



Title	日本の高等教育における相互評価に関する研究の動向と課題
Author(s)	沙, 華哲; Sha, Huazhe; 杉浦, 真由美 他
Citation	高等教育ジャーナル : 高等教育と生涯学習, 31, 1-12
Issue Date	2024-04
DOI	https://doi.org/10.14943/J.HighEdu.31.1
Doc URL	https://hdl.handle.net/2115/91705
Type	departmental bulletin paper
File Information	HighEdu.31_p01-12.pdf



Trends and Issues in Research on Peer Assessment in Japanese Higher Education

Huazhe Sha^{1)*}, Mayumi Sugiura²⁾ and Katsusuke Shigeta³⁾

1) Graduate School of Science, Hokkaido University

2) Institute for the Advancement of Graduate Education, Hokkaido University

3) Information Initiative Center, Hokkaido University

日本の高等教育における相互評価に関する研究の動向と課題

沙 華哲^{1)**}, 杉浦 真由美²⁾, 重田 勝介³⁾

1) 北海道大学大学院理学院

2) 北海道大学大学院教育推進機構

3) 北海道大学情報基盤センター

Abstract — This study examines the effects of and challenges in using peer assessment as formative and summative evaluations in Japanese higher education and proposes methods and possibilities for using peer assessment in digital educational settings. For this purpose, a review of research from 2013 to 2023, on peer assessment in Japanese higher education was conducted. The results showed that there were 59 studies on peer assessment with comments, 29 on peer assessment with scoring, and 42 on both. Furthermore, the studies implementing peer assessment were categorized into 3 types such as “studies focusing on the mechanism of peer assessment,” “studies verifying the learning effects of peer assessment,” and “studies on the integration of peer assessment and ICT.” Going forward, there will be a demand for classroom practices focused on supporting educators in implementing peer assessments, improving the quality of comments in formative peer assessments, and enhancing the convenience and reliability of peer assessments using ICT tools.

(Accepted on 13 February 2024)

*) Correspondence: Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan
E-mail: huazhe.sha.y9@elms.hokudai.ac.jp

**) 連絡先：060-0810 札幌市北区北 10 条西 8 丁目 北海道大学大学院理学院

1. はじめに

1.1 高等教育における教育評価

教育評価を定期的に行うことは、学生の能力向上や学習の持続性に効果がある (Barrio *et al.*, 2015)。また、教育評価の効果は、「単に個人個人の学力を測定することだけではなく、設定された教育目標に沿って学習者に対して学習の指導あるいは支援を行うため、教授者への情報提供をするという重要な役割を果たす」ことも示されている (永岡・赤堀, 1997)。このような定義に示したように、教育評価は異なる目的と実施方法によって、「学力を測定する」、「学習者に対して指導する」、「教授者への情報提供」といった効果がある。Scriven (1967) はカリキュラムの改善のための教育評価として、「形成的評価」と「総括的評価」を提唱した。その後、Bloom (1971) によって教育評価はカリキュラムや教授、学習改善のための評価として発展し、「診断的評価」が追加された。Bloom は教授活動における評価を考える際に、目的、時期および学習観に基づいて、「授業の開始時に生徒を適切に位置づけることと、授業の展開にあたって、生徒の学習上の難点の原因を発見することに用いる診断的評価」、「学習の途中に行われ、カリキュラム作成、教授、学習の3つの過程に用いる形成的評価」、および「1つの学期やコースのプログラムの終わりに、成績づけや認定、進歩の評価、カリキュラムや教育計画の有効性の検討に用いる総括的評価」に分けている。欧米では、Bloom による従来の形成的評価と総括的評価の概念区分が明確さを欠くと考えられ、教育評価の機能は、目的によって区別されるべきという考え方になっていく中、形成的評価の目的は「学習や指導改善」、総括的評価の目的は「資格や選抜、あるいはアカウンタビリティ」という論調が広がった (西岡他, 2015)。近年、教育評価に関する研究の多くは、形成的評価と総括的評価に焦点が当てられている (Harlen & James, 1997; Garrison & Ehringhaus, 2007; Dixon & Worrell, 2016)。例えば Harlen & James (1997) によると、形成的評価の有効性と実用性が重要であり、ほかの評価方法と比べ、学生は積極的に授業活動に参加する必要がある一方、総括的評価は、成績報告や学生の進捗状況を確認

する時点で行うもので、通常は科学的な評価基準が必要であるという形成的評価と総括的評価を実施する際の特徴が示されていた。これらのことから、教育評価の効果と課題は、評価の種類と実施方法によって異なることが明らかであり、教育評価の効果と課題を検討する際には、評価の種類と方法に基づいて検討する必要がある。

1.2 相互評価とは

1.2.1 相互評価の定義と実施方法

従来の教育評価の定義では、教員が評価者であり、学習者が被評価者であると捉えられる一方で (永岡・赤堀, 1997)、評価に参加する評価関係者には、教師や学校管理者だけでなく、学習者も評価の行為主体として考える必要がある (McTighe & Ferrara, 1998; Sadler, 1989)。「教育評価を評価の主体で分類すると、教員による評価と学習者による評価に分けられる。そして、相互評価は学習者による評価の一種と言える」と、相互評価が教育評価の一部であることを示している (藤原他, 2007)。

相互評価は、「学級集団において級友を評価させる方法」であり (東他, 1988)、「個人が同等の地位を持つ仲間の学習成果物や学習結果の量、レベル、一般的な価値、個人から見た価値、質、および成功を考える仕組み」と定義されている (Topping, 1998)。また、日本の高等教育において、相互評価の意味を含めた外来語として、「ピアレビュー」、「ピアアセスメント」、「ピアレスポンス」などがある。

これらの定義においては、相互評価の目的を述べているものの、具体的に授業でどのように相互評価を実施するのかは示されていない。Falchikov (2001) では、相互評価の実施方法を、「関連する基準を用いて、学生同士の作品やパフォーマンスをアセスメントする」と述べているが、「アセスメント (assessment)」という表現は、具体的にどんな行動に指すのかは示されていない。Topping (1998) によると、相互評価は「採点や評定」がある一方で、「オープン式の評価やフィードバック」もあると示している。このことから、相互評価には2種類の実施方法があると言えよう。それに対し、Liu & Carless (2006) は、相互評価の実施方法について検討し、「詳細なコメ

ントはあるが、正式な採点はないピアフィードバック」と、「正式な採点がある（コメントの有無を問わず）ピアアセスメント」の2種類に分けた。これらのことから、相互評価の実施方法は「採点の有無」と「コメントの有無」に着目していることがわかる。近年、相互評価に関連する研究で、「ピアレビュー」の場合はコメントがあり、「ピアグレーディング」の場合は採点があるという特徴が示されている（Sluijsmans & Prins, 2006；Vickerman, 2009；岩田, 2020；Gamage, 2021）。以上のことから、相互評価の実施方法は、「採点の有無」と「コメントの有無」によって区別され、「採点はあるが、コメントはない相互評価」（以下「採点のみの相互評価」）、「コメントはあるが、採点はない相互評価」（以下「コメントのみの相互評価」）、「両方ともある相互評価」と「両方ともない相互評価」の4種類に類型化される。また、本研究における「採点」とは、Liu & Carless (2006) を参考にし、具体的な成績付けを含み、評価基準に基づいて評定し、測定し、ランク付けする行為を指す。

さらに、相互評価は実施方法の違いによって、効果と課題も異なることが示されている。例えば、採点を中心とする相互評価では、成績の測定としては教員の負担を減らすことはできるが、学生による採点の信頼性が欠く課題があり（藤原他, 2007）、このような採点によって全面的に学生のスキルを測るとともに、学生の学習成果にも効果を与えることが求められる（Sadler & Good, 2006）。また、コメントを中心とする相互評価では、より早くかつより多くのフィードバックがもらえることが示され（Gibbs, 1999）、採点より学習に促進できると示唆されている（Liu & Carless, 2006）。これらのことから、相互評価の効果と課題を検討する際には、実施方法に基づいて検討する必要がある。

1.2.2 相互評価に関するレビューの動向

2000年以降、教育の情報化の進展により、デジタルツールを用いた相互評価が増加し、様々なツールに対して相互評価の有用性と課題をレビューする研究がある（Luxton-Reilly, 2009；Panadero & Alqassab, 2019）。レビューにおいて、デジタルツールの効果が報告されている一方で、多くのツールは普及していないことや、匿名で評価を行う困難さが課題とし

て挙げられている。これらのレビュー論文では、相互評価に影響を与える具体的な要素、例えばデジタルツールと匿名性に焦点を当ててレビューしているものの、高等教育における相互評価の実践を中心としたレビューは乏しい。さらに、相互評価に関するレビューは、主に海外の文脈に焦点が当てられており、近年の日本において相互評価に関する研究を整理したものは見当たらない。とりわけ、ここ10年は、モバイルネットワークを含めた教育の情報化の推進により、教室またはオンラインの学習環境における相互評価の役割が変わりつつあることから、近年の研究に関してレビューする必要がある。

さらに、相互評価の効果と課題は種類と実施方法の違いによって異なり、特に、相互評価の実施方法は「採点の有無」と「コメントの有無」に着目した整理が重要と考えられる。これらの点に着目してレビューすることにより、相互評価の現状と課題を整理できるとともに、高等教育における相互評価活動の改善に寄与できる可能性がある。

2. 方法

本研究では、スコوپングレビュー法のための報告ガイドラインである PRISMA Extension for Scoping Reviews (Tricco *et al.*, 2018) とその日本語訳（友利他, 2020）に沿って実施した。このプロセスでは、「①研究疑問を特定する；②重要な研究を特定する；③研究を選ぶ；④データを抽出する；⑤結果の収集、要約、報告；⑥結論」という6つのステップがある。

本研究では、日本の高等教育における「採点のみの相互評価」、「コメントのみの相互評価」、「両方ともある相互評価」、および「両方ともない相互評価」についての研究において、相互評価の活用に関してどのような効果と課題が示されているかを研究疑問とする。

高等教育における相互評価に関連する具体的な研究を整理するために、2013年から2023年までの10年間、CiNii Research（以下「CiNii」）に検索される論文をレビューする。

まず、CiNiiにおいて、この10年間、「大学」また

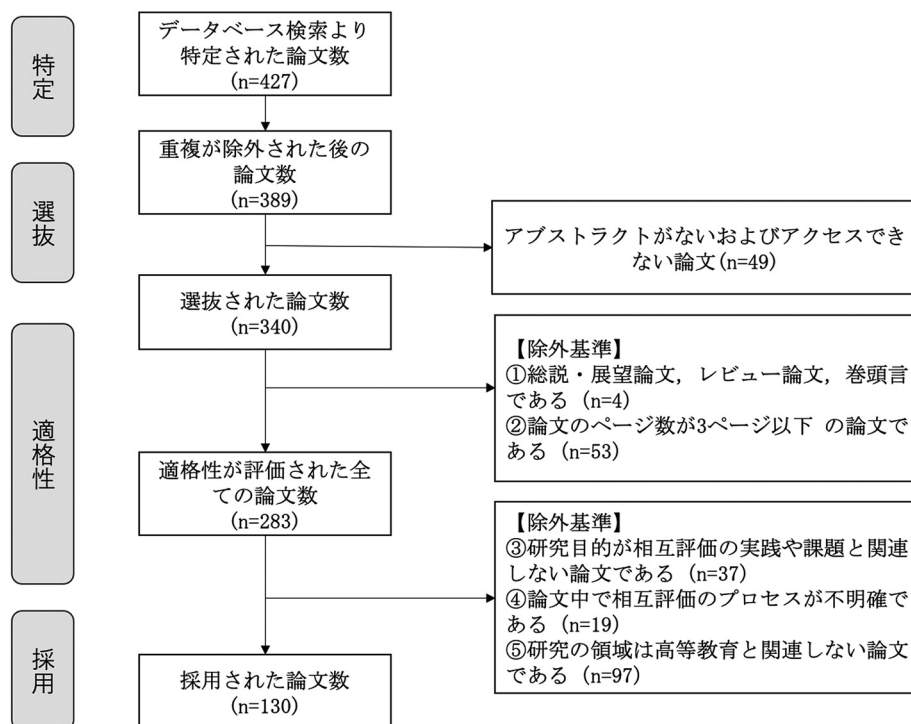


図1 文献選択のフローチャート

は「高等教育」を含み、かつ「相互評価」、「ピアレビュー」、「ピアレスポンス」または「ピアグレーディング」を含む論文を検索した。検索式は、「(大学 OR 高等教育) AND (相互評価 OR ピアレビュー OR ピアアセスメント OR ピアレスポンス OR ピアグレーディング)」であり、授業形態は問わなかった。その結果、合計 427 編の論文が検索された (2023 年 9 月 20 日時点)。本研究では、従来の PRISMA フローチャートに踏まえて、一次選別では、タイトルとアブストラクトにおいて、重複登録されているもの (該当 38 件) と、アクセスできないおよびアブストラクトのないもの (該当 49 件) を削除した結果、340 編の論文が抽出された。文献選択のフローチャートは図 1 で示す。

本論の目的に合致する論文を抽出するにあたり以下①～⑤の除外基準を設定した結果、210 編を除外した。

- ① 総説・展望論文、レビュー論文、巻頭言である (該当 4 編)
- ② 論文のページ数が 3 ページ以下の論文である (該当 53 編)
- ③ 研究目的が相互評価の実践や課題と関連しない論文である (該当 37 編)

- ④ 論文中で相互評価のプロセスが不明確である (該当 19 編)
- ⑤ 研究の領域は高等教育と関連しない論文である (該当 97 編)

以上の結果、本研究の分析対象として、130 編の論文が抽出された。筆者と相互評価の研究を行っている共著者 2 名が、抽出した 130 編の論文に対して、①著者；②発行年；③国、地域；④目的；⑤対象；⑥キーワード；⑦相互評価のプロセス；⑧研究結果という 8 つの項目を抽出した。

3. 結果

抽出した 130 編の論文の研究内容をレビューし、「コメントのみの相互評価」に関する研究は 59 編、「採点のみの相互評価」に関する研究は 29 編、「両方ともある相互評価」に関する研究は 42 編、「両方ともない相互評価」に関する研究は 0 編があった。これらの研究が対象とする授業形態には、対面授業、オンライン授業、およびブレンド型授業が含まれていたが、MOOC 等の一般向けオンライン講座に該当する授業形態は含まれていなかった。

さらに、日本の高等教育において相互評価を実施する研究を、研究内容に基づいて3つのタイプに分類した。

タイプ1は「相互評価のメカニズムに関する研究」であり、ほかの評価方法との比較、相互評価の方法の開発と改善などを検討している研究が含まれる。

タイプ2は「相互評価による学習効果を検証する研究」であり、実際の教育現場で相互評価を実施し、相互評価による学習効果を検討している研究が含まれる。

タイプ3は「相互評価とICTの融合に関する研究」であり、ICTを利用して相互評価を行う、あるいは新たな相互評価機能を含めたシステムを開発している研究が含まれる。

3.1 コメントのみの相互評価に関する研究

コメントのみの相互評価に関する研究は、合計で59編であった。その内訳は、タイプ1が8編、タイプ2が29編、タイプ3が22編であった。

タイプ1の相互評価のメカニズムに関する研究には、中西(2013)、重田(2016)などがある。中西(2013)は、大学の授業レポートの評価方法を改善するため、「二回提出」方法を導入し、学生同士がピアレビューを通してレポートの質を向上させることを目的とした。その結果、二回提出を行うことで、レポートの質は向上し、また、他者のレポートを見ることで、自分のレポートの欠点に気が付くことができると報告された。一方で、提出のチャンスが二回あることで、学生が手を抜くという問題も少数発生した。また、重田(2016)は、情報科学実習の授業において、学生のプレゼンテーションスキル向上を目的とし、ピアレビューを活用して相互評価基準を明確化する取り組みを行い、コメントをKJ法によって分類した。その結果、「発表内容に関する」「発表態度に関する」「スライドの出来に関する」「質疑応答・その他」の4つのカテゴリーにまとめた上で、学生間の相互評価で高評価を得るためには、発表内容の精査やスライドのデザイン性が必要であることが示された。なお、コメントのみの相互評価において、評価でのコメントの質に対する定義を検討する研究が見当たらなかった。コメントはフィードバックとし、

次の改善に活かせるために、どのようなコメントが求められるのかを検討する必要がある。

タイプ2の相互評価による学習効果を検証する研究には、河野他(2014)、澁井(2023)、藤田(2017)、松本・山本(2018)などがある。河野他(2014)は、エンジニアリングデザインを学ぶ授業において、プロジェクトベースの学習を通じてエンジニアリングデザインの基本的プロセスを学ぶことを目的とし、学生間の相互評価と相互添削を通じて、プレゼンテーションスキルとレポート作成能力を向上させるアプローチの効果を検証した。また、澁井(2023)は、教員養成課程の学生を対象に、小学校「外国語の指導法」の授業において、振り返りを充実させ、学生のメタ認知を高め、自己評価と模擬授業における相互評価を行うことで、学生の学習への動機付けを促すという4ステップの振り返りサイクルを用いて、学生の「気づき」を言語化し、自己の学びをコントロールすることを目的とした。その結果、学生の学習意識と学習の主体性が向上したことが明らかとなった。教員養成課程において、相互評価を導入し、学習効果への促進を検証する研究は、他にも5編あった。例えば藤田(2017)と松本・山本(2018)は、音楽科の教員養成授業において、相互評価を導入し、相互評価でのフィードバックはその後自分の取り組みに効果があることを示した。

タイプ3の相互評価とICTの融合に関する研究には、倉田他(2019)、米満・古賀(2023)、岩田(2022)、荻田(2023)などがある。倉田他(2019)は、非対面での教育・学習を対象に、携帯端末向けの動画アノテーションシステムを開発し、マーカープリセット機能、D&Dアノテーション機能およびマーカー重畳再生機能の3つの機能を実験に通じて検証し、それぞれが評価内容整理支援、指摘伝達支援および振り返り支援での効果が示された。相互評価機能を含めたシステム開発に関する研究は、他にも7編あった。(例えば、山下他, 2013; 浅井他, 2015; 森本, 2019; 間瀬他, 2020; 加藤他, 2005)。また、新たにシステムを開発するのではなく、既存の通信設備とICTツールを用いて、相互評価を活用する研究もあった。例えば、米満・古賀(2023)は、2020年コロナ禍の影響に対応するため、対面授業をオンライン授業に再設計し、動画配信による講義とリアルタ

イムの Web 会議システムを活用したグループワークを導入し、さらに 2022 年はオンライン提出したレポートを相互評価するピアレビューを導入した結果、学生の自己省察を深め、多様な視点から事象を捉える能力を育成できる可能性が示された。相互評価に用いる ICT ツールとして、ZOOM や Moodle を用いた相互評価に関する実践報告もあった(例えば、岩田, 2022; 荻田, 2023)。

3.2 採点のみの相互評価に関する研究

採点のみの相互評価に関する研究は、合計で 29 編であった。その内訳は、タイプ 1 が 10 編、タイプ 2 が 13 編、タイプ 3 が 6 編であった。

タイプ 1 の相互評価のメカニズムに関する研究には、岩間 (2023), 宇都・植野 (2018), グエン他 (2018), 堀越・田村 (2018) などがある。岩間 (2023) は、評価者特性を考慮した項目反応モデルに対して、モデル母数の推定精度から評価数を効果的に削減できる可能性を探求するため、大学の授業において相互評価を取り入れ、教育評価のための情報を得る取り組みを行った。その結果、評価数を抑えても、課題の特性や評価者の特性、能力に関する有益な情報が得られ、教育改善に役立つことが示された。項目反応理論に関する研究は、他にも 2 編あった。宇都・植野 (2018) は評価者特性パラメータを付与した新たな項目反応モデルを提案した。また、グエン他 (2018) では、グループ構成の仕方に焦点を当て、グループ内だけでなく、各学習者に対して少数のグループ外評価者を割り当てる外部評価者選択手法を提案した。その他、相互評価の信頼性と妥当性に関する研究もあった。堀越・田村 (2018) は相互評価における評価プロセスデータと評価行動の履歴を収集し、分析した。その結果、学生同士の相互評価は学習効果を高める可能性がある一方で、相互評価の信頼性や妥当性に関する課題が依然として存在することが示された。

タイプ 2 の相互評価による学習効果を検証する研究には、中村他 (2016) などがある。中村他 (2016) は情報特別講義において、アクティブ・ラーニング手法を取り入れ、学習の振り返ることで知識の定着を促すことを目的とし、学生が課題を自らのものと

して意識しながら相互評価の活動に取り組んだ。その結果、提示資料の見栄えなどの項目では学生と教師の間に有意差が見られない一方で、内容の理解に関連する項目では学生と教師の間に有意差があることが明らかとなり、相互評価を行う際に、ループリックの明示が必要であることが示された。また、このことから、学生同士がループリックを用いて評価することが自己学習能力を高める可能性がある一方、内容の理解に関わる項目に対して教員の指導や助言が必要であろう。そのほか、外国語教育や医学教育において、相互評価を取り入れて効果を検証する研究があった(例えば、水谷他, 2019; 鈴木・越智, 2016; 武末, 2019)。

タイプ 3 の相互評価と ICT の融合に関する研究には、植野 (2014), 桑原他 (2014), 高橋他 (2013) などがある。植野 (2014) は、多様な熟達レベルの学習者が参加する学習コミュニティにおいて、知識やスキルの発達を共通目標とし、「学び方」を学ぶことができ、学習成果を共有できることを目標とし、学習コミュニティの特性を支援できる e ポートフォリオ推薦システムを開発した。このシステムでは、類似の学習プロセスを持つ学習者を推薦することができるその結果、このシステムは他者からの学ぶための有用なツールと報告されている。また、既存の ICT ツールを用いて相互評価の効果を検討している研究もある。例えば、桑原他 (2014) では、初年次キャリア教育科目において Moodle を用いた相互評価を実践し、受講生の進路選択自己効力の変容を調査した。その結果、授業終了から約 2ヶ月後、授業に受講したことが受講生の進路選択自己効力に影響を与えていることが明らかになった。一方、高橋他 (2013) では、教材企画書を作成するオンライン授業で、チェックリストに基づく学生間の相互評価を行った。その結果、相互評価によって、教員の添削負担が軽減することが示唆されたものの、教材制作の初心者にとって、目標設定と評価には困難があることも示されている。これらのことから、採点のみの相互評価がオンライン授業で活用する際に、学生の主体性が鍛える一方で、学習者の特性に応じた教員サポートも求められている。

3.3 両方ともある相互評価に関連する研究

コメントと採点両方ともある相互評価に関連する研究は、合計で 42 編であった。その内訳は、タイプ 1 が 5 編、タイプ 2 が 15 編、タイプ 3 が 22 編であった。

タイプ 1 の相互評価のメカニズムに関する研究には、杉本 (2020)、岩田・田口 (2020) などがある。杉本 (2020) は、キャリア教育科目「秘書実務演習」授業において、人前での電話応対実践において学生同士の相互評価方法を探求すること目的とし、評価基準を明らかにしつつ、記名式ピアレビューと無記名式ピアレビューを実施し、教員の評価に近い評価結果を導き出す方法を調査した。その結果、無記名式ピアレビューがより公正な評価を促すことが示された。その他、相互評価の手法に関する研究について、岩田・田口 (2020) はパフォーマンス課題において、自己評価とピアからの評価の得点やその根拠が異なった箇所について、異なった理由を協議するワークを取り入れた相互評価活動を開発し、他者の視点を取り入れて自己評価を見直す機会を提供することで、自己評価力を高める可能性を検証することを目的とする。その結果、協議ワークを取り入れた相互評価活動は自己評価力の定着に効果がある可能性を示した。

タイプ 2 の相互評価による学習効果を検証する研究には、杉本 (2017)、小久保 (2017)、出雲・久村 (2023) などがある。杉本 (2017) は、ビジネスマナー教育において、基本技能到達度テストとグループワークによる相互評価を組み合わせた教育プログラムを提案した。その結果、学習者が自己評価と相互評価を通してキャリア形成に向けた改善点を把握し、教員には今後の指導に役立つ情報を得る可能性が示された。また、小久保 (2017) は、保育士養成教育において、相互評価と視覚教材を組み合わせた実習記録の指導を行った結果、学生が実習の記録を他者と評価し合うことによって、自分の課題を明確にするのに役立つことが示された。さらに、出雲・久村 (2023) では、教員が介入せず、学生同士の相互評価が制作活動の主体とする効果を検討した。その結果、学生がグループでの共同制作意識が向上し、アカデミック・ライティングの知識もよく身に付け

た。これらの研究では、教員の支援がない協同学習プログラムの効果が検証された。しかし、実際の授業では、学生の協同学習意識や、グループ内の各メンバーの能力、知識理解の度合いなどが、最終制作物に影響を与える。学生が主体とする協同学習活動を行う際に、事前学習支援の整備と教員が学生の知識把握状況を知ることが求められる。

タイプ 3 の相互評価と ICT の融合に関する研究には、植木・大西 (2016)、岸 (2018)、本田 (2017) などがある。植木・大西 (2016) は、相互評価の結果を可視化にするシステムを開発し、学習者に気づきを促し、プレゼンテーション能力の向上を目指した実験を行った。その結果、このシステムを用いて相互評価を繰り返し行うことが、学習能力の向上に寄与することが示された。また、岸 (2018) の研究では、学生が制作した動画を YouTube で互いに評価し、事前に作成したルーブリックに基づいて採点し、コメントすることを行なった。その結果、学生の評価能力の向上に寄与したことが示唆された。一方、多くの作品を評価する際に、入力が煩雑になることも課題として指摘された。本田 (2017) では、講義型授業に LMS を用いることでオンライン協働学習の実現が可能になり、自己決定理論を活用したモチベーション管理には有効であることを示した。既存の ICT ツールを活用する研究から、LMS、YouTube を通じて、相互評価の形態および方法が豊富になったことが明らかになった。

4. 結論

本稿の目的は、過去 10 年間の高等教育における相互評価に関連する研究のレビューを行い、コメントのみの相互評価および採点のみの相互評価の効果と課題を分析することであった。この目的を達成するために、はじめに、抽出された論文を、コメントのみの相互評価と採点のみの相互評価に分類した。次に、分類した研究を「相互評価のメカニズムに関する研究」、「相互評価による学習効果を検証する研究」、「相互評価と ICT の融合に関する研究」の 3 つのタイプに分類し、相互評価の効果と課題を検討した。

4.1 相互評価の効果と課題

コメントのみの相互評価に関しては、レポートの質の向上、プレゼンテーションスキルの育成に有効であることが多くの研究で検証されていた（中西 2013；河野他 2014）。また、ICT ツールを組み合わせた相互評価は、学生間のコミュニケーションと評価の質を向上させるために役立つことが示されていた（倉田, 2019）。学生は相互評価によってコメントとアドバイスを受け取ることによって学習効果とスキルが向上し、教員の負担が軽減する一方、評価プロセスが学生の負担になる可能性があり、一部の学生は相互評価に対して抵抗感があることも示唆されている。

これらのことから、コメントのみの相互評価を行う際に、教員が学生の異なる理解度とスキルへの対応、および一部学生の抵抗感への適切な支援、さらにコメント質の向上への工夫が必要であると言える。また、国外でも相互評価への苦手意識を持っている学生が存在するという指摘があり、3 割以上の学生がコメントのみの相互評価に対して中立的な態度を持っていることが示されている（Vickerman, 2009）。学生同士で相互評価を行う際に、各学生が知識に対する理解度や、相手に意見を表明するスキルが異なるため、教員の適切な支援が必要となる。例えば、学生が相互評価のグループで受けたコメントの質が低い場合、教員は適切な指導や意見を提供するだけでなく、学生が相互評価で見逃している点を発見し、同時に問題を発見する能力を向上させることが考えられる。

採点のみの相互評価に関しては、学生同士による採点が学習の定着や理解の促進に効果的であることが報告されていた（中村, 2016）。また、一部の研究では相互評価の結果の信頼性が検証されている一方で、採点のみの相互評価の採点と教員評価の採点との間には依然として差異が存在することも指摘されている（堀越・田村, 2018）。

これらのことから、相互評価での採点の結果の信頼性が不十分であることが伺える。また、国外においても、相互評価での採点が最終成績に関連する場合、その信頼性と妥当性がより不明瞭になることが指摘されている（Sridharan *et al.*, 2019）。

採点のみの相互評価において採点の信頼性を高める方法として、学生同士による採点および教員による採点の差を控えるには、ループリックといった評価基準の質を高める必要があると言える。また、評価の公平性を高めるために、藤原他（2007）が指摘した通り、評価者の選択に工夫することで、「お互い様効果」を控える必要がある。さらに、ICT の活用とシステム開発の観点において、評価者の匿名性を重視し、評価回数を増やすことで、採点のみの相互評価の信頼性を高めることもできると示されている（Bostock, 2000）。これらのことから、採点のみの相互評価を行う際に、使用する道具やデジタルツールの選択、および活動全体の流れの設計を考慮することが、採点のみの相互評価の信頼性を高めることに寄与できると言える。

4.2 本研究の限界

本研究の限界として、以下の 2 点が挙げられる。第一に、本研究で対象とした論文は過去 10 年間の高等教育における相互評価に関連する研究のレビューに限られており、初等中等教育や生涯学習など、他の教育分野における相互評価の可能性と課題を示していない。加えて、本研究で対象とした論文は日本語で本文またはアブストラクトが記述されたものに留まる。国内外における相互評価に関する研究の動向を横断的に分析することで、日本における相互評価の特徴と課題を明らかにすることが今後の課題である。

文献

- 浅井 宗海・中井 秀樹・佐藤 修 (2015). ジェネリックスキル育成のための大学間ピアアセスメントを取り入れた学習システム構築に向けて（＜特集＞情報教育の新展開） 日本情報経営学会誌, 36 (2), 3-11.
- 東 洋・梅本 堯夫・芝 祐順・梶田 叡一 (1988). 現代教育評価事典 (p.377) 金子書房
- Barrio, M. I. P., Escamilla, A. C., García, M. N. G., Fernández, E. M., & de la Rosa García, P. (2015).

- Influence of assessment in the teaching-learning process in the higher education. *Procedia-Social and Behavioral Sciences*, 176, 458–465.
- Bloom, B. S. (1971). *Handbook on formative and summative evaluation of student learning*. (B.S.ブルーム 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)
- Bostock, S. (2000). Student peer assessment. *Learning Technology*, 5, 245–249.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55, 153–159.
- Falchikov, N. (2001) *Learning together: peer tutoring in higher education* (London, Routledge Falmer).
- 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), 77–85.
- 藤田 光子 (2017). 音楽科模擬授業における学生相互の評価活動について—指導法の学習の質を考える— 別府大学短期大学部紀要, 36, 47–57.
- Gamage, D., Staubitz, T., & Whiting, M. (2021). Peer assessment in MOOCs: *Systematic literature review*. *Distance Education*, 42, 268–289.
- Garrison, C., & Ehringhaus, M. (2007). *Formative and summative assessments in the classroom*.
- Gibbs, G. (1999). Using assessment strategically to change the way students. *Assessment matters in higher education*, 41, 41–53.
- グエン ティエン・宇都 雅輝・植野 真臣 (2018). ピアアセスメントにおける項目反応理論を用いたグループ構成最適化 電子情報通信学会論文誌 D 情報・システム, J101-D (2), 431–445.
- Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, 4, 365–379.
- 本田 直也 (2017). 講義型授業での授業外オンライン協働学習の試み コンピュータ&エデュケーション, 43, 24–29.
- 堀越 泉・田村 恭久 (2018). 相互評価における評価プロセスデータを用いた評価行動分析 日本行動計量学会大会抄録集, 46, 196–199.
- 岩間 徳兼 (2023). 評価数を減らした相互評価に基づく教育評価 日本教育工学会論文誌, 47 (2), 237–248.
- 岩田 員典 (2022). Moodle ワークショップを利用した学生による相互評価の実践 Com=コム: 愛知大学情報メディアセンター紀要, 31 (1), 37–50.
- 岩田 貴帆 (2020). 協議ワークを取り入れたピアレビューによる学生の自己評価力向上の効果検証 大学教育学会誌, 42 (1), 115–124.
- 岩田 貴帆・田口 真奈 (2020). パフォーマンス課題における自己評価力を高めるための協議ワークを取り入れた相互評価活動の開発 日本教育工学会論文誌, 43 (Suppl.), 173–176.
- 出雲 俊江・久村 希望 (2023). 必修科目における相互評価を中心とした コラボラティブ・ライティングの実践—金城一紀『映画篇』を題材として— 広島女学院大学人文学部紀要, 4, 15–29.
- 加藤 浩・波多野 和彦・高比良 美詠子・坂元 昂・森本 均 (2005). 教育ナレッジの蓄積・流通・再利用に関する教員の意識調査 日本教育工学会論文誌, 28 (suppl.), 41–44.
- 岸 誠一 (2018). YouTube を活用した相互評価システムの試行について 中国学園紀要, 17, 135–140.
- 小久保 裕美 (2017). 視覚教材と学生の相互評価を用いた実習記録指導の効果 東海学園大学教育研究紀要, 2 (1), 63–68.
- 河野 昭彦・斉藤 博嗣・佐々木 大輔・平澤 一樹・須田 達・鶴谷 奈津子 (2014). 学生の相互評価によるアクティブラーニング型授業 工学教育, 62 (6), 62–67.
- 倉田 伸・藤木 卓・室田 真男 (2019). 視覚的な動画アノテーションを用いた相互評価による協調学習を小型携帯端末上で支援するシステムの開発とその機能の検証 日本教育工学会論文誌, 43 (1), 31–41.
- 桑原 千幸・喜多 敏博・合田 美子・根本 淳子・鈴木 克明 (2014). 初年次キャリア教育科目における相互評価学習の実践と進路選択自己効力の

- 向上 日本教育工学会論文誌, 38 (2), 79-89.
- Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher education*, 11, 279-290.
- Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education*, 19, 209-232.
- 間瀬 皓介・丸山 浩平・宮寺 庸造・森本 康彦(2020). 社会ネットワーク分析を用いた相互評価活動支援システムの開発と評価 教育システム情報学会誌, 37(2), 143-148.
- 松本 哲平・山本 双葉(2018). 音楽表現科目における相互評価活動—M-GTAによる学修プロセスの分析より— 駒沢女子短期大学研究紀要, 51, 33-42.
- McTighe, J., & Ferrara, S. (1998). *Assessing Learning in the Classroom. Student Assessment Series*. NEA Professional Library, Distribution Center, PO Box 2035, Annapolis Junction, MD 20701-2035.
- 水谷 幸嗣・則武 加奈子・鶴田 潤・關 奈央子・近藤 圭子・片桐 さやか・竹内 康雄・秋月 達也・塩山 秀裕・青木 章・和泉 雄一・岩田 隆紀・荒川 真一・荒木 孝二(2019). 診療参加型臨床実習への多職種連携の新規導入による臨床的効果および教育的効果 日本歯周病学会会誌, 61 (3), 148-156.
- 森本 康彦(2019). 社会ネットワーク分析を用いて相互評価を学習状況に応じて支援するシステムの開発と評価 JSiSE Research Report, 33 (7), 63-68.
- 永岡 慶三・赤堀 侃司(1997). 「教育評価」の研究動向 日本教育工学雑誌, 20 (4), 199-206.
- 中村 佐里・遠藤 敏喜・波多野 和彦(2016). 特別講義のための情報倫理教育の開発 生活大学研究, 2 (1), 64-70.
- 中西 裕(2013). ピアレビューを活用した授業レポートの「二回提出」方式の効果—アクティブラーニングの試みとして— 就実論叢, 43, 247-261.
- 西岡 加名恵・石井 英真・田中 耕治(2015). 新しい教育評価入門—人を育てる評価のために— (pp.52-54) 有斐閣
- 荻田 朋子(2023). ロールプレイ学習におけるアンテーションソフトを活用した相互評価の実践的研究 関西学院大学高等教育研究, 13, 79-90.
- Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & evaluation in higher education*, 44, 1253-1278.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18, 119-144.
- Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11, 1-31.
- Scriven, M. (1967). The Methodology of Evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation Chicago* (pp. 39-83). Rand McNally.
- 澁井 とし子(2023). 教員養成課程での自己調整学習の試み—振り返りによる自己評価と模擬授業による相互評価を通して— 東京福祉大学・大学院紀要, 13 (1-2), 57-66.
- 重田 崇之(2016). 情報リテラシー教育におけるプレゼンテーションのピアレビュー分析—医療福祉系短期大学における事例— 川崎医学会誌一般教養篇, 42, 33-42.
- Sluijsmans, D., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, 32, 6-22.
- Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education*, 77, 853-870.
- 杉本 あゆみ(2017). キャリア教育科目における効果的なフィードバック教育および相互評価についての考察 千葉経済大学短期大学部研究紀要, 13, 37-50.
- 杉本 亜由美(2020). キャリア教育科目におけるピア・レビューに関する調査報告 *International Journal of Human Culture Studies*, 30, 21-27.
- 鈴木 光代・越智 希美子(2016). 英語での論理展開

- スキルの向上を目指して—映像英語より Complaint Speech Act を分析— 映画英語教育研究：紀要, 21, 73.
- 高橋 暁子・根本 淳子・鈴木 克明 (2013). 教材設計基礎科目における『教材企画書チェックリスト』に基づいた受講者相互評価の分析 日本教育工学会研究報告集, 13 (3), 53-60.
- 武末 祐子 (2019). L2 としてのフランス語によるプレゼンテーション活動から考察するリスニング・スピーキングスキル 西南学院大学フランス語フランス文学論集, 62, 67-92.
- 友利 幸之介・澤田 辰徳・大野勘太・高橋香代子・沖田勇帆 (2020). スコーピングレビューのための報告ガイドライン日本語版：PRISMA-ScR 日本臨床作業療法研究, 7 (1), 70-76.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, 169, 467-473.
- 植木 勇太・大西 圭 (2016). 自己組織化マップを用いた相互評価可視化システムのプレゼンテーションスキル育成への応用 日本知能情報ファジィ学会 ファジィ システム シンポジウム講演論文集, 32, 407-412.
- 植野 真臣 (2014). 過去の学習者履歴データを利用した e ポートフォリオ・システム 情報知識学会誌, 24 (4), 414-423.
- 宇都 雅輝・植野 真臣 (2018). ピアアセスメントにおける異質評価者に頑健な項目反応理論 電子情報通信学会論文誌 D 情報・システム, J101-D (1), 211-224.
- Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & evaluation in higher education*, 34, 221-230.
- 山下 淳・加藤 浩・葛岡 英明・鈴木 栄幸 (2013). 顔画像認識によるユーザ選択機能を実装した相互評価システムの実装と評価 (学習支援システムの新展開/一般) 電子情報通信学会技術研究報告, ET, 教育工学, 113 (229), 81-86.
- 米満 潔・古賀 崇朗 (2023). オンライン授業でのグループワーク後の自己省察へのオンラインピアレビューの導入と評価 佐賀大学全学教育機構紀要, 11, 53-65.

