14	5	4.4 ± 0.5	3.9 ± 0.7	4.4 ± 0.5
	avg		4.0 ± 0.4	4.0 ± 0.7	3.9 ± 0.5

(AVG±SD)

3. 1. 2. 各アンケートの関係性

「総合評価」の得点を自己評価アンケートと相互評価アンケートの得点によりグループ分けしたものが図2である。

自己評価アンケートと相互評価アンケートの得点がほぼ同値であったグループは実線、自己評価アンケートが相互評価アンケートよりも高値を示していたグループを破線、自己評価アンケートが相互評価アンケートよりも低値を示していたグループを二重線で示した。

自己評価アンケートと相互評価アンケートの得点がほぼ同値であったグループは、振り返りの得点もほぼ同値となる傾向であった。自己評価アンケートが相互評価アンケートよりも高値を示していたグループは、振り返りアンケートが自己評価アンケートよりも低値を示す傾向であった。自己評価アンケートが相互評価アンケートよりも低値を示していたグループは、振り返りアンケートが自己評価アンケートよりも高値またはほぼ同値を示す傾向であった。

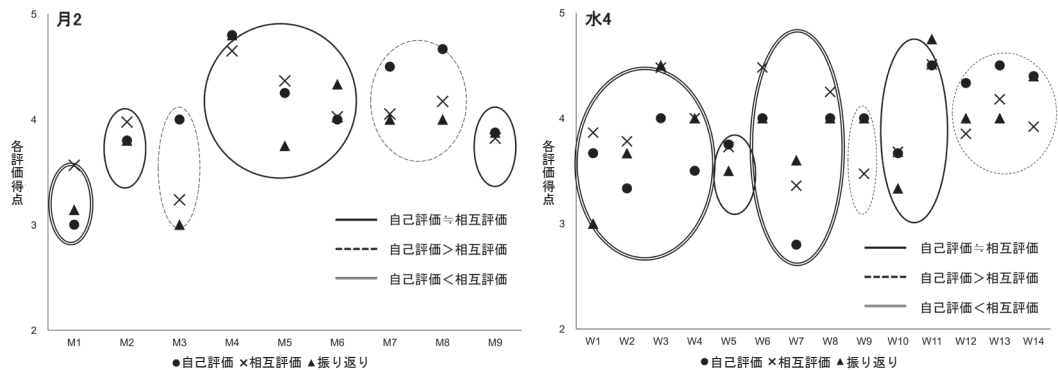


図2 自己評価、相互評価、振り返りの関係 (総合評価)

3. 2. 自由記述からの分析

3. 2. 1. 高評価グループの特徴 (ワードクラウド)

ワードクラウドは、単語の頻出スコアが高い単語を複数選び出し、その値に応じた大きさでイメージ化したものを図示するものである。各アンケートで高得点を示したグループW3のワードクラウドを図3に示した。中央に大きく表示された「チアリーディング」という文字と肯定的な単語が目立つ。

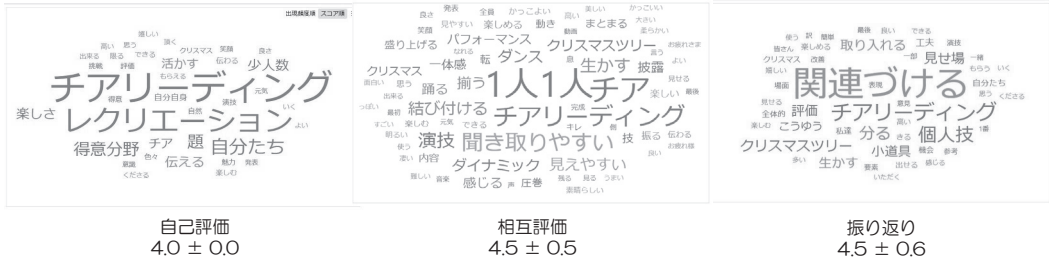


図3 高評価グループのワードクラウド (W3)

3. 2. 2. 高評価グループの特徴 (ポジネガ)

ポジネガは、ポジティブな感情とネガティブな感情の文章の存在比図示したものである。各アンケートで高得点を示したグループ W3 のポジネガを図4に示した。自己評価アンケートでは、ネガティブな記述は0.1%とほとんど見られなかった。相互評価アンケートでは11.4%のネガティブな記述がみられたが、振り返りアンケートでは0.1%とほとんど見られなかった。

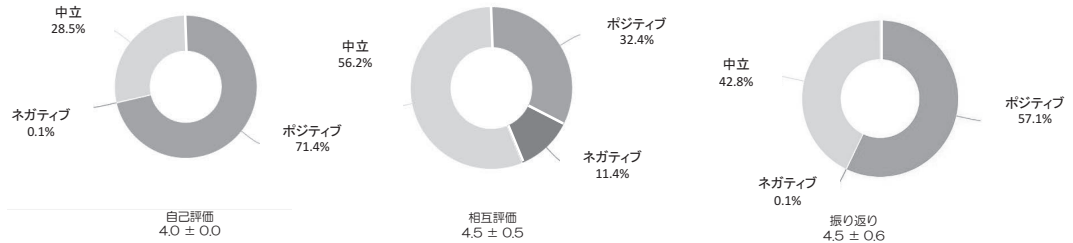


図4 高評価グループのポジネガ (W3)

3. 2. 3. 低評価グループの特徴 (ワードクラウド)

各アンケートで低得点を示したグループ M1 のワードクラウドを図5に示した。中央に大きく表示された「マジック」という文字と否定的な単語が見られる。



図5 低評価グループのワードクラウド (M1)

3. 2. 4. 低評価グループの特徴（ポジネガ）

各アンケートで低得点を示したグループ M1 のポジネガを図 4 に示した。自己評価アンケートでは、ポジティブな記述は 30.8%であったが、相互評価アンケートではポジティブな記述は 14.9%で、ネガティブな記述が 29.9%となった。振り返りアンケートではポジティブな記述は 0.1%とほとんど見られなかった。

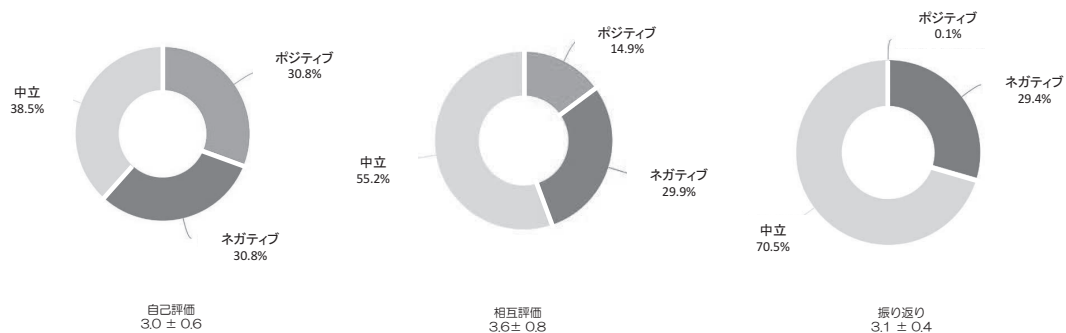


図 6 低評価グループのポジネガ (M1)

3. 3. 動画視聴時間

グループ毎の動画視聴回数、平均視聴時間、平均再生率を表したものが表 3 である。両時限とも視聴回数は、授業参加者数を超えてはいるが、動画時間の平均は全体で 3 分 51 秒であるのに対し、平均視聴時間は 1 分 51 秒であった。動画時間に対する平均視聴時間を平均再生数として求めたところ、全体では 50.5% であった。

表3 各グループの動画視聴回数、平均視聴時間、平均再生率

グループ	動画時間	視聴回数	平均視聴時間	平均再生率	グループ	動画時間	視聴回数	平均視聴時間	平均再生率
月曜2限 (学生数 40名)					水曜4限 (学生数 56名)				
M1	5'12"	51	2'10"	41.6%	W1	4'31"	66	1'57"	42.1%
M2	6'45"	48	2'18"	34.2%	W2	3'21"	62	1'42"	50.9%
M3	3'17"	59	1'24"	42.9%	W3	3'05"	84	1'55"	62.3%
M4	3'29"	54	1'59"	56.8%	W4	3'04"	66	1'22"	44.5%
M5	6'06"	65	2'06"	34.3%	W5	3'30"	74	1'39"	47.3%
M6	4'02"	58	1'51"	46.1%	W6	3'31"	63	1'46"	50.3%
M7	2'46"	59	1'35"	57.0%	W7	4'54"	69	2'03"	41.8%
M8	3'42"	54	1'27"	39.1%	W8	5'52"	71	2'51"	48.6%
M9	2'14"	55	1'16"	56.8%	W9	3'13"	66	1'35"	49.0%
M10	3'31"	52	1'33"	43.9%	W10	3'18"	68	2'08"	64.4%
avg	4'06"		1'46"	45.3%	W11	4'20"	62	2'15"	51.9%
全体平均 (学生数 96名)					W12	3'01"	81	1'33"	51.5%
	3'51"		1'51"	50.5%	W13	2'15"	68	1'45"	77.2%
					W14	2'31"	67	1'46"	70.3%
					W15	4'47"	67	2'42"	56.5%
					avg	3'40"		1'56"	53.9%

4. 考察

4. 1. 各アンケート

藤原ら（2007）は、学生に相互評価をさせる場合、互いに評価をする場合は平均点が高くなることを報告している。今回、評価する相手に自身も評価される条件で評価を行った結果、相互評価アンケートの平均は 4.0 ± 0.7 となった。5点満点中の4点と高得点を示していたことは、過去の報告と同様の傾向であったと考えられる。また、藤原ら（2007）は、評価する相手に自身も評価される場合よりも、評価する相手に自身は評価されない場合のほうが、教員の評価と比較した場合は、教員の評価に近く適正な評価になるとしていることから、今後、学生間で評価を行う場合には、次の発表者の評価のみ行うなど完全な相互評価にならないよう実施することで、より正確な評価が得られるものと推察された。また、鈴木ら（2017）は相互評価を行う際、質問紙はループリックを用いない5点満点絶対評価よりループリック評価表を用いた方が、より信頼度が高いとしていることから、今後の相互評価アンケートのループリック化を検討していくべきであると考えられた。

自己評価アンケートと相互評価アンケートの得点がほぼ同値であったグループは、振り返りの得点もほぼ同値となる傾向であり、自己の実力を正しく評価できていたと考えられる。しかし、自己評価アンケート結果と相互評価アンケート結果に乖離があったグループは、自己の作品を過大に評価したり、または極端に低く評価していた。いずれも自己の実

力を正しく認識できていなかったものと思われる。菊地（2017）は、学生間の相互評価をおこなった後、事後アンケートをおこない新しい気付きとして「他者からの多様な知識理解」「自己評価・認識」「思考力の幅の拡大」などが得られたとしているが、本研究では、新しい気付きがあったかという点までは分析をしておらず、今後の課題となった。

4. 2. ワードクラウド、ポジネガ

近年、キーワードの集計からブランド・イメージの可視化（豊田ら 2003）がおこなわれるようになった。テキストマイニングの手法は、ブランド連想の相違を際立たせるという目標には適しており、マーケティング分野では成果を上げている（黒岩 2005）。本研究においても、ワードクラウドによって得られたイメージが、それぞれの演技イメージを表していると考えて良いだろう。

高評価を得たグループの演技は、ワードクラウドからはネガティブな表現はほとんど見られない。実際、ポジネガでもネガティブな表現は 11.4% ととても少ない。一方で、低評価のグループは、ワードクラウドにもネガティブな文字が見て取れ、ポジネガからはネガティブな表現が 29.9% と 3 割近くある事がわかっている。

一般的に、制作された作品を評価させる場合には、点数法で評価するが多いが、今回の様に自由記述で評価しても、ワードクラウドやポジネガを活用することにより、演技をイメージ化し、好意的または批判的な意見が多いか少ないかについても数値で確認できる事がわかった。

4. 3. 動画視聴時間

相互評価として動画を YouTube で視聴する際、平均再生率が 50% 程度であった。YouTube が持つ機能に再生速度を可変させる機能がある。現在は 0.25 倍から 2 倍で再生することができるが、平均再生率は再生速度に影響を受けないので、2 倍速で視聴したとしても、平均再生率は低下しない。従って、平均再生率低下の原因は、動画を最初から最後まで視聴することなく、ある程度視聴した段階で評価を決めていたか、または、内容がわかるところのみ視聴した事が原因と考えられる。

NARA ら（2002）は、大学生を対象にアンケート調査をおこない、日常生活において道徳的に物事を考えられる人は、インターネット上でも倫理的な振る舞いができる。しかし、インターネット利用頻度が増すに従ってインターネット上の非道徳な行為と関連性が高くなるとしている。現在のネットワーク環境に順応することにより、情報機器の利用が増加し、適切な判断ができなくなっている（植田 2006）といわれている。日常生活においては道徳的ではあるが、近年のネット空間においてはまるで別の世界であるかのように、学生たちの倫理観は少しずつずれてきているものと考えられる。今回のような、動画を少しだけ見て評価をつけることも、特に問題と感じていないのかもしれない。佐渡島（2014）

は、論文やレポートを「コピペ」する学生には罪悪感はないとしており、今回の問題も、現代の学生たちを象徴するような行為だとも考えられる。

ICTを利用した授業では、オンライン授業の宿命でもある、学生の性善説に立った教育の限界についても今後検討する必要があるだろう。

5. まとめ

学生間の相互評価を実施する場合、お互いに評価をおこなわせると、評価点が高くなる傾向があるため、評価の正確性の向上を目指すには、評価対象者と評価者を一致させないなどの対応が必要であることが分かった。

アンケートの自由記述をテキストマイニングにより解析した結果、作品のイメージ化が可能となり、賞賛や批判がグラフ化できることから、わかりやすい評価が可能となった。

ICTを活用した相互評価をおこなうためにYouTubeにアップロードした動画視聴をおこなわせたが、動画の平均再生率が低く、動画全てを見て評価しているわけではないことがわかった。学生は、全ての動画を見ずに評価をしている可能性もあるため、動画の全編視聴の注意喚起と指導の必要性が示唆された。

謝辞

本研究は、帝京大学高等教育開発センター「SoTLプロジェクト2021」の補助金交付を受けて行いました。ここに厚く感謝いたします。

参考文献

- 石橋潔 (2010) 「レポート相互評価表—大学における授業実践の試み」『久留米大学文学部紀要情報社会学科編』第5号,17-33
- 植田義幸 (2006) 「道徳教育としての情報モラル教育の可能性 (2) - 共分散構造分析 (SEM) による行動志向性と情報行動の関連構造分析 -」『四天王寺国際仏教大学紀要』(43),111-127
- 佐渡島紗織 (2014) 「アカデミック・ライティング教育と情報リテラシー: 《情報を再定義》し意見を構築できる学生を育てる」『情報の科学と技術』64巻1号,22-28
- 大家千枝子 (2015) 「コロナ禍における学生ボランティアによるレクリエーション支援: オンライン会議システムと事前収録動画を用いたオンライン音楽交流会の実践」『自由時間研究』46巻1号,64-69
- 岡本敦、青山有理、田口由香 (2015) 「保健体育科教育法 (体操・器械運動) における IPAD の活用」『東海学園大学教育研究紀要』第1号,3-12
- 菊地達夫 (2017) 「学生相互評価を取り入れた授業方法の工夫と効果: 保育内容環境の授業を通じて」『北翔大学短期大学部研究紀要』55巻,35-47
- 黒岩祥太 (2005) 「ブランド・イメージと消費者接点の関連についてのテキストマイニング」、『マーケティングジャーナル』25巻1号,38-50

- 腰山豊（2004）「短大保育科における実践的指導力の形成と授業改善（7）紙芝居の文化論的検討と保育利用」『聖園学園短期大学研究紀要』 34,1-14
- 志柿光浩（1997）「第9章 ビデオ記録による教授・学習活動のフィードバックの実際」『放送教育開発センター研究報告』 103,119-127
- 鈴木伸子、石川奈保子、向後千春（2017）「大学院のオンライン授業におけるレポート相互評価の実践 - ルーブリック活用が評価の信頼性・妥当性におよぼす効果の検討 -」『コンピュータ & エデュケーション / CIEC 会誌創刊準備号編集委員会 編』 43,43-48
- 豊田裕貴・森永聡（2003）「企業におけるマーケティング分野でのテキスト活用事例：ブランド・イメージ調査へのテキストマイニング技術の適用」『情報処理』 44(10),1028-1031
- 中井聖（2015）、田中真紀、吉中康子、足立名津美「ウェブ会議システムを利用したオンライン・アイスブレイキング・ゲームのレクリエーション活動としての有用性と実施上の要点」『自由時間研究』 46 巻 1 号 ,21-30
- 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』 31,125-134
- 深田昭三、中村純、岡部成玄、布施泉、上原哲太郎、村田育也、山田恒夫、辰己丈夫、中西通雄、多川孝央、山之上卓（2013）「大学生の情報倫理にかかわる判断と行動」『日本教育工学会論文誌』 37 巻 2 号 ,97-105
- 山本朋弘、池田幸彦、清水康敬（2004）「体育「跳び箱運動」指導における動画コンテンツ活用の効果」『日本教育工学雑誌』 27 巻 ,153-156
- Yumiko NARA and Minako YOSHII (2002) : Internet Information Ethics and Ordinary Morals : An Empirical Study of University Students, J. Home Econ. Jpn. Vol.53(12)1167-1175.