Quantifying the Performance of Workflows

M. H. Jansen-Vullers, P. A. M. Kleingeld and M. Netjes
Department of Technology Management, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract  Business process redesign is one of the most powerful ways to boost business performance and to improve customer satisfaction (Limam Mansar & Reijers, 2005). A possible approach to business process redesign is using redesign best practices. A previous study identified a set of 29 different redesign best practices (Reijers, 2003). However, little is known about the exact impact of these redesign best practices on workflow performance. This study proposes an approach that can be used to quantify the impact of a business process redesign project on all dimensions of workflow performance. The approach consists of a large set of performance measures and a simulation toolkit. It supports the quantification of the impact of the implementation of redesign best practices, in order to determine what best practice or combination of best practices leads to the most favorable effect in a specific business process. The approach is developed based on a quantification project for the parallel best practice and is validated with two other quantification projects, namely for the knockout and triage best practices.

Keywords  business process redesign, business process simulation, best practices, performance measurement

Introduction

The domain of business process redesign can roughly be divided into two different approaches: the revolutionary and the evolutionary approach. In the revolutionary approach, a redesign starts from a clean sheet. In the evolutionary approach, the existing business process is taken as a starting point. An example of this approach is the application of redesign best practices. Reijers provided an overview of all best practices currently encountered in literature (Reijers, 2003); a short summary is given in the Appendix. Further, a rough qualitative estimation of the expected impact was given (Reijers & Limam Mansar, 2005). However, quantitative research is necessary to determine a more concrete impact of one or more redesign best practices on the performance of a workflow. Although not much is known about the impact of redesign best practices on the performance of a workflow, some papers have been found that are based on a quantitative study. These studies include several best practices: knockout best practice (Aalst, 2001), extra resources best practice (Goverde & Reijers, 1998), specialist-generalist best practice (Goverde & Reijers, 1998; Netjes, Aalst, & Reijers, 2005), flexible assignment best practice (Netjes et al., 2005) and task composition, triage and case types best practice (Zapf & Heinzl, 2000).

The main shortcoming of the abovementioned literature is that none of the authors, with the exception of Aalst (2001), provided guidelines for the redesign of workflows: what best practice should be applied in what situation, process, or setting? Other deficiencies are the lack of a general approach to quantify the impact of best practices, the limited number of different dimensions of performance, and the limited number of aspects per measured dimension. Further, none of the authors, with the exception of (Netjes et al., 2005), quantified the impact of the simultaneous implementation of more than one best practice.

In our research, we aimed to quantify redesign best practices on as many dimensions as possible. This paper provides an overview of possible performance dimensions and related performance measures. These performance measures have been applied in a simulation study to quantify the impact of a redesign best practice, i.e., the parallel best practice. In the parallel best practice, one considers whether tasks may be executed in parallel.
The setup of the paper is as follows. First, the dimensions of performance are summarized. Then, the quantification approach is introduced, including the setup of the simulations, the approach when comparing different variants, and the statistical analysis. We carried out three simulation projects; one to develop the approach and two to validate it. The results of these simulations (i.e., the impact on the identified performance measures) for the best practices involved are shown, followed by a discussion of the results.

**Performance Measurement**

This study focused on the quantification of the impact of a redesign best practice on the performance of a business process. Subject of study was the business process that is being redesigned, in contrast to, for example, the performance of individual employees or the optimization of all processes within an organization which compete with each other for resources.

In the last 20 years, a variety of performance measurement systems has been developed. We assessed the literature on this subject to see what dimensions of performance the authors discerned and which are suitable for measuring business process performance. The following six systems have been considered:

- performance pyramid (Cross & Lynch, 1988/1989);
- performance measurement matrix (Keegan, Eiler, & Jones, 1989);
- results/determinants matrix (Brignall, Fitzgerald, Johnston, & Silvestro, 1991);
- balanced scorecard (Kaplan & Norton, 1992);
- devil’s quadrangle (Brand & Kolk, 1995); and

The assessment resulted in five dimensions of performance: time, cost, external quality, internal quality, and flexibility. These dimensions are all present in the devil’s quadrangle. Furthermore, the other performance measurement systems do not provide additional relevant dimensions. An extensive overview and validation of the dimensions, the relevant measures per dimension and their operationalization can be found in (Jansen-Vullers, Looschilder, Kleingeld, & Reijers, 2007). Here, we suffice with a brief overview.

**The time dimension.** Time has been described as both a source of competitive advantage and a fundamental measure of performance. Based on the information on time measurements found in the literature, we derived a set of performance measures for the time dimension, specifically for workflows, consisting of lead time and throughput time.

Lead time is the time it takes to handle an entire case. Throughput time is the time between the moment a task is completed, and the moment the next task is completed. Throughput time is composed of: service time, queue time, wait time, move time, and setup time.

**The cost dimension.** The cost dimension is closely related to the other dimensions. For example, long lead times can result in a more costly process, low quality can lead to expensive rework, and low flexibility can also result in a more costly process execution. Focusing on the direct costs of running a process, we discerned running costs (for labor, machinery, and training), inventory costs, transport costs, administrative costs, and resource utilization costs.

**The quality dimension.** The quality of a workflow can be judged from at least two angles. External quality is defined from the customer’s side, i.e., the person or organization that initiates the workflow and will receive the output. Internal quality is defined from the worker’s side.

External quality can be measured as client satisfaction with either the product (output) or the process. Satisfaction with the product is the degree to which the customer feels that the product is according to specification or feels satisfaction with the delivered product. The satisfaction of a customer with the process relates to the way a workflow is executed (Reijers, 2003). Literature has been found on both the quality of a product, and the quality of a process. Quality of the output takes into account product performance, conformance and serviceability; whereas, quality of the process considers information availability and bureaucratic language simplification. These measures were included in our study.

Internal quality can be seen as the quality of a workflow from an operator’s perspective. Research in the area of work psychology has identified characteristics of jobs and tasks that are inherently motivating and satisfying and thus reflect high internal quality. These include: (1) a whole and identifiable piece of work is completed; (2) a variety of skills need to be used; (3) the work has a substantial impact on the lives or work of others; (4) substantial autonomy is provided; and (5) direct and clear feedback about performance effectiveness is available [e.g., Hackman & Oldham (1976)]. In addition to these job characteristics, group factors (e.g., cohesiveness and communication quality) and leader factors (e.g., leadership style) influence an operator’s motivation and job satisfaction (Kinicki, McKee-Ryan, Schriesheim, & Carson, 2002).

**The flexibility dimension.** Flexibility is the least noted criterion to measure the effect of a redesign effort. Flexibility can be defined as “the ability to react to changes.” It appears that flexibility can be identified for individual resources, for individual tasks, and for the workflow (process) as a whole. Five types of flexibility can be...
distinguished. Mix flexibility is the ability to process different kinds of cases (per resource, task, or workflow). Labor flexibility reflects the ability to perform different tasks (per resource or per workflow). On the workflow level we further distinguished routing flexibility (the ability to process a case by using multiple routes, i.e., the number of different sequences in the workflow), volume flexibility (the ability to handle changing volumes of input) and process modification flexibility (the ability to modify the process, e.g., the number of sub flows in the workflow, complexity, number of outsourced tasks, etc.)

Operationalization. Operationalization of the time, cost, and flexibility dimensions is quite straightforward. Measuring internal and external quality in a workflow model is less straightforward than measuring time or costs because many different factors influence and determine quality. For example, with respect to internal quality differences among people moderate how they react to the complexity and challenge of their work (Hackman & Oldham, 1976). To settle this, we decided to list (measurable) aspects of those dimensions and consider them proxies (e.g., the number of executed tasks and case types per resource, and the number of authorized decisions): a change in one or more of the aspects will have some impact on the quality dimension. However, the exact extent of impact cannot be determined in a simulation model.

Quantification Approach

Based on the quantification project performed for the parallel best practice, a generalized quantification approach was developed. This approach starts with a redesign quantification plan, based on (Law & Kelton, 2000) and (Mehta, 2000). The plan consists of 8 steps, of which steps 1 to 4 are mainly general steps in a simulation study: (1) project definition, (2) definition and building of a model of the original situation, (3) validation of this model, and (4) definition and building of a model of the redesigned situation. Step 5 (design of the experiments), step 6 (execution of the simulation runs), and step 7 (analysis of the output) are more specific for this kind of quantification projects. Finally, in step 8, conclusions are drawn.

The Redesign Best Practices Quantification Plan

Step 1: Project definition. The main objective of a quantification project is the collection of evidence to reject or support a proposition. In this case, the impact of the implementation of a certain redesign best practice was quantified. Literature can be used to set the objectives. The work of Reijers (Reijers, 2003) can be used as a literature guide.

Step 2: Definition and building of a model of the original situation. We created a high-level Petri net model of the original situation in CPN Tools (Jensen, 1997), which could be used as a starting point for the simulations. The model can be used directly or changed where necessary in order to measure the impact of a certain best practice. The basic model consists of six identical sequential tasks with an exponentially distributed service time with a mean value of 40 minutes and an equally distributed setup time with a mean value of 2 minutes. The model is very flexible and easy to adapt and also includes monitors for the specified operational performance measures.

Step 3: Validation of the model. Our basic model was validated through a comparison of the results of the simulation with the analytical outcomes of mathematical queuing models (Mehta, 2000). The mathematical model is a network of queues, i.e., a Jackson network and consists of 6 M/M/s queues (Kulkarni, 1999). With the formulas of Kulkarni (Kulkarni, 1999) a number of performance measures could be calculated: utilization of the resources, expected number of cases in the queue, expected queuing time, and expected time of a case in the system. After simulation of the CPN model, the results were collected and analyzed, and the 95% confidence intervals were calculated.

Step 4: Definition and building a model of the redesigned situation. Based on the model of the original situation, a redesign was created. Again, the work of Reijers (Reijers, 2003) could be used as a literature guide to acquire detailed insight. The CPN model of the original situation can be adapted to benefit from the structure and monitors already available.

Step 5: Design of the experiments. The design of the experiments is a very important part of the project, because the correct setup of the simulations is essential for the success of the simulation project. This step consists of five sub steps that should be followed before the actual simulation runs can be executed. The first two sub steps concern the selection of introducible variations. The parameters of the simulations are calculated in the remaining sub steps. We describe each sub step in more detail: below.

Step 5a: Choice of variations. Variations are introduced in the simulation models of the original and redesigned situation, to test the impact of a specific best practice under different settings. Variations in arrival rates, resource classes, number of resources, service times, and resource skills are examples of introducible variations. The selected variations and degrees of variation are different for every best practice. A profound knowledge of the redesign best practice—obtained from the literature—is required to be able to select the correct variations and to be as complete as possible in the selection. The types and degrees of variation should be chosen in such a way that eventually conclusions can be drawn about the impact of the implementation of the best practice in different situations.

Step 5b: Specification of model variants. Model variants specify what combinations of variations are used. An
example of a model variant is a model with a high arrival rate, low service times, and two resource classes. The number of variations and model variants determines the number of simulation runs.

**Step 5c:** Calculation of the warm-up period. The warm-up period is the amount of time a model needs to come to steady state. In this study, the time series method was used to calculate this. This was done based on a pilot run of 20 replications and the calculation of the WIP costs (Work In Progress) in relation to the model time (Mehta, 2000). This resulted in a warm-up length of 4800 minutes (= 2 simulation weeks).

Because CPN Tools resets the model after each replication, the initial state does not represent the normal working conditions of the actual system, and every replication has to start with a warm-up period (Mehta, 2000). This is the amount of time a model needs to come to a steady state.

**Step 5d:** Determination of run length. CPN Tools resets the model after every replication. We assumed that the seed of the random generator in CPN Tools produced independent number streams and that the results thus were independent. We used a run length of 10 working weeks. As the warm-up length was 4800 minutes, there were 19,200 minutes remaining for data collection.

The length of one single run must be long enough for the resulting data to be independent. The commonly used Von Neumann ratio would not be appropriate for this study as CPN Tools resets the model after every replication. Therefore, the model must warm-up before every single replication. Based on Law and Kelton (2000), data were plotted in a scatter diagram to investigate dependency with lead time of the cases. This resulted in a run length of 10 working weeks (24,000 minutes). As the warm-up length is 4800 minutes, there are 19,200 minutes remaining for data collection.

**Step 5e:** Calculation of the number of replications. Due to the very nature of random numbers, it is imprudent to draw conclusions from a model based on the results generated by a single model run (Mehta, 2000). We adopted the approach proposed in (Law & Kelton, 2000) to calculate the number of replications based on a pre-specified precision of the collected data. As a result, 21 replications were used in this study.

**Step 6:** Execution of the simulation runs. In this step, all original and redesigned models are created and simulated and the results are recorded and stored. The simulations are set up according to the parameters (calculated in the previous step) and all performance measures (specified in step 1) are measured. One should bear in mind that simulation of the models of all model variants in CPN Tools requires a lot of time and computer power.

**Step 7:** Analysis of the output. Before the actual analysis of the output data can be done, the comparisons between the different model variants are determined. It is decided what model variants need to be compared in order to comply with the objectives. For example: two model variants with equal resource setups and service times but different arrival rates can be compared, if one of the sub-objectives is to determine what the impact of a certain best practice is on systems with different arrival rates. The selected comparisons form the basis of the analysis of the output data.

When comparing results of simulated real systems, equality of variance cannot be assumed. Therefore a separate-variance-test such as the Welch test is recommended as it is more reliable and conservative (Law & Kelton, 2000). Thus, the hypothesis \( H_0 \) was tested against \( H_1 \) for every performance measure by means of the Welch approach, in order to see what performance measures change significantly in the redesigned model. The hypotheses are:

\[
H_0 : \bar{X}_1 = \bar{X}_2 \\
H_1 : \bar{X}_1 \neq \bar{X}_2,
\]

with \( \bar{X}_1 \) being the mean of the measure in the original model; and \( \bar{X}_2 \) being the mean of the measure in the redesigned model.

When comparing more than two alternatives and calculating several confidence interval statements simultaneously, the individual confidence levels of the separate comparisons have to be adjusted upwards, in order to reduce the number of Type 1 errors (rejecting the null hypothesis when it is true). For this purpose, the Bonferroni equality can be used (Law & Kelton, 2000; Miller Jr., 1981).

It implies that when making some number \( c \) of confidence interval statements, it is needed to make each separate interval at level \( (1 - \alpha)/c \), so that the overall confidence level associated with all intervals’ covering their targets will be at least \( (1 - \alpha) \) (Law & Kelton, 2000; Miller Jr., 1981).

Then the confidence intervals for all differences between the original model and the redesigned model (the Welch confidence intervals with the Bonferroni corrected values for \( \alpha \)) are calculated and this is repeated for all setups and all variants. When the confidence intervals of two or more setups overlap, it can be concluded that the difference between these setups is not significant. Conclusions can be drawn both within and between different model variants.

**Step 8:** Conclusions. Finally, conclusions are drawn based on the analysis and the sub-conclusions of the model variants. Furthermore, a reflection on the quantification is made by comparing the quantitative results and conclusions of the simulation project with the qualitative results of the research of Reijers and Limam Mansar (Limam Mansar & Reijers, 2005; Reijers & Limam Mansar, 2005) and possibly with earlier quantification efforts found in the literature.
Validation of the Quantification Approach

The quantification approach consists of three elements: (1) the set of performance measures; (2) the quantification plan; and (3) auxiliary files to support the execution of the quantification plan.

The basis of the approach is the redesign best practices quantification plan, which should be followed step by step in the simulation process. The auxiliary files (several MS Excel sheets, CPN Tools simulation models and user guides) were created for use in combination with the quantification plan. The files and models are created to increase the consistency of the project, to increase the usability and to save time when quantifying redesign best practices. This holds true for the design of the model, but especially for the monitors in the model that automatically measure all operationalized performance measures. Together, these tools, the performance measures, and the redesign best practices quantification plan form the quantification approach. The complete approach is shown in Figure 1.

The approach has been developed with the simulations of the parallel best practice and validated with the quantification of the knockout and the triage best practices. The setups and results of these quantification processes can be found in (Loosschilder & Jansen-Vullers, 2007a, b, c). The validation showed that the developed approach is suitable for the quantification of other best practices. The iterative nature of steps 5, 6, and 7 is stressed, as is the difficulty of measuring internal and external quality. Some of the results of the simulation projects are reported in the next section.

Results of Quantification Projects

The quantification approach was developed based on a simulation project for the parallel best practice (see item 8 in the appendix) and validated based on simulation studies for the knockout and triage best practices (items 6 and 9 in the appendix). Due to space limitations, only the main results for these best practices are reported here. Each project included about 150 simulations, i.e., 150 • 21 replications. In each subsection, the best practice is described shortly, followed by a number of observations of when the best practice could be applied. This mainly depends on the intensity of the arrival of cases, the assignment of resource classes to particular tasks, and service times of tasks.

![Figure 1. Validated version of the quantification approach.](image-url)
Quantification of the Parallel Best Practice

The parallel best practice runs as follows: consider whether tasks may be executed in parallel. The obvious effect of applying this best practice is that the throughput time may be reduced considerably. The applicability of this best practice in workflow redesign is large. When analyzing existing workflows in organizations we noted that tasks were mostly ordered sequentially without the existence of hard logical restrictions prescribing such an order. A possible disadvantage of introducing more parallelism in workflows with checks is an increase in costs or decrease in flexibility because in parallel, all checks are executed while the result of one check might have been enough. Furthermore, the management of workflows with concurrent behavior becomes more complex, which may introduce errors (quality) or restrict run-time adaptations (flexibility).

For the quantification of the parallel best practice a process model with at least five tasks, that is, a start task, three tasks in parallel, and an end task, is needed. The original model we used for this study consisted of a process with six tasks, named A to F, in a sequence. From this model we created two redesign models: one model with two tasks, B and C, in parallel, and one with three tasks, B, C, and D, in parallel. Further, we came up with several variations to test under which conditions a process would benefit from the application of the parallel best practice. We will elaborate on one of the variations in more detail and then present the results for other variations.

We assumed it would make a difference whether the parallel tasks would be performed by the same resource class or by different resource classes, and this became one of the variations we investigated. Table 1 shows the output data resulting from the simulation of this model variant for the model in which tasks B and C are in parallel. The variant consisted of four resource setups (ABC-DEF, AD-BC-EF, AC-BD-EF, and ACE-BDF). In this context ABC-DEF, for instance, means there were two resource classes, the resources in the first class were able to execute tasks A, B, and C, while the resources in the second class executed tasks D, E, and F. Tasks B and C were put in parallel, so for this setup these tasks shared their resources. Except for the resource classes, settings were the same for each setup. Table 1 shows the lower bounds (LB) and the upper bounds (UB) of the confidence intervals of the relative differences between the original model and the four redesigns for eight performance measures. From these confidence intervals, it can be seen that the implementation of the best practice in this example decreased the lead time and the WIP costs. The wait time of the cases increases in all situations, but this effect does not result in lower lead times since these waiting times are put in parallel. All other measures had insignificant differences with the original situation, as their intervals included 0. This means that these measures were not affected by the implementation of the parallel best practice.

Another comparison that can be made with the output data from different resource class variations is between the various redesigns. It allows for the selection of the best redesign alternative. Figure 2 graphically depicts the confidence intervals for two measures: lead time and volume flexibility. From the graphs of lead time it can be seen that the decrease in lead time of ABC-DEF was significantly higher compared to the other setups. The difference in lead time between setup AD-BC-EF and AC-BD-EF was nonsignificant, because the confidence intervals of both setups overlapped. The decrease in lead time of ACE-BDF was significantly lower than the decrease of the other setups. From the graph of volume flexibility it can be concluded that this measure was not affected by the redesign effort. An automated MS Excel sheet was created to generate this output.

Next to the resource classes we also varied the arrival rate and the service times. The variations in arrival rate showed that the observed positive impact on performance only held for processes with a low arrival rate. The positive result became smaller or even nonsignificant when the arrival rate increased. With a low arrival rate the positive impact of the parallel best practice was higher for
tasks with equal parallel service times than for tasks with completely different parallel service times. In both situations, implementation of the parallel best practice led to a decrease in lead time and WIP costs and therefore appears to be advisable. However, the differences in impact between the two service time variants decreased or even became nonsignificant when the arrival rate increased. Concluding, we advise to implement the parallel best practice when the arrival rate is low. Further, the improvement will benefit from involved tasks sharing resources and having equal service times.

Implementation of the parallel best practice changed the number of parallel tasks, which is a proxy of external quality and process modification flexibility. An increase in the number of parallel tasks led to a more complex workflow, which can result in slightly lower external quality and lower process modification flexibility. The other proxies of external quality and the remaining measures of the flexibility dimension remain unchanged with the implementation of the parallel best practice. Putting tasks in parallel does not change any of the proxies of the internal quality dimension. It is expected that the parallel best practice does not affect the internal quality of a workflow.

**Quantification of the Knockout Best Practice**

A typical part of a workflow is the checking of various conditions that must be satisfied to deliver a positive end result. Any condition that is not met may lead to a termination of that part of the workflow, the knockout. The knockout best practice comprises three possible redesigns.

- **Swapping tasks rule.** If it is possible to choose the order in which the various conditions are checked, the condition that has the most favorable ratio of expected knockout probability versus the expected effort to check the condition should be pursued.
- **Combining tasks rule.** If two tasks are executed by the same resource class, the combination of two tasks into one larger task is considered. As a result, this task can be executed by one resource without interruption.
- **Parallel tasks rule.** Putting tasks in parallel reduces the total flow time. The flow time is minimized by putting as many tasks in parallel as possible. However, if one of the parallel tasks returns NOK, the result of the other task is no longer relevant.

The process of the original situation consists of six sequential knockout tasks A – F. All tasks in the original situation are knock-out tasks, with their own reject probabilities, setup times and service times. All tasks have exponentially distributed setup and service times and it is assumed that all resources have equal setup and service times per task.

It is assumed in this research that the KO tasks have no fail probability. Therefore a task is always completed successfully. It is chosen to only model pure working time. This means that 1 week in the model consists of 40 hours (40*60 = 2400 minutes). Because of this, it is assumed that overtime, part time work, and shifts do not take place in the original situation and are therefore left out of consideration. Various variants of the original model have been used as a starting point for the different redesign possibilities. This is described per redesign rule in the following three subsections.

**Swapping Tasks Rule**

For every case and setup, the KO ratios (reject probability/process time), described by heuristic 1 and 2 of (Aalst, 2001), of every possible combination (e.g. AEDFBC) have

**Figure 2.** Confidence intervals for the lead time and volume flexibility.
been calculated. This resulted in 720 different ratios for every setup. A number of stable (none of the utilizations exceeds 100%) combinations is chosen from the sorted list of ratios and has been compared to the outcomes of the original situation, for several variations in service times, resource classes and allocation.

Applying the swapping tasks rule to processes with knockout tasks results in lower, more balanced utilization and lower WIP costs, both leading to a less costly process execution. In addition, also labor flexibility and volume flexibility increase, which positively influences the performance of the workflow as well. In most processes, implementation of the swapping tasks rule results in a decrease in lead time. However, when the arrival rate is too low to cause queues, or the utilizations of the resource classes are too unbalanced for the rule to balance them, implementation of the swapping tasks rule does not result in a reduction of lead time. External quality, internal quality, process modification flexibility, or any of the other measures are not affected by the swapping tasks rule.

Combining Tasks Rule

The possibility of combining tasks has been investigated for different setups. The rules of heuristic 3 and 4 (Aalst, 2001) have been used to determine what tasks to combine. In this redesign it is assumed that there is no possibility of putting tasks in parallel and that there are no constraints with respect to the order of execution of the tasks and the possibility of combining tasks.

In contrast to the model variants of the swapping tasks rule where the processing times varied, one case has been developed for this rule, because the processing times are constant for all model variants. This case is the starting point for the simulation of all model variants. The different scenario’s are based on changes in arrival rate, resource classes and allocation, and setup time ratios.

Implementation of the combining tasks rule leads to a considerable decrease in lead time. In some settings it also has a positive impact on the utilizations, the WIP costs, labor flexibility and volume flexibility. The combination of two or more KO tasks into one task can lead to too large tasks, which reduces the external quality and the process modification flexibility. The number of tasks and the scope of a task are proxies for internal quality. The number of executed tasks for one case per resource is reduced by the combining tasks rule. This would indicate lower internal quality. However, these tasks will have a larger scope, which would indicate higher internal quality. Overall, internal quality is expected to remain approximately the same.

Parallel Tasks Rule

The redesigned model is a model in which tasks B and C are executed in parallel. Different variations and setups have been simulated for the parallel tasks rule. In this redesign it is assumed that there is no possibility for combining tasks into a composite task and that it is not possible to swap tasks.

Also for the parallel tasks rule different variations have been introduced in order to quantify the impact of the implementation of the rule. Four types of variations have been introduced: Variations in arrival rate, variations in service times, variations in resource classes and allocation and variations in reject probabilities.

Putting sequential KO tasks in parallel leads to a decrease in lead time and to lower WIP costs. The highest positive impact can be expected when the following conditions are satisfied: (1) the service times of the parallel tasks are of the same order of magnitude; (2) the parallel reject probabilities are small; (3) the arrival rates are low; and (4) none of the resource classes are overloaded as a result of putting tasks in parallel. The positive impact of the parallel tasks rule decreases and some measures are even negatively affected when one or more of the conditions are not satisfied.

The increase in number of parallel tasks is a proxy of lower external quality and lower process modification flexibility because the complexity of the workflow increases. Internal quality increases because the number of executed tasks per resource increases, which is a proxy for internal quality.

Quantification of the Triage Best Practice

The main interpretation of the triage best practice is: consider the division of a general task into two or more alternative tasks. Its opposite (and less popular) formulation is to consider the integration of two or more alternative tasks into one general task. When applying the best practice in its main form, it is possible to design tasks that are better aligned with the capabilities of resources and the characteristics of the case, which improves the quality of the workflow. Distinguishing alternative tasks also facilitates a better utilization of resources, with obvious cost and time advantages. On the other hand, too much specialization can make processes become less flexible and efficient, and causes monotonous work with repercussions for quality. These problems are removed by the alternative interpretation of the triage best practice. A special form of the triage best practice is to divide a task into similar instead of alternative tasks for different subcategories of the case type. For example, a special cash desk may be set up for clients with an expected low processing time. The triage best practice is related to the
task composition best practice in the sense that it is concerned with the division or combination of tasks. Note that the best practice differs from the task composition best practice in the sense that alternative tasks are considered.

The triage best practice has been quantified based on the comparison of the original model with a redesign. The original model is again a sequential process with six tasks A – F. In the redesigned model, task B has been divided into two alternative tasks B1 and B2. B1 is a task that is specialized for easy cases and B2 for hard cases. To quantify the impact of the triage best practice, we introduced and simulated four types of variations: (1) arrival rate, (2) service times, (3) arrival ratio easy and hard cases, and (4) resources and resource class setups.

The replacement of a general task into multiple alternative tasks as a result of the implementation of the triage best practice, in all tested situations leads to lower utilizations and higher volume flexibility, which are both positive for the performance of the workflow. In contrast, labor flexibility decreases and labor costs increase in the same situations, which indicates a negative impact on workflow performance. Whether it is advisable to implement the triage best practice depends on the importance of the individual measures. The lead time and WIP costs decrease in models with an arrival rate that is low enough to prevent the occurrence of queue times. The impact on both measures becomes smaller or can even become negative when the arrival rate increases.

The expected external quality is higher, because more specialists are working on a case. The internal quality dimension is negatively affected, since the number of tasks a resource executes decreases, which results in more monotonous work and lower internal quality. In addition, the training of generalists to become specialists induces one-time training costs.

Replacing two or more alternative tasks with one general task leads to opposite results. In most situations it is not advisable to use this variant of the triage best practice.

Most of the best practices, as summarized in the appendix, can be analyzed in the same manner as the parallelism, knockout and triage best practice. This holds true for each of the best practices classified as task best practices, routing best practices and integral process best practices. Further, it holds true for some of the allocation best practices (ASSIGN, FLEX, and CENTR) and resource best practices (NUM, XRES, and SPEC). For the four best practices in these two classes for which this is not the case we found that: (1) specific resource requirements need to be modeled; and (2) the impact on the quality dimension is more severe. In the class of best practices for external parties, only one best practice can be quantified following the same approach (REDUC). The remaining five best practices focus on the entire workflow and/or the exchange of information between parties. The data perspective is not sufficiently included in our current simulation models to quantify these best practices.

Quantification of the three best practices in this research project resulted in some unexpected, counterintuitive outcomes, which are different from the qualitative evaluation results of (Reijers & Limam Mansar, 2005). This may be due to differences in the level of detail of these studies. The qualitative results of (Reijers & Limam Mansar, 2005) were based on expectations and rules of thumb. The predicted impacts were mostly averages, which were based on one measure supplemented with some possible extreme impacts. In contrast, the impacts in this study are the result of employing a complete set of measures for all dimensions, using a simulation model. More measures have been used per dimension, and a more precise impact has been provided. The impacts of the best practices have also been quantified in models with different settings, to obtain a good view of the impact of implementation in different situations.

From a comparison of Van der Aalst’s study on knock-out processes (Aalst, 2001) and this study, it can be concluded that most of the findings of (Aalst, 2001) are supported by the results of this research project. The results of this study also identify situations in which some best practices do not hold true or in which the conditions for the application of the best practice are different. In addition, more aspects of performance have been included, which can be seen as an extension of (Aalst, 2001).

To obtain a complete view on the impacts of the total set of redesign best practices identified by Reijers (Reijers, 2003), the exact impact of the remaining best practices and combinations of best practices should be executed in a future research project. This would support the identification of the correct choice when selecting a redesign best practice to improve a specific performance dimension. Further, the approach should be applied to a real life redesign project to test its applicability to real life data. In this test, the results of individual best
practices should be used to determine what redesign best practice could provide the most favorable results. With respect to generalizability, an interesting research topic would therefore be the relationship between the complexity of a business process and the applicability of the presented approach.

**Practical Implications**

Our approach to the quantification of the impact of a business process redesign project has been focused on the performance of the best practices in general. We used a basic original situation and created some straightforward redesigns based on one best practice. The approach, however, may also be used by practitioners to test their redesigns created with one or more best practices starting from a real business process. Exactly the same steps as defined in the approach, see Figure 1, may be applied for a redesign effort. First, the redesign project is defined by the organization and goals for the redesign, e.g., reduction of throughput time, are set. Then, the process that is currently being executed in the organization is modeled by extending the basic model provided with the approach. Process measures, at least the ones related to the redesign goals, should be included. If such measures are not available the organization first has to collect these. After this, the simulation results of the model are compared with the performance of the actual process to validate the model. Then, one or more redesigns are developed and modeled. We advise to use the best practices (Reijers, 2003) for the creation of redesigns. If the robustness of the redesigns, is important, additional simulations are necessary. In this situation, robustness means that the process still produces acceptable results under varying circumstances. Additional simulations for these variations are required to show this behavior. The simulation settings may be derived from literature. Finally, simulations for the redesign models are performed and the output is analyzed. The performance of the different redesign models is compared with the original model and with one another and the best redesign is selected. Following this approach in business process improvement provides a good indication whether a promising redesign would perform as expected without having to implement the actual changes to find out.

A weakness of the approach is that it cannot quantify the impact of a business process redesign effort on the external and internal quality of a process. Other methods that can be used to quantify the impact on these dimensions should be found. Drawing valid conclusions about the impact of a redesign on internal quality in actual practice calls for a quasi-experimental approach in which operators’ perceptions of work characteristics and social relations are obtained via questionnaires both before and after the implementation of the redesigned business process. If feasible, data from a control condition in which the redesign was not implemented could be used to further support conclusions about the redesign’s impact (Cook & Campbell, 1979). A similar quasi-experimental approach is proposed for measuring and analyzing the quality of the output and the process as perceived by customers. Whether these methods are suitable for the quantification of the impact on both quality dimensions should also be investigated in a subsequent research project.

**Author Bios**

Monique Jansen-Vullers, PhD, is an assistant professor in Business Process Management at the Faculty of Technology Management at the Eindhoven University of Technology. She holds a Master in Computer Science and obtained her PhD in Industrial Engineering and Management Science. She is interested in Process Modeling, Configurable and Executable Reference Models, Business Process Redesign and Simulation.

Ad Kleingeld, PhD., has a PhD in Industrial Engineering and Management Science from Eindhoven University of Technology in The Netherlands. He is an assistant professor at the Human Performance Management group of the Department of Technology Management at this university. His current research in the area of work motivation focuses on managing the performance of individuals and teams through performance measurement, feedback, goal setting and reward interventions.

Irm. Mariska Netjes is a Ph.D. candidate in Business Process Management at the Faculty of Technology Management at the Eindhoven University of Technology. She holds a Master degree in Industrial Engineering and Management. Her main research interests are Business Process Redesign, Business Process Management, Simulation and Petri nets.

**References**


Appendix

BPR Best Practices (Reijers, 2003)

Task best practices: focus on optimizing single tasks within a business process.

1. Task elimination (ELIM): delete tasks that do not add value from a client’s viewpoint.
2. Task addition (ADD): check the completeness and correctness of incoming materials and check the output before it is sent to clients.
3. Task composition (COMPOS): combine small tasks into composite tasks and divide large tasks into workable smaller tasks.
4. Task automation (AUTO): introduce technology if automated tasks can be executed faster, with less cost, and with a higher quality.

Routing best practices: try to improve upon the routing structure of the business process.

5. Resequencing (RESEQ): move tasks to more appropriate places.
6. Knockout (KO): execute those checks first that have the most favorable ratio of expected knockout probability versus the expected effort to check the condition.
7. Control relocation (RELOC): relocate control steps in the process to others, e.g., the client or the supplier, to reduce disruptions in the process.
8. Parallelism (PAR): introduce concurrency within a business process to reduce lead times.
9. Triage (TRI): consider the division of a general task into two or more alternative tasks.

Allocation best practices: involve a particular allocation of resources to activities.

10. Case manager (MAN): make one person responsible for the handling of a specific case.
11. Case assignment (ASSIGN): let workers perform as many steps as possible for single cases.
12. Customer team (TEAM): consider assigning teams out of different departmental workers that will take care of the complete handling of specific sorts of cases.
13. Flexible assignment (FLEX): assign resources in such a way that maximal flexibility is preserved for the near future.
14. Resource centralization (CENTR): treat geographically dispersed resources as if they are centralized.
15. Split responsibilities (SPLIT): avoid assignment of task responsibilities to people from different functional units.
16. **Numerical involvement (NUM):** minimize the number of departments, groups, and persons involved in a process.

17. **Extra resources (XRES):** if capacity is not sufficient, consider increasing the number of resources in a certain resource class.

18. **Specialist-generalist (SPEC):** consider making resources more specialized or more generalized.

19. **Empower (EMP):** give workers most of the decision-making authority and reduce middle management.

Best practices for external parties: tries to improve upon the collaboration and communication with the client and third parties.

20. **Integration (INT):** consider the integration with a process of the client or a supplier.

21. **Outsourcing (OUT):** relocate work to a third party that is more efficient in doing the same work to reduce costs.

22. **Interfacing (INTF):** consider a standardized interface with clients and partners.

23. **Contact reduction (REDUC):** combine information exchanges to reduce the number of times that waiting time and errors may show up.

24. **Buffering (BUF):** subscribe to updates instead of complete information exchange.

25. **Trusted party (TRUST):** replace a decision task by the decision of an external party.

Integral process best practices: applies to the business process as a whole.

26. **Case types (TYPE):** determine whether tasks are related to the same type of case and, if necessary, distinguish separate processes and case types.

27. **Technology heuristic (TECH):** try to elevate physical constraints in a process by applying new technology.

28. **Exception (EXCEP):** design processes for typical cases and isolate exceptional cases from normal flow.

29. **Case-based work (CASEB):** get rid of constraints that introduces batch handling may significantly speed up the handling of cases.
Copyright of Information Systems Management is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.