Trade-offs between reliability, validity and utility in the development of human factors methods

Patrick WATERSON[1], Chris W CLEGG[2] and Mark A. ROBINSON[2]

[1]*Loughborough Design School, Loughborough University, UK*
[2]*Socio-Technical Centre, Leeds University Business School, UK*

Abstract. We describe our experiences in developing human factors methods for a variety of purposes over the last couple of decades. In particular, we focus on recent discussions centred on the Fitts List (e.g., de Winter and Dodou, 2011) and accident investigation methods (e.g., the degree to which they are practice-focused, theoretically robust, valid and reliable). We draw on earlier work by Karl Weick (1979) concerning the trade-offs involved in developing theory within the social sciences and apply these to method development within human factors and ergonomics. The final section of the paper outline a set of issues for future work (e.g., satisficing trade-offs, using methods in combination, ways of improving the research-practice feedback loops).

Keywords. Human factors methods; trade-offs; reliability; validity; practical utility.

1. Introduction

The development of practical methods for applying human factors and ergonomics (HFE) in the workplace represents a core part of the discipline as a whole and has a long history (Stanton et al., 2013). In the last few years there have been a number of discussions amongst researchers and practitioners regarding the reliability, validity and utility of HFE methods (e.g., Kanis, 2014). Many of these discussions touch on wider debates regarding the status of HFE as a science, engineering discipline or craft (Moray, 2007) and the nature of the research-practice gap (Chung and Shorrock, 2011). We draw on a set of three brief case studies to reflect on our own efforts to build and evaluate HFE methods for a variety of purposes including function allocation, accident investigation and design for child safety. We argue that there is a need to acknowledge the role played by a range of trade-offs in the practical use of HFE methods and at the same time, to avoid exclusively scientific criteria when assessing the success/failure of a specific method. Those of us who develop methods need to be sensitive to these trade-offs and in some cases, relax our adherence to exclusively scientific and traditional ways of carrying out our method evaluation.

2. Case studies

2.1 Case study 1: Development of a method for function allocation

During the later 1990's the first two authors were based at Sheffield University and involved in developing a new method for function allocation for use in command and control systems in the UK Navy. Function allocation typically refers to the allocation of tasks and responsibilities between human and machines. In 2002 we published a journal paper in Human Factors which described 7 stages in using the method (from 'formation of the overall view of the system' to 'proposed allocations'), alongside a set of eight categories of social, organisational and technical decision criteria which could be used to allocate functions between humans, machines and teams. Each of the 7 stages was accompanied by a set of tools for recording and evaluating function allocations. Whilst a number of criticisms were made of the tools by Navy personnel (e.g., the need for an assessment of cost-effectiveness within 'live' projects), it was generally well received by our clients

In the years that have followed the publication of the paper we have monitored the number of times it has been cited (currently 37 times – Google Scholar, 2014). Comparing this with other, older attempts to develop methods for function allocation (e.g., Jordan, 1963; Fitts List - de Winter and Dodou, 2011), this rate of citation is modest and in many respects disappointing. Our method clearly has a long way to go in terms of popularity and use (especially given that citation rates are likely to be a poor proxy for actual usage). One explanation for why the method has proved to be unsuccessful is that it was too complex and unwieldy for HFE practitioners. They are likely to prefer methods which are more adaptable and tailorable to their work and immediate needs. Ironically, this was one of the key lessons from an earlier piece of work carried out by Clegg et al. (1996) at Sheffield University.

2.2 Case study 2: Systemic accident investigation methods

In the last few years, one of the authors (Waterson) has been involved in a programme of work on the use of systemic accident investigation (SAA) methods in a number of domains including rail and aviation (Underwood and Waterson, 2013). SAA methods are intended to cover a range of contributory factors leading up to an accident (e.g., regulatory, organizational, team, individual and technical factors). Examples of these types of methods include AcciMaps (Rasmussen, 1997), STAMP (Leveson, 2012) and HFACS (Shappell and Wiegmann, 2000). The findings from this work showed that SAA methods, although widely championed within academic settings, are rarely used in practice. A number of individual, organisational and industry-related factors impact upon the degree to which accident investigators are aware of SAA methods and/or decide to adopt and use them. These include: perceptions that the methods are too theoretical or conceptual; that they are time-consuming and impractical; the existence of an well established model which was in use within a specific sector or industry; the need to demonstrate accountability or fallibility in investigations; the previous experience of the analyst; the amount of training required to learn to use the SAA method.

2.3 Case study 3: Design for child safety: from guidelines to tools

The final example is taken from recent work involving the design of a set of tools to support the work of product designers designing safe products (e.g., toys, playground equipment) for young children (i.e., children aged 5-11). The tools are based on an earlier set of guidelines for the design of warnings for young children (Waterson and Monk, 2014) and include information on the physical, cognitive and emotional characteristics of different age groups in the form of cards, resource packs and web-based animations. Part of the work has involved interviews with product designers in order to assess their information needs and options for presenting materials and the content of the tools.

One of the main challenges with the project has been identifying who the target audience for the tools should be. Design within this context is a multidisciplinary and extended process with no clear starting point. Some stages of the design process (e.g., conceptual design) will be driven as much by the needs of marketing personnel as by designers themselves. Rather than there being one point in the whole design process where the tools could be used, there are several. Moreover, the tools need to be designed with these various audiences and specialisms in mind. The key lessons we have learnt from developing the tools is the need for flexibility – the tools need to be able to fit into a practical context and communicate in a simple, open-ended, as compared to prescriptive manner.

4. Discussion

In many respects the conclusions from the three case studies are nothing new or surprising. The discipline of human factors and ergonomics has always adopted a pragmatic approach to the analysis and design of work environments. The last few years however, have seen numerous calls for methods to be reliable, valid and useful (see Kanis, 2014 and the subsequent commentaries on that paper – e.g., Stanton, 2014). The questions are – is this always possible and is it always necessary? Here, an analogy might be drawn with some observations made by Karl Weick on the trade-offs involved in developing theory within the social sciences (Weick, 1979). Weick (drawing on earlier work by Thorngate, 1976) uses the metaphor of a clock face (figure 1) to argue that it is impossible for a theory of social behaviour to be simultaneously general, accurate and simple. You can seek any two of these three characteristics, but not achieve all three. Thus the more general and simple your theory, for example, the less accurate it will be in predicting specifics. Attempts to secure any two of the 'virtues' of a theory will mean that the third will be sacrificed. 'Two o'clock theories' for example, are general and accurate, but they will not be simple.

This analogy using the clock face might well be usefully applied to human factors methods. If we take what is arguably one of the most well-known HFE methods, task analysis, we might say that it fits the category of a 'two o'clock method (general and accurate). Likewise, we might argue that another successful (if heavily criticized) method, the Fitts List, is a typical 'ten o'clock method'. Thus the Fitts List is very simple to use and offers general applicability, however, it is far from accurate, not least since the types of automation available today are radically different from the those of the 1950's and 60's.

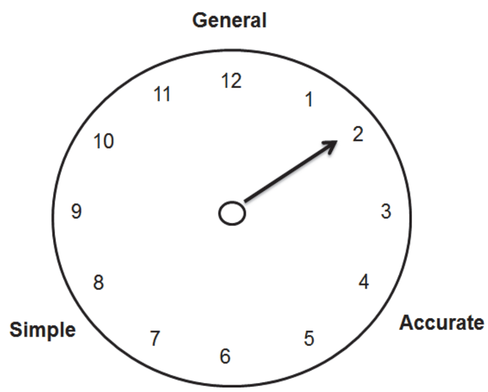

Figure 1: 'Weick's clock'- trade-offs between generality, simplicity and accuracy (Weick, 1979)

We are not arguing here that every HFE method will fit neatly onto such a clockface (though it would be an interesting exercise to attempt a critical review of this kind). Rather, the important point to note here is that achieving good levels of reliability, validity and utility (how useful a method is perceived to be) is likely to involve fundamental trade-offs. Many of these reflect the day-to-day realities of HFE in practice and should not be ignored. One important trade-off is the context in which they are used and the level of risk assessed by the method. In developing a method to assess safety in a safety-critical content (e.g., a nuclear reactor), simplicity might have to be sacrificed to accuracy. In other cases, simplicity and generality may be more important than accuracy (e.g., in the case of some types of dynamic risk assessments). The scientific community does not need to give up on stressing the importance of reliability, validity etc. In many cases it is vital that these are properly assessed. A number of tools for assessing patient safety culture demonstrate poor psychometrics (e.g., statistical reliability) for example (Waterson, in press). This stands in comparison to much more well-established safety domains (e.g., rail, oil and gas) where in some case safety culture tools are more reliable and valid.

5. Summary and conclusions

Our experience of developing and evaluating HFE methods is that many are too time consuming and complex whilst also proving expensive and difficult to use (Waterson et al., submitted). The first case study demonstrated that this can happen even when one is aware of the pitfalls of building over-complexity into HFE tools. Case study 2 showed that a wide-range of trade-offs govern the extent to which HFE tools will be used by practitioners. Finally, case study 3 goes some way to showing that the methods are part of a larger design supply-chain in which HFE competes alongside other product development priorities (e.g., marketing, branding).

Part of our response has been to develop set of tools where practitioners can pick and choose which tool they think best fits the particular task they are trying to complete. We suggest that there needs to be wider acknowledgement that reliability, validity and other measures of method ‘success’ may need to be seen within the larger context of trade-offs within HFE practice. This stands in contrast to the views of others who appear to regard them as immutable ‘gold standards’. In short, the use of ‘hard’ scientific criteria for the evaluation of HFE methods is necessary, but in other cases it may not be sufficient. Similarly, we currently know quite a bit about the problems of human-factors integration (HFI), but curiously little about the actual ways in which practitioners use methods. If we are to stand any chance of narrowing the gap between research and practice in HFE (Chung and Shorrock, 2011; Kirwan, 2012), then a pressing need to is to engage with and further understand some of the trade-offs and constraints within HFE practice we have mentioned in this paper.

References

Chung, A.Z.Q. and Shorrock, S.T. (2011), The research-practice relationship in ergonomics and human factors - surveying and bridging the gap, Ergonomics, 54, 5, 413-429.

Clegg, C.W. et al. (1996), Tools to incorporate some psychological and organizational issues during the development of computer-based systems. Ergonomics, 39, 3, 482-511.

Jordan, N. (1963), Allocation of functions between man and machines in automated systems. Journal of Applied Psychology, 47, 55-99.

Kanis, H. (2014), Reliability and validity of findings in ergonomics research. Theoretical Issues in Ergonomics Science, 15, 1, 1–46.

Kirwan, B. (2012), Time for some human factors intelligence. IEHF Institute Lecture, Blackpool, 2012.

Leveson, N (2012), Engineering a Safer World: Systems Thinking Applied to Safety. London: The MIT Press.

Moray, N. (2007), The human factors of complex systems: a personal view. In D. deWaard. G.R.J. Hockey, P.

Nickel, and K.A. Brookhuis (Eds). Human Factors Issues in Complex System Performance. Maastricht, Netherlands.

Rasmussen, J. (1997), Risk management in a dynamic society: a modelling problem. Safety Science, 27, 183-213.

Shappell, S. and Wiegmann, D., (2000). The human factors analysis and classification system – HFACS. DOT/FAA/AM-00/7. Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration.

Stanton, N. A. et al. (2013), Human Factors Methods: A Practical Guide for Engineering and Design. Farnham: Ashgate.

Stanton, N.A (2014), Commentary on the paper by Heimrich Kanis entitled 'Reliability and validity of findings in ergonomics research': where is the methodology in ergonomics methods? Theoretical Issues in Ergonomics Science, 15, 1, 55-61.

Thorngate, W. (1976), "In general" vs. "it depends": Some comments on the Gergen-Schlenker debate. Personality and Social Psychology Bulletin, 2, 404-410

Underwood, P. and Waterson, P.E. (2013), Systemic accident analysis: examining the gap between research and practice. Accident Analysis and Prevention, 55, 154-164.

Waterson, P.E. (Ed., in press, 2014), Patient Safety Culture: Theory, Methods and Application. Farnham, UK: Ashgate.

Waterson, P.E., Older-Gray, M. and Clegg, C.W. (2002), A sociotechnical method for designing work systems. Human Factors, 44, 3, 376-391.

Waterson. P.E., Robertson, M.M., Cooke, N.J., Militello, L, Roth, E. and Stanton, N.A. (under review), Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety. Ergonomics.

Waterson, P.E. and Monk, A. (2014, in press), The development of guidelines for the design and evaluation of warning signs for young children. Applied Ergonomics. http://dx.doi.org/10.1016/j.apergo.2013.03.015

Weick, K. (1979), The Social Psychology of Organizing. Reading, Mass.: Addison-Wesley.

de Winter, J.C.F. and Dodou, D. (2011), Why the Fitts list has persisted throughout the history of function allocation. Cognition, Technology and Work, August 2011, doi: 10.1007/s10111-011-0188-1