

An Exploratory Study of Reliability of Ranking vs. Rating in Peer Assessment

Authors : Yang Song, Yifan Guo, Edward F. Gehringer

Abstract : Fifty years of research has found great potential for peer assessment as a pedagogical approach. With peer assessment, not only do students receive more copious assessments; they also learn to become assessors. In recent decades, more educational peer assessments have been facilitated by online systems. Those online systems are designed differently to suit different class settings and student groups, but they basically fall into two categories: rating-based and ranking-based. The rating-based systems ask assessors to rate the artifacts one by one following some review rubrics. The ranking-based systems allow assessors to review a set of artifacts and give a rank for each of them. Though there are different systems and a large number of users of each category, there is no comprehensive comparison on which design leads to higher reliability. In this paper, we designed algorithms to evaluate assessors' reliabilities based on their rating/ranking against the global ranks of the artifacts they have reviewed. These algorithms are suitable for data from both rating-based and ranking-based peer assessment systems. The experiments were done based on more than 15,000 peer assessments from multiple peer assessment systems. We found that the assessors in ranking-based peer assessments are at least 10% more reliable than the assessors in rating-based peer assessments. Further analysis also demonstrated that the assessors in ranking-based assessments tend to assess the more differentiable artifacts correctly, but there is no such pattern for rating-based assessors.

Keywords : peer assessment, peer rating, peer ranking, reliability

Conference Title : ICALT 2017 : International Conference on Advanced Learning Technologies

Conference Location : Paris, France

Conference Dates : October 19-20, 2017