

# 相互評価プロセスログを用いた「評価にかけた時間」の分析 Analysis of “Evaluation Time” in Peer Assessment based on Evaluation Log

堀越 泉<sup>\*1</sup>, 田村 恭久<sup>\*2</sup>

Izumi HORIKOSHI<sup>\*1</sup>, Yasuhisa TAMURA<sup>\*2</sup>

<sup>\*1</sup>上智大学大学院 理工学研究科

<sup>\*1</sup>Graduate School of Science and Technology, Sophia University

<sup>\*2</sup>上智大学 理工学部

<sup>\*2</sup>Faculty of Science and Technology, Sophia University

Email: izumihorikoshi@eagle.sophia.ac.jp

あらまし：本稿では、相互評価中のプロセスログを取得し、1回の授業において連続評価を実施した時の評価行動に与える影響を分析した。連続評価の実施により評価に対する動機付けに影響を与える可能性があるという仮説のもと、「評価にかけた時間」について分析を行った。結果として、連続評価により「評価にかけた時間」が単純に短くなることはないこと、評価者によって評価にかける時間が長い場合や短い場合があることが明らかになった。

キーワード：相互評価, 評価行動, Learning analytics

## 1. はじめに

近年のアクティブ・ラーニング型の授業の普及に伴い、学習者にプレゼンテーションをさせ、学習者同士で相互評価させるという形式の授業が増えている。しかし、この学習者による相互評価は信頼性・妥当性・公平性などの問題が懸念されている<sup>(1)</sup>。

多くの従来研究では基本的に相互評価の評点に着目して議論されることが多い。例えば、信頼性は学習者同士の評点の一致度、妥当性は教員による評点と学習者による評点の一致度で議論されている<sup>(2)</sup>。これに対し本研究では、相互評価のプロセスデータを取得し、各評価項目の評価タイミングに着目して分析を行うことを提案し、相互評価の質保証に対する有効性を検討してきた。

相互評価の質の議論においては、学習者の詳細な評価行動を分析した研究はほとんど存在しない。一方で、社会調査の分野においては、Web調査の普及に伴い、アンケートの回答にかけた時間や中断といった回答行動の履歴を取得し、調査の回答の質の検証に用いる研究が始まっている<sup>(3)</sup>。そして、例えば、「短時間に回答する行動」は、調査が継続していくに伴って正確に答える動機が欠如していくことの表れである<sup>(4)</sup>、という知見が蓄積している。

本稿では、ある1回の授業において連続した相互評価を実施した時の評価行動に与える影響を分析する。前述の通り、アンケートの回答行動の研究では、アンケート調査の継続に伴い、回答者の動機づけの低下が観察された。このため、相互評価においても連続評価の実施により、評価に対する動機付けに影響を与える可能性があるかと推測されるためである。

## 2. 方法

### 2.1 開発

相互評価中の各評価項目の評点と評価タイミングを取得するため、相互評価用ツールを開発した。こ

れはオンラインフォームであり、評価項目のリストと評点に対応したラジオボタン（1点–5点）、送信ボタンを備える。評価者は送信ボタンを押すまで何度でも評価を変更可能であり、評点の選択、変更、送信ボタンの押下をトリガーとして、履歴がサーバに送られる。取得した主なログ項目は、評価日時、評価者学生番号（記入者）、被評価者学生番号（発表者）、評価項目番号、評点である。

### 2.2 実験

上記のツールを用いて実験を実施し、相互評価の履歴を取得した。被験者は上智大学開講の全学共通科目「情報リテラシー（情報検索）」受講者のうち、被験者として同意を得た者とした。実験対象回では、プレゼンテーション作成の演習を行い、その発表に対し相互評価を課した。プレゼンテーションは発表10分、質疑応答4分からなり、グループA～Fの計6グループが発表を行なった。実験実施日は2017年7月10日、被験者は72名であった。

## 3. 結果と考察

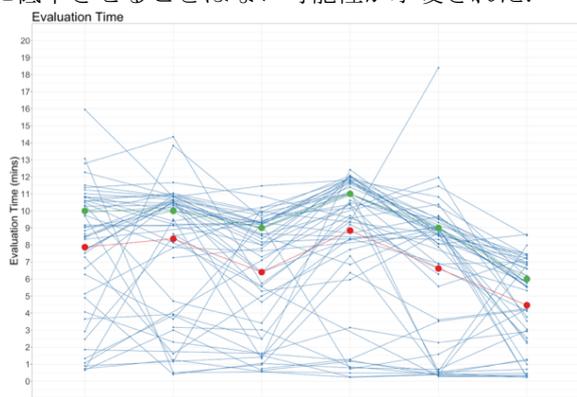
### 3.1 仮説1：『評価にかけた時間』は、連続して評価を行うことにより短くなっていく

繰り返し相互評価を実施することにより、評価に対する動機付けに影響を与える可能性があるという推測をもとに、『評価にかけた時間』は、連続評価を行うことにより短くなっていく」という仮説を立てた。これを検証するため、「評価にかけた時間」を以下のように定義し、各評価者について算出した。

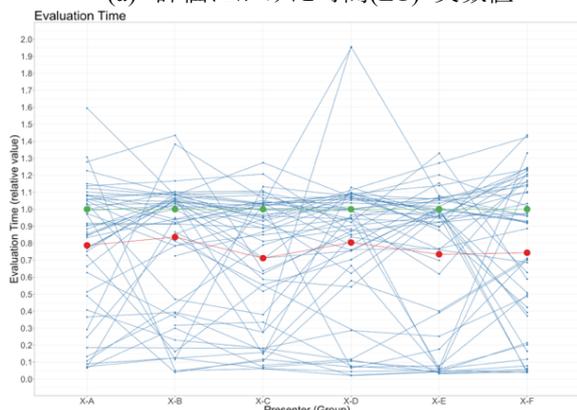
**評価にかけた時間(ET: Evaluation Time)**：最初の評価項目のラジオボタンをクリックしてから最後の評価項目をクリックするまでの時間

図1(a)は、6グループに対する評価におけるETの推移を可視化したものである。小点は各学習者のETを表し、下側の大点は平均、上側の大点は当該グル

ープのプレゼンテーションの長さを表す。図1(a)は、一見、連続して評価を行うとETが短くなっているように見える。しかし、プレゼンテーションの長さを表す点に着目すると、ETの傾向は、プレゼンテーションの長さの際の影響を大きく受けていることがわかる。この影響を除いてETを議論するために、各グループのプレゼンテーションの長さに対するETの比を算出し、可視化したのが図1(b)である。図1(b)では、図1(a)で見られた連続した評価に伴う下降傾向が消滅している。以上より、仮説に反した以下の2つの結果が得られた。すなわち、(1)「評価にかけた時間(ET)」はプレゼンテーションの長さの際の影響を大きく受け、また、(2)ETは連続して評価を行ってもクラス全体として単純に短くなっていくことはなかった。この結果より、相互評価においてはアンケートへの回答行動の場合と異なり、連続した相互評価の実施が評価に対する動機付けを単純に低下させることはない可能性が示唆された。



(a) 評価にかけた時間(ET) 実数値



(b) プレゼンテーションの長さに対する ET

図1 評価にかけた時間(ET)の推移

### 3.2 仮説2：『評価にかけた時間』の傾向は、評価者ごとに異なる

前節では、仮説に反し、評価にかける時間(ET)は連続して評価を行っても単純に短くなっていくことはないことが明らかになった。このため、『評価にかけた時間』の傾向は、評価者ごとに異なる」という仮説を立てた。これを検証するため、各評価者の

ET 推移グラフに対して時系列クラスタリングを行い、その結果に基づき4クラスに分類した。図2は図1(b)をクラスター別にプロットしたものである。図2より、全体的に評価に時間をかけるクラス、全体的に評価に時間をかけないクラス、最後のグループを短時間で評価するクラス、最初のグループを短時間で評価するクラスに分けられたことが分かる。これより、評価者によって評価にかける時間が長い場合や短い場合があることが明らかになった。また、授業の最初や最後のグループについては何らかの原因により、評価にかける時間が短くなる評価者が現れることも明らかになった。

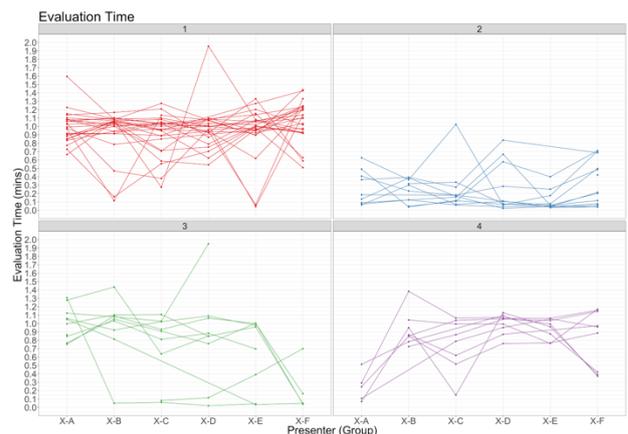


図2 クラスター別の評価にかけた時間(ET)

## 4. 結論

本稿では、連続して相互評価を実施することにより、評価に対する動機付けに影響を与える可能性がある、という仮説のもと、『評価にかけた時間』について分析を行った。結果として、相互評価においてはアンケートへの回答行動の場合と異なり、連続した相互評価の実施により『評価にかけた時間』が単純に短くなることはないことが明らかになった。また、評価者によって評価にかける時間が長い場合や短い場合があることが明らかになった。今後は、評点や評価タイミングのばらつきなど評価に対する動機付けを表すと考えられる他の特徴量や授業条件との関係の解析を進めたい。

### 参考文献

- (1) 藤原康宏, 大西仁, 加藤浩. “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, 第4巻, 第1号, PP.77-85, (2007)
- (2) 深澤真. “スピーチにおける生徒相互評価の妥当性. 全国英語教育学会紀要”, 第21巻, pp.181-190, (2010)
- (3) Mick P. Couper, “Web surveys: A review of issues and approaches”, The Public Opinion Quarterly, Vol.64, No.4, pp.464-494, (2000)
- (4) Ting Yan, and Roger Tourangeau, “Fast times and easy questions: the effects of age, experience and question complexity on web survey response times”, Applied Cognitive Psychology, Vol.22, No.1, pp.51-68, (2008)