

Rate it Twice: Using the Wisdom of Many in One Mind to Improve Performance Evaluation

Meir Barneron

Hebrew University of Jerusalem and National Institute for Testing and Evaluation

Avi Allalouf

National Institute for Testing and Evaluation

Ilan Yaniv

Hebrew University of Jerusalem

Abstract

The wisdom of crowds refers to the idea that judgmental estimation can be improved by averaging the estimates of different judges. Recent research has suggested that combining multiple estimates made by the same judge also yields accuracy gains, evidencing what is called “the wisdom of many in one mind.” The present study extends the use of the wisdom of many in one mind to performance evaluation, specifically the evaluation of essays written as part of standardized college admissions tests. The participants in our field study were professional raters who were asked to evaluate a set of essays twice. The findings suggest that combining evaluations (within raters) is beneficial in that such combinations are more accurate than single evaluations in terms of squared errors and correlations. The within-rater combinations were also compared to combinations of pairs of independent raters. The independent-rater combinations were more accurate than the within-rater combinations. Notably, the within-rater combinations realized two thirds of the accuracy gains obtained from combining independent raters. That performance evaluation can benefit from the wisdom of many in one mind should be of interest to theorists as well as professionals in applied fields, such as human resources, education, and testing & assessment organizations.