

グローバル MOOC における相互評価の信頼性に関する検討

The Reliability of Peer Grading in Global MOOCs

—早稲田大学における事例から—
-In the Case of Waseda University-

石井 雄隆*	アダム ゴードン**	平賀 純*
Yutaka Ishii*	Gordon Adam**	Jun Hiraga*
永間 広宣*	大浦 弘樹*	森田 裕介***
Hironori Nagama*	Hiroki Oura*	Yusuke Morita*

早稲田大学大学総合研究センター*
早稲田大学アカデミックソリューション**
早稲田大学人間科学学術院***
Center for Higher Education Studies, Waseda University*
Waseda University Academic Solutions Corporation**
Faculty of Human Sciences, Waseda University***

<あらまし>早稲田大学は、2015年9月にグローバル MOOC プラットフォームの一つである edX に加盟し、2016年1月から3月まで第一弾講座「Tsunamis and Storm Surges: Introduction to Coastal Disasters」を開講した。本発表では、早稲田大学におけるグローバル MOOC の開発・運営とオンライン相互評価課題のデータ解析について報告する。主にオンライン相互評価の信頼性と学習者の主観的評価に焦点を当てる。

<キーワード> グローバル MOOC オンライン相互評価 高等教育 eラーニング

1. はじめに

高等教育では、MOOC (Massive Open Online Courses) を活用した教育内容の公開や新たな教育手法の開発などが盛んになされている。重田 (2016) は、オープンエデュケーションが高等教育に与える影響として、教育現場での MOOC 活用、デジタルラーニングの普及、大学間教育連携の促進などを挙げており、その利活用の在り方は盛んに検討されている。

荒ほか (2014) では、東京大学で開講した MOOC コースにおける受講者データについて報告している。具体的には、受講者の分布や属性及び講義ビデオ視聴者数の推移や最終成績などについて分析している。また永田ほか (2015) では、MOOC における受講者のコンテンツ遷移を分析し、学習履歴データに現れる受講者の学習様態について検討した。その結果、学習者を4つのパターン (コースに登録するだけの学習者、講義動画のみを閲覧する学習者、テストのみに回答する学習者、

講義動画・テストの双方に取り組む学習者) に分類できることが報告されている。

早稲田大学は、2015年9月にグローバル MOOC のプラットフォームの一つである edX に加盟し、第一弾講座「Tsunamis and Storm Surges: Introduction to Coastal Disasters」を開講した。本稿では、そのコースの概要及びオンライン相互評価課題のデータ解析について報告する。その中でも、オンライン相互評価の信頼性と学習者の主観的評価に焦点を当てる。

2. 講座

2.1. 講座の概要

本講座は、グローバル MOOC のプラットフォームの一つである edX より開講した。講座の内容は、海岸工学に関するもので津波と高潮に関する研究を中心として扱った。受講期間は、2016年1月19日から3月21日までの6週間であった。

講座の構成は、大きく分けると、授業動画

とクイズから構成されており、3週目と6週目にオンライン相互評価課題を行った。

2.2. 授業動画

授業動画は、全6週（導入も含むと7週）にわたって収録された。各セクションによってバラつきがあるが、10本前後のコンテンツにより、一つのセッションが構成されている。

表1 授業動画の構成

週	セクションの名前	ビデオ数
1	Introduction and Welcome	1
2	1. Tsunami Disasters	17
3	2. Storm Surge Disasters	10
4	3. Basic Mechanisms	7
5	4. Mechanisms of Disaster	10
6	5. Numerical Prediction Methods and Future Predictions	7
7	6. Disaster Mitigation and the Protection of Residents	7

2.3. 成績の評価

成績の評価は、クイズ40%、ディスカッションへの投稿30%、中間課題（ORA: Open Response Assessments）5%、期末課題（ORA: Open Response Assessments）15%、事前・事後アンケートへの解答10%となっており、60%以上の得点を獲得することで修了となった。

2.4. オンライン相互評価

前節で述べたように、中間及び期末課題としてオンライン相互評価課題を出題した。今回の講座では、5名の受講者の解答を評価すると、受講者自身のレポートを評価してもらえるように設定を行った。

植野・ポクボン・岡本・永岡（2008）によると、相互評価には下記7つのメリットが存在する。

- (1) 相互評価をすることにより、学習者により自律的にさせ、学習動機を高める
- (2) 他者からの意見は、テスト得点以上に学習者の内省を促進する
- (3) 他者を評価することにより、他者の成

果から学んだり、自己の内省を促すことができる

- (4) 経歴が似た学習者同士からのフィードバックは、理解しやすい
- (5) 教師の負担を軽減するとともに、教師不在でも有用な学習者へのフィードバックが促進される
- (6) 教師が考えつきもしない有用なフィードバックを与えることや、フィードバックのバラエティが広がる
- (7) 成人学生の場合、教師一人で採点を行うよりも、多人数で評価を行った方が信頼性が高くなる。

またMOOCにおけるオンライン相互評価は、渡邊・向後（2015）では、相互評価への親和性が高い受講者ほど学習継続意欲や講座評価が高いことも指摘されている。

一方で、渡邊・森・向後（2015）によると「負担の大きさ」「評価の難しさ」「評価人数の多さ」に関するコメントが多く挙げられることも指摘されている。

2.5. 中間課題・期末課題の内容

中間課題と期末課題の具体的な内容について述べる。中間課題は、「Study the history of coastal disasters in your country or a country you have an interest. Describe the nature of the disasters and provide details of the historic events. Discuss what, if any, changes there were toward disaster mitigation. Provide details of the type of disaster mitigation measures with detailed examples. You should provide references to support your arguments.」であり、自国の沿岸災害について論じる問題であった。評価の観点は、下記の通りである。

- (1) Does this response describe the nature of historical coastal disasters? (Not at all, Briefly, In detail の3段階評価)
- (2) Does this response explain how approaches to coastal disaster mitigation changed? (Not at all, Very briefly, Briefly, In detail の4段階評価)

(3) Does this response provide evidence to support the arguments? (Not at all, Very briefly, Briefly の 3 段階評価)

また期末課題は、図 1 に詳細を示しているが、ある地方において津波が発生した際に、4つのスタート地点の中から 2つを選んで、避難ルートを考えるタイプの問題であり、評価の観点は下記の通りであった。期末課題においては、観点 5・6 は、二段階評価であったため、今回の分析からは除外した。

< 観点 1・2 > Does the response for ANSWER A/B adequately describe the specific evacuation point so that it is identifiable on the elevation map? (Not at all, Briefly, In detail の 3 段階評価)

< 観点 3・4 > Does the response for ANSWER A/B provide a suitable reason for the chosen evacuation point? (Not at all, Briefly, In detail の 3 段階評価)

< 観点 5・6 > Does the response for ANSWER A/B explain why another location closer to the chosen evacuation point or starting point was not suitable? (Not at all, In detail の 2 段階評価)

3. 調査対象・調査方法

本講座の受講者は、2279名であり、受講者の年齢層は、26歳から40歳の間が45%を占めていた。また127カ国の受講者が登録をしており、全体の19%がアメリカの受講者であり、15%が日本の受講者、6%がインドの受講者であった。このように幅広い年齢層の多様なバックグラウンドを持った受講者が受講していた。

本研究では、3人・4人・5人から評価された中間・期末課題の解答255名分を抽出した。また人数別・観点ごとの信頼性について級内相関係数（二元変量効果モデル）を用いて検討した。級内相関係数とは、ある測定の検者間または検者内信頼性の指標（対馬 2008）であり、今回のデータを分析する際に適していると判断した。級内相関係数の判定基準には表 2 などが存在するが、0.7 以上であれば信頼性が高いと判断することが可能である（対馬 2008）。

表 2 級内相関係数の判定基準（Landis and Koch 1977, 対馬・石田 2013）

ICC の値	判定
0.00-0.20	わずか / slight
0.21-0.40	まづまづ / fair
0.41-0.60	適度 / moderate
0.61-0.80	十分 / substantial
0.81-1.00	ほとんど完全 / almost perfect

4. 結果と考察

4.1. 相互評価における評定者間の信頼性

表 3 が中間課題における観点別の級内相関係数であり、表 4 が、期末課題における観点別の級内相関係数である。中間課題においては、平均測定値を踏まえると、5名以上で評価することで、0.7 以上の十分な信頼性があった。一方で、期末課題においては、3名以上の評価で 0.7 以上の十分な信頼性があった。これには、一緒に学んでいる受講者同士の相互評価やフィードバックおよびコンテンツの学習を重ねることで、受講者間の評価の信頼性は向上する可能性が理由として考えられる。

一方で、中間課題は期末課題に比べて主観的な要因が入りやすかったということも要因として考えられる。受講者間の評定の揺らぎを抑えるためには、客観的な基準に基づくルーブリックを用いる必要があるということも考えられる。

4.2. 相互評価に対する受講者の評価

相互評価に対する受講者の評価について調査した。SurveyMonkey を用いて、事後アンケートを行った。5段階（1:poor から 5:excellent）で中間課題と期末課題の効果について評価をしてもらった。表 5 の通り、中間課題、期末課題における受講者の評価はおおむね高かった。

4.3. 相互評価に対する受講者のコメント

また上記の質問紙項目の自由記述欄の受講者の解答をいくつか紹介する。肯定的なコメントとしては、下記のようなコメントが記入された。

図1 期末課題の内容

Situation

You are visiting a Japanese coastal city, popular with tourists for the maritime sports and traditional picturesque Japanese scenery during peak season. Typically 15,000 tourists alone visit per day and the city is very crowded. During your visiting a large earthquake hits offshore and generates a tsunami. You must decide a suitable place for evacuation.

Tsunami conditions

Inundation height: 10m at the shoreline.

Tsunami arrival time: the first wave will reach the shoreline 22 minutes after the earthquake.

Run-up height: potentially 15m.

Other Conditions

You will be informed about the tsunami two minutes after the earthquake via loudspeakers.

You injure your ankle during the earthquake. You can walk but not run.

Assignment

Refer to the provided elevation map and data on the evacuation centers. Choose two of the four starting locations (S1~S4). Decide where you would evacuate to in the case of the above scenario.

Write down two separate answers for each of your chosen starting locations.

In your answers be sure to:

- Separate your answers for the two locations by clearly writing "ANSWER A: Location (S1~S4)" and "ANSWER B: Location (S1~S4)"
- Specify exactly where you would evacuate to so that it is easily identifiable on the map.
- Explain why you choose the evacuation point. You should specify details about each of the following: time, evacuation capacity, and height
- Explain why other locations closer to the starting point are not suitable evacuation points. You should use one or more of the following categories: time, evacuation capacity and/or height as a justification.

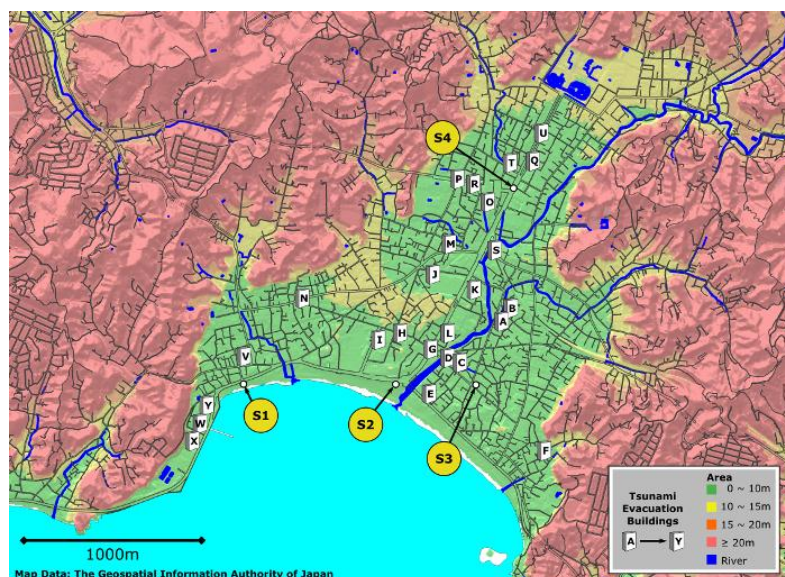


表3 中間課題における観点別の級内相関係数

評定者数	測定値	Disaster Origin	Disaster Mitigation Changes	Supported Arguments
3名 (n=43)	単一測定値	0.16	0.36	0.25
	平均測定値	0.36	0.63	0.50
	標準誤差	0.85	0.83	0.87
4名 (n=45)	単一測定値	0.42	0.29	0.26
	平均測定値	0.75	0.61	0.58
	標準誤差	0.45	0.45	0.45
5名 (n=32)	単一測定値	0.32	0.43	0.33
	平均測定値	0.70	0.79	0.71
	標準誤差	0.88	1.01	0.93

表4 期末課題における観点別の級内相関係数

評定者	測定値	Specified Location A	Specified Location B	Suitable Evacuation Point Reasoning A	Suitable Evacuation Point Reasoning B
3名 (n=43)	単一測定値	0.58	0.58	0.57	0.57
	平均測定値	0.80	0.80	0.80	0.80
	標準誤差	0.85	0.83	0.87	0.84
4名 (n=52)	単一測定値	0.46	0.51	0.47	0.50
	平均測定値	0.77	0.80	0.78	0.80
	標準誤差	0.45	0.45	0.45	0.44
5名 (n=40)	単一測定値	0.57	0.49	0.59	0.53
	平均測定値	0.87	0.83	0.88	0.85
	標準誤差	0.88	1.01	0.93	0.99

表5 相互評価に対する受講者の評価の度数分布表

質問項目	1	2	3	4	5	平均
The effectiveness of the Mid-term Report (peer assessment)	8 (4%)	14 (7%)	46 (25%)	48 (25%)	68 (37%)	3.84
The effectiveness of the Final Report (peer assessment)	8 (4%)	11 (5%)	34 (18%)	64 (33%)	76 (39%)	3.98

Very nice assessment! It made me really think how I would or should act when there is a storm surge in The Netherlands.

本講座は、自国で海岸災害が生じた際に、どのような行動をとるべきかについて教えることがコースの大きな目的の一つであった。特に期末課題は、中間課題及びコンテンツで学習したことを踏まえた上で、回答する必要

があり、この受講者は、この課題を通して、コースの目的を達成することができたのではないかと考えられる。

一方で下記のようなコメントも存在した。
Some reviewer's don't have any background knowledge on local contents.

I did not like the grading system for this because everyone has their own opinion

and this might effect your grade on how you write.

中間課題は、受講者が生活している地域における状況を記す必要があり、そのレポートを評価される受講者があまり詳しくない地域を評価する場合には、評価の信頼性に影響を与えたのではないかということが推察される。また、宗教・国・人種などそれぞれの多様なバックグラウンドを持った受講者が参加するグローバル MOOC においては、受講者個人のビリーフや価値観が評価に影響を与えた可能性も否定はできないと考えられる。これらの観点については、学歴や年齢など評定者の属性情報と紐づけたより深い分析が検討される必要がある。

5. 結論

複数の受講者によって評価されることで、相互評価の信頼性は保たれる。またその人数は、受講者同士の学び合いやタスクの特徴などによって左右される。具体的には、一緒に学んでいる受講者同士の評価やフィードバックの経験を重ねることで、受講者間の評価の信頼性は向上する可能性が示唆された。

今後の課題としては、信頼性だけでなく妥当性の検討もなされる必要がある。相互評価に対する受講者のコメントでも述べたように、学習者自身のビリーフや価値観などが、評価に影響を与える可能性を本研究では検討できていないため、専門家による評価との一致率なども検討する必要が存在する。

グローバル MOOC における教育実践・研究は、オンキャンパスへの新たな教育手法の導入についても応用が可能であり、今後もさらなる検討が必要とされる。

参考文献

- 荒優, 藤本徹, 一色裕里, 山内祐平 (2014) MOOC 実証実験の結果と分析—東京大学の 2013 年の取り組みから—。東京大学大学院情報学環紀要情報学研究, 86 : 83-100
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* : 159-174
- Luo, H., Robinson, A. C., and Park, J-Y. (2014) Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects. *Journal of Asynchronous Learning Networks*, 18(2) : 1-14
- 永田裕太郎, 村上正行, 森村吉貴, 椋木雅之, 美濃導彦 (2015) MOOC における大規模学習履歴データからの受講者の学習様態獲得。人工知能学会第 73 回先進的学習科学と工学研究会 : 25-30
- 重田勝介 (2016) オープンエデュケーション開かれた教育が変える高等教育と生涯学習。情報管理, 59(1) : 3-10
- 対馬栄輝, 石田水里 (2013) 医療系データのとり方・まとめ方—SPSS で学ぶ実験計画法と分散分析。東京図書。
- 対馬栄輝 (2008) SPSS で学ぶ医療系多変量データ解析, 東京図書, 東京
- 植野真臣, ソンムアン・ポクポン, 岡本敏雄, 永岡慶三 (2008) ピアアセスメントにおける評価者特性を考慮した項目反応理論。電子情報通信学会論文誌, J91-D(2) : 377-388
- 渡邊文枝, 森裕樹, 向後千春 (2015) JMOOC の講座における相互評価に対する受講者アンケートの分析, 人間科学研究, 28(2) : 237-245
- 渡邊文枝, 向後千春 (2015) JMOOC の講座における受講生の e ラーニング指向性と相互評価指向性が学習継続意欲と講座評価に及ぼす影響, 日本教育工学会研究報告集, JSET15-4, pp.87-94