

Title: Consistency in teachers' judgments of the final high-school examination mark

N. Bonini, L. Canal, R. Micciolo  
University of Trento

This abstract reports preliminary results of an experimental study of consistency between the teachers' rating of weights of the five attributes for the final examination mark of candidates at the high-school, and the weights of the same attributes estimated using constrained multiple regression on the teacher's final examination marks.

Two samples of 55 experienced high school teachers were interviewed. Both groups were presented with 64 profiles of students to be evaluated for the final examination (task 1). One group of teachers (usual profile,  $n = 32$ ) was presented with typical student profiles to be evaluated. The other group (unusual profile,  $n = 23$ ) was presented with atypical student profiles to be evaluated. Also, they were asked to rate the relative weight of five attributes that the teachers have to consider in order to make the final mark (task 2). The same teacher did the two tasks within one/two weeks-interval period; 26 teachers performed task 1 before task 2 (PW) and 29 performed task 2 before task 1 (WP). To evaluate concordance between the final mark given by each teacher ( $Y_t$ ) and the mark derived from the teachers' rating of weights ( $Y_w$ ) and between the final mark given by each teacher ( $Y_t$ ) and the mark estimated using multiple regression ( $Y_e$ ), the concordance correlation coefficient ( $r_c$ ) was used (Lin L.I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45, 255–268, 1989); this index is the correlation between two measures that fall on the 45° line through the origin.

An highly significant difference between the mean values of  $r_c$  between  $Y_t$  and  $Y_w$  in relation to the type of profile (usual vs. unusual) was found ( $F=52.6$ ;  $p<0.001$ ), while no overall difference was found between PW and WP ( $F=0.09$ ;  $p>0.5$ ); however a significant interaction between the type of the profile and the order of presentation was found ( $F=4.31$ ;  $p=0.043$ ). An highly significant difference was also found between the mean values of  $r_c$  between  $Y_t$  and  $Y_e$  in relation to the type of profile ( $F=22.6$ ;  $p<0.001$ ), while neither the overall difference between PW and WP nor the interaction between the type of the profile and the order of presentation were significant ( $F=0.39$ ;  $p>0.5$  and  $F=0.38$ ;  $p>0.5$  respectively). The  $r_c$  values between each of the five teachers' rating of weights and the corresponding weight estimated from multiple regression were the following: 0.239, 0.176, 0.358, 0.300, 0.052.

These results show that (i) the consistency of teachers presented with the set of unusual profiles was much lower than the consistency of teachers presented with the set of usual profiles and (ii) the consistency between the teachers' rating of weights and the weight estimated from multiple regression was very low.