Experimental Research into Usability and Organizational Impact of Workflow Software

Peter Boersma[1]
Stef M.M. Joosten[2]
Gerrit C. Van der Veer[2]
{boersma,joosten}@cs.utwente.nl

Abstract

The University of Twente has started an Open Workflow Laboratory to facilitate *experimental research* in the area of workflow software, since this has been a hiatus in workflow research so far. This article reports the design and results of the first series of experiments conducted in this laboratory: In a simulated office environment seven groups of 5 subjects used both a non-workflow and a workflow application to deal with instances of a loan application process. We show that the transfer to workflow software results in an increase in productivity and spare time, and has mixed effects on consultation times. Recommendations regarding future experiments and the Open Workflow Laboratory are given and a scenario for future workflow experiments is proposed.

1 Introduction

The University of Twente has introduced the idea of an Open Workflow Laboratory. This will be a place facilitating workflow research and education by and for students, researchers, buyers and manufacturers of workflow tools. It will provide several workflow tools operating on multiple hard- and software platforms in company-like environments. The purpose of the Open Workflow Laboratory is to acquire and maintain live expertise and experience with workflow tools and to make this expertise available in the public domain.

The experiments reported here were conducted in the Open Workflow Laboratory context. Workflow and usability researchers from universities and the industry form the intended audience of this article.

Structure This report starts with an indication of the problem underlying the reported experiments: The lack of experimental workflow research. A problem analysis and hypothesis formulation are performed in section 3, followed by the design of the experiments, the results and the conclusions. The conclusions are preceded by a description of the validity analysis that was performed.

2 Problem Definition

From the literature on workflow tools and their usage it becomes clear that the full effects of introducing a workflow tool into an organization, automating one or more business processes, are unknown. Predictions of productivity increase and complexity decrease are based on assumptions, sales talk and single-case successes. This is caused by the fact that no sound research in the area of workflow tool usability and impact has been conducted as

[1]Design Methodology Research Group, Computer Science Department, University of Twente
[2]Centre for Telematics and Information Technology, University of Twente, The Netherlands

yet.

The goal of this research project is to conduct a series of multidisciplinar experiments in the context of the Open Workflow Laboratory. The research topic will be the effect of the introduction of a workflow tool into an organization on the tasks of the members of the organization, so recommendations for designers and introduction specialists regarding this topic can be expected. To reach the aforementioned goal the following research question was formulated:

> *What are the effects of introducing a workflow tool into a simulated organization on productivity, consultation time, and spare time?*

The specific topics were chosen after conducting a literature survey, of wich the outcome indicated that *"... there is a strong motivation to chart the various changes so that cause - effect relations can be identified, wich would enable system designers and policy makers to predict the kinds of systems that will be most beneficial"*([Preece et al., 1994], page 188). Among the expected 'various changes' are changes in the following variables:

productivity Kirkbride [Kirkbride, 1992] expects the possibility to concentrate on primary tasks, and several sources expect an increase in productivity (see [Joosten et al., 1994]).

consultation time Palermo and McCready [Palermo and McCready, 1992] expect a decrease in complexity, and consultation time is expected to decrease as complexity decreases.

spare time a lightening and shortening of tasks is expected by the writers of the Ovum report on workflow management software [Hales and Lovery, 1991].

3 Analysis and Hypothesis Formulation

The research question has to be operationalized to allow measurements to be taken and processed to results: The independent and dependent variables have to be declared, and hypotheses have to be formulated.

Variables The research question indicates that introducing a workflow tool into an organization will change times spent on certain activities. To detect these changes *two situations* have to be created: One in which activities are supported by a non-workflow tool and one in which a workflow tool is present to support activities. Therefore the *independent variable* is declared as follows:

> *The presence of a workflow tool to support activities in a business-process.*

The *dependent variables*, as introduced in section 2, are declared as shown in table 1. All variables that refer to a situation where a workflow tool is present, have a '$_w$' subscript to them.

Dependent variables		
without	with	description
P	P_w	Productivity, the number of tasks performed per situation
C	C_w	Consulting ratio, the seconds of consultation with fellow workers and the experimenter, related to the seconds of primary activity
W	W_w	Waiting ratio, the seconds of spare time per situation, related to the seconds of primary activity
'Primary activity' is defined as: the time the situation is valid, minus time spent on consulting. This means that spare time is regarded as being part of the primary activities.		

Table 1: Dependent variables and their meaning

Hypotheses In [Wasser, 1992], Jurgen Wasser

states that productivity increases of 40 to 60% are not uncommon when workflow tools are introduced. This gives rise to the following hypothesis:

> *Hypothesis (1): Productivity will increase when using a workflow tool.*

Or, in terms of P:

$$H_0 : P_w = P$$
$$H_1 : P_w > P$$

In [Palermo and McCready, 1992], Ann Palermo and Scott McCready state that the complexity of tasks will decrease when workflow tools are introduced. Since consulting is expected to decrease with decreased complexity, this gives rise to the following hypothesis:

> *Hypothesis (2): The ratio of consulting time vs. time spent on primary activities will decrease when using a workflow tool.*

Or, in terms of C:

$$H_0 : C_w = C$$
$$H_1 : C_w < C$$

In [Hales and Lovery, 1991], Keith Hales and Mandy Lovery state that the total elapsed time of a process that has been automated by a workflow tool will decrease significantly (they speak of 'several hours' after automation in stead of 'severals days, possibly weeks' before). This gives rise to the following hypothesis:

> *Hypothesis (3): The ratio of spare time vs. times spent on primary activities will increase when using a workflow tool.*

Or, in terms of W:

$$H_0 : W_w = W$$
$$H_1 : W_w > W$$

4 Design of the Experiments

The design of the experiments included the design of an organization of which subjects would be members, the design of a business-process they would have to work on, the design of the instructions, and the programming of the applications subjects would have to work with. Before actually conducting the experiments, they were evaluated to ensure internal and external validity (as advocated in [Jarvenpaa et al., 1985]). These design steps will be presented here.

Organization The simulated organization is the Group for Short-term Personal Loans of the Social Services Department in the town of Berfelo, The Netherlands. This is a (fictional) town that, according to the instructions that were handed out to the subjects, started the Group for Short-term Personal Loans when it found out that many of its inhabitants at some point in time were in need of money to pay for bills that are only refunded after they have been payed in front (e.g. hospital bills). The Group consists of five civil servants who receive loan requests and try to answer the request as soon as possible. The civil servants have different roles, related to different tasks and responsabilities.

Process The outline of the business-process is described below. In the first situation of the experiment a database program is used to carry out the activities in this process. On transferring to workflow software a few minor changes could be made to this process, including electronically checking and routing of forms. No extensive Business Process Redesign was performed.

Requests for a short-term loan enter the organization by paper mail and are placed in a mailbox. After that, 4 steps with the following activities are necessary to complete the process (see figure 1):

1. The data from the request form are en-

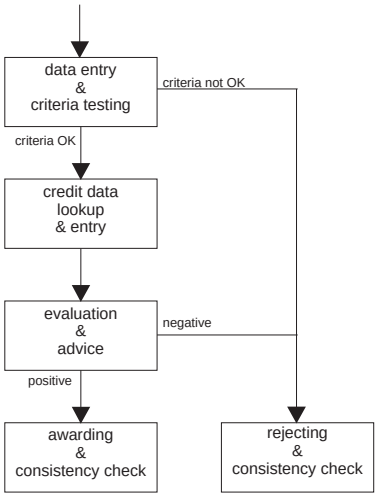

Figure 1: The loan application process

tered into the computer system. Some simple tests are conducted to see if the request meets the official criteria. If the request fails to meet these criteria it is directly transferred to the rejection officer (step 4).

2. On acception of the request, the official credit data of the requestor are looked up in a database which is, in this experiment, implemented as an old-fashioned chest of drawers containing the credit information of all requestors. The credit information is filled out in a second (electronical) form.

3. The next step is evaluating the request and the credit information. In order to be able to give an objective judgement for every request, a list of criteria is present. This is a simplification of reality, since in the real world these evaluation criteria will probably be vague or even partly undefined. To ensure a little reality the criteria are placed in a random order and formulated in different ways. The result of this evaluation is a positive or negative advice.

4. The last step is the filling out of a letter noticing the requestor of awarding or rejection. A final check on data consistency is performed.

This process has the following characteristics that make it suitable for these experiments:

Comprehension The process could be a real life example, and is easy to learn. Anyone with a little fantasy can imagine he or she is a civil servant in Berfelo. Comprehension can be enhanced by adjusting the work environment to the chosen process.

Activities The process provides different activities for several roles. These activities call for interaction and consulting.

Tool support The process can be completed manually, with standard database programs or with a workflow tool. In all cases, the process can be adjusted to the functionalities offered by the supporting tool up to the point were these functionalities can be analyzed (e.g. efficiency or control).

Instructions and Observations Every subject was responsible for two or three steps/activities of a role. Table 2 shows how activities were assigned to the roles. For every step there was an instruction that explained what tasks were to be carried out and how they could be carried out. The data entry subjects and the evaluation subjects were given their list of criteria. The subjects and their immediate surroundigs were observed by two videocamera's during the experiments; afterwards the tapes were analyzed by the experimenter with the help of observation forms.

Subjects Subjects particpated in groups of 5, and the predefined sets of tasks were randomly assigned to the subjects. Selecting the subjects was done by placing an advertisement in the local University Newspaper calling all readers to apply for this subject position. They

Assignments		
role	code	activities
1	de/cr	data entry, credit
2	de/cr	data entry, credit
3	de/cr	data entry, credit
4	cr/ev	credit, evaluating
5	ev/a/r	evaluating, awarding, rejecting

Table 2: Assignment of activities to roles

were offered a reward of Dfl 25,- (approx. 15 US$) for 4 hours of attendance. Both men and women responded and a total of 34 attended the seven experiments. Most of them were university students, but they differed in years of study and in faculty.

To analyze the experience with certain groups of software products all subjects were asked to fill out a questionnaire with which they could scale their experiences. The results from these questionnaires are in table 3.

Questionnaire Results					
	Experience (34 subjects)				
software	– –	–	+/-	+	++
typing	0	6	8	11	9
(GW)BASIC	10	8	10	4	2
Pascal/Modula	4	6	11	7	6
ASSEMBLY	20	7	6	0	2
other low level	17	9	3	3	1
DOS	0	8	7	9	10
UNIX	16	7	6	4	1
other OS	23	4	2	3	2
WINDOWS	1	6	5	15	7
drawing tools	4	6	10	11	3
mouse or. tools	0	7	5	12	10
wordprocessors	0	1	5	17	11
databases	11	9	9	3	2
groupware	28	5	0	1	0
workflow tools	30	3	0	0	1[4]
– – = none, – = some, +/- = reasonable + = much, ++ = very much					

Table 3: Scaling of subjects' experiences

Software Two situations were set up which differed in their use of software and the applications built with this software. In the *first situation*, where no workflow software is used, the data is entered into a NoWorkFlow[4] application. On the computer screen two to three icons represent activities that have to be carried out by a subject. A case of the loan request process is started by a data entry officer taking a request out of the mailbox and clicking with a mouse on the data entry-icon. This results in NoWorkFlow showing an electronic input form on the screen that can be filled out by the subject. It is possible to fill out several forms in a batch-like mode, but it is also possible to close a form on completion. Closing a form allows the subject to start other steps. With NoWorkFlow it is possible to share files among users, so every request form that is entered is almost immediately available to other users who might be able to complete the next step of the process.

In the *second situation*, the workflow tool YesWorkFlow[5] is used to complete the process. Here a main window is shown along with icons representing the tasks to be performed. The main window enables data entry officers to start processing a case. The icons represent work queues and show case descriptions of the tasks to be performed. Since these icons indicate that a user has work to do, it is no longer necessary to circulate the request forms; after the data-entry step, the paper form can be filed.

In both situations, the electronic forms were designed to look like the original forms as much as possible. Both situations were simulated for 30 minutes.

Early Validation In any experiment it is necessary to check internal and external validity, where internal validity means that the experiment itself is designed in a valid way so that *"... the observed effect is due to the experimental treatments and not due to some other extraneous variable"* ([Christensen, 1980], page 92), and external validity means that the results from the ex-

[3] This represents a fellow researcher stepping in for a missing subject

[4] This is an alias

[5] Another alias

periment are applicable in the real-world (provided that the experiment is designed to reflect part of the real world). In these experiments, the following extraneous variables were controlled to ensure *internal validity*:

Personal variables None of the subjects had any previous experience with workflow tools. They were given a chance to practice with both applications: At the start of a half hour situation, one case of the process was started and all appropriate steps of the process were carried out.

The process was designed to be quite simple but had enough reality in it to be treated as a real world example.

Subjects acted as their own control in regard to the independent variable.

Situational variables The presence of other persons in the experiment room was brought to a minimum. The experiments were conducted in a simulated office environment, including filing cabinets and in- and out-boxes that were functional to the tasks. The experimenter had instructions to restrain himself to giving only necessary comments.

Chronological variables As a result of the setup of the experiments, the only logical order was 'no workflow → workflow'.

Subjects were asked to fill out a Software Usability Measurement Inventory (SUMI) form (see section 6 and [Porteous et al., 1994]) at the end of both situations.

All subjects were also asked to scale their experience with several categories of software like database programs, operating systems and generations of programming laguages. These scalings are being analyzed to check *external validity*.

5 Results

After conducting the experiments the raw result material consisted of 68 observation forms (34 with observations of the non-workflow situation, 34 with workflow observations), 34 experience scaling forms, 68 Software Usability Measurement Inventory (SUMI) forms, and transcripts of interviews with subjects. The observations, experience scalings, and SUMI forms were used to produce the tables below, the interviews were used to check conclusions and to improve the experiment.

5.1 Productivity

By how much did the transfer to workflow increase productivity? In table 4 the difference in productivity between the two situations is shown. The last column shows the ratio of the average number of steps carried out *with* the workflow tool vs. the average number of steps carried out *without* the help of the workflow tool, that is P_w/P.

From these tables it is clear that in these experiments productivity was roughly 1.80 times higher when the workflow tool was present to support the process. In table 5 it is shown that this increase in productivity is not the effect of learning, since the numbers in this table do not increase as time flows from 0 to 30 minutes (NoWorkFlow), nor do they increase from 30 to 60 minutes (YesWorkFlow). This provides evidence that the experimental process is a rather simple process that can be learned in one or two practice-rounds.

Productivity			
Activity	P	P_w	P_w/P
entered	11	21	1.9
not accepted	3	7	2.3
credit lookup	8	14	1.8
evaluated	8	14	1.8
awarded/rejected	10	19	1.9

Table 4: Productivity results (average over all experiments)

5.2 Consulting

Since hypothesis (2) states that consulting time will decrease in favor of time spent on primary

Requests over time					
Timespan					
from	to	de	cr	ev	a/r
0–5		2	-	-	-
5–10		2	3	-	-
10–15		2	2	2	2
15–20		3	1	2	3
20–25		2	1	2	3
25–30		0	1	2	2
Introduction of workflow tool					
30–35		3	-	-	-
35–40		3	4	3	3
40–45		5	3	2	1
45–50		4	2	3	4
50–55		2	3	3	3
55–60		4	2	3	3
('-' = no requests available)					

Table 5: Requests over time (average over all experiments)

activities, table 6 was drawn up. It shows the time spent on consulting other subjects and the experimenter. The ratio of consulting time vs. time spent on primary activities, C and C_w are shown in the last column. In the rows the subjects (1 to 5) are shown. Because of the way the observation forms were designed, the ratios can not be given for every task seperateley, but have to be aggregated to subjects. From this table several observations can be made:

- In general, there was hardly any time spent on consulting others or the experimenter.

Summarizing, we state that introducing workflow to support simple tasks (like subjects 1, 2, 3 and 5 had) causes a decrease in the need for consulting, while workflow support for complex tasks (like that of subject 4) allows for an increase in consulting

5.3 Waiting

As indicated before, some subjects had to wait up to several minutes before they could actually start working. When this is observed in a company, it is usually an indication of a bottleneck of some sort. In table 7 the average

Consulting times (in seconds)				
code	consulting	C	consulting$_w$	C_w
de/cr	166	0.09	45	0.03
cr/ev	2	0.00	4	0.00
ev/a/r	2	0.00	4	0.00

Table 6: Consulting times (average over all experiments

spare times of these experiments' subjects are shown.

From these tables several observations can be made:

- Subjects 1, 2 and 3 worked a lot and hardly took time off. This is probably because they were the starters of cases. They must have (and admitted this during the interviews) felt obliged to work on, to 'feed' the other subject work.

- The work of subject 4 could be done quicker with the help of the workflow tool, which caused longer spare times, 1.4 to 2.8 times as long as without the workflow tool.

Waiting times (in seconds)				
code	waiting	W	waiting$_w$	W_w
de/cr	80	0.05	0	0.00
cr/ev	1395	0.87	1442	0.87
ev/a/r	1482	0.89	1372	0.80

Table 7: Waiting times (average over all experiments)

6 Validity

As was stated in section 4, these experiments were designed with validity in mind. In the following two subsections the analysis of the internal and external validity will be discussed.

Internal Validity In this experiment many variables were kept constant by letting the same subjects do the same work in both situations. Still some variables might have cor-

rupted the results. Others, like he learning effect, were shown to be non-confounding (see table 5).

A possibly disrupting variable might have been the difference in preference for the two applications. If the subjects would have preferred working with YesWorkFlow over working with the NoWorkFlow application this might have had some effect on the results. That is why the Software Usability Measurement Inventory (SUMI) test was administered, as it gives clues regarding the usability of software or, as the designers say: *"The SUMI measures how usable a software product is according to the perceptions and attitudes of users. It produces a set of valid and reliable numbers which are indicators of the usability of the software being rated."*([Porteous et al., 1994], page 1).

SUMI has a Global scale for usability wich has a standard-normal curve around the median $\mu = 50$ and a standard deviation $\sigma = 10$. Furthermore there are five seperate scales for efficiency, affect, helpfulness, control, and learnability, all with the same curve.

First some ratings according to the Human Factors Research Group, authors of the SUMI User Handbook, will be given:

- Specialized ratings below 50 indicate poor usability and ratings on or below 40 call for remedial action.
- Below average software products will score below 50 on the Global scale and below 50 for most of the specialized scales.
- Above average products will have a Global rating of 50 – 60 and specialized ratings on or above 55.
- If a vendor wants to sell state of the art software he must be sure that his product scores above 60 on the Global scale and above 60 on most of the specialized scales.

Table 8 shows the Global ratings of subjects for both NoWorkFlow and YesWorkFlow. It is clear that there are some differences but that both applications score a Global rating in the 50 – 60 category of above average products. Table 9 shows the five specialized scales and their ratings according to these experiments' subjects. Since the ratings for the 'affect' scale in average do not differ a lot, this variable is considered to be non-confounding.

SUMI Global Ratings	
NoWorkFlow	YesWorkFlow
48	58

Table 8: Ratings for the Software Usability Measurement Inventory (SUMI) Global scale (average over all subjects)

SUMI Specialized Ratings									
effic.		affect		help.		control		learn.	
N	Y	N	Y	N	Y	N	Y	N	Y
46	63	52	58	41	58	49	56	43	58
N = NoWorkFlow, Y = YesWorkFlow									

Table 9: Ratings on SUMI's specialized scales

The previous results show that it is possible to ensure and check *internal validity*. The question whether the results are applicable to the real world can be answered by checking the *external validity*.

External Validity External validity, meaning that the results from these experiments are applicable in the real world, can be tested by checking the representability of the subjects. The random selection of subjects and assignment of roles ensure this partly; a comparison of the subjects' experiences with those of a group of real-world civil servants is underway.

7 Conclusions

The experiments reported here were the first to be conducted in the Open Workflow Laboratory context. Their aim was to reveil the effects of introducing a workflow tool into an

organization, concentrating on usability, productivity, consultation and spare time.

Regarding these Experiments During the discussion of the results, the following conclusions regarding these experiments were proposed:

1. Productivity is roughly doubled (a factor of 1.8 is reported here) when a workflow tool is used to carry out *the same* process that was used before the introduction of the workflow tool.

2. Consulting times were only small percentages of the total times (ratios of 0.00 to 0.10 are common). This may have been the result of designing a rather simple process.

3. Consulting times for *simple tasks* will decrease when supported by a workflow tool. Subjects with *complex tasks* supported by a workflow tool take more time for consulting. This means that the initial hypothesis is neither rejected, nor supported. This conclusion calls for further research to support the following new hypothesis:

   > *Consultation of fellow workers will increase when complex tasks are automated with workflow tools, and decrease when simple tasks are automated with workflow tools.*

4. Users whose job it is to start cases of processes tend to keep working to 'feed' other workers.

5. Workflow tools may cause longer spare times between cases of the same process (increase with a factor of 1.4 to 2.8 is reported here). This can be positively refrased into: The total elapsed time to complete a process will be shortened by introducing a workflow tool.

Regarding Future Experiments From the interviews with subjects, and the discussion of the results, the following conclusions regarding future experiments can be drawn:

1. The loan application process used in these experiments is a suitable process for the purpose. It should however be evaluated to discover the factors that increase complexity in a single case. This way, complexity can be varied in some given amount to determine its effect (see Conclusion regarding these experiments 3).

2. The design of these expriments may also be used in future experiments, since it was conceived as 'natural' by subjects and its results were of use to researchers.

3. The Experimental Research Approach, as descibed by Christensen in [Christensen, 1980], has proven to be useful in designing and conducting experiments, and may serve as a basis for future experiments.

4. The NoWorkFlowapplication needs repair to receive the same SUMI-ratings on the 'affect' scale. This way the software-variable will be constant and therefore non-disruptive.

8 Further Research

The concept of an Open Workflow Laboratory is important, since todays literature shows a lack of, and a call for experimental workflow research to chart the expected changes of introducing a workflow tool into an organization. As long as the full effects of worklow automation are unknown, experimental research will be necessary, and therfore the Open Workflow Laboratory will have to exist.

The conclusions of this project call for further research: More, sounder, and complexer *experiments* uncovering more effects of workflow tools should be conducted and reported to make sure that the right kind of programs

are introduced to the right kind of organizations with predefined effects. To prevent reinventing the wheel a generic scenario for future experiments will have to be designed. A first draft of this scenario is presented in [Boersma, 1994], and describes a method to set up a workflow experiment and techniques to support steps of the method. These steps include: justification of the research goal, choosing a research method, defining variables, ensuring validity, designing the experiment, and testing of hypothesis.

Empirical research should be conducted to report the effects that actually occur in real business environments to back up experimentally acquired knowledge and to search for effects that are not yet understood. The result of one empirical research project is the WA-12 report ([Joosten et al., 1994]).

Normative research should be conducted, focussing on methods and techniques for the development and validation of workflow tools (e.g. by formalization, see [Geertsma, 1994]). The ultimate goal should be a workflow methodology.

References

[Boersma, 1994] Boersma, P. (1994). Experimental research into usability and organizational impact of workflow software. Master's thesis, University of Twente, Enschede, The Netherlands.

[Christensen, 1980] Christensen, L. B. (1980). *Experimental Methodology*. Allyn and Bacon, Boston, Mass.

[Geertsma, 1994] Geertsma, W. (1994). Some formalizations of the Action Workflow approach. Master's thesis, University of Twente, Enschede, The Netherlands.

[Hales and Lovery, 1991] Hales, K. and Lovery, M. (1991). Workflow management software: the business opportunity. Technical report, Ovum Ltd., 7 Rathmore Street, London W1P 1AF, England.

[Jarvenpaa et al., 1985] Jarvenpaa, S. L., Dickson, G. W., and DeSanctis, G. (1985). Methodological issues in experimental IS research: Experiences and recommendations. *MIS Quarterly*, 9(2):141 – 156.

[Joosten et al., 1994] Joosten, S., Aussems, G., Duitshof, M., Huffmeijer, R., and Mulder, E. (1994). WA-12: an empirical study about the practice of workflow management. Technical report, University of Twente, Enschede, The Netherlands. Research monograph.

[Kirkbride, 1992] Kirkbride, L. (1992). Meeting the software support challenge with groupware technologies. In Coleman, D. D., editor, *Groupware '92*, pages 450–455, San Mateo, CA 94403, USA. Morgan Kauffman Publishers.

[Palermo and McCready, 1992] Palermo, A. M. and McCready, S. C. (1992). Workflow software : A primer. In Coleman, D. D., editor, *Groupware '92*, pages 155–159, San Mateo, CA 94403, USA. Morgan Kauffman Publishers.

[Porteous et al., 1994] Porteous, M., Kirakowski, J., and Corbett, M. (1994). *Software Usability Measurement Inventory User Handbook*. Human Factors Research Group, University College, Cork.

[Preece et al., 1994] Preece, J. et al. (1994). *Human Computer Interaction*. Addison-Wesley, Workingham, England.

[Wasser, 1992] Wasser, J. (1992). Workflow: A new look on automation (in Dutch: Workflow: een nieuwe kijk op automatisering). *VIP*, pages 59 – 60.