INTERNATIONAL DESIGN CONFERENCE - DESIGN 2004
Dubrovnik, May 18 - 21, 2004.

DESIGN 2004

The Lure of the Measurable in Design Research

Claudia Eckert, P. John Clarkson and Martin Stacey

Keywords: design research methodology, design process, design methodology, systems theory.

1. Introduction: the lure of the measurable

For good reasons, engineers like rational criteria for decisions and hard, precise data to base them on. They like measurements. When engineering companies are considering introducing new tools and procedures into their design processes, they want evidence that they will bring improvements. Accordingly some views of how to do design research place success criteria and measurements front and centre; the DRM methodology involves specifying success criteria and metrics at the beginning of the study [Blessing and Chakrabarti, 2002].

However focusing on quantitative methods and, indeed, on success criteria both devalues major branches of design research, and devalues qualitative analyses of research findings. Our own view of the strategies and methodologies required for design research emphasises the need to integrate different types of research, many of which are aimed at increasing understanding, and only very indirectly at achieving improvements [Eckert, Clarkson and Stacey 2003]. The purpose of this paper is to analyse the role and appropriate uses of success criteria in design research, and point out the pitfalls of using quantitative analyses in design research.

2. The scope of design research: understanding and changing design

Research into design has two objectives: understanding designing as a complex human activity, and finding ways to improve the effectiveness and efficiency of design processes. (Some design research focuses on understanding the nature of designed artefacts themselves, but the forms of descriptions and models are intimately linked to their human purposes.) In our experience, understanding-directed research often suggests opportunities for improving tools, methods and processes, as well as further research questions more directly aimed at finding improvements. We view studies aimed at enhancing our understanding of design as both worthwhile in their own right and a crucial part of the academic contribution to design process improvement. Introducing new computer systems or procedures into complex human activities like design is dangerous: without a subtle understanding of how a design process really works, making changes risks disrupting important aspects of designing, or simply failing to meet designers' real needs. Among software engineering and information systems practitioners and researchers this is a truism, founded on the bitter experience of dozens of multi-million-dollar software projects that have failed because the developers have failed to understand the users really do their jobs and what their needs really are [for instance Bronzite, 2000]. Well managed software projects specify a wide variety of functional, non-functional and usability requirements, some of which can be given quantitative form; and progressively revise their requirements specifications.

We have argued elsewhere [Eckert et al, 2003] that applied design research should encompass eight distinct types of research objective – the eightfold path: (1) empirical studies of design behaviour; (2) evaluation of empirical studies; (3) development of theoretical understanding; (4) evaluation of theory;

(5) development of tools and procedures; (6) evaluation of tools and procedures; (7) introduction of tools and procedures into industrial use; (8) evaluation of the dissemination of tools and procedures. Individual projects may only cover one or a few of these types of activity – and can begin with empirical research, theorising, tool development, or making changes to industrial practice. But any project should be grounded in a clear view of how it fits into the context formed by other types of research. In practice, these different types of research are often carried out in parallel. While DRM [Blessing and Chakrabarti, 2002] encompasses all these activities, it is very narrowly focused on research aimed at the development of tools and methods, and prescriptive about which research objectives a study should include. Accordingly we regard it as only relevant to a limited subset of the research relevant to design process improvement.

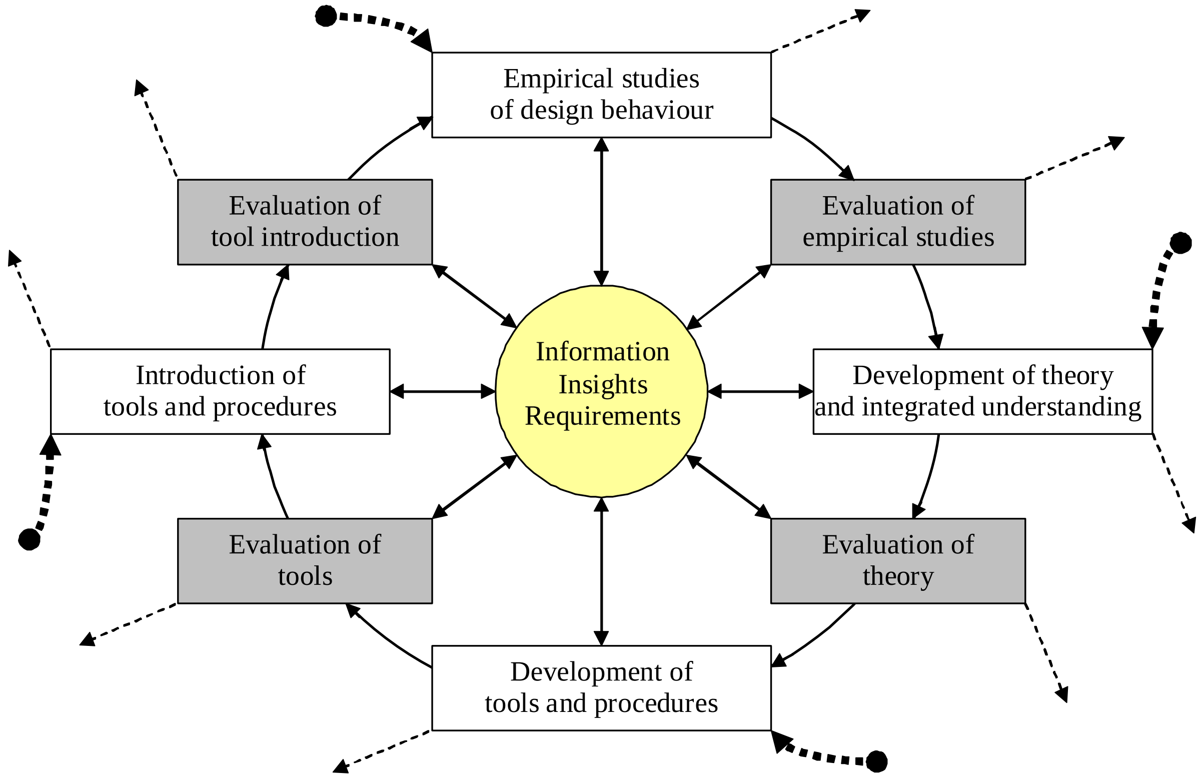

Figure. The Spiral of Applied Research: the eight types of research objective

3. The scope of design research: a complex human activity

Design, especially large-scale engineering design, is a complex activity that can be studied at several different scales, using the research questions, theoretical constructs, methodologies and critical standards of a variety of contributory disciplines, including cognitive psychology, social psychology, sociology, and organisation theory, and employing conceptual tools drawn from philosophy, artificial intelligence, mathematics, systems theory and complexity theory, as well as the design disciplines themselves. So design research has no single methodology or characteristic form of knowledge.

These disciplines give us tools to understand layers or aspects of design, such as the thought processes involved in conceptual design, or the types of information expressed in design meetings. But as design researchers we are especially concerned with understanding and making changes to complex and highly structured systems of human activity. Solving a design process problem means dealing with the complex interaction of a variety of causal influences operating at the different levels studied by different academic disciplines [for example, Eckert, 2001]. We have advocated documenting understanding of design processes by mapping these causal influences [Stacey et al., 2002]; similarly

Blessing and Chakrabarti [2002] recommend identifying the network of influencing factors as part of the descriptive study stage of DRM, and planning and assessing change in terms of changes to them.
The *human activity systems* that create designs involve a variety of human participants with different roles, skills, and perspectives, plus a variety of objects that represent and convey information, and act to structure activity. But here again design is not unique. There is now a large amount of research, theoretical analysis and practical experience of how to understand such human activity systems and plan and make changes to them. Developing systems theory to deal with human activity systems, Peter Checkland [1981] and his collaborators have developed *Soft Systems Methodology* as a rigorous philosophically grounded procedure for developing understanding and planning changes. SSM stresses the need to understand the different actors' views of the systems they participate in, and avoiding treating particular views of roles, organisational structures, et cetera, as objectively true. SSM and other qualitative sociological methods are widely used in software development projects in developing the accounts of activities and needs from which requirements can be derived.

4. Assessable objectives versus measurable criteria

Checkland [1981] argues that a methodologically rigorous analysis of a human activity system will be essentially qualitative, and quantitative metrics are of very little value. And some of the various academic disciplines that contribute to understanding design are purely qualitative, using qualitative methods to assess the validity of causal hypotheses and models.
Others such as cognitive psychology use metrics to support essentially qualitative analyses: to assess the relative importance of different contributory factors, or to assess whether some factor influences the phenomenon or not, to see whether a model accounts for the data. While there can be value in exploratory studies based on the expectation that a particular kind of data will prove interesting, it is usually only possible to devise appropriate metrics or ways to gather quantitative data when one has a good understanding of what one is looking for. This requires either an analysis of research into related phenomena (in our case, other design processes) or observations of the phenomenon under study (in our case, some design process) that suggest hypotheses about how it works or how it will respond to changes. Similarly, we think that formulating success criteria for changes, that are specific enough to be useful, needs to follow detailed analysis.
So we think that beginning design research by choosing metrics will only be appropriate or feasible in a very narrow range of situations. Our view is that studies should begin with *questions*, and that these guide the formulation of objectives. Thus researchers should think carefully about what sorts of results would constitute answers to their questions, and how the validity and value of these results can be assessed. Typical results for our own work are explanations of communication problems as networks of causes and effects [Eckert, 2001], accounts of how companies handle change, models of change propagation [Eckert, Clarkson and Zanker, in press], and techniques for assessing the risk of changes propagating to other parts of a system [Clarkson, Simons and Eckert, in press].
But in our experience, the results produced by design research – especially observational case studies of design processes – can be very different from the results expected. We find that the causal processes that influence particular design processes are often different from those we expected to see, and, also, that engineers sometimes find value in using tools and methods in ways different from those we anticipated. So empirical design research can – and should – involve reconsidering what results are there to be achieved, what these will look like, and how they can be assessed.

5. Metrics for the success of action research

The primary use for quantitative metrics envisaged by DRM – and that pragmatic engineering designers and managers want to see – is seeing whether a change to a design process has resulted in an improvement. However we think that action research in design – introducing new tools, methods or procedures into a process and seeing what happens – should not be driven by success metrics. Metrics for success can be valuable, but they can also be dangerous, and we don't regard them as a necessary component of good research. Sometimes such numbers may be both clear and valid, but the use of success metrics risks over-interpreting results that may be unreliable or the product of factors the researchers have not fully considered.

And focusing on one or a few broad success criteria, and making measurements to assess them, may distract attention from what is really going on. Understanding the subtle and varied consequences of introducing new methods into complex processes may need serious research. The consequences of making changes to design processes – good and bad – will inevitably go beyond the scope of a small set of measures like time-to-completion (as Blessing and Chakrabarti [2002] emphasise). New methods and procedures thus have a corresponding variety of requirements for success, whether these are recognised or not: that these manifold consequences should be beneficial or at least not significantly harmful. The range of potential changes that should be recognised may require a variety of different qualitative or quantitative evaluation techniques. The lessons from large-scale software development are that many qualitative and quantitative success criteria need to be specified and considered in parallel, and that detailed qualitative analyses of the structures and causal influences that shape the processes may be required to choose appropriate metrics, or to recognise qualitative changes to design processes caused by the introduction of tools and procedures.

Analyses of the effects of changes to design processes need to recognise that the design processes for individual products differ in ways that researchers cannot easily control – the problem is different, or the designers are different, or the designers have more experience. Studies of small projects done by students can observe enough very similar cases to control for the major sources of variation and obtain statistically significant results, but this cannot be done with large-scale real-life designing in industry.

Psychologists are familiar with the Hawthorne Effect, first observed when dramatic improvements in the productivity of an experimental group of assembly workers at the Western Electric plant at Hawthorne, Illinois, turned out to be due to the improvements in morale and motivation resulting from the attention researchers and managers gave to changing their working conditions, not from the changes themselves [Roethlisberger and Dickson, 1939; Homans, 1950]. The productivity of groups of workers depends crucially on motivation and social relationship factors, that can be disrupted in unanticipated ways by new procedures [see for instance Schein, 1988].

Our view is that action research should aim for insight first rather than evaluation: if the consequences of changes are *understood*, useful evaluations will result. Our approach to assessing the utility of new tools and methods is to look for acceptance in industry: how are the tools used, and for which purposes; who finds value in them; whose working practices are affected, and in which ways. For instance our change propagation prediction tool has been championed by individuals within our partner company; it gets used in review meetings to check the completeness of the review, which we had not anticipated [Jarratt, Eckert and Clarkson, submitted]. We have sought to assess the validity of the method by seeing how many of the changes it would have predicted in historical cases.

6. Some dangers in using quantitative metrics

Disciplines such as psychology that rely on numerical measures for testing hypotheses, assessing the applicability of models, and estimating numerical parameters are aware of the methodological problems they pose. The pitfalls of measuring complex social systems are especially tricky, so metrics should be used with caution.

- *Validity*: Does the metric correspond sufficiently closely to the aspect of the phenomenon one wishes to measure? For many kinds of psychological and sociological study, what a metric is a measure *of* is a problematic issue; for instance what intelligence *is* and how is it related to IQ tests has been a controversial issue for decades [for instance Anastasi, 1976].
- *Reliability*: Is the metric sufficiently stable to be a useful measurement of anything? Physicists and engineers can perform sensitivity analyses, to see if the value of the metric changes sharply with small changes in the input parameters. Psychologists can administer a test again or repeat the experiment to assess the variation due to random factors. Design researchers who can do neither should be careful with estimating the reliability of their results.
- *Generality*. How general is the relationship between the value of a metric and the factors that determine it? What predictions can be made, and with what confidence, about the behaviour of

- *Coarse measure*: The metrics that are used as success criteria for action research in design, for instance in DRM, usually measure some very broad, large scale parameter, such as time to complete a design task. The values of such metrics are influenced by a wide range of factors, not all of which have been controlled or considered. Thus they may tell one something of the overall success of the task but little about what happened and why. Design researchers should beware of underestimating the range of significant success factors that a metric represents.
- *Narrow measure*: A metric that is governed by only one or a few aspects of what really constitutes success for the process. Design researchers should beware of overestimating the range of significant success factors that a metric represents, or the extent to which a metric is indicative of overall 'goodness'.
- *Causal measure*: When the producers of good scores on the metric are rewarded, behaviour is distorted towards achieving good scores at the expense of other kinds of success.

Above all, design researchers like others should beware of the lure of the measurable: attaching undue weight to causal factors that are easy to measure versus factors that are hard to measure, or attaching undue value to benefits that are easy to measure versus benefits that are hard to measure. We have observed causal factors in design processes, such as the designers of complex engineering products lacking an overview over the product and the process, that appear to be widespread and to cause serious problems, but which we do not know how to measure meaningfully [Jarratt et al, submitted].

7. Conclusions: some methodological principles for action research in design

The use of success criteria for action research in design, as advocated by Blessing and Chakrabarti [2002] in DRM, has a number of problems, that lead us to the conclusions, first, that designing studies around success criteria is a mistake, and second, that qualitative metrics and success criteria are of limited utility in evaluating the success of introducing new tools, methods and procedures into design processes in industry.

- A design process is a human activity system that combines aspects that can be described using a variety of conceptual frameworks and methodologies. Success criteria and metrics may obscure the complex causal influences that shape the process. Understanding these causal influences requires developing rich qualitative analyses.
- The real requirements for a successful change to a design process are many and varied. The consequences of making changes need to be recognised and assessed in ways appropriate to their nature, which may be qualitative.
- Success criteria and other metrics need to be valid and reliable to have any value. In action research in design, achieving reliability is an especially difficult problem. This may make using apparently-valid metrics for specific success criteria infeasible in practice.
- Using metrics may lure researchers into overestimating the validity of their metrics as measures of success, and overvaluing aspects of design processes that can be measured at the expense of aspects of design process than cannot.

Our own experiences of empirical design research lead us to some methodological conclusions.

- Research aimed purely at understanding design is an essential part of the academic contribution to design process improvement. It frequently suggests ways to improve design processes with new tools and methods, as well as more improvement-directed research questions.
- Applied design research can begin and end with any of the eight types of research objectives (see Figure). But it should be grounded in the context formed by related research in the other categories.
- Empirical research in design should begin with *questions*, and a consideration of objectives: what kind of results the study should produce, and how the validity of the results can be evaluated.

- Empirical research into design processes produces surprises, and findings that weren't anticipated at the beginning of the study. Researchers can, and should, ask new questions and reconsider their objectives as they learn more, and take opportunities to investigate different aspects of design.
- Empirical research should focus on developing understanding. Evaluations of the consequences of introducing tools, methods and procedures into design processes will follow from understanding what those consequences are.
- The most useful criterion for success is the perception of value in new procedures and methods by design practitioners in industry.

Quantitative measures have useful roles in design research, including assessing the efficiency of tightly defined procedures, and how well software tools meet specific requirements. But criteria and methods for evaluating success depend on the research questions and methodology. Many types of research that are important to understanding design are purely qualitative.

Acknowledgements

This research is funded by an EPSRC Innovative Manufacturing Research Centre grant to the Cambridge University Engineering Design Centre.

References

Anastasi, A., 1976, "Psychological Testing", Macmillan, New York.

Blessing, L.T.M. and Chakrabarti, A., 2002, "DRM: A Design Research Methodology", Proceedings of Les Sciences de la Conception, INSA de Lyon, Lyon, France.

Bronzite, M., 2000, "System Development: A Strategic Framework", Springer, London.

Checkland, P.B., 1981, "Systems Thinking, Systems Practice", Wiley, Chichester, UK.

Clarkson, P.J., Simons, C. and Eckert, C.M., in press, "Predicting Change Propagation in Complex Design", Journal of Mechanical Design.

Eckert, C.M., 2001, "The Communication Bottleneck in Knitwear Design: Analysis and Computing Solutions", Computer Supported Cooperative Work, 10, 29-74.

Eckert, C.M., Clarkson, P.J., and Stacey, M.K, 2003, "The Spiral of Applied Research: A Methodological View of Integrated Design Research", Proceedings of the 14th International Conference on Engineering Design, The Design Society, Stockholm.

Eckert, C.M., Clarkson, P.J. and Zanker, W., in press, "Change and Customisation in Complex Engineering Domains", Research in Engineering Design.

Homans, G., 1950, "The Human Group", Harcourt, Brace, New York.

Jarratt, T.A.W., Eckert, C.M. and Clarkson, P.J., submitted, "The benefits of predicting change in complex products: application areas of a DSM-based prediction tool", submitted to Design 2004.

Roethlisberger, F.J., and Dickson, W.J., 1939, "Management and the Worker", Harvard University Press, Cambridge, MA.

Schein, E.H., 1988, "Organizational Psychology", 3rd ed., Prentice-Hall, Englewood Cliffs, NJ.

Stacey, M.K., Eckert, C.M., Earl, C.F., Bucciarelli, L.L. and Clarkson, P.J., 2002, "A Comparative Programme for Design Research", Proceedings of the Design Research Society 2002 International Conference: Common Ground, Brunel University, London.

Claudia Eckert, John Clarkson
Engineering Design Centre
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge CB2 1PZ, UK.
Email; {cme26, pjc10}@eng.cam.ac.uk

Martin Stacey
School of Computing
De Montfort University
The Gateway
Leicester LE1 9BH, UK
Email: mstacey@dmu.ac.uk