

Developing formative evaluation for complex interaction

ROAST, Chris <<http://orcid.org/0000-0002-6931-6252>> and URUCHURTU, Elizabeth <<http://orcid.org/0000-0003-1385-9060>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/25449/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

ROAST, Chris and URUCHURTU, Elizabeth (2019). Developing formative evaluation for complex interaction. In: NUNES, M.B., ISAIAS, P., POWELL, P., RAVESTEIJN, P. and ONGENA, G., (eds.) IADIS International Conference Information Systems 2019. IADIS - International Association for the Development of the Information Society, 35-42.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

DEVELOPING FORMATIVE EVALUATION FOR COMPLEX INTERACTION

Chris Roast and Elizabeth Uruchurtu

*Department of Computing and Computing & Communication Research Centre
Sheffield Hallam University, Sheffield, United Kingdom*

ABSTRACT

This paper reports on the development, refinement and use of a design instrument for supporting the exploration of interactive tool design. The distinctive focus of the instrument is that it is intended for interactive tools that go beyond the aim of enabling direct interaction. It is intended to tools that by necessity mediate access to rich complex functionality. We argue that the majority of evaluative instruments that HCI designers have at-hand are not particularly useful in this context. We proposed that alternative concepts are of more relevance, specifically the concepts developed in the Cognitive Dimensions Framework. We have developed an instrument to enable and support the critical assessment of alternative complex interaction designs that is motivated by these concepts. Having described the origins of the tool, the paper then describes how it has been applied in three real world design and development settings. The paper concludes with reflections upon how best to refine, encapsulate and further improve the instrument.

KEYWORDS

Interaction Design, Ease of Use, Evaluation, Cognitive Dimension

1. INTRODUCTION

Approaches to evaluating interactive systems are wide and varied and can be broadly judged in terms of the value of their outputs and the effort required in obtaining such an output. Hence, the continued popularity of discount usability techniques, as first proposed by Nielsen (1994), where evaluation outputs can be easily managed, and the skills required perform evaluation are easily developed. Other more design led approaches have also gain credibility, as discussed in McCarthy and Wright (2004) where evaluation is an integrated element of a continuous collaborative process. Despite the success of these user experience and usability focused approaches they are limited in their domain of application. They are commonly applied in context where the functionality accessed is of a relatively limited scope, when compared with the functional power offered by information systems in general.

The technique described in this paper is formative and lightweight in character, and specifically developed to address the usability of a class of interactive system that are intrinsically challenging for users. The challenge arises when users are given access to a wealth of functionalities that imply a range of task beyond those that any individual user is likely to have. In general, the functionality and features of a system are a product of user needs that have informed it, and what can easily designed and developed. Hence the overall functional may not always match what is explicitly needed.

An over simplified example would be a user need to see a say select a file within a group, and also a user need to see a file's details. Faced with these two, a solution may fulfil both and also both in combination (selecting and showing the file details at the same time). Technically offering both is relatively simple their combination may never have been validated with users. A more substantive illustration would be a contemporary WYSIWYG word processor – its functionality may be relatively easily scoped by different contexts of use. However, the impact of complexity can be illustrated when we consider the provision of user defined macro facility in a word processor. According to good design practice, such a feature would be a response to a user need. However, it is unlikely that a user need will have defined the macro language with the level of detail that the feature requires. It is more than likely that the extent of the marco language, its concepts and its control features will be been developed with limited consideration of specific user needs. When treated

as whole, the needs that could be fulfilled by the language exceed those ever evidenced. Evaluating the word processor's support for user tasks, becomes impractical when the macro facility is considered: (i) the scope of user tasks working with macros would be difficult to establish, and (ii) in general user performance would be influenced dynamically by which ever macros are already defined.

Hence, the challenge of evaluating complex interaction is one of having to assess the quality of interaction when the functionality offered is beyond that established through the principle of good user centred design. In this context the Cognitive Dimensions Framework is of value since it is a framework designed to examine human factors associated with human tasks such as programming and thus the tools and environments designed to enable programming. The factors focused upon by the framework are evident but rarely addressed by traditionally in user interface design (Green, 1989 and Blackwell et al, 2001). The framework captures features of complex interaction by characterizing recurring features of interaction when it is unrealistic to define a set of realistic user tasks. Although the dimensions, as such, do not offer an objective means of measuring or assessing the quality of interactive behaviours, they do indicate how common issues are inter-related and have been addressed in other systems. Blackwell (2006) summarizes the range of contexts in which the framework has been used. (For concrete cases, the series psychology of programming workshops (see PPIG n.d.) provide many examples.)

Because of the higher-level perspective adopted by the Cognitive Dimensions, they are designed to be applicable across levels and systems, such as: they are presented in a concise and accessible manner; they are un-biased in the sense of not promoting specific styles of solution. Nine of the dimension concepts are summarized in table 1.

Table 1. A brief summary of some of the dimensions used to motivate the method developed in the paper

Dimension name	Explanation
Abstraction	Ability to introduce new abstractions, degree to which the notation requires abstractions in order to work
Consistency	Internal consistency of notational elements
Error Proneness	Susceptibility of users' to error when using the notation
Hard Mental Operations	Requirement for users to perform complex operations 'in the head'
Premature commitment	Degree to which the notation imposes a particular order of working
Provisionality	Ability to 'sketch' things out
Secondary notations	Ability to make notes to self, add comments
Viscosity	Ease or difficulty of making changes to previous work
Visibility	Ease or difficulty of finding parts of the notation and comparing them

In the rest of this paper we analyze the concept of complex interaction in more detail, along with the rationale for our approach to operationalizing an analysis method motivated by the Cognitive Dimensions Framework. The method developed, termed the tabular approach, is then described. We go on to describe how the approach has been used and refined in three real world case studies. The work concludes with reflections upon how best to ensure the such instruments can be employed effectively and productively in interaction design.

2. COMPLEX INTERACTION AND USABILITY

The concept of complex interaction outlined in the introduction is of interest because of the number of systems, such as enterprise systems, that in principle provide access to extensive functionality and the wide range of users exposed to that functionality. A simple characterisation of complex interaction is when the functional space that the designer makes available to users is sufficiently broad that the designer resorts to offering the users intermediate (mediating) notations for controlling the functional space. In the introduction, we treated a macro language as such a mediating notation. Alternatively, a configuration panel on a desktop app represents

a similar mediating notation. In simple user interfaces, such notations could be viewed as a case of designers wrongly discharging their responsibility by exposing a user to new concepts that are indirectly relevant to their tasks. In general, it would be argued that a good user interface would hide users from that detail.

However, when the functionality supported prevents a coherent and bounded set of user goals to be identified and directly supported, the appropriateness of a mediating notation becomes a reasonable option. Despite best intentions, such notations may never have a corresponding "user need" during design, or "user story" during development. As a consequence, it will be likely that the notation will be novel to a user. This naturally represents a tension for traditional user centred design since the introduction of "un-needed" notations and concepts is presumptive and risks introducing unwarranted complexity.

From a user perspective such intermediate notations demand that they use new concepts beyond that of their primary task (see figure 1). More often than not, an intermediate notation exposes the user to new concepts of controlling information, processes or a computation. In addition, there is the risk of mistakes with such notations having significant unwanted impacts. For instance, in the case of a macro facility, an arbitrary macro is capable of radically altering a document in unintended ways.

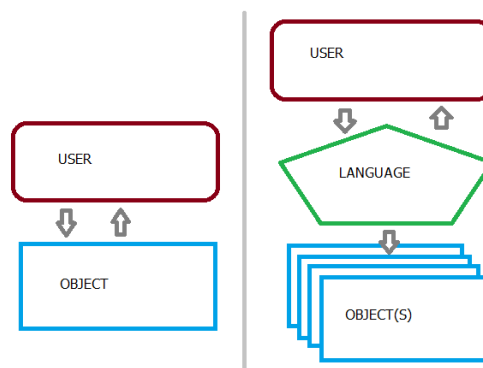


Figure 1. Becoming complex, a conceptual illustration of the impact of moving to a complex interaction view

2.1 Assessing Complex Interaction

Conventional evaluation methods tend to be oriented to user feedback and behavior gathered as a product of conducting realistic tasks that a system is designed to support. For complex interaction this is an unrealistic model since:

- The scale of realistic tasks that the mediating language entertains is extensive. The task space implied by a mediating notation could not be assessed by traditional methods.
- The mediating notation becomes the entity the user manipulates; as such it needs to be a focus for usability evaluation itself along with its constructs, abstractions, etc., it offers.
- With the presence of a mediating notation, knowledge and experience with a notation becomes an important factor.

Together these factors undermine conventional evaluation techniques. By contrast the Cognitive Dimensions Framework has motivated evaluation methods that are in principle better suited for complex interaction, such as API design (Clarke, 2001) and formal specification notations (Triffitt and Khazaei, 2002). Blackwell (2006) offers one systematic method for employing the framework is assessing an existing system. It involves identifying elements of the system assessing each in terms that relate to the dimensions and then iteratively assessing interactive components of a system. More generally, the frameworks common informal use is predominantly analytic and abductive, helping draw associations between different interaction designs. The popularity of this informal speculative use of the dimensions is to some extent evidenced in a recent review of Cognitive Dimensions research (Hadhrawi, et al. 2017).

3. INSTRUMENT DESIGN

In contrast to traditional methods, our approach was motivated by the desire to encourage the exploration of the different perspectives that the framework encourages and thus provide more formative feedback. Both of these points implicitly promote opportunity for the framework to expose designers and users to new ideas and new perspectives. The governing principles for the instrument design were:

- to maintain the generality and reduce the risk of the approach being considered inapplicable;
- to ensure a lightweight flexible process that can be easily managed and adopted;
- to support formative assessment through a descriptive process capable of encouraging reflection.

Our approach has been led by ensuring that the problem owner frames the context for the evaluation in their own terms. Clearly considering the range of potential settings where complex interaction may become important or evident, an untailored generic framework risks appearing irrelevant. To support this we use framing concepts to contextualise the assessment. The framing concepts are developed in collaboration with the problem owner.

To encourage reflection, and minimize implicit bias, the framing concepts are used to instantiate a number of questions. Permuting the use of the concepts the instantiated questions presents a number of comparable and thought provoking cases. As a toy example, if the evaluative question were, say:

"How easy is it to find the **A** for a specific **B**?"

and the scoping concepts were "time" and "train", both of the following questions should be considered:

"How easy is it to find the time for a specific train?"

and

"How easy is it to find the train for a specific time?"

For convenience these permutations of evaluative questions are presented in a tabular format, this helps convey the simple focus upon exploring and contrasting answers to questions, and thus encourage reflection.

The evaluative questions were derived from a series of questions corresponding to the different dimensions developed in Blackwell and Green (2007) and used in Roast and Khazaei (2006).

By considering permutations of framing concepts, the number of question instances to be presented can grow rapidly. To ensure the approach was lightweight the number of scoping concepts was limited to only three. While this weakens the reach of the evaluation, it does reduce the risk of temporal evaluation bias.

Temporal evaluation bias: when evaluation procedures are sufficiently resource demanding that the consistent performance of activities overtime becomes unrealistic.

Over all there are 21 tabular form questions addressing nine of the cognitive dimensions. Eight of the questions are 'binary' in style (concerning the inter-relationship between framing concepts) as illustrated in figure 2. Notice that the framing concepts are indicated by placeholders A, B and C. The other main type of question is 'unary' in the sense that they do not concern the explicit relationship between framing concepts but they do implicitly encourage comparative assessment of some concept, illustrated in figure 3. In this case the same concept placeholders A, B and C are used, but there is the opportunity to consider them singularly or as collections. The other four questions are general and cannot be easily associated with the framing concepts. One example of such as question is: "What's the best way of preparing to use the tool?". We summarise the range of questions, their tabular form and associated dimensions in table 2.

The current form of the tabular assessment framework is publicly available Roast and Uruchurtu (2019).

What are the dominant / common ways in which these concepts are shown together or reached from one another?

to from	A	B	C
A			
B			
C			

Figure 2. A generic tabular question to examine how easy it is for the user to navigate (Corresponding to “Visibility” in the Cognitive Dimensions Framework)

Indicate the relative ease of changing or editing instances?

A(s)	B(s)	C(s)

Figure 3. A unary generic tabular question to examine ease of change (Corresponding to “Viscosity” in the Cognitive Dimensions Framework)

Table 2. A collation of the Cognitive Dimension based questions and their question types

Dimension name	Single questions	Unary questions	Binary questions
Abstraction	-	2	2
Consistency	-	2	-
Error Proneness	1	-	-
Hard Mental Operations	2	1	1
Premature commitment	1	-	2
Provisionality	-	2	-
Secondary notations	-	1	-
Viscosity	-	1	1
Visibility	-	-	2

4. USE AND REFINEMENT

Here we summarize the uses of the approach and its refinement using three real world case studies

4.1 A Rule Based System for Media Production Management

The initial application of the tabular approach was in a software development company working in digital media management. They were developing a media production management tool to reduce the effort and errors encountered with the bespoke crafting of media assets for when they are localized for individual languages and regions. When performing localization, assets (such as posters, blu-ray box covers, blu-ray menus, etc.) are re-designed to reflect regional requirements and sensitivities. Hand crafting such re-designs is time consuming and requires extensive quality control. The tool developed by our collaborator focused upon improving the

efficiency of this process by allowing layout rules to be defined, so from one original, localized versions could be generated in a consistent and approved form. For example, a rule might be to scale a film title to fit the space available on a poster, if its translation lengthens it. The rule based approach was similar in capability that of web technology style sheets (CSS), but presented and manipulating through a GUI interface.

Despite the power and promise of the tool, the principle challenge for users was that the rules constituted a new intermediate notation. In keeping with our notion of complex interaction, the rules distanced users from their traditional design domain.

A comparative assessment of the rule base tool was conducted contrasting two versions of it. The versions differed in terms of how rules were presented: *version A* was asset oriented showing the rules that influenced asset components when they were selected; and *version B* presented all the rules highlighting those influencing any selected asset component.

In keeping with our proposed tabular approach, we initially established framing concepts. Following interviews with users and stakeholders, we settled upon: "asset components", "styling rules" and "target language". Using these, the tabular form was used with developers and product owners within the company to contrast the two versions.

The outcomes highlighted the core importance of rules and issues focused around: the differentiation between rules; the abstract level of rules; and, the inability to annotate rules with their intended purpose. Overall the outcomes identified version B as being more fit for purpose.

Post engagement reflection on the use and appropriateness of the tabular form was conducted with academic peers. The tractability of the tool as a means of gaining coverage of a number of views of potential usability issues was seen as a benefit. However, the limit of three scoping concepts clearly biased stakeholder engagement. Despite that it was noted that the framing concepts limited scale of the activity although respondents used the concepts flexibly.

4.2 Authentication Monitoring and Management

A second case study employing the tabular form concerned a tool for the monitoring of online resource authentication within higher education. The context was one of overseeing and analysing the numerous differing online library subscription services. By way of illustration, a library manager may wish to identify whether, say, computing students use the ACM Digital library (www.acm.org/dl) on a comparable basis to IEEE Explore (ieeexplore.ieee.org), or whether Sciencedirect (www.sciencedirect.com/) is used in preference to both. Since such services are individually paid for analyzing their use and value is important for financial planning and management.

The complexity of such monitoring arises from: the different service subscription models used, the different authentication protocols used by them, and interpreting the data in terms value for money. The tool evaluated, called RAPTOR, was designed to support such monitoring activity by integrating authentication service logs, providing aggregate views of accesses over time.

The decision to employ the tabular approach in assessing this tool came from a cursory review of RAPTOR. One of the main functionalities on offer was the ability build a search specification, using a number of given types-of-search. Within each search-type parameters could be set by the user, such as: authentication-type; the date range; filters and data presentations.

Based on discussions with the tool developers and prospective users, tailored searches were judged to in effect be a novel intermediary notation demanding a significant change of perspective if they were to be used effectively by intended end users.

The tabular form approach was adopted to analyse and evaluate the tool's user interface collaboratively with a group of intended end users. As with the previous case study, scoping concepts were established through initial discussions and then used to instantiate questions to lead subsequent discussions. The relative novelty of the tool for the end users resulted in the scoping concepts being more domain based than tool based. In this case: "search specification", "data set" and "publisher service" were the chosen concepts.

Limited experience with the tool meant that the mode of discussion was often more about the users' needs than the appropriateness of specific tool functionality. The outcomes were collated reported to the developers. A number of issues were sufficiently specific to be recorded in an issue tracking system. Over the subsequent months many of issues raised were addressed by the development team.

Reflection upon the use of the tabular form in this case study focused upon how it could be used more effectively when engaging end users. An intrinsic feature of complex interaction is that end users may lack confidence and familiarity with complexities that are the motivation for this type of evaluation. Compared with the initial case study the outcomes were closer to user requirements, than formative design improvements. This insight was used to help inform how best to prepared and engage different types of evaluators when using the tabular form.

4.3 Data Replication Management

The final case study concerns the design and configuration of a comprehensive distributed big data replication and management service. The commercial service is founded on unique intellectual property for governing replication management and is highly integrated with existing data and cloud computing services, such as AWS, Hadoop, Azure, Google Cloud, etc..

The company was aware of issues that customer encountered during installation and management of their product. Despite the technical context of the product, technical staff tasked with installing and managing the service still required significant support. Specifically, the company's customers generally worked in sectors where service interruptions could have significant financial implications.

Based on our experience with previous case studies the prospect of conducting an evaluation was judged to be appropriate for the tabular approach due to the domain's complexity. The configuration and operation of the data replication is reliant upon integrating and linking numerous services to data nodes, and specifying the replication rules across groups of nodes for different data and differing geographic jurisdictions.

In contrast to the previous case studies, groups of staff teams including designers, developers and customer staff were encouraged to engage by the company. This resulted in three half-day workshops with largely different cross disciplinary groups. Groups involved between eight and ten people. To support this more collective use of the instrument, the workshops were planned to involve: individual introductions; an introduction user centred evaluation concepts, and; then using the tabular form. With each workshop framing concepts were established collaboratively and thus they differed between each workshop.

Analysis and reporting was a challenge simply because the authors were insufficiently familiar with the product and related services. As the participants were highly familiar with the product, the tabular form questions frequently led to discussions of product history and detailed aspects of its functionality. Within this context participants were encouraged to record their findings and observations. The company has reported that a number of innovations from the workshop have been recorded for inclusion in future versions of the product.

As a workshop the analysis and evaluation activity was less focused and demanded additional effort to keep participants focused and "on board". The relatively open format of the tabular approach needed to be more structured and supportive of participants. This led to consideration of pre-tailoring the instrument in future uses so as to prioritize the questions likely to be more productive.

5. CONCLUSIONS

The evaluation approach discussed in this paper is a lightweight reflective method motivated by concepts taken from the Cognitive Dimensions Framework. We have motivated the specific approach and described its development. Its use in a three case studies of design that vary in nature and character has been described and the lessons learnt summarized. Overall the experience with the tabular method has informed our way of working with it in different context - the general tabular form has not varied.

Future work with the instrument will include more accurately documenting ways of working with it and, more significantly, attempting to develop a means of pre-assessing an evaluation context in order to prioritize specific evaluative questions from the instrument.

ACKNOWLEDGEMENT

This work was partially supported by the Technology Strategy Board, JISC and European Regional Development Funds. We are indebted to the organizations involved and their staff for their cooperation.

REFERENCES

- Blackwell, A.F., Britton, C., Cox, A. Green, T.R.G., Gurr, C.A., Kadoda, G.F., Kutar, M., Loomes, M., Nehaniv, C.L., Petre, M., Roast, C., Roes, C., Wong, A. and Young, R.M. (2001). Cognitive Dimensions of Notations: Design tools for cognitive technology. In M. Beynon, C.L. Nehaniv, and K. Dautenhahn (Eds.) *Cognitive Technology 2001* (LNAI 2117). Springer-Verlag, pp. 325-341.
- Blackwell, A.F. (2006). Ten years of cognitive dimensions in visual languages and computing. *Journal of Visual Languages and Computing*, 17(4), 285-287.
- Blackwell, A.F. & Green, T.R.G. (2000). A Cognitive Dimensions Questionnaire optimised for users. In A.F. Blackwell & E. Bilotta (Eds.) *Proceedings of the Twelfth Annual Meeting of the Psychology of Programming Interest Group*, 137-152.
- Blandford, A.E., Green, T.R.G. and Connell, I.W. (2004). Formalising an Understanding of User-system Misfits. *Proceedings of EHCI-DSVIS 2004, the 9th IFIP Working Conference on Engineering for Human-Computer Interaction (Jointly with the 11th International Workshop on Design, A and Verification of Interactive Systems)*, Hamburg, 11-13 July 2004.
- Blackwell A. and Green, T. (2007) A Cognitive Dimensions Questionnaire. <https://www.cl.cam.ac.uk/~afb21/CognitiveDimensions/CDquestionnaire.pdf>. Accessed Jan 2019.
- Clarke, S. (2001). Evaluating a new programming language. In G. Kadoda (Ed.) *Proceedings of the Thirteenth Annual Meeting of the Psychology of Programming Interest Group*, 275-289.
- Green, T. R. G. (1989). Cognitive dimensions of notations. In *People and Computers V*, A Sutcliffe and L Macaulay (Ed.) Cambridge University Press: Cambridge., pp. 443-460.
- Mohammad Hadhrawi, Alan F. Blackwell, Luke Church (2017) A Systematic Literature Review of Cognitive Dimensions. *PPIG 2017*: 3 (see PPIG).
- McCarthy, J. and Wright, P. (2004) *Technology as Experience*, The MIT Press, ISBN: 9780262633550
- Nielsen, J. 1994. Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*, Catherine Plaisant (Ed.). ACM, New York, NY, USA, 413-414. DOI: <https://doi.org/10.1145/259963.260531>
- PPIG The Psychology of Programming Interest Group. Workshop archive: <http://www.ppig.org/> (Last accessed: December 2018).
- Roast, C. and Khazaei, B., (2006) *An Investigation into the validation of formalised Cognitive Dimensions. Design, Specification and Verification of Interactive Systems (DSVIS)*, Springer Verlag.
- Roast, C. and Uruchurtu, E. (2019). A tool for the mediated exploration of complex digital interaction design. SHU Research Data Archive (SHURDA). <http://doi.org/10.17032/shu-180010>.
- Triffitt, E. and Khazaei, B. (2002). A Study of Usability of Z Formalism Based on Cognitive Dimensions. In J. Kuljis, L. Baldwin and R. Scoble (Eds), *Proceedings of the Fourteenth Annual Meeting of the Psychology of Programming Interest Group*, pp. 15-28.