

## Evidence relating to Object-Oriented software design: A survey

John Bailey  
David Budgen  
Department of Computer Science  
Durham University  
U.K.  
david.budgen@durham.ac.uk

Mark Turner  
Barbara Kitchenham  
Pearl Brereton  
Stephen Linkman  
School of Computing & Mathematics  
Keele University  
U.K.  
m.turner@cs.keele.ac.uk

### Abstract

**Context:** *There is little empirical knowledge of the effectiveness of the object-oriented paradigm.*

**Objectives:** *To conduct a systematic review of the literature describing empirical studies of this paradigm.*

**Method:** *We undertook a Mapping Study of the literature.*

**Results:** *138 papers have been identified and classified by topic, form of study involved, and source.*

**Conclusions:** *The majority of empirical studies of OO concentrate on metrics, relatively few consider effectiveness.*

## 1 Introduction & Background

Although the Object-Oriented paradigm (henceforth abbreviated to ‘OO’) has its roots in the 1960’s and has become a highly pervasive influence upon imperative programming languages and design practices, the literature provides surprisingly little in the way of widely-known empirical assessment of its effectiveness in particular roles<sup>1</sup>. The study described here has emerged from our recent work investigating the *evidence-based* paradigm that seeks to aggregate primary empirical studies in an objective and unbiased manner in order to create evidence that can support or refute particular research hypotheses. It is motivated by the experience of one of us (DB) when seeking suitable forms of evidence to be cited in a book on software design [2].

The core tool of the evidence-based paradigm is the *Systematic Literature Review* (usually abbreviated to ‘Systematic Review’) [3, 4] which provides a framework for systematically searching the literature, extracting the data, and

<sup>1</sup>The same comment applies equally to other, earlier, approaches to design of course (for example, ‘Structured Design’).

performing the necessary analysis. A precursor to a Systematic Review, that is sometimes undertaken as a separate study used to identify the extent and form of the literature on a particular topic, is referred to as a *Mapping Study*. Here we describe its use to help identify those primary studies that evaluate aspects of object-oriented design in any way and hence to determine what forms and issues have been studied, as well as by what means.

As with previous empirical studies we have undertaken, we first created a *protocol* to describe our plans for the study [3]. A key element of the protocol is the research questions that the review sets out to address (examples of other elements are the search strings, inclusion/exclusion criteria and data extraction plan). Our research questions included:

- Which journals include papers on software design?
- What are the most investigated OO design topics and how have these changed over time?
- What are the most frequently applied research methods, and in what study context?

This short paper is a preliminary report that outlines our data collection process, presents a summary of the results from this, and provides an initial analysis of the outcomes.

## 2 The Mapping Study

A mapping study (sometimes termed a *scoping review*): “involves a search of the literature to determine what sorts of studies addressing the systematic review question have been carried out, where they are published, in what databases they have been indexed, what sorts of outcomes they have assessed, and in which populations” [4].

While the planning phase is similar to that of a systematic review (although the resulting protocol will generally

be much shorter), the focus of a mapping study is upon the first three stages of the second phase of a review, namely:

- identification of research (searching);
- selection of primary studies (inclusion/exclusion);
- study quality assessment (bias/validity).

The data extraction stage is generally much broader than that for a systematic review, and is aimed mainly at classification and categorisation (as here).

## 2.1 Conduct of the Study

The study is being conducted by two research assistants (JB and MT), with other members of the team acting as reviewers where necessary. This paper reports on the outcomes of the first two stages only.

For the searching stage, the general *scope* of the study was identified as being:

**Population:** Published scientific literature reporting OO software design.

**Intervention:** Empirical studies involving software design practices, techniques and processes.

**Outcomes of relevance:** Quantity and type of evidence relating to various OO design techniques and processes.

**Experimental design:** Any scientific experiment or empirical study.

The search itself encompassed a wide range of computing journals and conferences (with the initial lists being generated by using the references from [2]) and the major digital libraries (including IEEE and ACM) accessed via a range of search engines that included IEEE Xplore, Google Scholar, Science Direct and Web of Science). No limits were placed on date of publication.

The search terms used were selected from a trial with a candidate set, and the three finally used were:

- “object oriented” design “empirical evidence”
- OO empirical design
- “software design” OO experimental

These were used with six search engines/sources: ACM; IEEE Xplore; Google Scholar; CiteSeer; ScienceDirect; and Web of Science. Collectively these addressed the main digital libraries considered to be appropriate to the study.

The study employed the following *inclusion* criteria:

- books, papers, technical reports and ‘grey literature’ describing empirical studies regarding OO software design;
- where several papers reported the same study, only the most recent was included;

- where several studies were reported in the same paper, each relevant study was treated separately;

and the following *exclusion* criteria:

- studies that did not report empirical findings;
- literature that was only available in the form of abstracts or *Powerpoint* presentations.

We should note that no quality assessment was performed at this stage, in order to ensure maximum coverage.

## 2.2 Outcomes

Two rounds of searching have been performed, with some refinement of search terms being used for the second round. Details of the papers found have been entered into a MySQL database for ease of analysis. The initial figure of 157 publications was reduced to 138, with 19 being filtered out because the study was reported in multiple articles or lacked empirical data. 244 individual authors were listed.

The initial classification of paper type presented here is informal and based upon title and abstract. Table 1 shows the 138 papers classified by the form of ‘intervention’ used.

Form of Intervention	Number
OO versus non-OO	19
Abstraction (e.g. modelling)	32
Design Patterns	10
Metrics	62
Design Comprehension	15

**Table 1. Papers classified by intervention**

Experimental Form	Number
Case-control Study	2
Case series	3
Case Study	28
Document Analysis	4
Experience report/case study	1
Laboratory Study	52
Literature Review	2
Observation	28
Randomised experiment performed in an artificial setting	13
Semi-structured Interview	1
Survey (convenience sample)	3
Survey (random sample)	1

**Table 2. Papers classified by form of study**

A further classification using the form of empirical study involved has been performed, and is shown in Table 2. As

can be seen, the great majority of studies used case study, laboratory study and observational forms (as classified by the analysts, although not necessarily by the original authors). Table 3 gives more detail about the 19 comparison papers, and publication forms are shown in Table 4.

Comparison form	Number
Usability/Readability	9
Modularity	5
OO Fault Detection	1
General Comparison	1
Maintainability	3

**Table 3. Forms of comparison**

Publication Type	Number
Journal	68
Conference forms	59
Technical Report et.	5

**Table 4. Papers classified by publication form**

### 3 Analysis & Conclusions

This work is ongoing and hence the analysis presented here is a preliminary one—focussing chiefly on the evidence ‘clusters’ and ‘deserts’. First though, we provide a short assessment of the possible *threats to validity*.

These threats can be identified as: whether we have truly identified all relevant publications; and whether our initial classification is robust enough for analysis.

The former is dependent upon both our search criteria and the scope of the search. These are influenced by the limitations of the search engines (see [1]) as well as the search terms used. Our chief benchmark has been to calibrate against a known set of references using [2]. (One question is whether such papers might appear under other related headings, such as *design patterns* and *aspect-oriented design*. However, cross-checking with a mapping study being performed on design patterns did not indicate that we were missing any relevant papers.) So, we are confident that our processes have identified most of the relevant papers. The one possible exception is the ‘grey literature’ which may be covered less well by the search engines used, and for which we have no readily-available benchmarks to use as checks.

The question of classification and analysis is one that we plan to extend. In [5] Wieringa and Heerkens have proposed a conceptual framework that we intend to employ for further analysis. However, for reporting work in progress, the scheme adopted should suffice to provide some initial answers to our research questions.

What conclusions can be drawn from this work so far? The initial answers to our research questions are as follows.

1. *Sources of papers on OO design*. These are almost equally split between journals and conference-style events with no dominant forum within these.
2. *Most investigated topics*. Table 1 clearly shows a dominant focus upon metrics (both in terms of devising and using), with a relatively small number that assess the *effectiveness* of the paradigm in any way. Table 5 shows some of the main themes for the metrics papers.
3. *Research methods used*. The use of ‘laboratory’ studies is quite extensive and we are investigating whether this is linked to the emphasis upon metrics as a topic.

Topic	Number
Fault detection/proneness	17
Maintenance characteristics	14
Design characteristics	10
Inter-relationship of metrics	10

**Table 5. Major themes for the metrics papers**

So, overall, we suggest that the (very preliminary) analysis presented here has identified a somewhat worrying lack of empirical evaluation of a paradigm that has become so dominant in software development. There should also be concern that there appears to be a rather significant ‘gap’ in terms of the number of studies that assess its effectiveness when compared to other design paradigms.

### Acknowledgements

This work was performed by the EPSRC-funded study on *Evidence Based Software Engineering* (EP/C51839/X).

### References

- [1] O. Brereton, B. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the Systematic Literature Review process within the Software Engineering domain. *Journal of Systems & Software*, 80(4):571–583, 2007.
- [2] D. Budgen. *Software Design*. Addison Wesley, 2nd edition, July 2003.
- [3] B. Kitchenham. Procedures for undertaking systematic reviews. Technical Report TR/SE-0401, Department of Computer Science, Keele University and National ICT, Australia Ltd, 2004. Joint Technical Report.
- [4] M. Pettitcrew and H. Roberts. *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing, 2006.
- [5] R. J. Wieringa and J. M. G. Heerkens. The methodological soundness of requirements engineering papers: a conceptual framework and two case studies. *Requirements Engineering*, 11:295–307, 2006.