

G Gibbs and M Coffey The impact of training on university teachers' ...

I have had 15 years experience as a statistical referee. This paper is the worst manuscript I have seen.

Under an important code of ethics for experiments on human subjects, the Nuremberg Code (1947), the study is unethical and should not have been carried out. The design ensured that no useful results could be obtained. The control group is blatantly different from the experimental group. Critical discussion of the value of the outcome measures is needed. Therefore time and money has been wasted, including the time of the participants.

Data from this bad design is poorly summarized and incorrectly analyzed. No basic summary statistics are given on the countries, the type of universities, prior experience of teaching, sex or age. The authors did not even mention what country the control universities were in, and what type of universities they were. There is substantial dropout (96/235 is only 40%), the potential impact of which is ignored. The numbers presented in the Tables are not consistent with the first paragraph of the results. The cluster structure is ignored, which will almost certainly result in under-estimation of the standard deviations and hence inflation of test statistics. The range of the scores is not given, so the changes in mean cannot be evaluated for practical importance. All changes are small. The figure has a truncated axis, which visually exaggerates the changes and no indication of the precision of the points is given.

The discussion is almost entirely disingenuous, if not dishonest. Numerous statements are made which are not justified, even by the results presented. For example, the change between student cohorts are tiny (crude mean difference of 1 unit, ignoring problems of dropout and incorrect analysis). This change is attributed to a difference in teaching, rather than differences between the intake, or the courses. We are not told whether the teachers taught the same courses, nor whether courses had been revised: the phrase "where possible" implies that this was not always the case.

Some claims are simply false, e.g. "the Control Group showed evidence of negative changes". A major weakness of the analysis - the failure to consider the variation attributable to the various characteristics of teachers, training programmes and universities - is claimed to be a strength: findings "involve aggregated data".

It is interesting to know that a Professor at the Open University produces work of this quality.

Dr J L Hutton, 15 March 2002
Senior lecturer,
Department of Statistics
The University of Warwick
Coventry, CV4 7AL
Email: J.L.Hutton@warwick.ac.uk