Optimizing Return-On-Investment (ROI) for Empirical Software Engineering Studies

Lutz Prechelt

Institut für Informatik, Freie Universität Berlin

The notion of ROI

Return-on-investment (ROI) is a concept from the financial world. In the dynamic view, ROI describes the periodically recurring profits (returns) from fixed financial capital (investment). In the static view, ROI describes the one-time income or saving (return) realized as a consequence of a one-time expenditure (investment). In this case, if the return does not occur within a short time, later parts of the return may be discounted for interest. For our purposes, we will use the static view and ignore discounting. We will call the return *benefit* and the investment *cost*.

The notion of ROI can be generalized to any domain (in particular engineering) in which both cost and benefit can be quantified on a rational scale and both scales use the same units. Most commonly, the scale is either money or time.

ROI can be used for retrospective analysis of the performance of an investment (controlling) or as an aid for decision-making about potential investments (planning).

In the former case, both cost and benefit may be known and the ROI is a single fixed number. Values greater than 1 indicate successful investments, values smaller than 1 indicate failed ones. In the latter case, however, there is usually some uncertainty about the cost and almost always significant uncertainty about the benefit; both expected cost and expected benefit are best described by probability distributions and hence the ROI is also a distribution.

In many cases, the uncertainty can be factored into a base case plus a number of identifiable risks, each of which increase cost or reduce benefit.

ROI of empirical studies: Industrial view

The view of an industrial organization onto an empirical study in software engineering is well compatible with the definition of ROI as stated above: The study is considered an investment that is made in order to produce a return; the expected ratio of benefit to cost determines whether a suggested study will be performed or not.

When considering how to optimize ROI we may therefore list the kinds of costs and kinds of benefits that typically occur and collect suggestions for improving each of these.

The major kinds of costs are:

- C1. Effort for determining the precise research question.

- C2. Researcher effort for designing the study.
- C3. Researcher effort for performing the study.
- C4. Industrial participant effort for performing the study.
- C5. Researcher effort for evaluating and reporting the study.
- C6. Indirect cost (for instance losses in time-to-market, product quality, or developer motivation).

Note that C2 may involve learning about empirical methods first (in particular for industrial researchers) or learning about the application domain first (in particular for academic researchers).

The major kinds of benefits are:

- B1. Process effectiveness impact: Improved process capabilities with respect to quality or cycle-time.
- B2. Process efficiency impact: Improved process capabilities with respect to cost.
- B3. Morale impact: Improved staff motivation and cooperation, e.g. due to resolution of a dispute.
- B4. Indirect benefit: Improved process capabilities for further process improvement (for instance via further empirical studies).

Note that B1 and B2 apply in particular to technology evaluation studies. They break down into (a) better estimates for the cost and risk of applying the technology, (b) better understanding of its benefits and limitations, (c) initial technology training for subsequent change agents, and (d) improved chance of successfully adapting the technology if it is valuable and rejecting it if it is not.

There may be secondary benefits beyond those listed, such as the publicity if the study results in a scientific publication.

Optimizing ROI in the industrial view

We can now use the above partitioning of cost and benefit into elements as a guide for formulating a few general hints how to reduce cost or increase benefit in an empirical study. The various hints work in different ways: each may target a reduction of base cost, an increase of base benefit, the management of a specific class of risk factors, or may just generally reduce the variance of study results (also in the sense of risk management).

- C1. Plan whole research programmes at once rather than individual studies.
- C2. Reuse or adapt the design of a previous study.
- C3. Reuse existing infrastructure (such as instructions, checklists, automation) for data collection.
- C4. Use continuous data evaluation in order to stop using further participants as soon as the study result becomes sufficiently clear; avoid collecting non-essential data; use automation to minimize the work and disruption of the participants; embed the study in the standard software process and adapt your manipulations so as to become indistinguishable from normal project management.

- C5. Use checklists for data evaluation and report in standardized formats.
- C6. Include the empirical study as a risk factor in your project risk management.
- B1. Favor studies whose results can be turned into process improvements easily. Assess the major threats to validity and confounding influences explicitly during the study.
- B2. Ditto.
- B3. Perform empirical studies specifically in areas that have both high relevance and high disagreement; make sure everybody accepts the study design as unbiased. Carefully explain study results and discuss their implications with the stakeholders.
- B4. Introduce explicit experience management.

These hints could be formulated more concretely and in more detail if we discussed different types of research methodology separately (e.g. experiments, surveys, case studies, post-hoc studies, benchmarks)

ROI of empirical studies: Scientific view

From a scientific point of view, most of what has been said above about cost does apply, however the benefit structure is completely different.

The major kinds of benefits from a scientific point of view are:

- B1. The specific, immediate insight gained from the study.
- B2. New conjectures that help in formulating preliminary theories and corresponding research programs.
- B3. Impact on theory: Partial validation or refutation of a previous conjecture or theory.
- B4. Impact on research agenda: Improved ideas and empirical capabilities for future studies.
- B5. Indirect benefit: Improved trust with industrial partners (or at least an extended track record) and thus increased likelihood of opportunities for empirical studies.

The above formulations use the term *theory* in a rather broad sense; most statements about mechanisms and relationships that are somewhat abstract qualify as theories.

Optimizing ROI in the scientific view

- B1. Make sure the study design provides insight even if the expected results do not occur.
- B2. Observe sufficiently many different variables or phenomena at once so that you will probably be able to formulate a new conjecture. In particular, do not unnecessarily restrict the study to quantitative observation only.
- B3. Always formulate some kind of expectation, even if the study is mostly exploratory.

- B4. Perform a post-mortem for an empirical study much like you should do for a software project; keep good ideas and identify mistakes.
- B5. Work on human relationships along with working on the study.

Conclusion

While the concept of ROI can be directly applied to individual empirical software engineering studies if an entirely industrial view (focussing on cost efficiency) is used, it should be obvious that ROI in a strict sense cannot be applied to scientific considerations of research benefits, as this would require assigning a monetary value to insights (and do so before you even had them!), which does rarely make any sense.

Nevertheless, the above partitioning of cost and benefit into components and subsequent reflection on ways for optimizing each component suggest that ROI considerations can provide useful guidelines for optimizing research efficiency. Concrete examples for such optimizations can be found in many published studies and are waiting to be reused.

Acknowledgements

Thanks go to the members of the breakout group whose work provided input for this writeup: Frank Houdek, Barbara Kitchenham, Jürgen Münch, Dietmar Pfahl, Austen Rainer, and Laurie Williams. Barbara, Dietmar, Austen, Laurie, and Giovanni Cantone helped significantly in shaping this summary.