Creating Real Value in Software Engineering Experiments

James Miller

Undertaking empirical work to understand the software engineering process is an increasingly common activity. However, due to the nature of most software engineering studies, drawing reliable conclusions from a single study is inherently dangerous. A single study undertaken by a small group of practitioners, in a single location, with a single group of subjects, will possess a large number of parameters, some controlled, some completely unconstrained. Any one of which can potentially cause a significant change in the result of the study; and hence their effect must be controlled, eliminated or understood, if reliable results are to be derived. The normal method of achieving this objective is to repeat the study, often while changing some of the parameters, to see if the original result is stable with regard to repetition and alteration of some of its components. Without this confirming power, empirically based claims in software engineering should be regarded as having at best limited value and at worst with suspicion and mistrust. Hence, despite their now significant volume, software engineering experiments impact has been rather limited upon industrial software processes. The potential problems with single software engineering empirical studies are three-fold: low statistical power; large number of potential covariates with the treatment variable; and the verification of the process and products of the study.

The normal way to tackle these limitations is by replication of the study. Perhaps a more relevant alternate, to the Software Engineering community, is to create a single experiment, which internally contains multiple versions of the experimental theme all based around a common sub-set of hypotheses. This would allow the experiment to amass enough “evidence” to be convincing (if successful) based upon its direct solution of the three aforementioned issues. But while the design of such meta-experiments might seem obvious, the logistical problems associated with their successful implementation are massive. Solutions needing to be found to allow the successful collaboration of such large, and geographically disperse, network of researchers include: (A) the co-ordination of the information flow and experimental design; (B) ensuring suitable knowledge, training and skill levels across all subjects; (C) understanding any, and all, culture issues which may exist between sites; (D) how to obtain funding for such cross jurisdictional projects; and the resolution of clashing ethical, privacy; (E) and general human rights issues between the various sites and jurisdictions.