Study of the validity of the Self-Assessment as an educational evaluation method: use of bootstrapping

Universidad de Jaén

Samuel P. León

Departamento de Pedagogía

@SamuelPLeon

INTRODUCTION

- Since the works published by Boud and Falchikov in 1989, Self-assessment has taken a great interest both, in the field of research and in the educational practice.
- The heterogeneity in the conception of self-assessment translates into a wide range of results on its validity as an evaluation tool
- The main goal of this work is to show empirical evidence on the validity of Self-assessment as a method of evaluation in University education.
- Additionally, since the sample is a natural class group, a bootstrapping method will be used to infer the population effect.

METHOD

Participant

- 64 students from the University of Jaén (Me 22.48 years)

Apparatus

- Final Exam (six open-ended questions), 0 to 10 range score.
- Self-Evaluation question with evaluation criteria.
- Evaluation rubric for the teacher.

Procedure

- After the exam, the professor blinded the exams and evaluated them applying the previously established rubric.

RESULTS and DISCUSSION

- Self-evaluation, in general terms, can be a valid evaluation technique in university education.
- However, this technique loses validity when applied to the extremes of the population.
- Specifically, it might result in an over-evaluation of actual grades by the less skilled students or in a underestimation of them by the most capable students

Self-Assessment is accurate for average students, less competent students tend to overestimate, and the most competent to underestimate.

Scan me to download the full PDF

Extra Table and Figures

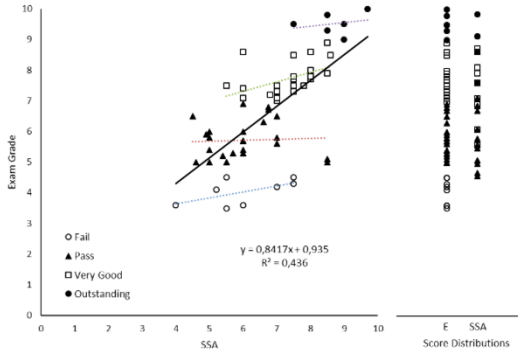

Figure 1. In the vertical axis, the scores obtained by the students in the exam (Exam Grade) on the horizontal axis are Self-Assessment (SSA) scores. The left part of the figure shows the linear regression between SSA and Exam Grade. The regression equation is shown at bottom right and the overall line of best fit is shown in black. Scores are shown by group (Fail = open circles, Pass = black triangles, Very Good = open squares, and Outstanding = black circles) with the lines of best fit within groups shown in dashed lines (Blue for Fail, red for Pass, green for Very Good and purple for Outstanding). The distributions of the scores for the variable Exam (E) and for the variable Self-assessment (SSA) are shown to the right of the figure

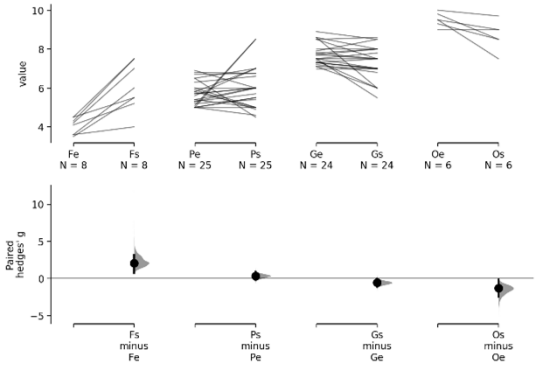

Figure 2. The paired Hedges' g for 4 group (F: Fail, P: Pass, V: Very Goods, and O: Outstanding) comparisons are shown in the Cumming estimation plot. The subscript e indicates Exam score; subscript s indicates SSA score. Each paired mean difference is plotted as a bootstrap sampling distribution. Mean differences are depicted as dots; 95% confidence intervals are indicated by the ends of the vertical error bars.