Human-Computer Interaction -- INTERACT'03
M. Rauterberg et al. (Eds.)
Published by IOS Press, (c) IFIP, 2003, pp. 367-374

The Misapplication of Engineering Models to Business Decisions

Gitte Lindgaard

HOT Lab, Carleton University, Colonel By Drive, Ottawa, Canada

gitte_lindgaard@carleton.ca

Abstract: The HCI community has long been accused of delivering ‘common sense’, ‘useless’ information, and to be ignorant of business needs. HCI experts are also criticized for failing to provide or apply theory-based techniques. This paper shows that the two goals may be incompatible. It discusses one case study in which HCI data intended for one purpose were inappropriately applied to support another. Theory-driven GOMS (Goals, Operators, Methods, Selection rules) models generated to predict performance in two competing applications were subsequently used for making a business decision. Three similar data-driven studies designed to inform a business decision are then presented. Findings from all these studies demonstrate that the parameters on which the business decision based on GOMS data was made were largely irrelevant to that decision. It is argued that HCI experts must learn to relate their findings to business needs and values if HCI practice is to progress.

Keywords: engineering models, GOMS, business applications

1 Introduction

The HCI community has long been accused of providing ‘too little, too late’ in the development process, of delivering results that are mere ‘common sense’, and of ‘adding no value’ to software development and evaluation. Recent additions to the list of sins include the accusation that usability experts focus too narrowly on research, testing, and critiquing, and that they do not understand business needs, objectives, or values (Olsen, 2002). It is true that the HCI community tends to concentrate on usability activities without regard for, or apparent understanding of, broader business goals, even when assessing the costs and projecting the potential benefits of their own work. The link between HCI-related findings and the implications of these for the business bottom line is rarely made, with some notable exceptions (e.g. Karat, 1994; Donahue, 2001; Donoghue, 2002). When this is done, the resulting figures are usually staggering and very convincing for ‘selling’ and justifying HCI activities (Bevan, 2002; Lindgaard, 2001).

At times, such justification comes from simple observations of user activity. One recent field study in which the activities of a specific group of workers in a large Canadian grocery distribution centre were observed and carefully documented showed that two thirds of these were unnecessary from the business point of view. The study also showed that the removal of certain automated performance-based work statistics applied to another group of workers in the same workplace could completely eliminate the organizational bottlenecks causing this waste. The cost of these activities amounted to the equivalent of one full-time person in the small part of the workplace alone in which this finding emerged (Lindgaard & Madore, 2003). As was true in this study, user observations tend to be easy to turn into dollar statements that are readily substantiated even though the information may not always be directly related to improving the *usability* of applications.

User observations generally lead to data-driven models of human-system behaviour. While such models may be excellent for informing business decisions, they tend to lack generality and have little or no theoretical basis. This is of concern to the HCI community but not to business decision makers.

By contrast, the GOMS concept was specifically developed to provide theory-based engineering models predictive of human-system behaviour that could be applied without further empirical validation (Card, Moran & Newell, 1980; 1983). The main motivation behind GOMS modeling was to reduce the amount of user testing by applying engineering models wherever feasible, thereby freeing the HCI community from collecting unnecessary performance data and businesses from incurring unnecessary costs. The intention behind GOMS modeling was thus *not* to *replace* user testing, but to reduce it. GOMS models were derived from an applied information-processing psychology "based on task analysis, calculation, and approximation" (Card et al., 1983, p. 44). Yet, approximations may not assist management in making business decisions.

Tension thus exists between HCI goals and business goals. On the one hand, HCI analysts are encouraged to employ more theoretically based techniques, and on the other, they are also expected to generate data to support 'big picture' business decisions. They rarely have an opportunity to fulfill both of these demands.

One consequence of this tension between HCI and business management is the potential misapplication of HCI data to solving business problems for which they are unsuited or irrelevant. HCI folks may be ill informed about business procedures, values, or about what might or might not affect the bottom line, but business leaders also often misunderstand the implications of HCI data. Sometimes interpretation of HCI findings goes well beyond the data; at other times, the very questions asked of HCI people are shown to be irrelevant to the business.

The purpose of this paper is to demonstrate how a business decision may be made inappropriately on the basis of data that are unfit for such a purpose. In the next section, several versions of a GOMS study in which findings were used to make a business decision are reviewed critically. Section 3 discusses two studies conducted in a similar environment but relying on user observations instead of GOMS. A third study providing comparable, but more detailed data, also based on user observations and performed in a similar environment, is presented in Section 4. A general discussion draws the studies together in Section 5, and a conclusion is reached in Section 6.

2 Predictive Models: GOMS in Business

The GOMS case study in question compared a proposed application with an existing workstation used by Toll and Assistance Operators (TAOs) in a Telephone company in the United States. The TAO's job is to handle collect calls and person-to-person calls. The initial GOMS models were produced from logged calls for the existing workstation, and from the manufacturer's documentation for the proposed workstation.

The manufacturer of the proposed workstation had predicted that it would reduce the average call-completion time by 2.5 sec per call (Gray, John & Atwood, 1990), or 2 sec per call (Gray, John & Atwood, 1992), depending on which version of the study one reads. This reduction translated into a saving of $7.5 million per year according to the 1990 paper. In a later version it is said that an average decrease of 1 sec per call would translate into a saving of $3 million per year (John & Kieras, 1996). So, a saving of 2 seconds per call should amount to $6 million per year, and 2.5 seconds to $7.5 million. A large scale field study was conducted at the same time as the GOMS models were constructed to generate a set of benchmark tasks. The models were then modified to reflect the differences in design between the two workstations, which included different keyboard and screen layouts, keying procedures, and system response times. In concert with the GOMS concept, the models predicted expert, error-free call-handling time for the 15 call types benchmarked for each of the two types of workstation. When combining all the call types into one calculation, the resulting GOMS model predicted that the proposed workstation would be 0.63 sec slower than the existing workstation, which was equal to 104% task-completion time in the current workstation. This finding that tasks took longer in the proposed workstation was a big surprise to all, as the vendor had predicted the opposite.

The GOMS models were produced in CPM-GOMS (Cognitive, Perceptual, Motor, or Critical Path Method). The critical path is the sequence of tasks that determines the fastest route to call completion. CPM-GOMS determines the duration of the task. The proposed workstation used 10 fewer operators, representing two keystrokes, than the current application. The analysis showed that the 10 operators occurred at the start of a call, at a point at which the work is controlled by the conversation with the customer, not by keystrokes. Because neither of the saved keystrokes was in the critical path, they did not affect the predicted completion time for the relevant portion of the task. However, one of the two keystrokes eliminated at the beginning of the call occurred later in the call, thereby moving from being off the critical path to being on it. From the papers reporting this study, it is unclear how the customer side of the interaction was calculated and incorporated into the model. However, relying on independent research in speech generation as well as on estimates of operator execution times provided by previous GOMS research shows how the duration of standard utterances may be calculated and measured

(John, 1990). For example, the greeting phrases "New England Telephone. May I help you?" was predicted to take 1300 msec and found to take 970msec (-34%), whereas the phrase "Hello" was predicted to take 340 msec and found to take 460msec, or 26% more than predicted. John (1990) provides a list of phrases typically spoken by the TAOs during call completion and finds an average percent of error across the whole sample to be –7%. That is, observed performance is slightly better on average than predicted performance.

The TAO carries out a range of more than 20 different, albeit very similar, tasks. GOMS models assume that customers utterances are brief, precise, clear and articulate, that they never need to repeat a number, that the operator can hear, and does not confuse, any digit, letter, or word the customer pronounces (e.g. 'five' and 'nine'; 's' and 'f', 'd' and 'b'). For the purpose of generating GOMS models, such assumptions are necessary and appropriate, as they provide an approximation rather than a record of actual performance.

2.1 The 1990 Version

In the earliest paper (Gray, John & Atwood, 1990), a two-second increase in performance time was predicted for the proposed workstation from CPM-GOMS models for 18 call types accounting for 90% of all calls taken by the TAOs. The time increase in the proposed application was attributed to an additional keystroke. Empirical data confirmed this prediction, albeit at a questionable level of significance $F(1,46) = 3.14$, $p < 0.10$. Careful reading of this paper shows that the difference was due to an increase in System Response Time (SRT) rather than to the additional keystroke. Since SRT is a constant in GOMS, and because the initial GOMS models were generated from the vendor's documentation rather than from logged calls, it is unlikely that the GOMS models could have predicted this SRT difference. Thus, the early GOMS models would appear to have made no contribution to the analysis of the differences in operator performance whatsoever. Yet, the authors concluded that "GOMS is an important and valuable tool that can be used to predict the effectiveness of a given design in a given environment on human performance" (p. 55).

2.2 The 1992 Version

In the 1992 version of the study (Gray, John & Atwood, 1992), the authors say that the 20 categories (instead of 18) of calls analysed account for 88.33% (instead of 90%) of all calls taken. One tenth of the total of 78,240 calls were analysed (i.e. 7,824 calls), but another five call types had now been removed due to their infrequent occurrence.

The GOMS analysis, this time modeled on 15 call types for each application, showed that the proposed system would take 0.8 seconds (3%) longer per call than the same work carried out on the current workstation. The field study confirmed this with a difference between the workstations at the 0.05 level of significance, reflecting a 4% increase in overall performance on the proposed workstation.

It is never quite clear just how long these calls take. In one publication, John (1990) mentions a performance time of 24,040msec/call in a study that seems to refer to the same dataset. The performance increase, the author claims, is "both statistically and financially significant" (p 308), because the 0.8 seconds translates into a cost of $ 2.4 million per annum. If 0.8 seconds equal 3% of the call time, each call must take 26.67 sec on average, as opposed to John's 24.04sec.

It is unclear whether the 10% of 78,240 calls that were analysed were selected at random or by some other process and whether the different call types represented in the final calculation reflect their naturally occurring distribution. Further, we do not know whether the logging software records pauses between calls, or only the actual call duration time. From the point of view of producing GOMS models, between-call pauses don't matter, but from the point of view of assessing the operators' job for making business decisions, it should be entered into the equation. If 20 call types represent 88.33% of all TAO work, then 15 call types should represent 66.25% of the work, provided these are equally distributed across the workload. Thus, the business decision would have been made roughly on the basis of two thirds of the entire TAO workload. However, it is impossible to verify this from the data provided. The differences between the three versions of the study are shown in Table 1.

From the data presented in the papers (Gray et al., 1990; 1992; 1993), it is impossible to discern how call types were defined and thus what distinguishes one from another. It is also unclear which specific call categories were excluded from the final analysis. In fairness, the emphasis of all the papers is on the

Year	Performance Decrement	No. of call Categories	Predicted Reason	State Reason
1990	2 sec	18	Extra keystroke	Increase in SRT
1992	0.8 sec	15	Keystroke in critical path	Keystroke in critical path
1993	0.65 sec	15	Keystroke in critical path	Keystroke in critical path

Table 1. Comparison of predictions in the three studies

generation/application of GOMS models, not on business decisions based on any data analysis. By the same token, one is entitled to question the degree to which this continued refinement of the GOMS models and the repeated data re-analyses are scientifically justifiable. Finally, it is unclear how, or whether, the customer input was entered into the equation, seeing that the focus is on the 'performer', in this case, the TAO.

The authors do state that the biggest, most time-consuming factor was TAO/customer interaction, and in terms of providing an 'approximation' of user performance, this part of the interaction is irrelevant. However, from a business point of view, it is reasonable to question the representativeness of the TAOs' work reflected in both the resulting models and the performance measures with which the models were compared. It appears that the GOMS models became *the* decisive factor underlying the client's decision. That is, rather than providing a mere approximation, the models became the basis for an important business decision. It is important to note that the HCI analysts were hired to model and predict the TAO-system performance, and not to make business decisions.

2.3 Operator-Customer Interactions

In the above study, performance calculations were based on the arithmetic mean of call-completion times. In order to comment on the reliability and validity of the mean as a basis for comparing different software solutions, three studies of telephone operator performance are reviewed next. Two of these were conducted in Directory Assistance (DA) and one in a pager company. Unfortunately, data for TAOs were not available for this comparison, but several features are shared across calls involving telephone operators, computer applications, and customers acting as third parties in the human-computer interaction. First, call duration times depend to a large extent on how well customers state their question – how articulate they are, how briefly they can convey their query, how clearly they pronounce names and digits, and the quality of the telephone connection. Second, some call-logging software tends to show the beginning, but not the end-time of calls. Finally, management relies on 'average call times' and number of calls per operator for judging application and operator performance.

3 Evaluative Models: Directory Assistance

Data from two studies in (DA) are considered here. Like TAO calls, DA calls are standard, simple and quick to perform. In a Canadian study, McEwen and Bergman (1993) found that interacting with the calling customer took up 35% of the average call duration time (18.7 seconds). This observation led to a creative reorganisation of DA calls by inserting a recorded voice announcement into the very start of the call, saving a clear 2 seconds per call. In their subsequent trial, they found that customers adjusted easily to this change, and that they as well as the operators liked it better than the old method in which they were greeted by a person rather than by a voice recording. The sample comprised 6,700 videotaped calls. The median call duration time was 14.5 seconds, or 4.7 seconds below the mean. Some three percent of calls exceeded 60 seconds. So, the most rarely occurring calls each amounted for more than 300% of the average call time.

In Australia the average DA call duration time was 19 seconds at the time of the study, compared with 18.7 seconds in Canada. A set of videotaped DA calls collected for an ergonomic analysis of keyboard usage (Lindgaard & Caple, 2001) was re-analysed to enable a comparison with the Canadian data. Due to thee different purpose of the original data collection, the available sample is rather small (n = 356 calls). These yielded a mean of 16.6 seconds, or 2.4 second below the population mean, μ, which is probably due to the small sample size. Figure 1 shows the distribution of the analysed calls by call duration times.

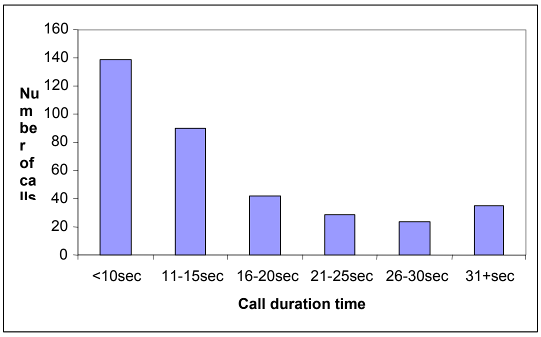

Figure 1: Distribution of DA calls (n = 356)

As in the McEwen and Bergman study, the median call duration time of 12.20 seconds was lower than the sample mean (16.6sec). Calls exceeding μ (19sec) amounted to 24.16%, or nearly one quarter of the calls. The average duration of calls in this group was 37.43 seconds, or just under 200% of the population mean. Contrary to McEwen and Bergman who found that 3% of sampled calls exceeded 60 seconds, this percentage was 8.57% in the Australian study, suggesting either that the sample was too small to be truly representative or that a high percentage of calls are well below the population average. This would offset the relatively high percentage of longer calls. Common to both the Canadian and the Australian datasets is the observation that infrequently occurring calls lie in the most time-consuming end of the distribution. Unfortunately, it is impossible to know whether the infrequently occurring calls removed from Gray et al.'s GOMS models also represented this end of the spectrum.

4 Pager Services

A pager company answers the telephone and forwards messages to clients. The company in question took some six million calls per week at the time the study was conducted, with each call taking on average 30 seconds (Lindgaard, 1995). Management's objective was to reduce the average call completion time by one second. Applying the same logic as in the Gray et al. study, this would amount to a saving of several million dollars per year. Some 2,344 calls were sampled over a period of eight working days observing 14 highly experienced operators for half a day each in two major Australian cities. The call centres operate 24 hours per day, but calls were sampled only during the day and the evening to capture traffic at the busiest times. The distribution of call duration times and average pauses between calls are shown in Table 2 by time of day. A pause starts when the operator hangs up and ends when she presses the 'accept new call' button on her telephone. The Table shows that both call times and pauses varied substantially at different times of the day.

Time of day	Call time (sec)	Pause (sec)	Call time + pause
10 - 12	22.23 (19.78)	7.23 (4.35)	29.46
12 - 14	24.24 (15.40)	8.99 (3.78)	33.23
14 - 16	20.56 (13.62)	6.27 (5.44)	26.83
16 - 18	22.12 (15.51)	4.86 (2.39)	26.98
18 - 20	18.70 (14.33)	10.78 (8.24)	29.48
Overall	21.57	7.63	29.20

Table 2: Distribution of call and pause duration by time of day (Standard deviations are given in brackets)

One interesting observation is that, in the two DA studies, none of the observed call duration times came close to the average of 30 seconds. However, when adding the average pause time to call duration in the present study, this combined average comes to 29.20 seconds overall – very close to the 30-second population average. Inspection of the software logs showed that the call start times, but not call end times, were logged, giving no indication of pauses between calls. Additional observations from several different operational settings involving telephone operators suggest that pauses between calls that are beyond the control of operators do occur even at extremely busy times (Lindgaard, 1992; 2001). Pauses were not measured in the Australian DA calculations shown earlier, but it is possible that they may contribute to the estimates of population mean call duration times. The distribution of call duration times is shown in percentages in Figure 2.

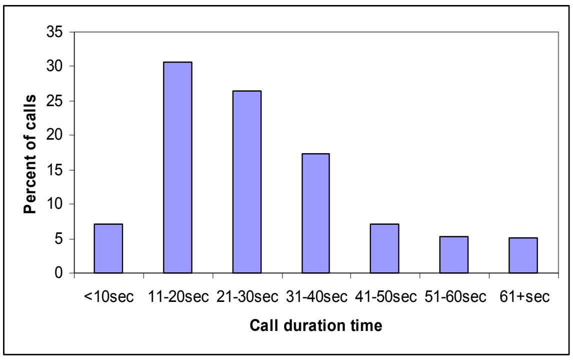

Figure 2: Percentage of calls by call duration times (N = 2344)

As in the two DA studies, the median of 22.53 seconds was lower (by 7.57 seconds) than the average population call duration time. Just over 5% of sampled calls took more than twice the average time. Taken together, the data in Table 2 and Figure 2 suggest that 'average call duration' is not a very accurate indicator of operator activities in this workplace. In addition, although management does distinguish between four different levels of pager service, and hence between four call types, the fact that these involve substantial differences in complexity and in call handling times is not reflected in the average call duration times. The distribution of average call duration (not including pauses) for each of the four call types is shown in Table 3 below.

Type	Mean (sec)	S.D. (sec)
1	16.24	(10.63)
2	17.76	(8.96)
3	24.31	(18.26)
4	26.97	(22.58)

Table 3: Average call duration times for call types 1-4

The data suggest that call handling times increase incrementally with the complexity of the call type. The difference between the fastest, call type 1, and the slowest, call type 4, is more than 60% on average. The relative complexity associated with these calls stems from differences in the level of service provided, which, of course, is reflected in the price clients pay. Operator responses to a 'level 1' call are limited to "<company name>. Your message, please", whereas, to a 'level 4' call they are customised according to the client's wishes. Greetings are invariably longer, typically something like: "Good morning. This is <company X>, Sarah speaking. May I help you.", and customisation may include transferring the caller to a number at which the client is currently available, making it seem to the caller that the call centre operator is, in fact, an employee in the client's company.

A measure of the number of calls that could be classified as 'unprofitable' showed that these amounted to 36.21% of all calls sampled. 'Unprofitable' calls were those in which the caller decided not to leave a message, had dialed the wrong number, or had hung up before the operator was able to answer the call. Removal of these calls would amount to an average reduction of 4.13 seconds per call, or over four times the management objective. A simple change in the answering procedures, for example, by inserting a Recorded Voice Announcement (RVA) before a call is passed on to an operator, would eliminate most of these types of call. To make this solution viable, the 'level 4' calls would need to be directed to specific numbers circumventing the RVA. This kind of solution could not be derived from the automatic data collection on which the average call time was based, and nor would it have emerged in models of standardized calls. It is acknowledged that, since this study was carried out, RVAs have become extremely disliked. It is therefore possible that this solution would be unacceptable. However, the purpose of the example in the present context is to allude the reader to the richness of data logged in the work context and compare them with data generated by predictive models, rather than to provide a viable business solution.

5 Discussion

The above results all suggest that the mean call-completion time is not an accurate reflection of the way telephone operators spend their time while on the job. In both the DA and the pager studies, the population average was higher than the sample means and medians. It is possible that the small samples may account for the lower mean call times. However, inspection of software logs in the pager company showed that call onset times, but not offset times were shown, suggesting that pauses between calls may be hidden in the automatic data. In the case of the pager company, the observed figures did come close to the average call time when pauses between calls were added to the actual call times.

As Gray and his colleagues argued, and consistent with all results shown here as well as others reported elsewhere (Lindgaard, 1992; 2001), the biggest call component involves greeting and obtaining information from the customer. In all similar situations we have encountered in the past 22 years, as well as in the samples discussed above, that component comprises 30% - 40% of an entire call. In calls exceeding the average call time, the increase can invariably be attributed to the operator-caller interaction the control of which, as Gray (1992) and his colleagues correctly noted, is beyond the operator.

Call completion times associated with the four different paging call types suggest that fluctuations in call duration by call type are likely to be substantial. This observation was probably made in

Gray et al.'s (1992; 1993) 15 call types too. Likewise, call duration times also vary substantially by the time of day. Indeed, calls were not taken during the quiet times between eight in the evening and ten in the morning. Yet, all of these times are also included in the population grand mean in the pager study. Operators will, at least at certain times, have more idle time than at others which, just like the human-human interaction times, are beyond their control. The above data suggest that time of day should be factored into the calculation of call duration times. For all of these reasons, we would contend that average call duration times are a poor and unrepresentative measure of operator performance. All the above findings also call into question the meaningfulness of projected savings that are based exclusively on reductions in *average* call completion times. That is, management objectives for achieving higher efficiency and for making business decisions to that effect should be based on far more detailed evidence than an average call time can possibly provide. Evidently, call centre management desperately needs to be educated about the value of carefully performed field observations and tedious analyses of videotaped calls.

This is not a 'march against GOMS', and nor is it, strictly speaking, a criticism of the research performed by John, Gray and Atwood on TAOs. Sure, various details from their study would have been valuable, and, yes, they do seem to have refined both the GOMS models and the dataset to which the models were compared iteratively which, scientifically would be indefensible. Details of the definition of a call type, the criteria by which logged calls were selected for analysis, reports of actual call lengths by type, particularly for those calls that were eliminated, and a clarification of the actual percentage of TAO work represented by the GOMS models, would have been helpful. Such data would have enabled a judgment of the extent to which their data could have provided the client with information that may have led to a different business decision.

The client hired the researchers to compare the relative performance of two competing applications, and that is precisely what they did. Their focus was on task modeling, not on the relevance of those to business decision making. Unquestioningly, the authors fulfilled the goals of the contract and of the engineering models, namely to provide an 'approximation'. They cannot take responsibility for the way a client might choose to employ the findings. As the researchers were not employees of the client organisation, they can hardly be reprimanded for failing to scrutinise the field data in a manner similar to the way it was done in the other studies reported above, as this would have added substantially to the time frame of the project. It is unlikely that the client management would have been receptive to budget overruns.

6 Conclusion

The absence of detailed field data serving to educate management points to one of the major dilemmas for the HCI community. HCI specialists are rarely given an opportunity to do their work in a way that allows them to maximise its value to the hiring organisation. Moreover, there is a discrepancy between delivering value to a project and to an organisation. Project schedules do not, as a rule, allow for additional activities such as delving into observation data in greater depth, or even collecting such data. Such data usually do not help the existing project, as they would require a complete revamping of the project goals, software design, operator procedures, and business strategies. In the case of the TAOs, the business decision was apparently based on incomplete understanding of the variations in TAOs' jobs, which led management to ask the wrong question. The investigation should have been directed at uncovering the details of the TAO tasks. Indeed, an analysis of the human-human interaction may have led the company to consider a very different solution to merely replacing one application with another. The McEwen and Bergman (1993) example of DA tasks showed just how user data may be applied creatively to consider alternatives to aiming at perfecting the software without consideration of people involved in the task.

The HCI community needs to find ways to demonstrate the true value to businesses beyond calculating the costs and benefits associated with performing standard usability tests or designing interfaces based on best guesstimates. GOMS modeling may provide one set of tools to streamline design and evaluation of certain types of application, but we must prevent the misapplication of engineering models to business decisions and find ways to educate business management.

References

Bevan, N. (2002). *Cost-effective user-centred design based on ISO 13407*, IFIP WCC 2002

Tutorial Notes, IFIP World Computer Congress, Montreal.

Card, S. K., Moran, T.P. & Newell, A. (1983). *The psychology of human-computer interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Card, S.K., Moran, T.P. & Newell, A. (1980). Computer text-edition: An information-processing analysis of a routine cognitive skill, *Cognitive Psychology*, 12, 32-74.

Donahue, G.M. (2001). Usability and the Bottom Line, *IEEE Software*, January/February.

Donoghue, K. (2002). *Built for use: Driving profitability through the user experience*, McGraw-Hill, Toronto.

Gray, W.D., John, B.E. & Atwood, M.E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance, *Human Computer Interaction*, 8, (3), 237-309.

Gray, W.D., John, B.E. & Atwood, M.E. (1992). The precis of project Ernestine or an overview of a validation of GOMS, in *Proceedings CHI '92 Striking a balance Human Factors in Computing System,*, Monterey, CA, 307-312.

Gray, W.D., John, B.E. & Atwood, M.E. (1990). An application and evaluation of GOMS techniques for operator workstation evaluation, *Proceedings 13th. International Symposium Human Factors in Telecommunication*s, HFT, Turin.

John, B.E. & Kieras, D.E. (1996). Using GOMS for user interface design and evaluation: Which technique?, *ACM Transactions on Computer-Human Interaction*, 3 (4), 287-319.

John, B.E. (1990). Extensions of GOMS analyses to expert performance requiring perception of dynamic visual and auditory information, *CHI '90 Proceedings Human Factors in Computing Systems*, ACM Press, New York, 107-115.

Karat, C-M. (1994). A business case approach to usability cost justification, in R. Bias & D. Mayhew (Eds) *Cost-justifying usability*, Academic Press Inc., Boston, MA.

Lindgaard, G. & Madore, S. (2003). The impact of interactive technology on worker efficiency: A job assessment, *Proceedings of the 4th World Congress on e-commerce management*, Hamilton, Ontario, Canada, 15-17 January.

Lindgaard, G. (2001). From the ashes of disaster into a human factors boom: The legacy of large databases, *Proceedings HCI International'01*, New Orleans, 5-10 August

Lindgaard, G. & Caple, D. (2001). A case study in iterative keyboard design using Participatory Design techniques, *Applied Ergonomics,* 32, 71-80.

Lindgaard, G. (1995). Analysing customer services for Link Telecommunications: An initial field study, Technical Report LT-0040695.II

Lindgaard, G. (1992). Evaluating user interfaces in context: The ecological value of time-and-motion studies, *Applied Ergonomics*, 23, (2), 105-114.

McEwen, S. & Bergman, H. (1993). Automating directory assistance service: A human factors case study, in *Proceedings 14th. International Symposium Human Factors in Telecommunications*, Darmstadt, Germany, May 1993.

Olsen, G. (2002). The emperor has no lab coat, *Interactions*, 9 (4), July-August, 13-17.