# Peer Assessment Dataset

**Andrii Vozniuk**
EPFL, Switzerland
andrii.vozniuk@epfl.ch

**Adrian Holzer**
EPFL, Switzerland

**Denis Gillet**
EPFL, Switzerland

**ABSTRACT**: Peer assessment is seen as a powerful supporting tool to achieve scalability in the evaluation of complex assignments in large courses, possibly virtual ones, as in the context of massive open online courses (MOOCs). However, the adoption of peer assessment is slow, due in part to the lack of ready-to-use systems. Furthermore, the validity of peer assessment is still under discussion. In this paper, in order to tackle some of these issues, we present a dataset containing the assessment of student submissions by student peers and by instructors during our Social Media course with 60 Master's level university students. The dataset allows for training and testing algorithms that predict the grades of instructors based on the grades of student peers.

**Keywords**: Learning analytics, dataset, peer assessment, social media

## 1    INTRODUCTION

Peer assessment can be defined as an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners (Topping, 1998). Peer assessment was already being deployed as an evaluation tool in university settings at the beginning of 19th century (Topping, 1998) and it has recently gained more attention in the context of massive open online courses (MOOCs) (Cooper & Sahami, 2013), where it can serve as a powerful tool to achieve scalability in the evaluation of complex assignments with large student cohorts. It can also be argued that peer assessment can reduce the workload for instructors, saving time in the process if the setup and implementation are not too time consuming. Even though peer review and assessment have advantages, they have not become a mainstream evaluation feature in universities. Pearce, Mulder, and Baik (2009) argue that this is in part because of the difficulty of usage and price of dedicated peer review and assessment systems. They point to online tools as promising solutions to this problem but they deplore the fact that so few of them exist.

Furthermore, the validity of peer assessment is still under discussion. In this paper, in order to encourage the development and adoption of peer assessment algorithms, we present a dataset from a case study of peer assessment using Graasp, a social media platform (Bogdanov et al., 2012), in a Social Media course taught at the Swiss Federal Institute of Technology in Lausanne (EPFL) with 60 Master's

level students and 4 instructors. It should be highlighted that in the context of this social media course, the objective of peer assessment was twofold. First, it aimed at assessing the work of students. Second, it aimed at introducing students to the challenges and opportunities related to ratings in social media platforms. Our previous work provides additional details about this case study (Vozniuk, Holzer, & Gillet, 2014). Below we present more detail about the dataset itself.

## 2    THE DATASET

### 2.1    Creator and Owner

The dataset was created by the REACT Group at EPFL and is owned by EPFL.

### 2.2    Access Details

The dataset is available for download from http://go.epfl.ch/peer-dataset and consists of three files: two with grades and one with student-report mapping. The following files are available in the tab-separated values (TSV) format:

- **grades_studens_all.tsv** contains grades assigned by the students

- **grades_instructors_all.tsv** has the grades assigned by the instructors

- **students_reports_mapping.tsv** contains the mapping between student ids and their report ids.

All participants of the assessment were instructed to use a 5-point Likert-type scale for grading the reports. Since the course was about social media, the grade scale was inspired by the 5-star rating system often found in social media (e.g., used for product review on Amazon.com): 5 stars – I love the report; 4 stars – I like the report; 3 stars – The report is OK; 2 stars – I don't like the report; 1 star – I hate the report. In total there are 1346 grades: 1106 by the students and 240 by the instructors.

### 2.3    Version and Date

This is the first public release of the dataset, which is dated 30 May 2015.

### 2.4    Format and Schema

Each of the two files with the grades has three columns. The first column contains a unique identifier of the person grading the report, i.e., a student id in **grades_studens_all.tsv** and an instructor id in **grades_instructors_all.tsv**. The second column contains a unique identifier of the report and the third column contains the grade assigned to the report (from 1 to 5).

The file with the mapping has two columns. The first column contains the student identifier and the second contains the identifier of the report authored by the student.

### 2.5    Restrictions of use

The dataset is free to use and distribute for any purpose; citation of Vozniuk et al., 2014 is appreciated.

## 2.6     Provenance, extraction, maintenance

The peer assessment was performed in the 2013 spring semester during the Social Media course at EPFL with 60 Master's level students in Computer Science and 4 instructors supervising the course. During this course, each student had to hand in a two-page project report to show a long-tail effect in the social media platform of his or her choice.

Proper instructions are a key success factor for peer evaluation. We made clear to students at the beginning of the course that peers would evaluate their reports, and that they would have to evaluate the work of 20 peers. As the reports discussed the long tail effect in various social media platforms, we also made it clear that their reading of these reports was a key learning activity. Finally, we told them that, in order to ensure fairness, their peer evaluations would only be taken into account if they did not deviate significantly from the evaluation carried out by the instructors.

To conduct the peer assessment, we randomly assigned students to review 20 reports of their peers. The instructors assessed all the reports. The grades were submitted via Graasp, a social media platform that supports the peer assessment process. Afterwards the dataset with the grades was manually extracted from the database.

## 2.7     Ethical and privacy considerations

The released dataset does not contain personally identifiable information.

## 2.8     Limitations

The dataset is based on peer assessment of 60 students and contains 1347 grades, but it still represents peer assessment in a single class over single semester and hence could be subject to some factors specific to this setting. To overcome this limitation, we are planning to extend the dataset with peer assessment results from the Social Media course in subsequent years.

## REFERENCES

Bogdanov, E., Limpens, F., Li, N., Helou, S. E., Salzmann, C., & Gillet, D. (2012). A social media platform in higher education. *Proceedings of the IEEE Global Engineering Education Conference (EDUCON '12)*, 1–8. http://dx.doi.org/10.1109/EDUCON.2012.6201105

Cooper, S., & Sahami, M. (2013, Feb.). Reflections on Stanford's MOOCs. *Communications of the ACM, 56*(2), 28–30. http://dx.doi.org/10.1145/2408776.2408787

Pearce, J., Mulder, R., & Baik, C. (2009). *Involving Students in Peer Review: Case Studies and Practical Strategies for University Teaching.* Melbourne, Austraila: Centre for the Study of Higher Education. Retrieved from http://peerreview.cis.unimelb.edu.au/resources-3/publications/

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. http://dx.doi.org/10.3102/00346543068003249

Vozniuk, A., Holzer, A., & Gillet, D. (2014). Peer assessment based on ratings in a social media course. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 133–137. http://dx.doi.org/10.1145/2567574.2567608