

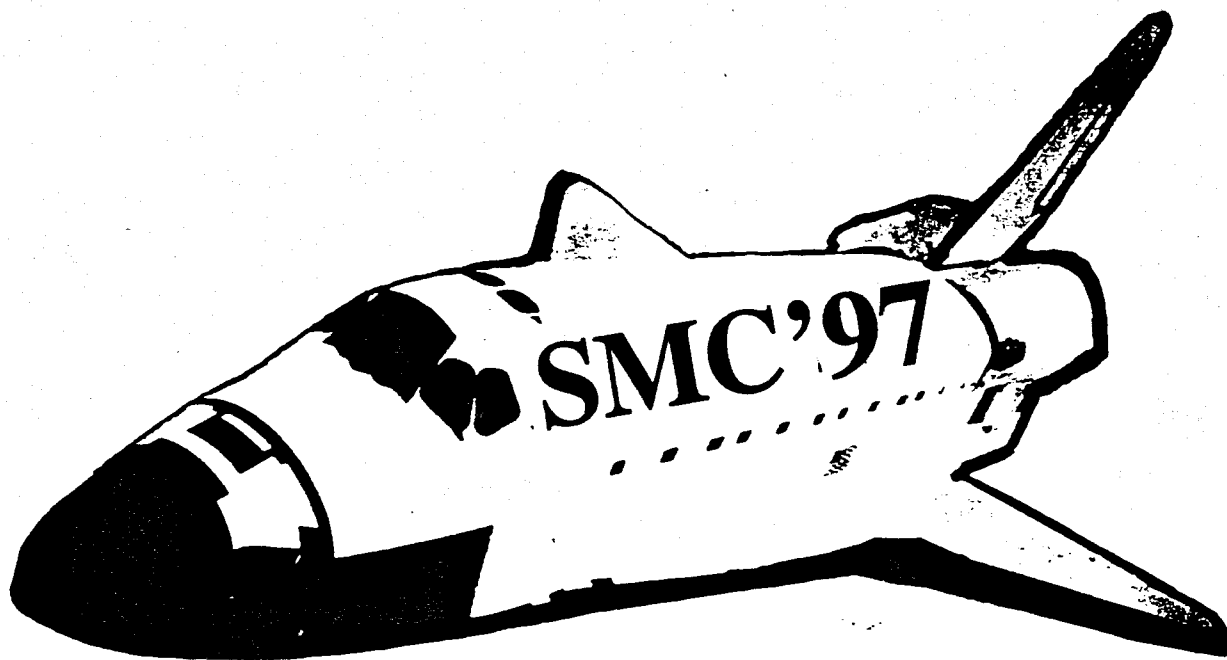
SMC '97 Conference Proceedings



*1997 IEEE International Conference on
Systems, Man, and Cybernetics*

Conference Theme

Computational Cybernetics and Simulation



*Volume 3 of 5
97CH36088-5*

*October 12-15, 1997
Hyatt Orlando
Orlando, Florida, USA*

*Sponsored by the Systems, Man, and Cybernetics Society of
the Institute of Electrical and Electronics Engineers, Inc.*

**THE IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS
HYATT ORLANDO, ORLANDO FLORIDA, USA, OCTOBER 12-15, 1997
THEME: COMPUTATIONAL CYBERNETICS AND SIMULATION**

**MEASURING SUCCESS IN NON-TRIVIAL WAYS; HOW CAN WE KNOW THAT A DSS
IMPLEMENTATION HAS *REALLY* WORKED**

**Lia Di Bello Ph.D. Principal Investigator
Workplace Technology Research Group
CUNY Graduate School
25 West 43 Street Suite 620, New York NY 10036
Phone: 212-642-2903/fax 212-719-2488
ldibello@broadway.gc.cuny.ed**

With significant budget reductions, many “public” industries, such as transportation and power generation are finding they must reexamine the way they do business. With the increasing threat of outsourcing, those serious about maintaining their existence are looking more closely at totally new approaches that emphasize the efficient use of funds and labor resources with the goal of maintaining high performance. Many are investigating computerized Decision Support Systems (DSS’s), often borrowing technologies developed for related private industries and tailoring them for new purposes.

Unfortunately, most implementations fall short of expectations and several million dollars are usually spent before the failure is certain. In addition, the causes of failure are often still unknown when the system is finally unplugged. Specifically, this paper deals with the problematic issue of measuring success and/or sensitively monitoring an implementation in order to identify early trouble spots. As will become clear, research methods from the social sciences can be successfully adapted for the purpose of measuring. Designed to unpack the “why” of human behavior and practice, such methods can render more sensitive measures while at the same time make better use of increasingly detailed data made available by the DSS’s themselves.

Specifically, this paper details a three-level analytic method for measuring a decision support implementation and identifying trouble early in the roll-out. Originally developed for basic research on the development of expertise among users of DSS’s, (e.g., DiBello, 1996; 1996b, Chamberlain & DiBello, 1997) the approach has proven to increase the chances of identifying small – but eventually fatal – problems with the deployment, the design, or subtle user misuses. The method incorporates: 1. Domain-specific cognitive profiles of target users, 2. Measures of user navigation through

both legacy systems and the target DSS and ; 3. Ways of tracking the financial impact of the system at a level of detail normally not found among typical management reports. There are a number of unique features of each of these sub-methods. For example, the cognitive profiles measure the similarity of the user’s spontaneous mental model of workflow to the underlying logic of the target DSS *while at the same time* measuring the lingering impact of legacy systems. Therefore, these profiles show movement (or the lack of) between paradigms during implementation. The user log analyses measure complexity of navigation through DSS’s using system logs, and identify problems that would not be apparent by simply watching to see if users spend time at terminals. The financial impact measures provide an alternative to the site’s usual measures while at the same time identifying lingering problem areas. This paper details each method in turn, and gives example data from actual sites. The paper concludes by showing how the results of each sub-method can be linked together, showing, for example, how paradigm shifts in users are affecting financial performance of shops.

**UNDERSTANDING USERS WITH COGNITIVE
PROBE INTERVIEWS**

One of profound features of Decision Support Technologies is that – in essence – they are actually computerized instantiations of formal theories with clearly defined relations between conceptual objects. A consequence of this is that they are idealized and rigid. In fact, one of the complaints about DSS’s that they “assume a perfect world”, and in fact, they do. However, talented users who understand the “real world” better than the computer does find ways to use the systems’ rigidity to their advantage. For example, many use a DSS system as a constant for evaluating real

world occurrences; e.g., if an MRP (Manufacturing resource planning) system reliably predicts a steady lead time for a lot size of material to be processed in a metal press, an unexpected variance with that prediction might alert workers that the machine itself is slowing down and developing mechanical difficulties.

In order to use the system’s rigidity, however, -- as should be clear from the example above -- users must implicitly understand the formal theory represented by the DSS before they can do anything creative or even useful with the system. Just as someone needs to understand the system of arabic numerals within the system of arithmetic before using either of them to make and balance a household budget, a user must understand the formal theory of a DSS before he or she can creatively use it to change practice. This is not a new idea. However, the difficulty has been knowing how a user “knows” something. The kind of knowing we are describing has been called by others “intuitive expertise” (e.g., Polanyi 1986, Dreyfus & Dreyfus 1986, Dreyfus 1997; Klein & Hoffman, 1993; Orasanu & Connolly 1993) and yet eludes easy measurement. For example, many who perform well on written tests covering verbalized concepts about various systems are not effective or even competent users (DiBello & Glick, 1992) The kind of knowing that is required from innovative DSS users is not so much a quantitative difference in what is known about, but rather a qualitative difference in how the knowledge is mapped onto practice.

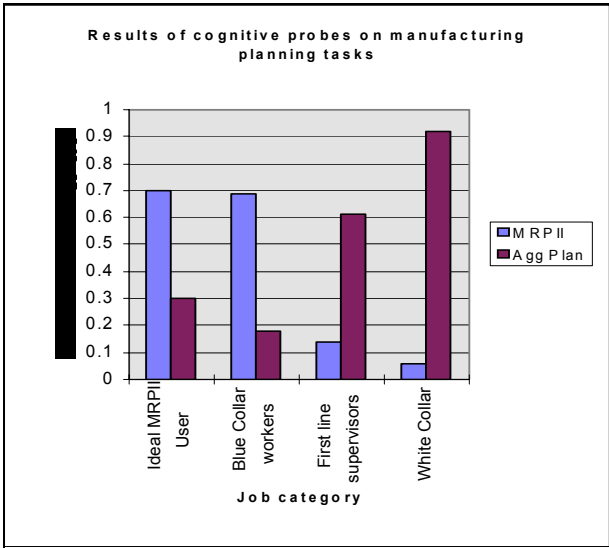
Since we are interested in such levels of knowing, we have designed domain specific “probe batteries” that get at how a person is thinking within a given domain without relying on verbal report or even conscious awareness of preferred strategies.

Discussions of the preparation and skill required to develop a probe battery are beyond the scope of this paper, but examples exist in the cognitive psychology literature (e.g., DiBello, 1996; 1996b Lesgold 1993) . What follows here is a discussion of its basic features and value to a DSS implementation.

The idea behind a probe battery is that one cannot really understand thinking as related in post hoc narrative. Rather, thinking is best understood as an activity and must be observed in action. Also, knowledge cannot be understood in an historic vacuum or out of context (Olson 1994; Spender, in press). Therefore, a probe battery seeks to characterize an individual’s knowledge of both the guiding formal principles of the target DSS and either legacy technologies or practices (or both). In that sense, the probe attempts to characterize an individual’s readiness for the paradigm shift needed to make the DSS implementation successful.

The battery consists of a number of carefully designed “toy tasks” that can be successfully solved by either using old methods, or by using the logic and approach underlying the target DSS. For example, if the implementation involves a cycle based preventive maintenance system, the interviewees would be asked to look at histories and inspection reports and predict what problems are going to occur next with the equipment. The histories and inspection reports would contain information that could lead to identifying either “reactive” oriented problems (Chronic undiagnosed symptoms) or cyclical patterns of wear and tear that indicate a preventive repair is due soon. The interviewee may notice only the indicators of chronic problem, or only the indicators of cycles, or notice both kinds of data. In any case, analysis of the data used and strategy for using it helps to identify the dominant problem solving paradigm of the interviewee. The protocol is scored by counting the number of different strategies employed from each the legacy and target approach and comparing that list to a pre-determined checklist of all possible strategies for that task. A proportional score is then calculated, indicating “how much” of the person’s approach is legacy oriented vs. How much is in line with the planned changes.

For example, the chart below compares workers in various titles on their manufacturing planning approaches. From previous research, an “ideal” knowledge profile was identified and we wished to see which departments and job titles had skills most similar as to better plan system deployment and training .



When conducted on a statistically significant sample of individuals (about 40 or 50), the probe results can also reveal the dominant paradigm of an entire workplace. Our work has shown that – in a very real sense – the results of probe batteries reveal more about the decision

culture of a workplace than they do about individuals per se. That is, workers in the same job title from the same workplace show significant homogeneity in their decision profiles.

Because of this, we have found it useful to “pre-probe” and “post-probe” a sample of individuals to measure the change in workers’ thinking. In fact, this method has been very useful to measure the impact of training. (See Chamberlain & DiBello 1997; DiBello & Spender 1996 for discussions and results of pre and post probes).

MEASURING PERFORMANCE USING SYSTEM ARCHIVES

There has been significant discussion that DSS’s don’t work unless a large number of workers who are “close” to the work being done influence the data being entered. Various approaches have been used to ensure data integrity. One of these – the use of data entry clerks to type in worker’s comments – has met with particularly disappointing results. In a study of four sites (DiBello and Kindred, in preparation) it was found that little or no correspondence existed between worker’s handwritten reports and their electronic counterparts. In one site, workers had noticed the lack of veracity in system records, and had stopped handing in accurate handwritten reports. Most striking, however, is the lack of good methods for measuring workers’ use of the system and data integrity.

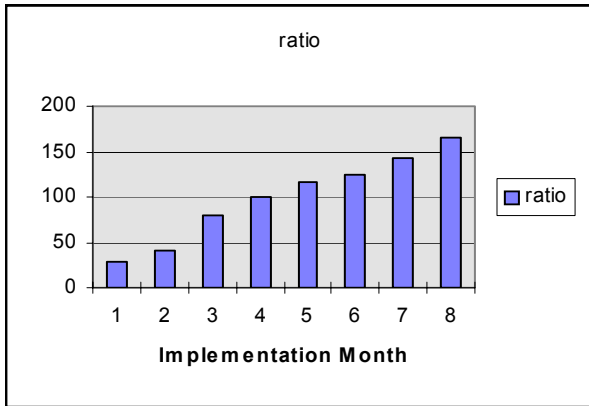
We have found that a great wealth of information is contained in the transaction logs already built into most network resident systems. In order to manage the traffic of logins and transactions, handle billing (when network time is leased) and allocate user-specific access, most systems literally log every transaction that takes place. In many cases, this basic logging function can be modified to collect specific information about each transaction, such as functions used, queries made or specific field-dependent data entered. A carefully thought-out analysis of these data can reveal much about how a system is being used, by whom, and in what areas of the workplace.

For example, for a site implementing MRPII, we measured planners use of the “what if” simulation module to see if they were using it to make better decisions about master scheduling. We reasoned that if the planners were using the system properly, they would access the simulation module frequently at the beginning of the month (when plans are modified using new sales orders) and make fewer changes in the schedule throughout the month, once a plan had been decided upon. The problem at this site had been poor planning and frequent reactive changes to the planning spreadsheets. We downloaded information on the

number of accesses to the simulation module by planners by week, for eight weeks. We then downloaded transactions concerning changes in the schedule. These data were summarized for each month, calculating the mean degree of change for each month, in days. We found that both the number of changes and the extremity of the change decreased with increased use of the simulation module.

Another example involves shop floor mechanics collecting repair data for a preventive maintenance system. For this system, data detail and accuracy are critical so that previously undocumented wear and tear cycles could be identified in a large fleet of public transportation vehicles. In this site, mechanics did their own data entry using terminals placed next to workbenches and vehicle lifts. Very little or no handwritten records were kept. After some analysis of the work practices, we determined that the “component code” was the key to the system’s success. For example, in an engine with hundreds of parts, indicating that the engine as a “component” has a problem is not helpful. We needed mechanics to specify the code or stock number for exact part *within* the engine or other major component. One way we measured quality of input was by downloading all transactions in which a component is specified as having a problem. Then, the overall average component code frequency was measured for each person, for each month. For example, we found that users who entered accurate data used any particular code less than two times a month, even though they averaged about 10 to 20 transactions a day. Users who do not understand the importance of detail use the same few general codes over and over and, hence, their average frequency was very high

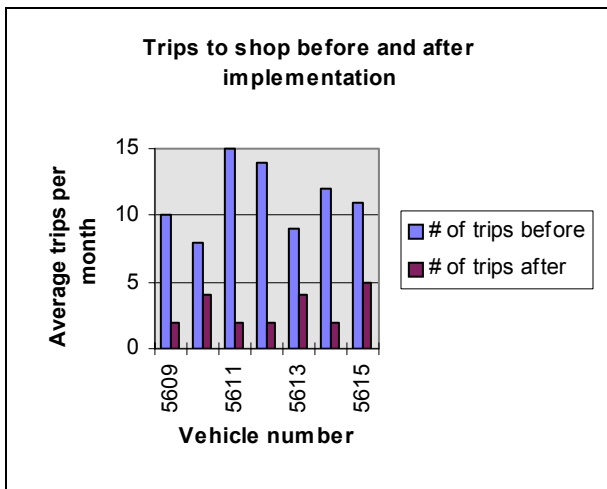
Obviously, as users increased their number of transactions, there was a limit to how much variability there could be in their codes (as a function of the finite number of components). We controlled for this by calculating the ratio of individual average frequency to number of transactions. As the user becomes more proficient at identifying problem parts at an appropriate level of detail, the ratio should continue to increase, even if the mean frequency begins to stabilize or increase. The chart below shows a well-running implementation. Within the first 8 months, users are showing a steadily increasing ratio of component frequency to number of transactions.



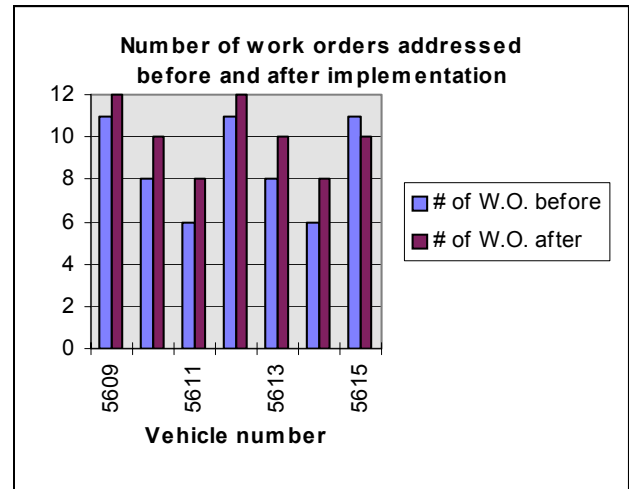
FINANCIAL PERFORMANCE MEASURES.

The entire reason for investing in an expensive Decision Support system is often hard to justify if measures of financial performance are not sensitive enough to show short term impact. Typically, DSS's are being implemented in industries with notoriously poor records and the implementation is often the first attempt to document what is going on. Again, using the system's own data (although not its pre-packaged reports) in creative ways can show both short term improvement or – more importantly – developing problems. Often it is just a matter of asking the right questions and understanding how the system collects data.

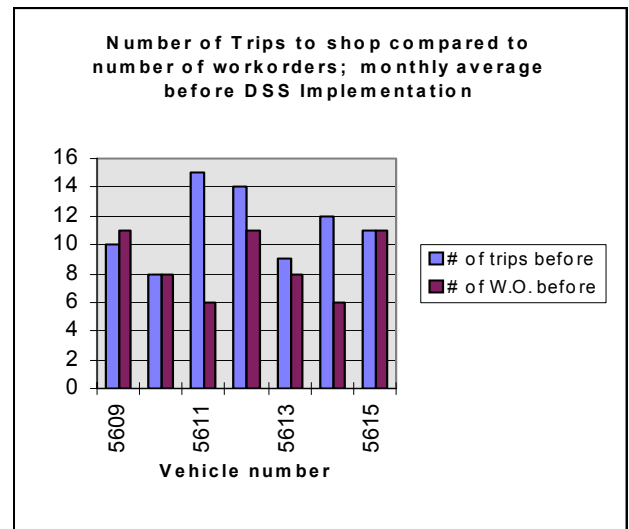
For example, at a northeastern public transportation site, one way to show short term impact was to show that the vehicles being maintained were making fewer trips to the shop and were, hence, spending more time collecting rider's fares. There was no pre-created report for doing this, but the system did record every work order created, by vehicle number and date. Therefore, we were able to graphically represent the number of days on the road Vs days in the shop, for each vehicle, at each location and during each month.



Likewise, we reasoned that if users were truly understanding preventive maintenance, they would soon realize that predictable maintenance or preventive replacement can be *coordinated in time*, resulting in *more work* being done during each visit to the shop. Comparing the chart above to the chart below illustrates this point. As can be seen, the same vehicles which visited the shop less frequently after implementation had as many or more work orders. The difference is that these work orders were coordinated to be addressed *during planned shop visits*.

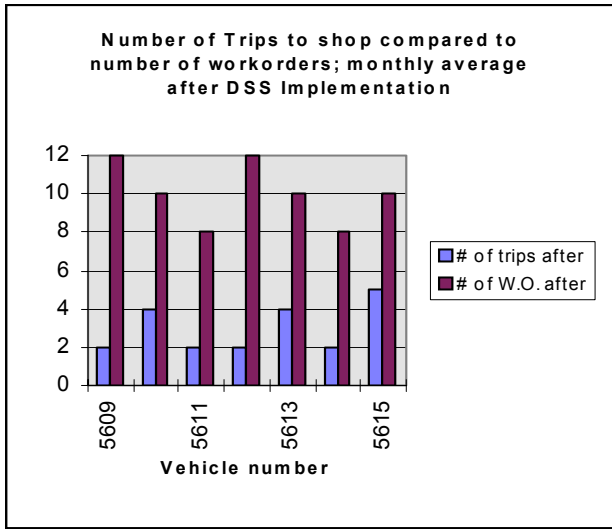


As can be seen, the total number of work orders actually increased, while the number of trips to the shop dramatically decreased for the same period of time. The two charts below compare the relationship of number of trips to number of work orders for the same vehicles, before and after implementation.

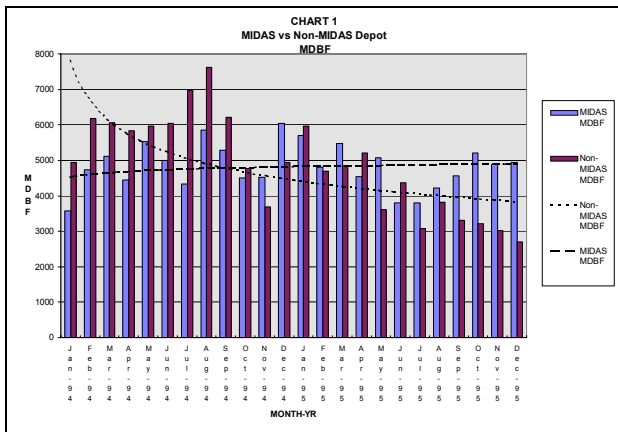


The first shows that the number of trips were equal to or exceeded the number of work orders satisfied, suggesting that vehicles had to return to the shop

multiple times for the same work order, possibly because parts were not available for unplanned work. The second chart shows that multiple jobs are being handled in a single trip, suggesting better planning and coordination.



To look at the effectiveness of the repairs – another way to examine the integrity of the information in the system – we examined mean distance between failures for the maintained equipment. If the work orders created from system information are addressing actual problems, MDBF should go up.



CONCLUSION

As should be clear, one of the weakest points in any implementation strategy – sensitive measurement – can be addressed by borrowing the methods developed in the social sciences to characterize human behavior. Since the successful deployment of Decision support technology is mainly an issue of changing practice and shifting paradigms, the adaptation of social science to assist technology implementation is a deceptively obvious match. What remains is linking the results of various measures to each other and developing ways to

diagnose at which level intervention is required when a technology implementation is not going well. Although we have found ways to make meaningful links between measurable paradigm shifts and financial impact (Chamberlain & DiBello, 1997), ways to use these links to diagnose trouble quickly are only now developing. For example, in one site, user knowledge had undergone radical shifts (as shown in cognitive probe results) but workers transferred new knowledge to work practice for only three weeks before returning to previous ways of working (as shown in system transactions). Further research revealed that implicit “punishments” for innovative thinking were in place in the structure of foreman/worker relationships. In this case we were able to intervene (DiBello & Kindred 1995), but more systematic ways of looking for links must be developed.

REFERENCES.

Chamberlain, E.S. III & Di Bello, L. (1997). Iterative Design and Implementation: A model of successful scheduled maintenance technology deployment. Conference proceedings of the Annual Transit Research Board

Di Bello, L. (1996). Exploring the relationship between activity and the development of expertise; paradigm shifts and decision defaults. In *Naturalistic Decision Making*, (C. Zsombok & G. Klein eds.) "

Di Bello, L. (1996). Providing multiple "ways in" to expertise for learners with different backgrounds: When it works and what it suggests about adult cognitive development. Special Issue of the Journal of Experimental and Theoretical Artificial Intelligence.

Di Bello, L. & Glick J. (1992). The relative roles of on-the-job experience and classroom training for employees learning to use MRP effectively. Paper presented at the Special Academic Session of the APICS 25th International Conference, sponsored by the APICS Educational and Research Foundation. October 18-19, 1992. Montreal, Canada.

Di Bello, L. & Kindred, J. (in preparation). Comparing the virtual to the real; what are the data like when workers input their own information directly and when they don't.

Di Bello, L. & Spender J-C (1996). Constructive learning: a new approach to deploying technological systems into the workplace. *International Journal of Technology Management*.

Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Kindred, J. & Di Bello, L. (1995). "Reframing in depth learning and development: Comparing constructive activities and knowledge "replacement" with traditional training and knowledge "acquisition". Presented as part of "Sociocultural studies of development and learning at work", Annual Meeting of the A.E.R.A. April 22, 1995. Hilton Hotel, San Francisco CA.

Klein, G.A. & Hoffman, R.R. (1993) Seeing the invisible: Perceptual-cognitive aspects of expertise. In Rabinowitz, M. (ed.) *Cognitive science foundations of instruction*. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Lesgold, A. (1993). An object-based situational approach to task analysis. In M. Caillot (Ed.), *Learning electricity and electronics with advanced educational technology* (pp. 291-302). NATO ASI Series F, Vol. 115. Berlin: Springer Verlag.

Olson, G.M. (1994) Situated cognition (pp. 971-973). In R.E. Sternberg (Ed.), *Encyclopedia of Intelligence*. New York: Macmillan.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Caderood, & C. Zsombok (Eds), *Decision making in action: models and methods*. Norwood, NJ: Ablex.

Polanyi, M. (1986). *The tacit dimension*. Garden City, NY: Doubleday.

Spender, J.C. (in press). Knowing, managing and learning: A dynamic managerial epistemology
Management Learning.