Focus, Fit, and Fervor: Future Factors Beyond Play With the Interplay

Gilbert Cockton
School of Computing & Technology, University of Sunderland, UK

This special issue advances our understanding and exploitation of evaluation methods. It continues to move studies from methods to evaluators and their customers. Thoughtful studies show when and why methods have specific impacts, supporting a focus on ways to better target method use. It is also the case that a focus on how evaluation customers use usability results reveals the fit between methods and customers. However, evaluation and its influence on redesign is not solely a matter of evaluators and designers or developers. Evaluation and redesign are activities within broader system development contexts. Improvements need commitment and support from management. Organizational fervor for user-centered design is key to effective evaluation. Without management support, usability evaluation cannot be well planned, properly conducted, or effectively exploited, leaving usability specialists to just play with the interplay between design and evaluation. This article reviews contributions in this special issue, noting where they address specific factors in the interplay between design and evaluation and identifying factors that require more attention in future research.

1. INTRODUCTION: FROM CRISIS TO CRISPNESS

Articles in this special issue address specific factors in the interplay between design and evaluation. They represent advances in effective usability evaluation, moving from a focus on assessment of usability evaluation methods (UEMs), in terms of finding usability problems, toward a broader view of the impact of evaluation within iterative development. They are further evidence of a sustained move away from the “Damaged Merchandise” crisis toward research that can and does have a positive practical impact on usability practice.

In 1998 a special issue of the journal Human–Computer Interaction comprised a critique of assessments of evaluation methods (Gray & Salzman, 1998), a series of commentaries on this article, and a rejoinder. The charge of Damaged Merchandise represented a culmination of concerns raised since 1993 on the quality of UEM as-
sessments (Muller, Dayton, & Root, 1993). Since then, methodological improvements have moved UEM assessment from crisis to crispness. Hertzum’s (this issue) article combines diary studies and action research to relate UEM effectiveness to both the UEM itself and its timing within development, and then it investigates correlations between severity, frequency, and other measures of likely problem impact. Law (this issue) presents methodologically sophisticated analyses of the impact of usability problem reports on development behavior and redesign quality. Høegh, Nielsen, Overgaard, Pedersen, and Stage (this issue) complements Law with a field experiment employing qualitative techniques, including Grounded Theory and opinion condensation within semistructured interviews. Blandford, Keith, and Fields (this issue) presents a case study on adapting Claims Analysis to a specific development. Their action research combines intervention and data collection. Troost, Venturi, and Jokela (this issue) presents results of a practitioner survey. This supports assessment of the ecological validity of the other articles via a broader view of usability within organizations.

No article will have the last word on methodological innovation. However, these articles better meet the criteria applied by Gray and Salzman (1998). Authors guard against statistical conclusion invalidity. Similarly, caution and candor guard against blatant conclusion invalidity. For case studies, internal validity will always be a challenge, but this is acknowledged and addressed to let readers make decisions on what can safely be generalized. Gray and Salzman treated construct validity as separate from internal validity (many see it as a type of the latter). It has been addressed through grounding in both problem report formats and redesign impact, but some critical causal constructs remain elusive. For example, Law (this issue) notes (a) the difficulty of accessing subjective factors that really drive decision making and (b) the impact of using more accessible objective surrogates. Case studies have high external validity for similar development contexts and hence bring some ecological validity, which Gray and Salzman did not cover separately.

Research methods applied in this special issue reveal when and why UEMs succeed and fail and provide support for best practice in UEM usage. There is no doubt that practitioners and researchers in 2008 will be far better served in selecting and using UEMs than in 1998. The evidence for this lies in the very focus of this special issue. For much of the last decade in human–computer interaction, the focus has had to be on the reliability and effectiveness of UEMs in finding usability problems (Lewis, 2001). However, as Wixon (2003), among others, pointed out, having found problems, they need to be understood and fixed. Still, normative positions on how evaluation should be related to redesign need to consider empirical studies of how they do relate. The articles in this issue go beyond the accuracy of problem lists to assessing their persuasive power and actual impact on both the practice and outcomes of redesign. In so doing, they continue the user-centered focus on evaluators in recent research (e.g., Cockton, Woolrych, Hall, & Hindmarch, 2003) and extend this to other stakeholders in the iterative steps from evaluation to redesign and beyond.

This summary article first reviews how more methodologically extensive, rich, and sophisticated assessments of UEMs improve research and practice. A more effective focus on the key factors in evaluation and redesign success is accompanied by assessment of fit between UEMs and development. Second, it balances praise
for progress in research and practice with caution on more powerful drivers, of not only management fervor but also the craft nature of software, which need not be designed and is generally made. All articles address focus to some extent. Only Blandford and colleagues (this issue) attempt fit: The other three case studies instead express some surprise at the lack of it! Only Troost and colleagues (this issue) directly address fervor, although all articles identify development mindsets as key to effective evaluation. I thus begin with a lengthy focus on focus! This has moved us beyond the methodological crisis of Damaged Merchandise to something that is at least Reconditioned (Cockton et al., 2004).

2. **FOCUS**

The five studies criticized by Gray and Salzman (1998) attempted to find the “best” UEM, or at least to find where each did well relative to others. How and why UEMs performed differently received little attention, but this underlays many highlighted validity problems. Had each study focused more on actual UEM usage, they may have been able to explain why UEMs perform differently. The naive implicit assumption was that usability specialists only need be directed toward the best UEMs. Over a decade later, the pride and advocacy of early UEM inventors is replaced with a concern for the effectiveness of fellow usability specialists. This concern is supported by a focus made possible by more sophisticated research methods, instruments, and measures. These shift attention from Gray and Salzman’s critique of simple textbook errors in experimental design and study reporting. The authors’ candid reflection on the effectiveness of innovative research instruments and measures should direct future critiques at collaborative improvement rather than methodological indignation.

Hertzum (this issue) concentrates first on the relative effectiveness of variants of user testing at different points in development. Understandings here were essential to ensure external validity for his second experiment on severity assessment. Hertzum recognizes that development factors could neutralize the effectiveness of the “best approach to severity” from controlled studies. He shows that the timing of a usability study determines much of its effectiveness. Several factors are responsible, including the inertia and resistance of software architectures once internal quality attributes are optimized, but most of all, the effort available for changes simply gets reduced toward the end of development. So, whatever the quality of the problem list and however severe the usability problems found late in development, there simply may be no resource available to make essential changes. Even were further funds available, developers may have been irreversibly committed to other work.

Aware of the consequences of when UEMs are applied, Hertzum next considers severity ratings. Clearly, we would expect the most severe usability problems to be prioritized for redesign—but what makes a problem so severe that it must be fixed? This practical matter has largely been overlooked by researchers (me and colleagues included) when assessing UEMs. Where severity is considered, it tends to be at a more strategic level of how we can relate severity ratings to business impact.
and achieved value (Cockton, 2005). In earlier usability work, severity has been considered, but the tendency has been to invent scales that have never been calibrated against actual problem impact and redesign effectiveness. Hertzum’s conclusions are sobering, arguing that, while we need one scale for severity, we currently have no basis for grounding such a scale in factors such as frequency, predicted persistence, task impact and business impact. Risks associated with unreliable severity ratings must be closely managed. Hertzum suggests involving users, which is worthwhile, but if ‘users’ is synonymous with ‘end-users’, then other stakeholders must be involved. Incidentally, best practice in some UK organizations already includes collaborative impact workshops with all development stakeholders.

Hertzum’s research toolbox lets him focus on the impact of when UEMs are used and how severity ratings are made, and how these relate to measures of frequency and impact. Law’s (this issue) wide research toolbox lets her focus on a broad range of interactions, keeping an open mind about specific interplays. Law could thus look for, but not find, significant correlation between the severity of a problem and the effort required to fix it. The interplay between design and evaluation is not a simple mapping from severities to fix effort and effectiveness. It is as realistic to fix a severe problem as it is to fix any other problem. However, in Law’s study, the e-learning system was developed within a collaborative research project and was not central to the development partner’s commercial strategy. This let developers avoid redesign, instead “irresponsibly” allocating responsibility to users (through training) or system administrators (through configuration?). This irresponsibility may however reflect developers’ perceptions of fix costs, which Hertzum noted are often due to software architecture factors that are independent of the severity of problems experienced by users.

The (unsurprising?) lack of simple correlations between severity and fix cost leads Law to focus on the influence of development effort on redesign outcome. This time she finds that effort and effect do not correlate, that is, there is no direct correspondence between the development resources applied to a fix and its effectiveness. Law finds this counterintuitive and discouraging, a surprising position in a naturalistic inquiry, where the aim cannot be to test hypotheses, as too many potential confounds cannot be controlled. Even independent variables such as severity ratings and developer effort are beyond control. The research approach thus needs to focus on uncovering relationships and then examining the data further to look for explanations. Discoveries can be made despite poor measures, which led Law to uncover more intricate interactions between problem frequency, elaborateness of report, and fix effectiveness. Verbose reports for infrequent problems were associated with ineffective fixes. In reports for frequent problems, evaluators progressively add detail for the same problem and, in the process, add credibility and causal insights that are vital to effective redesign proposals. Conversely, a low signal-to-noise ratio in padded reports will offer little insight or value to developers.

Although post hoc, such explanations are the main result of naturalistic inquiry, which let Law find interactions of both developer effects (commitment, motivation, irresponsibility) and evaluator effects (elaborateness of problem and redesign descriptions) in determining whether a fix was attempted and was effective. This is
despite the use of word count as a measure of problem and redesign “quantity.” Although few would hypothesize a causal link between verbosity and redesign insight, through using available measures, Law uncovered something that high standards of construct validity would have hidden. Law’s discovery clearly establishes the value of open-ended case studies that let a range of instruments and measures be applied, even when potential relationships appear to be counterintuitive. Whereas an experiment would not manipulate report length as an independent variable, quasi-experimental designs in case studies (naturalistic inquiry) harvest many variables, which, if investigated with an open mind, can open up new avenues of exploration for research. Practical advice also follows. Law thus recommends that claimed severe usability problems are “qualified with accurate identifications of major concerns,” that is, that details of specific user difficulties are provided (Cockton, Woolrych, & Hindmarch, 2004; Lavery & Cockton, 1997).

The discovery shifts Law’s attention to the quality of redesign recommendations. She suggests that users may have better ideas than evaluators. This is fine occasionally, but if it predominated, this would question evaluators’ design competence. When only developers and evaluators fix designs, this may turn out to be a key factor in evaluation effectiveness. Where there was little serious initial design, it is no surprise that evaluators can find many usability problems, but both evaluators and developers may not effectively redesign what was never designed. The ineffectiveness of several key fixes in Law’s study indicates that evaluators are not infallible, acknowledged in her position that redesign proposals are “sources of inspiration rather than prescriptions or fixes.” In the absence of adverse software architectures, developers may unwisely accept fixes uncritically.

Law’s extensive analysis of problem and redesign descriptions focuses on the impact of how evaluation results are communicated. For her developers, the availability of a redesign proposal was a significant predictor of their response. However, the least effective fixes were associated with multicausal problems where a major redesign may have been the only effective response. Presumably, the redesigns addressed design-specific user difficulties rather than returning to the initial solution and starting anew. Some interplay between evaluation and design may be counterproductive. Feature optimization and “point design” experiments (Czerwinski, 2004) cannot overcome the “big picture” problems of inappropriate designs (Dye, 2001).

Iterative design can be at its most effective when it throws major parts, if not all, of an existing design away. There is little evidence and no explicit discussion in this special issue of this ever happening, suggesting that the interplay between design and evaluation in these studies is locking developers into unrescuable “designs” via hopeless attempts at feature optimization. Still, far more obstructive in Law’s particular study was the commercial irrelevance of the project to the developing company. Given this, what was achieved through iterative development is a credit to the collaboration within the project. Clearly many problems were fixed, but issues of construct validity do arise over “persuasiveness.” Were the problem and redesign reports persuasive? The reports could have equally been “undemanding” advice that could be followed uncritically—not what is normally understood by “persuasion.” The study did not examine perceived difficulty to fix or the user ben-
benefit of a fix. These must surely be factors in persuading developers but were beyond the study. This is a risk in all naturalistic inquiry, as constructs must be anticipated to collect data that can be coded for them. Currently, we do not have a good enough grasp of impact to reliably collect relevant data on perceived benefit. However, collecting what we can does advance understandings. The extended problem report format used in my research with colleagues was known to have potential limitations (Cockton et al., 2003), yet it still revealed previous unknown relationships between problem discovery heuristics and prediction quality in usability inspection as well as suggesting interactions between evaluators' cognitive resources and prediction outcomes (Cockton et al., 2004).

The main research challenge is to move problem coding beyond the existing support for measures of thoroughness, validity, appropriateness, and similar problem list measures as the basis of UEM effectiveness. We must operationalize persuasiveness to measure it. Furthermore, internal validity for measuring fix efficacy is threatened if proposed fixes are just “sources of inspiration.” Before we can reliably measure fix efficacy relative to an actual design change recommendation, we must identify the developer resources that bridge from inspiration to the final fix.

Høegh and colleagues move Law’s focus on how evaluation results are communicated to developers beyond written reports. They compare the impact of written reports with direct observation. Developers have different responses to each. Both are needed to persuade developers to make changes and to guide them toward specific redesign options. In their second study, the written reports were seen as requiring some video support, but adding discussions was seen to improve developer understanding. Both studies identified gaps in problem reporting, such as insufficient causal analysis, but more social solutions based on joint meetings may overcome written report limitations. However, content could still have a stronger impact than presentation. Developers thus challenged who had been selected to test the system and what they had been asked to do, indicating that the wider issue of evaluation planning is a further consideration in the interplay between design and evaluation.

The three case studies only address some aspects of evaluation planning. All look at how evaluation results are communicated and factors here receive the most attention overall. Høegh and colleagues reveal that persuasiveness depends on who tests systems and what they are asked to do. Hertzum demonstrates that whether problems get fixed depends on when systems get tested, as well as considering where evaluations took place, how evaluations are conducted, and how data are analyzed for severity. However, we must address the full range of questions for evaluation planning, which must include the following:

- Why are we evaluating?
- Who will we ask to evaluate (both specialists and participants)?
- What information and design artefacts are given to evaluators, what resources are available for conducting the evaluation, and what will we ask specialists and participants to do?
- Where will we carry out the evaluation?
- When will we carry out the evaluation?
Which instruments will be used to record which measures, and which levels of performance and preference will indicate success or failure?

How will we conduct the evaluation overall (protocol), how will we analyze the data, and how will we communicate the results?

The three studies of user testing have mostly assumed, implicitly or explicitly, that the purpose (why?) of evaluation is formative for design, although Høegh and colleagues do note that very early evaluation may check assumptions about users, their wants, and their needs and not just the adequacy of detailed designs. No article focuses on the use of different instruments and measures. Future research needs a more balanced consideration of more aspects of evaluation planning, that is, questions of why, who, what, where, when, which, and how. All impact the quality of problem reporting and discussion, which is the starting point for all interplay between evaluation and design, unless, that is, the designer and the evaluator are the same person.

3. FIT

The three articles on empirical user testing and its impact all reveal competing concerns between the development context and the evaluation process. Blandford and colleagues set out in the full knowledge that misfits are inevitable with analytical methods, which must thus be adapted to fit development contexts. Insights and knowledge from the design of digital libraries were incorporated into a tailored version of Claims Analysis, which was structured to become a form of cognitive walkthrough with direct use of personas and scenarios. Full use of the method needed the skills of a human factors expert, thus sometimes making the designer and evaluator the same person.

The digital library developers were less interested in causes or any detail of the reasoning process behind problem discovery but preferred to move straight from a current poor version to an alternative. One conclusion may be that, although we expect complex interplay between design and evaluation, nothing as sophisticated may happen between development and evaluation. Evaluation can encourage and inspire developers to change their current systems, but little in the way of explicit reasoning can be expected. Blandford et al.’s “solution-focused” developers “did not perceive any value in identifying problems that could not be fixed” and “were not interested in pursuing the theoretical cause of a problem unless that contributed directly to fixing it.”

The main value of their new Claims Analysis appeared to lie in the use of personas, which helped developers identify incorrect assumptions about users. Similarly, when scenarios were used, these could identify incorrect assumptions about usage. However, once identified, the developers appear to move tacitly to “making” a new system without evident redesign.

Despite this, using personas and scenarios, complemented by a human factors expert who could construct claims, did lead to design changes, but the effectiveness of these changes was not systematically evaluated. However, the purpose of evaluation
need not be solely one of formative redesign but, as already noted, may instead serve
as a way to extend and deepen understanding of usage contexts. Thus UEMs can be
used to fix the development process as well as the developed product. However, de-
velopers cannot do this unaided. Management support is essential.

4. **FERVOR**

Troost and collaborators’ (this issue) survey reinforces the focus on usage context in
the digital library study. User-centered design (UCD) activities appear to be most
heavily concentrated early in development. UCD activities that are mostly late in
development may be evidence of engineering-dominated development that will
limit the interplay between any evaluation and redesign. The interplay between
late evaluation and design may not be leading companies now invest their
effort. The focus here is on contextual research and design, with evaluation largely
hidden within design activities. The survey uncovered difficulties in getting access
to test users as well as a preference for basing first designs on accurate understand-
ing of user wants and needs. Initial decisions may have already been taken by mar-
keting and business. Rapid prototyping and discount evaluation methods predom-
inate, allowing understanding of user wants and needs to change, not only because
of (marketing and business) misconceptions but also because users’ views on
wants and needs can change as they experience new technologies.

The pendulum must not swing too far, because new technologies can often re-
shape users’ needs and wants, and thus what was true for market research before
the existence of a full working prototype may no longer be true after extensive use
in field trials. Iteration needs to be driven by systematic evaluation. Up-front con-
textual research is unlikely to produce right-first-time designs. Even so, the inter-
play between contextual research and design may be more critical to product suc-
cess than the interplay between evaluation and redesign, especially if only
fine-tuning results from user testing. Evaluation’s interplay with contextual re-
search also is critical to high-quality evaluation planning. Only Blandford and col-
leagues address this within the context of a specific UEM.

Four articles focus on the persuasiveness of UEMs with developers. Troost and
colleagues indicate that persuading management is equally critical. I hope that re-
cent UEM improvement should improve persuasiveness, but we should not under-
estimate the challenges. The key evaluation planning question (i.e., “Why do we
evaluate?”) has received limited attention in this special issue, and yet the survey
article suggests that it may be the most important. If we can be persuasive (and
learn!) here, it will be worth investing more to provide better answers to the re-
main ing evaluation planning questions of who, what, where, when, which, and
how. The **why** question is the key rendezvous between UCD and sponsoring orga-
nizations. UCD does not yet have all the answers, nor can it. Answers have to come
from mutual understanding of what UCD can do and what sponsoring organiza-
tions need (Siegel & Dray, 2003). Some organizations have extensive UCD capabili-
ties, others do not. There are both internal and external reasons for this. Internal
brand considerations direct some business strategies towards usability (e.g.,
Philips Sense and Simplicity), whereas external compliance may require it in some markets (especially safety critical and high-dependency systems). Different contexts need different methods and processes.

5. FACTORS

Several factors determine the interplay between design and evaluation, and only a few are addressed in this special issue. Poor methodology results in inaccurate UEM assessments (Gray & Salzman 1998), which in turn must result in poor UEM usage. The impact of poor UEMs and poor usage on the interplay between design and evaluation can only be negative. Methodological progress in UEM assessment is thus absolutely fundamental to positive interplay between design and evaluation and is the dominant factor addressed in this special issue. Three articles apply a range of methods to shine new light on evaluation and its interplay with design. This results in detailed analyses, which are reflected in the proportion of this critique that covers research focus. This also reflects my own enthusiasm for new measures and instruments, rather than textbook experimental design alone, as the decisive enablers of research progress. In 1996, Darryn Lavery and I were focused on problem matching and reporting (Lavery & Cockton, 1996). We later moved onto problem extraction (Cockton & Lavery, 1999). More recently, Alan Woolrych, Mark Hindmarch, and I have developed falsification testing and extended problem reports (Cockton et al. 2004; Woolrych, Cockton, & Hindmarch, 2004), which have further extended the range of available research tools. Other researchers have made key contributions, for example, asymptotic testing (Hartson, Andre, & Williges, 2001). As a result of continued methodological innovation, we can now shine a more focused light on UEMs. There is a clear interaction between measurement and insight. Methodologically richer measures and instruments shine a light where none has previously been aimed. New observations can be made, although researchers may also find that they need a better torch! Still, once a new research instrument has been developed, the process of improving it is quickly established (e.g., Cockton et al., 2004). Even so, we have not, and most probably never will, reach closure on methodology. Each new concern with usability evaluation will present new challenges. However, the move toward “instruments-to-go” needs to be focused on making usability evaluation more effective. We are now seeing a sustained move from assessing problem discovery to assessing problem understanding and redesign, which should cheer up Dennis Wixon (2003)! It would be interesting to see how his favored RITE method (Medlock, Wixon, McGee, & Welsh, 2005) fares in systematic assessments.

Methodology is a common factor in all evaluation research. The fit factor is more specific to the interplay between design and evaluation. Unfortunately, the articles in this special issue shed little light on it. Rather, they have examined the interplay between development and evaluation. This is not hair-splitting or pedantic semantics. Design is explicitly not “making,” and most of what we now see as essential to design did not become so until mass production separated craft from manufacture.
In software engineering, craft and manufacture remain closely intertwined and may involve neither designers nor design. Høegh and colleagues’ (this issue) developers thus welcomed a previously denied opportunity to think about design. It should be clear that development contexts that make rather than design will severely limit the possible interplay with evaluation.

As quasi-experimental designs, each study had to “take what they got” as regards developers’ design skills and orientation, but the evidence is generally that these were unfavorable in UCD terms, or at least significantly restrained by organizational factors. Such contextual factors impact design and evaluation differently, resulting in common perverse situations where a sponsoring organization is committed to user testing but unable to exploit its results. Here, the process needs fixing. Blandford and colleagues ended up providing as much support for contextual research as for actual claims formation. Troost and colleagues’ survey indicates that this focus on early development stages matches current emphases in industrial UCD.

The commitment, or fervor, of sponsoring organizations is the third factor addressed in this special issue. The organizational context is critical, because this is where clear motivation for evaluating and designing must originate, and it thus constrains all possible interplays between the two. However, commitment must be underpinned by competence. Sponsoring organizations can be immensely naive about the nature of both design and evaluation and thus cannot effectively manage their interplay. Sponsoring organizations may expect evaluation to simply “find” the problems “in” the product, despite limited support for test planning and focus. They then also specify “fixes” with no dialogue about product goals, without which we cannot determine what is or is not a usability problem. Few project sponsors understand the knowledge requirements for UCD and evaluation. We could thus move from the days of Damaged Merchandise, where evaluators are let down by poor method assessment, toward the reverse, where UEMs are let down by their usage. There is a limit to how far UEMs can be fitted to their users and their customers. There are inescapable information, knowledge, and skills requirements for UCD and evaluation. Once sponsoring organizations combine commitment and competence, project team skills will form the overriding factor in determining the interplay between design and evaluation.

6. FUTURE

Focus, fit, and fervor are thus essential factors in the interplay between design and evaluation, but others may be even more critical. In future, these must be addressed. UEMs can be improved and improved, but the people who use them must be competent, as must people who collaborate with usability specialists to plan, enable, and exploit evaluation results. Case studies in this issue have mostly encountered developer skills and organizational contexts as negative factors in the interplay between design and evaluation.

We need a broader understanding of the nature and scope of evaluation in software development. Case studies largely assume that evaluation is formative and goes beyond problem finding and then explore the barriers to effective support for
redesign. Future research must expand on Høegh and colleagues’ vital observation that evaluation may correct errors and oversights in contextual research as much as to correct product defects. The very nature of evaluation itself is thus an overlooked factor in this issue. The very nature of design is another. This too must be a factor for future research, which should thus do the following:

Distinguish between design and making, and avoid watering down methods to fit development contexts that cannot support true design. We must challenge the craft production of software and employ design thinking that is grounded in the best practices of leading design organisations (Fast Company, 2005).

Establish a clear view of why we evaluate and what interplay we expect for both product and process fixes, guided by organizational fervor for quality in use, fit to context, and value in the world (Cockton, 2005). We should rightly relate UCD methods to the strategies of sponsoring organizations.

Develop UEM packages for learning organizations that will invest in staff skills and high-quality process support—such a package goes beyond method description to provide training and process integration.

In short, we must design UEMs for people, which, after all, is what UCD is all about! We must take several doses of our own UCD medicine, improving not only UEMs but also associated training and accreditation. There never will be walk-up-and-use UEMs, but neither should there be walk-away-and-leave usability reports! The interplay between design and evaluation is a team game where most need to be on the playing field for the whole match.

In summary, although improved research focus can improve UEM fit, we must be careful not to fit second-rate UEMs to ineffective development contexts. We need a wider focus that interrelates all contextual, design, and evaluation methods to what sponsoring organizations want to achieve with a new or revised digital product or service. All usability problems are relative to, and all true severity ratings are absolutely dependent on, the intended value for a digital product or service. The most instructive lessons about the interplay between design and evaluation can only arise in contexts that focus on real impact. Anything else is just play with the interplay.

REFERENCES


